UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
CURSO DE CIÊNCIA DA COMPUTAÇÃO

RAFAEL OLEQUES NUNES

A classification approach for estimating subjects of bills in the Brazilian Chamber of Deputies

Work presented in partial fulfillment of the requirements for the degree of Bachelor in Computer Science

Advisor: Prof. Dr. Carla Maria Dal Sasso Freitas
Co-advisor: Prof. Dr. Dennis Giovani Balreira

Porto Alegre
September 2023

“4, 8, 15, 16, 23, 42”

— *The Numbers*, LOST

AGRADECIMENTOS

Primeiramente, desejo expressar minha profunda gratidão a Deus pelo dom da vida e por Sua orientação constante em meu caminho. Também sou grato pela inspiração da Boa Mãe, que se tornou uma presença íntima e reconfortante, sempre iluminando meu caminho em direção ao Seu Filho.

Com profunda gratidão e carinho, dedico este trabalho à minha família. Seu apoio incondicional e dedicação foram pilares fundamentais que me permitiram trilhar esta jornada. Em especial, quero expressar minha gratidão aos meus pais, cujo incansável empenho e sacrifício moldaram a pessoa que sou hoje e continuam a guiar-me em direção ao meu potencial completo.

Minha sincera gratidão se estende aos amigos que acompanharam cada passo desta trajetória. Destaco meu padrinho e amigo, Victor Lindner, por sua presença e impacto enriquecedor em minha jornada. Expresso também minha gratidão ao movimento CLJ, assim como aos grupos *Ora et Labora* e Transfiguração do movimento de Emaús, que ofereceram apoio, acolhimento e amizade durante minha jornada em Porto Alegre.

A Professora Carla merece meu reconhecimento especial por sua paciência incansável, apoio contínuo e motivação inspiradora ao longo deste trabalho. Sua orientação magistral e dedicação foram essenciais. Agradeço também ao Professor Dennis, que orientou com dedicação e maestria, sempre aberto a ouvir sugestões, trocar ideias enriquecedoras e compartilhar momentos descontraídos. Minha gratidão se estende a André Spritzer, cujo profundo conhecimento legislativo impactou significativamente este trabalho. A Rodrigo Moni, o qual foi excepcional em meu crescimento profissional e influência motivadora na ideia deste trabalho, também expresso meu agradecimento.

Agradeço ao Henrique dos Santos e à Professora Renata Vieira por me introduzirem à inteligência artificial e processamento de linguagem natural. Minha gratidão se estende ao Professor Joel, à Professora Renata Galante e à Professora Viviane, sempre prontos a esclarecer dúvidas e contribuir com suas sugestões neste trabalho.

Com esses sentimentos em mente, é crucial reconhecer o papel essencial desempenhado por cada um mencionado em minha trajetória acadêmica e pessoal. Expresso minha profunda gratidão a todos por seu impacto duradouro. Vale ressaltar que diversas outras pessoas, ainda que não mencionadas nominalmente, também deixaram contribuições marcantes. A cada um, meu mais sincero e profundo obrigado.

ABSTRACT

A political-legal environment usually involves many documents and stages regarding laws and their processing route. Due to this large volume of data, a considerable amount of essential data, such as subject classification, keywords, and summary, is often missing for bills that are proposed. This issue increases the gap between citizens and politics, negatively affecting society. Considering the Brazilian Chamber of Deputies from 1991 to 2022, around 75% of the bills do not have subject classification included in their associated metadata. However, thanks to many bills in the corpus, this scenario suits machine learning and natural language processing approaches. This study proposes a new method for estimating subjects for the Brazilian Chamber of Deputies' bills. Our solution presents and compares two BERT models adapted for the Portuguese language using the summary information, referring to a brief description or overview of the main points of a political document. We obtained our best results using the BERTimbau model variation, achieving 78.94% of the weighted F1 score and 72.78% of the macro F1 score. To the best of our knowledge, this is the first work to propose a model for predicting the subjects of the Brazilian Chamber of Deputies' bills. Our approach encourages researchers to explore similar techniques for other legal documents. Our findings help political scientists perform a more robust data analysis, which was not possible with the previous data, directly impacting society.

Keywords: Multi-label classification. Legislative documents classification. Text mining. Language models.

Uma abordagem de classificação para estimar temas de proposições na Câmara dos Deputados do Brasil

RESUMO

O ambiente político-legal geralmente envolve diversos documentos e etapas relacionadas a leis e seu trajeto de processamento. Devido a esse grande volume de dados, uma quantidade considerável de informações essenciais, como classificação de tema, palavras-chave e ementa, frequentemente está ausente. Esse problema aumenta o hiato entre os cidadãos e a política, impactando negativamente a sociedade. Considerando a Câmara dos Deputados do Brasil de 1991 a 2022, cerca de 75% das proposições não contêm classificação de tema em seus metadados associados. No entanto, devido a muitas proposições no corpus, esse cenário é adequado para abordagens de aprendizado de máquina e processamento de linguagem natural. Este trabalho propõe um novo método para estimar temas nas proposições da Câmara dos Deputados do Brasil. Nossa solução apresenta e compara dois modelos BERT adaptados para a língua portuguesa usando as informações de ementa, que se referem a uma breve descrição ou visão geral dos principais pontos de um documento político, como um projeto de lei ou uma proposta. Obtivemos nossos melhores resultados usando a variação do modelo BERTimbau, alcançando 78,94% de pontuação F1 weighted e 72,78% de pontuação F1 macro. Até onde sabemos, este é o primeiro trabalho a propor um modelo para prever temas de proposições na Câmara dos Deputados do Brasil. Nossa abordagem aumenta a classificação dos temas das proposições e incentiva os pesquisadores a explorar técnicas semelhantes para outros documentos legais. Nossas descobertas auxiliam os pesquisadores em ciência política a elaborar análises de dados mais robustas, o que não era possível com os dados anteriores, impactando diretamente a sociedade.

Palavras-chave: Classificação multi-rótulo. Classificação de documentos legislativos. Mineração de Texto. Modelos de Linguagem.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

AI	Artificial Intelligence
BCE	Binary Cross-Entropy
BCoD	Brazilian Chamber of Deputies
BERT	Bidirectional Encoder Representations from Transformers
BOW	Bag of Words
FN	False Negative
FP	False Positive
ML	Machine Learning
MLM	Masked Language Model
NER	Named-Entity Recognition
NLP	Natural Language Processing
NSP	Next Sentence Prediction
TF-IDF	Term Frequency – Inverse Document Frequency
TN	True Negative
TP	True Positive

CONTENTS

1 INTRODUCTION

There is a joint effort to approximate society with politics since a society's values, beliefs, and behaviors can shape its political landscape. At the same time, political decisions and policies can profoundly influence society and its members. Although essential, this issue remains open due to an implicit gap between citizens and politics. One needs further efforts to communicate political decisions to the community. To decrease this breach, the field of information visualization contributes with techniques that citizens can use, helping them to understand the process, as presented by Silva, Spritzer and Freitas (2018) and Méndez, Moreno and Mendoza (2022).

However, even though visualization techniques can help, there is a lack of basic data regarding political data. For example, considering data obtained from the Brazilian Chamber of Deputies (BCoD) related to the bills presented from 1991 to 2022, around 75% of them miss their subject classification. This scenario discourages further studies to help understand bills' impact on the population. So, despite the efforts to achieve good results, the available data is often insufficient to give a clear overview of a political context due to missing excerpts. For instance, the subject *Direito Constitucional* corresponds to only 4% of the whole corpus. However, since this data corresponds to around 24% of the non-missing data only (Figure 4.3), how can we be sure that this situation correctly reflects our reality? In which areas should deputies increase their efforts to improve the population's quality? These questions are hard to answer, considering that only 1/4 of the bills have complete data. However, despite the significant missing subjects, the 24% existing data corresponds to 156,016 bills, suiting Machine Learning and Natural Language Processing approaches.

This study proposes to answer the following research question: "Can we reliably estimate the missing bill subjects of the BCoD dataset using BERT models?". In order to achieve this goal, we present and compare the performance of two Portuguese BERT models (SOUZA; NOGUEIRA; LOTUFO, 2020; POLO et al., 2021) applied to the summary information of BCoD, which contains a brief description of the document. We also evaluate our approach using standard Machine Learning methods for the existing data.

The two main challenges encountered in this work were highly unbalanced data and multilabel classification. The BCoD corpus presented a complex imbalance, with certain bill subjects being vastly overrepresented, whereas others were significantly underrepresented, demanding strategies to address this imbalance. Additionally, the intri-

cacies of multilabel classification, where bills can belong to multiple subject categories simultaneously, introduce complexity that necessitates customized machine learning approaches. These formidable challenges motivated us to use data analysis techniques to better understand the data and how to use them in preprocessing, select appropriate evaluation metrics, and comprehensively navigate these obstacles with a better understanding of the results in each subject domain. Our research highlights the significance of these challenges and proposes practical solutions to address them in similar contexts.

The remainder of this work is organized as follows. Chapter 2 introduces the main concepts for understanding this study, while Chapter 3 presents and discusses the most relevant approaches found in the literature. Chapter 4 details the necessary information regarding the Brazilian Chamber of Deputies bill's corpus and its internal structures. Our methodology is explained in Chapter 5, and our experimental evaluation in Chapter 6, followed by results and discussion in Chapter 7. Chapter 8 presents our final remarks, limitations, and future work perspectives. Additional material is included in the appendix.

2 BACKGROUND

In this chapter, we present the main concepts for comprehending this study. Section 2.1 provides a concise overview of the examined legislative domain. Section 2.2 presents the fundamentals of Artificial Intelligence (AI), whereas Section 2.3 elaborates on the realm of Natural Language Processing (NLP). The architecture of BERT, a pivotal component of this study, is introduced in Section 2.4. Section 2.5 outlines the metrics utilized in our analysis and discusses their applications in conjunction with cross-validation techniques.

2.1 Brazilian Legislative Domain

In Brazil, the structure of governmental authority adheres to the principles of the Separation of Powers theory (MONTESQUIEU, 1989), resulting in three distinct branches: Executive, Judiciary, and Legislative. Article 44 of the Brazilian Constitution (CONSTITUIÇÃO, 2006) designates the Chamber of Deputies and the Federal Senate as institutions entrusted with legislative powers. This division encourages thorough debates and establishes a process wherein one house initiates the legislative process, followed by a comprehensive review of the other. Beyond its procedural role, Legislative Power undertakes pivotal functions encompassing governance oversight, representation of the Brazilian populace, and providing a platform for discourse on matters of national significance.

The Brazilian Chamber of Deputies (BCoD) handles a considerable volume of bills annually within this framework. A bill constitutes a formal proposal that is subjected to thorough deliberation. Article 100 of the Internal Regulations of the Chamber of Deputies (DEPUTADOS, 1989) enumerates the various forms a bill can assume. These forms encompass proposals to amend the Constitution, projects, amendments, indications, requests, appeals, opinions, and proposals for supervision and control. The Chamber meticulously reviews and engages in debates surrounding these proposals as part of its legislative process.

The Chamber's Documentation and Information Centre employs thematic areas to categorize bills effectively. A human assigns one or more subjects to each bill, thereby ensuring proper classification. However, an information gap persists, as numerous bills lack a definitive thematic attribution. Table 2.1 illustrates this gap through four examples,

Table 2.1 – Examples of classified and non-classified bills with their respective summary and subjects. The different subjects are separated by semicolons.

Bill	Subjects
SUGERE AO PODER EXECUTIVO, POR INTERMEDIO DO MINISTERIO DA AERONAUTICA, AS PROVIDENCIAS NECESSARIAS NO TOCANTE A HORARIO DE ATIVIDADES DE POUSO, DECOLAGEM E MANUTENÇÃO DE MOTORES E TURBINAS DE AERONAVES, NO AEROPORTO DE CONGONHAS, ESTADO DE SÃO PAULO.	Viação, Transporte e Mobilidade.
Fica expressamente proibida a retirada de qualquer homenagem feitas a pessoas elencadas nesta Lei, pelo Poder Executivo e dá outras providências.	Arte, Cultura e Religião; Direitos Humanos e Minorias; Homenagens e Datas comemorativas.
SUGERE AO PODER EXECUTIVO, POR INTERMEDIO DO MINISTERIO DA PREVIDENCIA SOCIAL, O EXAME DA OPORTUNIDADE E CONVENIENCIA DE PROGRAMA DE INCENTIVO A PREVIDENCIA COMPLEMENTAR.	
Cria a política nacional de valorização da mulher no campo e dá outras providências.	

with the first two correctly classified and the last two lacking thematic assignments.

2.2 Artificial Intelligence

Artificial Intelligence (AI) is an area of Computer Science that branches into many subareas. Authors usually do not have a consensus on only one specific definition of AI. Russell (2010) presents four main perspectives: Thinking Humanly, Thinking Rationally, Acting Humanly, and Acting Rationally. Over the years, many types of AI have been developed, such as Specialist Agents, Problem Solving, Machine Learning Models, etc. In this study, we focus on Machine Learning (ML) approaches.

ML, a subfield of AI, is concerned with developing agents capable of learning from their interactions with the world, and subsequently improving their performance in future tasks (RUSSELL, 2010). This approach becomes particularly valuable when

designers are unable to anticipate all possible variations in a problem, allowing the agents to autonomously adapt and learn from their experiences.

Three primary types of learning exist in ML: supervised, unsupervised, and reinforcement learning (RUSSELL, 2010). Supervised learning entails providing the agent with input-output examples to learn solutions and discern general patterns in mapping inputs to specific output classes (RUSSELL, 2010; MANNING, 2009), which often takes the form of classification problems. On the other hand, unsupervised learning focuses on identifying patterns within the input data without the assistance of corresponding output labels, usually encompassing clustering or recognizing underlying associations among the data. Finally, reinforcement learning centers around the agent learning from its experiences, guided by rewards or penalties, to gradually master a given task. In this study, we will focus on supervised learning.

2.3 Natural Language Processing

A significant portion of online content exists in unstructured textual formats and is abundantly accessible. To address this issue, Natural Language Processing (NLP) provides an array of methodologies designed to extract knowledge from unstructured input. These techniques include text classification, information retrieval, and extraction (RUSSELL, 2010; MANNING, 2009).

In the NLP domain, a collection of texts is referred to as a *corpus*, and when multiple such collections exist, they are known as *corpora*. Beyond simply containing the text, a corpus can also be *annotated*. Annotations can take various forms, such as labeling texts for categorization problems or tagging individual tokens within a text with their corresponding part-of-speech information. In addition, annotations can be extended to semantic labels, such as identifying entities in the text for tasks like part-of-speech tagging or Named Entity Recognition.

2.4 BERT

Bidirectional Encoder Representations from Transformers (BERT) is a language representation model proposed by Devlin et al. (2018). As shown by these authors, BERT is state-of-the-art in some NLP tasks, such as text and token classification. In this section,

we briefly introduce the main concepts of the BERT models.

BERT uses Transformers architecture, proposed by Vaswani et al. (2017), and based on the attention mechanism. The attention mechanism's main point is to better understand the sentence and its words, providing better representations by analyzing how the surrounding words are relevant to a specific one. It allows the creation of different representations of the same word in different contexts, helping, for example, homonym words to have different embeddings, even though they have the same spelling. It is a gain compared to traditional representations, such as Bag of Words (BOW) and term frequency – inverse document frequency (TF-IDF), that every word has the same representation, regardless of its context.

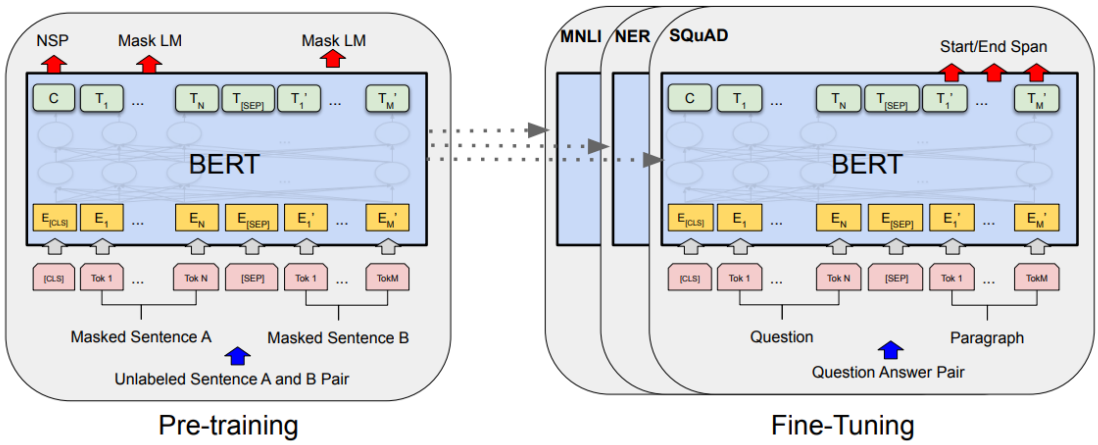
Unlike the previous unidirectional transformer models, with the limitation of only knowing the tokens that came before, BERT introduces a pivotal advancement by obtaining the context from both directions. This is possible using a mechanism called Masked Language Model (MLM). BERT employs an MLM during the pre-training phase using an unlabeled corpus for a pre-training task. In this process, BERT masks a random token and attempts to predict what the original token should fill in that position by learning using the left and right context.

Another technique used in pre-training is Next Sentence Prediction (NSP), in which the main objective is to predict whether sentence B follows sentence A. In this case, Devlin et al. (2018) pointed out that this task is trivial and can be extended for use in any monolingual corpus for a binarized NSP. In practice, 50% of the time is chosen as sentence B that does not follow sentence A, and in the other 50% of the time, it chooses a sentence that follows, each with a respective flag showing the relationship of being the next or not.

These approaches facilitate BERT in capturing richer contextual nuances, enhancing its ability to understand language intricacies and effectively update its parameters. The acquired knowledge of the general domain with these two tasks allows BERT to obtain good results in downstream tasks with fine-tuning, which we call Transfer Learning, using a general domain after learning about a specific domain. Figure 2.1 illustrates the process of pre-training and fine-tuning, showing how a unique pre-trained model can be used to generate many fine-tuned models.

About the fine-tuning step, all pre-trained parameters of BERT are tailored to the specific downstream task. This process involves providing labeled data to BERT and fine-tuning the parameters end-to-end. The outputs of the last layer are harnessed to a

Figure 2.1 – Overall pre-training and fine-tuning procedures for BERT.



Source: Devlin et al. (2018)

specific downstream task. The key advantage of fine-tuning is that the model acquired a robust understanding of word semantics during pre-training. So, adapting the model for a particular task becomes more efficient and expeditious (JURAFSKY; MARTIN, 2023). In the context of classification tasks, which is the focus of this study, the embedding of the special classification token [CLS] is used as input to the classification layer, which enables the model to make predictions based on the distinctive features learned during pre-training and fine-tuning.

There are two versions of BERT: Base and Large. The difference between them is the size: the Base version has 12 layers, 768 hidden states, 12 self-attention heads, and 110M parameters, while the Large version has 24 layers, 1024 hidden states, 16 self-attention heads, and 340M parameters. In addition, Google proposed BERT in English and in a multilingual version, each with a specific Base and a Large version. These variants allow researchers to choose an appropriate trade-off between the model complexity and task requirements. In addition, to use BERT in languages other than English, it is possible to use a multilingual or fine-tuned model from multilingual to a specific language, as in the case of BERTimbau (SOUZA; NOGUEIRA; LOTUFO, 2020) for Portuguese.

2.5 Model evaluation

In this section, we provide an overview of the metrics used in this study, their formulas, and their use. Furthermore, we present the validation technique used in this study

and how it works. This chapter serves as the cornerstone of our model evaluation process, equipping us with tools to measure, understand, and enhance our model's performance, providing a background for insightful analysis, and meaningful conclusions in the realm of machine learning.

2.5.1 Metrics

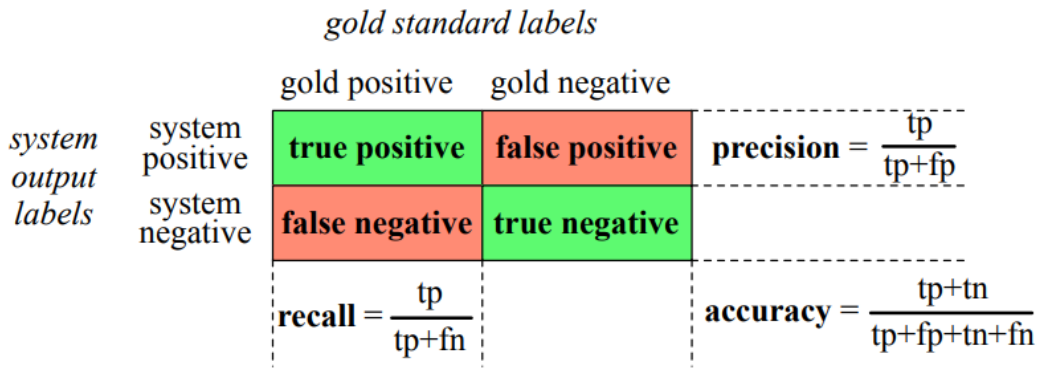
Several techniques can be used to analyze the performance of predictive models. In this study, we use well-established metrics to analyze NLP predictive models and other types of problems (KADHIM, 2019; JURAFSKY; MARTIN, 2023; MANNING, 2009). The main metrics used in this study are precision, recall, and F1 score, as well as macro and weighted variations. Moreover, we also show the results for accuracy and F1 micro.

2.5.1.1 Gold Standard Labels

To calculate the metrics, it is necessary to have a *golden standard set* (MANNING, 2009), i.e., a set of documents with *golden labels*, and a set of labels for each document classified by humans (JURAFSKY; MARTIN, 2023). The usual way to use these labels and the output labels to calculate the metrics is based on a *confusion matrix* (MANNING, 2009).

The confusion matrix, as depicted in Figure 2.2, provides a structured framework for assessing the performance of the classification model, based on the concepts of True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). A TP represents an instance that was correctly predicted by the system with a label that matches the actual label in the golden set. Conversely, an FP represents an instance incorrectly predicted by the system with a label that is not present in the actual instance. Similarly, TN corresponds to an instance in which the system correctly predicts the absence of a specific label. Finally, an FN indicates an instance not predicted by the system as having a specific label, despite being present in the golden set to the particular instance. These distinctions are fundamental for comprehensively evaluating the effectiveness of classification models.

Figure 2.2 – Structure of a confusion matrix.



Source: Jurafsky and Martin (2023)

2.5.1.2 Accuracy

Accuracy is a common metric that indicates the fraction of correct model predictions, as shown in Equation 2.1. It is not the main metric used in text classification (JURAFSKY; MARTIN, 2023; MANNING, 2009); generally, corpora labels are unbalanced hidden misclassifications to the minority class if the majoritarian has good results.

$$Acc_i = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \tag{2.1}$$

2.5.1.3 Precision

Precision is a fraction of how many documents labeled to the positive class are corrected (Eq. 2.2). It is a useful metric since the focus is on minimizing false positives in cases where the model incorrectly predicts the positive class. It is a good approach to classification models since the main objective is to indicate that the positive predictions are reliable and that fewer instances are wrongly classified as positive.

$$P_i = \frac{TP_i}{TP_i + FP_i} \tag{2.2}$$

2.5.1.4 Recall

Recall (or TP rate) measures the proportion of correctly predicted positive class instances relative to the actual number of positive instances (Eq. 2.3). The primary objective of this metric is to assess the model's effectiveness in identifying and predicting TP

instances within the entire set of actual positive labels.

The key distinction between recall and precision lies in their respective focuses. Recall concentrates on assessing the model's ability to correctly predict TP instances exclusively based on the knowledge of the positive class without factoring in the negative class. In contrast, precision considers both the positive and negative classes in its evaluation. This differentiation underlines the nuanced perspectives that these two metrics offer in evaluating the classification model performance.

$$R_i = \frac{TP_i}{TP_i + FN_i} \tag{2.3}$$

2.5.1.5 F1 Score

Precision and recall are paramount metrics in NLP tasks, particularly in text classification scenarios (JURAFSKY; MARTIN, 2023; MANNING, 2009). However, combining these two metrics effectively provides a more comprehensive understanding of model performance. Thus, F1 serves to strike a balance between precision and recall.

The F1 Score is calculated by computing the harmonic mean of the precision and recall, as shown in Equation 2.4. By taking the harmonic mean, the F1 Score gives equal importance to precision and recall while maintaining a harmonious trade-off. It is important to note that the F1 Score is a specific instance of the Fβ Score that accommodates varying weights of precision and recall by adjusting the value of β. For our purposes, where a balanced consideration of both metrics is desired, we chose β = 1. This ensures a balanced assessment of model performance by considering both the ability to avoid false positives and capture true positives effectively.

$$F1 = \frac{2 \times P_i \times R_i}{(P_i + R_i)} \tag{2.4}$$

2.5.1.6 Micro, macro and weighted Averaging

In micro averaging, the individual confusion matrices of all classes are consolidated into a single matrix, and the TP, TN, FP, and FN are summed to calculate the specific metric. However, it is important to note that the final value is skewed towards the majority class, making micro averaging more suitable for balanced classes or situations where specific class distinctions are less relevant.

A specific metric for each class is computed in macro averaging, and a simple

average of these individual values is calculated. This approach offers a distinct advantage in providing a more accurate reflection of the performance across all classes, particularly for minority classes. This distinction sets macro averaging apart from micro averaging, which does not adequately address imbalances in the corpora. Thus, macro averaging is particularly suitable for unbalanced corpora, where certain classes may have fewer instances.

A calculation similar to macro averaging is performed in weighted averaging, but the average is computed using weights that are determined based on the number of instances in each class. The distinction between weighted and macro averaging lies in the fact that the weighted approach assigns importance to each class proportionate to its instance count. This method is particularly beneficial when the distribution of examples for each class in the training corpus reflects the real occurrence distribution.

2.5.2 Stratified k-fold cross-validation

K-fold cross-validation is a widely adopted technique for evaluating ML models (JURAFSKY; MARTIN, 2023; FACELI et al., 2021). In this approach, the dataset is partitioned into k non-overlapping folds, with the common values for k being 5 or 10. Each iteration designated one fold as the test set, whereas the remaining k - 1 folds were utilized for training the model. In some cases, particularly when fine-tuning hyperparameters are not applied, a portion of the training data is reserved for validation, serving as a validation dataset. Repeating this process k times, the model was successively trained and evaluated for every subset of the dataset. This repetitive evaluation helps mitigate the inherent bias that could emerge from training and testing on isolated subsets. This approach helps prevent scenarios where a model performs exceptionally well on an easier test and produces an overly optimistic evaluation.

Stratified k-fold cross-validation was employed to ensure that each fold maintained a proportional representation of examples from all classes, aligned with the distribution in the original dataset (FACELI et al., 2021). This strategy is vital for maintaining consistency in the distribution of class instances across training, validation, and testing sets. It plays a crucial role in enabling the replication of the original dataset distribution patterns in the results.

3 RELATED WORK

Natural Language Processing (NLP) has found various legal applications, aiming at improving citizen understanding and automation of internal processes and professional workflows. In this chapter, we discuss part of the literature on NLP in the legal and legislative domains with a particular focus on the Brazilian Chamber of Deputies data.

Researchers have extensively explored text classification in legal and legislative domains. Chalkidis et al. (2019) worked on a multilabel approach for classifying English legislative documents from the EUR-Lex portal. These documents were organized into an annotated corpus, making them available for further research. Additionally, they experimented with different word-embedding models and classifiers, including BERT, for legislative text classification. In a follow-up study, Chalkidis et al. (2020) systematically investigated the application of BERT in the English legal domain. They explored strategies and hyperparameter tuning to adapt the BERT to legal-specific tasks and corpora.

Assogba et al. (2011) utilized classification techniques to visualize the thematic areas of bill sections, aiding citizens in comprehending legislative data. However, they focused on US federal legislation using techniques predating BERT, which is unrelated to BCoD data.

Regarding the Brazilian legal domain, Silva et al. (2021) addressed comment classification on bills by clustering them into semantic topics. They developed a specialized version of SBERT for the BCoD domain. Albuquerque et al. (2022) also works with BCoD data creating a corpus specialized in Named-entity recognition (NER) in this context. These studies provide valuable insights into the Brazilian legislative domain but do not directly focus on the summary of bills or multi-label classification.

Researchers have also explored classification within the context of legal texts in various domains. Aragy, Fernandes and Caceres (2021) focused on the Courts of Justice of Brazil, with manually labeled data of rhetorical roles for petitions, using Naive Bayes and Support Vector Machine classifiers with Bag of Words and TF-IDF representation. They use two variants of BERT classifiers, with a linear layer and a multilayer perceptron.

Serras and Finger (2022) and Polo et al. (2021) presented approaches for legal texts in Brazilian Portuguese using fine-tuned BERT models. They tested these models in classification tasks in the legal domain, especially in Brazilian court corpora.

Caled et al. (2022) applied a hierarchical multi-label classification approach to Portuguese Portugal using the EuroVoc corpus. They studied different architectures for

this approach, developing a classifier architecture to work with the three levels of the EuroVoc thesaurus, comparing the use of BERT embedding and traditional Word Embedding. In addition, they made available the resultant corpus with 220k documents labeled according to the EuroVoc thesaurus.

Similarly, Avram, Pais and Tufis (2021) adopted a multilingual approach in the context of a multilabel classification task involving legislative documents of the European Union, where EuroVoc thesaurus labels were utilized. The authors employed various monolingual and multilingual BERT models and fine-tuned them to specific legislative domains. This study incorporated the BERTimbau model into Brazilian Portuguese (SOUZA; NOGUEIRA; LOTUFO, 2020).

Additionally, Vianna and Moura (2022) explored various clustering techniques and word-embedding models in Brazilian Legal Documents, specifically from court decisions, including the use of BERTikal (POLO et al., 2021) embeddings to extract topics. Furthermore, this study provides insights into the use of full text and summary texts for bills. Although insightful, these studies did not directly address BCoD data or classify them into thematic areas.

Despite the wealth of literature on NLP for the Brazilian legal domain that provides valuable insights and methodologies, to the best of our knowledge, a gap remains in addressing the specific problem of classifying the subjects of bills in the context of BCoD data. In this study, we aim to fill this gap by leveraging BERT for multilabel classification, contributing to addressing specific challenges within the BCoD.

4 CORPUS

In this chapter, we discuss the selection and filtering criteria used to produce the final corpus and the data source, extracted from the Chamber of Deputies' Open Data API[1]. We also provide relevant statistics on the corpus size and bill's subjects distribution.

4.1 Description

In this study, we selected the bills processed (or being processed) by the BCoD between January 1991 and December 2022. Our corpus includes not only essential details such as the summary, subjects, and keywords but also additional information such as the presentation date, processing date, authors' names, and bill API ID (Table 4.1).

An important point regarding the bills in our corpus is that we opted not to use the bills' full texts. Bill texts are available in PDF format and do not always follow the same pattern. Moreover, many documents were scanned from printed sources. These factors generate noise, making it difficult to use the data, as pointed out by Aragy, Fernandes and Caceres (2021), who required manual noise removal from the 70 documents they analyzed. Given the size of our data, manual processing is impractical. Vianna and Moura (2022) showed that using only the text of the bill summary yields satisfactory results, as it condenses the essential content and does not require full-text processing. Therefore, we decided to use summary text exclusively, rather than the bill's full text.

[1] https://dadosabertos.camara.leg.br/swagger/api.html

Table 4.1 – Information on instances of bills being processed in the Brazilian Chamber of Deputies over the years.

Name	Description	Always exist	Total
id_API	ID of the bill obtained by API of the Chamber of Deputies.	Yes	647.105
summary	Summary of the bill in free text.	No	613.870
subjects	List of subjects of a bill.	No	156.027
keywords	List of keywords in a bill.	No	143.810
authors	List of authors of a bill.	No	646.878
year_query	Year in which the bill was being processed.	Yes	647.105
year	Year in which the bill was submitted.	No	408.613
submitted_date	The full date on which the bill was submitted.	No	647.027

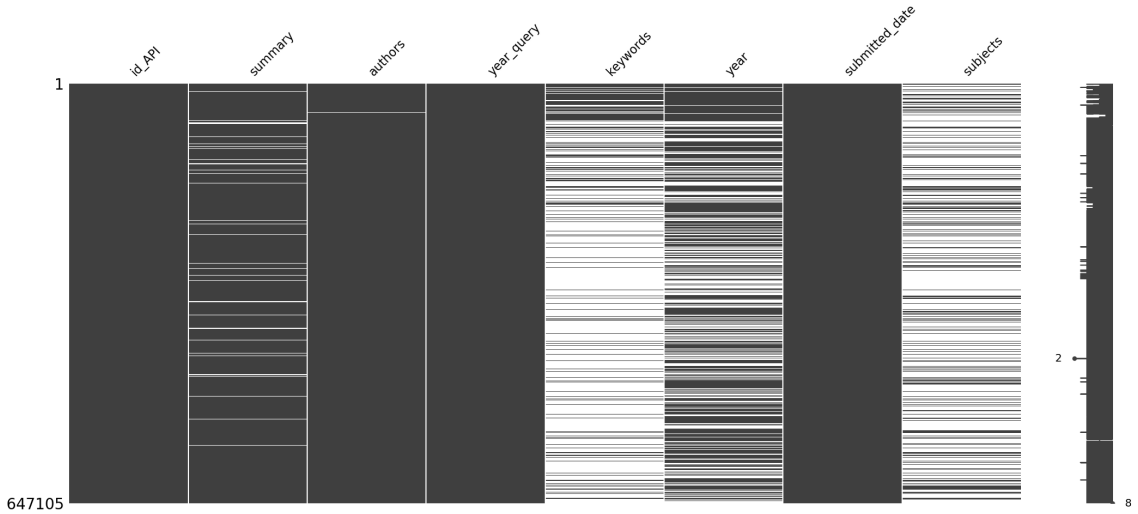


Figure 4.1 – A dense matrix showing the distribution of missing data in the collected data, considering the filter for removing duplicated bills.

4.2 Statistics

Initially, we collected 964,857 processed bills. However, we found that a copy-in-chamber dataset was created each time a bill passed through the processing stage, resulting in duplicates. We ended up with 647,105 distinct bills after deleting these duplicates. Additionally, because our classification approach depended on the summary of the bill, we added another filtering step to eliminate any bills that lacked this data. Consequently, 613,870 bills constituted the filtered dataset. Finally, we filtered out any bills that had no designated subjects to concentrate on bills with clearly defined thematic areas. The final corpus size after filtering was 156,016 bills.

Regarding missing data, Figure 4.1 shows a dense matrix illustrating the distribution of missing values across the collected data, considering the filter for removing duplicate bills. The matrix clearly highlights significant data gaps in attributes, such as *summary*, *keywords*, *subjects*, and *years*, with the latter three exhibiting particularly notable missing values. In addition, *authors* and *submitted_date* suffer from missing data, albeit to a lesser extent. Missing data in *submitted_date* are indistinguishable in this plot, with the lowest count of missing data at 45. For a comprehensive overview of the available data, refer to Table 4.1, which presents the total number of valid values for each attribute.

To help better understand our corpus, Figure 4.2 illustrates the distribution of the bills' summary lengths. We used a logarithmic scale to smooth the curve to improve the visualization. These numbers work together to examine corpus statistics thoroughly.

Figure 4.2 – Frequency of summary lengths in the bill corpus.

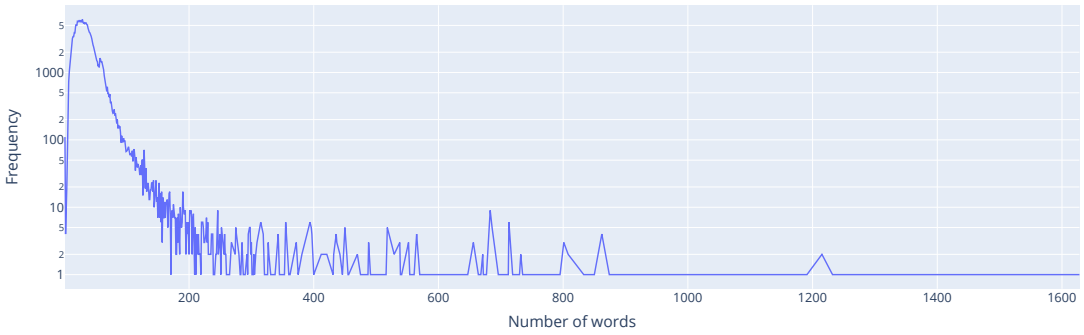


Table 4.2 shows the distribution of bills across different subjects in the corpus, presenting a numerical breakdown of the bill counts for each subject. Using a bar chart to further illustrate this point, we highlight the significant data imbalance among the 32 original classes, which becomes apparent when comparing the three most prevalent classes (20,805, 19,573, and 18,446 bills) to the three least represented classes (268, 7, and 5 bills). This problem requires special attention to the division of the training, validation, and test sets to maintain the original division of the corpus and not to further increase the imbalance between the classes by putting fewer examples of the minority classes and more of the majority. Thus, we adopted a stratified division approach, as outlined in Subsection 5.1.2. This technique ensures that each subset maintains a proportional representation of different subjects, allowing for more robust model training and evaluation.

Figure 4.3 provides an insightful visualization of the thematic information available in our corpus, excluding duplicate bills. Remarkably, over 74% of the bills did not assign thematic information, which is an intriguing aspect to explore in our research. The figure also reveals that the bills exhibit a range of thematic classifications, with the majority having one or two classes assigned. This observation underscores the diversity and complexity of the legislative proposals gathered in our corpus.

Figure 4.3 – Number of subjects per bill.

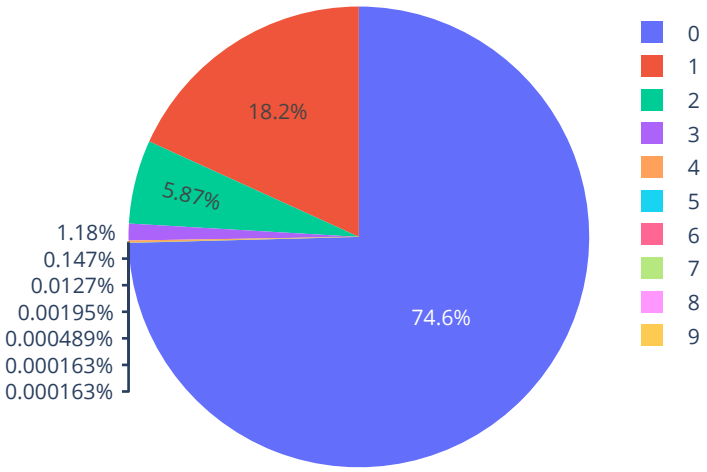


Table 4.2 – Number of bills for each subject.

No.	Subject	Total
1	Comunicações	20805
2	Administração Pública	19573
3	Saúde	18446
4	Educação	14483
5	Finanças Públicas e Orçamento	14393
6	Direitos Humanos e Minorias	13875
7	Trabalho e Emprego	11550
8	Economia	9403
9	Viação, Transporte e Mobilidade	8751
10	Defesa e Segurança	7206
11	Indústria, Comércio e Serviços	6726
12	Direito Penal e Processual Penal	6487
13	Cidades e Desenvolvimento Urbano	6326
14	Meio Ambiente e Desenvolvimento Sustentável	5695
15	Energia, Recursos Hídricos e Minerais	5690
16	Previdência e Assistência Social	5622
17	Direito Civil e Processual Civil	3892
18	Agricultura, Pecuária, Pesca e Extrativismo	3824
19	Relações Internacionais e Comércio Exterior	3381
20	Esporte e Lazer	3197
21	Homenagens e Datas Comemorativas	3094
22	Estrutura Fundiária	2985
23	Direito e Defesa do Consumidor	2953
24	Política, Partidos e Eleições	2548
25	Arte, Cultura e Religião	2538
26	Processo Legislativo e Atuação Parlamentar	2450
27	Ciência, Tecnologia e Inovação	1809
28	Direito e Justiça	833
29	Turismo	828
30	Direito Constitucional	268
31	Ciências Sociais e Humanas	7
32	Ciências Exatas e da Terra	5

5 METHODOLOGY

This chapter discusses the approach used in our study for the multi-label classification of legislative documents from the Brazilian Portuguese Chamber of Deputies. We describe the data preparation procedure, including data extraction and pre-processing, the cross-validation approach used to evaluate our models as well and the specific BERT models utilized in our studies.

5.1 Data preparation

Our data preparation involved two steps: (i) data extraction and (ii) the generation of cross-validation sets. It is important to highlight that transformer models can handle low levels of data cleaning (VIANNA; MOURA, 2022), whereas word representations such as TF-IDF and BoW require more attention in this aspect. Hence, we do not emphasize data cleaning because these models were not used in this study. Instead, we present our data cleaning phase within the data extraction step.

5.1.1 Data extraction

In the data extraction phase, we collected bills between 1991 and 2022 using the BCoD API (INFORMAÇÃO, 2022) provided by The Chamber's Directorate of Innovation and Information Technology. All data were provided as JSON files that were processed to collect information from bills, authors, and subjects. All information related to a bill is structured in a document of our corpus and stored in a MongoDB database, as described in Chapter 4.

During the sending of data to the database, the first preprocessing step is the removal of the special characters "\n", "\t" and "\r" in the text-type fields, which we replaced by a blank space. Subsequently, we replaced the spaces with only one space that may have been inserted in the previous processing step. In addition, we converted all keywords to lowercase to facilitate further processing. We find this approach appropriate because the order of the keywords does not have a specific semantic load.

5.1.2 Cross-validation sets

To ensure a robust evaluation of our models, we employed a 10-fold cross-validation strategy using the multi-label stratification algorithm proposed by Sechidis, Tsoumakas and Vlahavas (2011) and following the benchmark set by Caled et al. (2022). The implementation of this algorithm is available on GitHub[1]. We focused on retaining legislative text summaries (input) and subjects (output) for our classification task.

The cross-validation sets were preprocessed and stored for faster data loading and enabled training on different models using the same folds. In addition, it can facilitate future ensembles of multiple models trained on the same folds.

We utilize stratified k-fold cross-validation to maximize the amount of training data and to perform testing on a larger number of folds (CALED et al., 2022; LILLIS; NULTY; ZHANG, 2022; LIMSOPATHAM, 2021; AMBALAVANAN; DEVARAKONDA, 2020). This approach provides a more comprehensive and statistically robust evaluation of our models, ensuring reliable and unbiased results by testing a single set of folds. During each cross-validation iteration, nine folds were used for training, allocating 10% of the training set for validation, and one fold for testing. It is important to highlight that all models were trained and evaluated on the same folds, ensuring a fair and consistent evaluation across different experiments.

5.2 Models

This section presents the two different BERT models used in this study. Since we are dealing with a specific Brazilian Portuguese data domain, we conducted our experiments using two models: the first fine-tuned to Brazilian Portuguese in the general domain, and the second tailored to Brazilian legal content.

5.2.1 BERTimbau

BERTimbau is a pre-trained BERT model specifically fine-tuned for Brazilian Portuguese (SOUZA; NOGUEIRA; LOTUFO, 2020), which uses the brWaC corpus (FILHO et al., 2018) as the starting point. BERTimbau has 12 layers, 768 hidden dimensions, 12

[1] https://github.com/trent-b/iterative-stratification

attention heads, and 110M parameters. This model has been used to advance the state-of-the-art model for named entity recognition, sentence textual similarity, textual entailment, and other NLP tasks in Brazilian Portuguese (ZANUZ; RIGO, 2022; CONSOLI; VIEIRA, 2021; SOUZA; NOGUEIRA; LOTUFO, 2020). For this study, we utilized the base version of BERTimbau provided on the Hugging-Face Hub.

5.2.2 BERTikal

BERTikal is a BERT model that was explicitly fine-tuned for the Brazilian legal language, as proposed by Polo et al. (2021). BERTikal shares the same model specifications as the BERTimbau base case, including 12 layers, 768 hidden dimensions, 12 attention heads, and 110M parameters. The key difference between the models lies in the fine-tuning process of BERTikal, which specifically targets the Brazilian legal language using the BERTimbau base case model. This characteristic ensures that BERTikal was optimized for the unique linguistic characteristics and terminology found in legal texts, making it interesting for our experiments on classification tasks. The training data for BERTikal consists of documents from Brazilian courts, including publications, motions, and longer legal documents, predominantly from the Court of Justice of São Paulo (TJSP).

5.3 Predicting Subjects of Bills from their Summaries

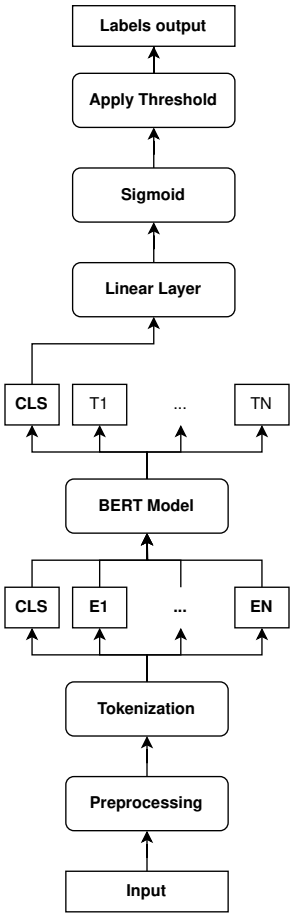
Firstly, this section presents the two corpora used in our classifier as variants of the corpus described in Chapter 4. Then, we describe the classifier architecture, represented in Figure 5.1.

5.3.1 Corpora

From the corpus described in Chapter 4, we created two corpora for the experiments. The first was our standard corpus, which is composed of all the 156,016 bills with a unique summary without repetition, as previously described. Both corpora were formatted in CSV files, enabling streamlined integration for use in the models and analyses. The array types were converted into strings, and the values were separated using a semicolon.

The second is the standard corpus without the classes *Ciências Exatas e da Terra*

Figure 5.1 – Our classifier architecture.



and *Ciências Sociais e Humanas* because these classes have seven and five bills, respectively. Such small classes are challenging because they do not allow the division to the stratified 10-fold cross-validation, which means the model cannot learn the patterns of these classes. Another problem is that the scarcity of data for these classes makes approaches such as data augmentation difficult because to have an acceptable number for training, we would have more than twice the amount of synthetic data. Therefore, it seemed reasonable to remove these classes.

5.3.2 Classifier

During the preprocessing phase, we exclusively utilized the “clean_bert” function for the BERTikal model, following the guidance of its authors Polo et al. (2021). This function was specifically employed to clean BERTikal inputs by removing special encoded characters such as "\n" and multiple spaces. Next, we split the labels of each bill in text format into an array using the semicolon character as the delimiter. We tokenize all

bill summaries and create a label array for each bill. The tokenizer used in this step corresponds to each model (BERTimbau and BERTikal). Finally, we created a Hugging-Face dataset to improve compatibility and optimization with the Hugging-Face Trainer API.

As previously mentioned, we used the Hugging Face Trainer API for the multilabel fine-tuning task. The Trainer discarded the pre-trained head of the selected model and randomly initialized a classification head to transfer the knowledge of the model and fine-tune it to our specific task. In this case, a linear layer receives the output of the classification special token [CLS] as the input. It takes a 1-dimensional array of 768 dimensions (the [CLS] vector) and maps it to a 1-dimensional array of 32 positions, representing the probabilities for each class.

The [CLS] token output (the final hidden state) captures the contextual representation of the entire input sequence, compressing all information into a fixed-size vector, what is commonly used to make predictions in classification tasks. The tokenizer automatically added a [CLS] token at the beginning of all the sentences during tokenization.

We used a sigmoid function to obtain the probabilities for each class. In multiclass classification tasks, a softmax function is commonly used, ensuring that the sum of the probabilities for all classes equals 1. However, in multilabel classification, such as in our case, where each bill can be associated with multiple thematic classes, the sigmoid function is more suitable for usage (CALED et al., 2022; TANG; TANG; YUAN, 2020; AVRAM; PAIS; TUFIS, 2021). The sigmoid function provides an independent conversion of the real number to probabilities that vary between zero and one for each class. Thus, we can obtain the probability of each class separately, which aligns with the requirements of a multilabel classification task, where multiple classes can be assigned to each input instance.

Once we obtained a list of probabilities, we used a threshold to classify the values. We set probabilities below the threshold to 0, and values above the threshold were set to 1, meaning belonging to the class or not belonging to the class, respectively. In this study, we used a fixed threshold of 0.5, following Tang, Tang and Yuan (2020). By applying this threshold, our model provides the final predictions for the multilabel classification task.

6 EXPERIMENTAL EVALUATION

In this chapter, we present the setup used to train and evaluate our models, the hyperparameters utilized in the fine-tuning process, and the metrics for evaluation. We conducted experiments to compare different models and subset corpora by employing the HuggingFace Trainer API for training with consistent hyperparameters across all models.

6.1 Setup

Our setup consisted of a computer with an Nvidia GeForce RTX 3060 GPU and 32.0 GB of memory, which we used for training and evaluating the models. We extracted data on a personal computer with an Intel i7-12700 CPU, 16.0 GB of memory, and 20 threads.

We also chose the Python 3.7.6 programming language due to the variety of libraries for Machine Learning and Natural Language Processing. MongoDB Atlas[1] was used to store the data in the cloud for easier access.

For text processing, we leveraged the Hugging Face Transformers package introduced by Wolf et al. (2019), along with its training API[2], to streamline the training development process and provide adaptability through various hyperparameters. In this study, our approach involves utilizing the pre-trained BERT model in Portuguese, BERTimbau (SOUZA; NOGUEIRA; LOTUFO, 2020), which is available on the Hugging-Face website. Furthermore, we adopted the BERTikal model (POLO et al., 2021), which involves fine-tuning BERTimbau for legal texts in Portuguese. Despite considering the verBERT model (SERRAS; FINGER, 2022), its accessibility was unavailable during the project's development phase.

The performance metrics we used come from the scikit-learn libraries[3]. Finally, to generate the visualizations and analyze the data, we choose ploty[4], missigno library[5] and transformers-interpret libraries[6].

[1] https://www.mongodb.com/atlas/database
[2] https://huggingface.co/docs/transformers/main_classes/trainer
[3] https://scikit-learn.org/
[4] https://plotly.com/
[5] https://github.com/ResidentMario/missingno
[6] https://github.com/cdpierse/transformers-interpret

6.2 Hyperparameters

We conducted a set of experiments to evaluate the models and corpora described in Chapter 5. For consistency, we maintained uniformity across all models by employing identical hyperparameters during training facilitated through the HuggingFace Trainer API. Our choice of hyperparameters was significantly informed by the findings presented in Tang, Tang and Yuan (2020), which correspond to multi-label classification in an unbalanced domain. In addition, the study by Sun et al. (2019) influenced our decision because of its insight into fine-tuning text classifiers across different scenarios, the specific parameters are shown in Table 6.1. Moreover, we draw inspiration from the approaches outlined in Aragy, Fernandes and Caceres (2021) and Aguiar et al. (2021), both of which operate within similar domains.

We used an AdamW optimizer (LOSHCHILOV; HUTTER, 2017), Adam β_1 of 0.9, Adam β_2 of 0.999, and a learning rate of 2e-5 to optimize our models. We used batch sizes of 8 and 4 epochs to fine-tune each model. Regarding the tokenizer, we initially set the maximum sentence length to 512, taking into consideration the diverse range of bill sizes within our corpus, as elaborated in Chapter 4. Additionally, using the BERTimbau tokenizer, we experimented with a maximum sentence length of 75, prompted by the observation that the average number of tokens per bill in our standard corpus was 45, with a standard deviation of 30. The data are presented in Figure 6.1, which illustrates the token distribution across the corpus. In both cases, we applied padding to reach the maximum sentence length while truncating excess tokens. We calculated the loss function using binary cross-entropy (BCE), which is the default loss function of the HuggingFace Trainer API for multilabel classification.

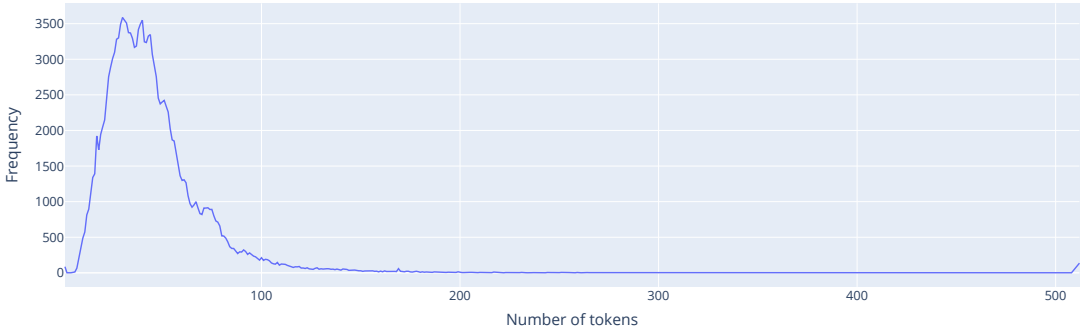
6.3 Metrics

The main metric used in this study was the F1 score in the macro and weighted variants. We chose these variants because of the nature of our domain and two assumptions: (i) all the subjects are important to the final classification because all are valid subjects that a bill can have, so macro F1 is important to address this global view; (ii) subjects with more bills can be more important than others, because some subjects have a small set of bills, such as (*Ciências Sociais e Humanas*), which have only seven instances, meaning that these subjects are infrequent and less likely to happen in the future,

Table 6.1 – Comparison of hyperparameters used in the experiment and similar studies

Hyperparameter	Tang, Tang and Yuan (2020)	Sun et al. (2019)	Aragy, Fernandes and Caceres (2021)	Avram, Pais and Tufis (2021)	Briskilal and Subalalitha (2022)	Our Model
Adam β_1	-	0.9	-	-	-	0.9
Adam β_2	-	0.999	-	-	-	0.999
Learning Rate	2e-5	2e-5, 5e-5	1e-5	6e-5	2e-5	2e-5
Batch Size	16	32, 24	4	8	16	8
Epochs	4	4	4	30	5	4
Loss Function	-	-	-	-	-	BCE
Type	Multilabel	Single label	Single label	Multilabel	Single label	Multilabel

Figure 6.1 – Frequency of number of tokens per bill in the corpus using the BERTimbau tokenizer.



so weighted F1 is important to obtain the impact of the more frequent classes.

In addition to macro and weighted F1 for the global result, we calculated micro F1, macro precision, macro recall, and accuracy. We also calculated the F1, precision, and recall for each class.

To obtain better confidence in the results, we used stratified k-fold cross-validation to compute the metrics. We choose stratification because we have not only multi-class and multi-label but also unbalanced classes, demonstrating the need of stratification to replicate the characteristics of the corpus in all folds.

7 RESULTS AND DISCUSSION

In this chapter, we explore our experiments' outcomes by delving into global and local results. The global performance of the two models, corpora employed, and number of tokens used are briefly presented in Table 7.1. These findings highlight BERTimbau's superior performance across all metrics for both the corpora. We attribute this success to the context of BERTikal's fine-tuning of Supreme Court data, which may include more specialized languages and formats that are not optimally suited for application within the BCoD data.

Consequently, our discussion primarily focuses on the results achieved by BERTimbau. Meanwhile, BERTikal's performance followed a comparable methodology but yielded lower results; the specific results for each test case in the global and local scenarios can be found in Appendix A.

7.1 Global results

The results presented in Table 7.1 show the BERTimbau model's consistent superiority over BERTikal in both macro and weighted F1 scores across diverse corpora and token configurations. The macro F1 scores attained by BERTimbau range from 68.06 to 72.78, while the corresponding values for weighted F1 vary between 78.72 and 78.99. These findings underscore the robust and dependable performance of BERTimbau and establish it as a promising choice for classification tasks in the legislative domain.

Furthermore, echoing the approach of Vianna and Moura (2022), a hypothesis regarding potentially missing subjects within the gold corpus introduces a compelling perspective. The recall metric gains prominence in this scenario, as it can potentially increase with the prediction of absent subjects in the original corpus. However, it is crucial to acknowledge that this assumption could lead to inaccurate predictions, adversely impacting the precision. Consequently, exploring this hypothesis serves as a reminder of the complexity of interpreting these metrics. Section 7.2 presents a practical example of classifier predictions validated by a domain expert, emphasizing the model's potential for identifying previously unseen subjects.

When evaluating the impact of token numbers, it is noteworthy that a marginal enhancement in F1-weighted scores with 512 tokens compared with 75 tokens might not hold substantial practical significance. For instance, considering the macro F1 score for

the standard corpus in BERTimbau, the performance was 78.99 ± 0.22 with 512 tokens, whereas it was 78.76 ± 0.25 with 75 tokens. However, these results imply that selecting 75 tokens offers computational advantages, particularly for real-time applications. This applies to both the prediction and training phases, as evidenced by the reduction in processing time from eight to two hours per fold.

The distinction between the standard corpora and the one without the *Ciências Exatas e da Terra* and *Ciências Sociais e Humanas* classes has a limited impact on the overall results. This exclusion primarily benefits the macro F1 score because classes with few examples are removed. Although this corpus variant may have improved the macro F1 scores, it did not necessarily enhance the overall results. It is important to differentiate between improved macro F1 and the overall model efficacy, especially when the class distribution varies significantly. The fact that the second corpus had a better result in macro F1 is unsurprising, given the exclusion of two classes that had 0% of F1, justifying that these classes did not have significant examples of running in ten folds cross-validation, wherein the first class comprises just five instances and the second, seven.

Additionally, Table 7.2 provides an insightful view of BERTimbau's performance across folds for the standard corpus with 512 tokens. The uniformity in metrics across folds and the average macro F1 score of 72.78 (with a standard deviation of 0.28) reaffirmed the reliability and stability of the model across different data partitions. This uniformity in the performance metrics across folds enhances the model's credibility and reinforces its consistent performance.

Table 7.1 – Global results using stratified 10-fold cross-validation to BERTimbau and BERTikal model into the (A) standard corpus and (B) the standard corpus without the classes *Ciências Exatas e da Terra* and *Ciências Sociais e Humanas*.

Model	*Corpus*	*Tokens*	$F1_{macro}$	$F1_{weighted}$	$F1_{micro}$	$Recall_{macro}$	$Precision_{macro}$	*Accuracy*
BERTimbau	A	512	68.25 ± 0.38	**78.99 ± 0.22**	**79.38 ± 0.21**	65.54 ± 0.41	72.21 ± 1.11	64.92 ± 0.25
BERTimbau	B	512	**72.78 ± 0.28**	78.94 ± 0.23	79.34 ± 0.25	**69.92 ± 0.32**	76.62 ± 0.49	**64.96 ± 0.28**
BERTimbau	A	75	68.06 ± 0.41	78.76 ± 0.25	79.14 ± 0.24	65.25 ± 0.46	71.78 ± 0.58	64.74 ± 0.23
BERTimbau	B	75	72.50 ± 0.26	78.72 ± 0.19	79.13 ± 0.19	69.49 ± 0.41	**76.72 ± 0.68**	64.76 ± 0.19
BERTikal	A	512	66.86 ± 0.31	77.82 ± 0.27	78.30 ± 0.25	63.94 ± 0.46	71.57 ± 0.96	63.74 ± 0.34
BERTikal	B	512	71.58 ± 0.30	77.92 ± 0.19	78.38 ± 0.19	68.50 ± 0.42	75.99 ± 0.40	63.82 ± 0.36
BERTikal	A	75	66.66 ± 0.38	77.65 ± 0.26	78.15 ± 0.25	63.41 ± 0.48	71.38 ± 0.59	63.68 ± 0.32
BERTikal	B	75	71.33 ± 0.28	77.73 ± 0.20	78.19 ± 0.20	68.06 ± 0.32	76.02 ± 0.68	63.72 ± 0.32

7.2 Local results

In this step, we present an overview of the performance metrics obtained from our multilabel classification model for legislative bills for each class, focusing on the

Table 7.2 – BERTimbau fine-tuned with the standard corpus without the classes *Ciências Exatas e da Terra* and *Ciências Sociais e Humanas* metrics to each fold, using 512 tokens as the maximum sentence length. In the last two rows, we have the average and standard deviation for each metric in the 10-fold.

Folds	$F1_{macro}$	$F1_{micro}$	$F1_{weighted}$	$Recall_{macro}$	$Precision_{macro}$	*Accuracy*
Fold 0	72.57	79.38	78.99	69.99	75.95	65.14
Fold 1	72.89	79.78	79.32	69.74	77.27	65.22
Fold 2	73.34	79.74	79.31	70.28	77.39	65.38
Fold 3	72.42	79.00	78.58	69.46	76.56	64.45
Fold 4	72.53	79.12	78.76	69.78	76.10	64.79
Fold 5	72.96	79.32	78.94	70.26	76.71	64.89
Fold 6	72.70	79.23	78.84	69.44	77.01	64.96
Fold 7	73.05	79.26	78.87	70.31	76.60	64.76
Fold 8	72.68	79.35	78.99	69.92	76.18	64.83
Fold 9	72.66	79.22	78.81	70.04	76.39	65.23
Average	72.78	79.34	78.94	69.92	76.62	64.96
Std. Dev.	0.28	0.25	0.23	0.32	0.49	0.28

BERTimbau model using the standard corpus without classes *Ciências Exatas e da Terra* and *Ciências Sociais e Humanas*, using 512 tokens as the maximum sentence length. Table 7.3 contains the key evaluation metrics that are widely used to assess the model's classification capabilities, such as precision, recall, and F1-score. The classes are listed in the table based on their respective F1 scores, with the highest-performing classes ranked at the top. We obtained 23 classes with F1 results greater than 70% and two classes with F1 values greater than 90%. In addition, 28 classes had F1 scores greater than 50%, indicating that the model does not work as a random classifier in most classes.

By analyzing the results, we expected the F1 of classes *Comunicações*, *Educação*, and *Saúde* to be greater than 90%, once these three classes were among the four classes with more bills, with 20,805, 14,483 and 18,446, respectively. A large amount of data helped achieve better results of 95.96%, 92.63%, and 86.83% with a smaller standard deviation. However, by performing a syntax analysis on the summary of these data using lemmatization at the unigram level, we discovered that these classes had significant words in their domains with a high frequency.

Figure 7.1 show that the most frequent word to *Comunicações* is *radiodifusão* a specific word about the transmission used in radio, television, etc. In the same way, we have specifics words to *Educação*, like *educação*, *ensino*, *escolar* and *universidade*. The same is true for *Saúde* with *saúde* and *covid-19*.

On the other hand, we also have classes with significantly less data and good results, such as *Esporte e Lazer* with 2,548 and *Turismo* with 828 bills. *Esporte and*

Table 7.3 – BERTimbau fine-tuned with the standard corpus without the classes *Ciências Exatas e da Terra* and *Ciências Sociais e Humanas* local metrics to each class. Each value is in the format "xx ± yy", which is the average and standard deviation of the stratified 10-fold cross-validation.

Order	*Target*	*F1*	*Precision*	*Recall*
1	Comunicações	95.96 ± 0.33	96.36 ± 0.49	95.57 ± 0.38
4	Educação	92.63 ± 0.50	93.25 ± 0.78	92.03 ± 0.66
3	Saúde	86.83 ± 0.52	88.20 ± 0.49	85.51 ± 1.11
20	Esporte e Lazer	86.80 ± 1.46	88.30 ± 2.14	85.42 ± 2.35
15	Energia, Recursos Hídricos e Minerais	85.26 ± 0.70	85.44 ± 1.00	85.11 ± 1.75
12	Direito Penal e Processual Penal	84.09 ± 0.58	88.05 ± 1.79	80.53 ± 1.58
21	Homenagens e Datas Comemorativas	84.06 ± 1.50	80.54 ± 1.83	87.94 ± 2.27
19	Relações Internacionais e Comércio Exterior	83.61 ± 1.46	84.65 ± 1.44	82.63 ± 2.12
24	Política, Partidos e Eleições	83.59 ± 1.62	82.13 ± 1.49	85.17 ± 2.86
14	Meio Ambiente e Desenvolvimento Sustentável	80.88 ± 1.26	83.24 ± 1.64	78.72 ± 2.57
7	Trabalho e Emprego	80.20 ± 1.03	84.36 ± 1.33	76.44 ± 1.27
29	Turismo	79.68 ± 3.47	79.95 ± 4.30	79.46 ± 3.31
9	Viação, Transporte e Mobilidade	78.82 ± 0.73	79.38 ± 1.25	78.28 ± 0.96
5	Finanças Públicas e Orçamento	77.29 ± 0.98	82.88 ± 1.27	72.41 ± 1.22
26	Processo Legislativo e Atuação Parlamentar	76.45 ± 1.09	79.40 ± 2.94	73.80 ± 1.38
16	Previdência e Assistência Social	75.97 ± 1.74	81.21 ± 1.86	71.42 ± 2.49
18	Agricultura, Pecuária, Pesca e Extrativismo	75.47 ± 1.76	80.92 ± 1.56	70.82 ± 3.24
13	Cidades e Desenvolvimento Urbano	75.12 ± 0.95	81.48 ± 1.51	69.76 ± 2.14
17	Direito Civil e Processual Civil	74.51 ± 1.35	79.98 ± 2.16	69.78 ± 1.71
6	Direitos Humanos e Minorias	73.33 ± 0.74	78.64 ± 1.01	68.72 ± 1.14
10	Defesa e Segurança	71.40 ± 1.49	75.17 ± 2.18	68.01 ± 1.39
22	Estrutura Fundiária	70.50 ± 1.86	73.78 ± 1.69	67.60 ± 3.30
25	Arte, Cultura e Religião	70.07 ± 2.38	75.34 ± 3.47	65.56 ± 2.43
2	Administração Pública	69.84 ± 0.86	76.08 ± 0.86	64.56 ± 1.23
8	Economia	68.43 ± 1.39	75.27 ± 1.19	62.77 ± 2.08
27	Ciência, Tecnologia e Inovação	64.68 ± 3.06	69.45 ± 5.74	60.75 ± 2.80
11	Indústria, Comércio e Serviços	57.46 ± 1.28	64.59 ± 1.99	51.81 ± 1.95
23	Direito e Defesa do Consumidor	56.73 ± 2.23	64.14 ± 2.05	51.03 ± 3.68
28	Direito e Justiça	23.75 ± 4.09	46.30 ± 5.18	16.08 ± 3.34
30	Direito Constitucional	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00

Lazer have *esporte*, *saúde*, and *academia* as examples of specific words, when *Turismo* contains words such as *turismo*, *turístico*, and *São Paulo*, as shown in Figure 7.1b.

These examples help us to understand the model's decision-making process and alignment with domain-specific knowledge. The results show that a large number of examples is important; however, the specific vocabulary of the domain of each class is an essential feature of the classifier.

It also clarifies the lower results of *Direito e Defesa do Consumidor*, *Direito e Justiça* and *Direito Constitucional*, as shown in Figure 7.2 to *Direito e Defesa do Consumidor* and *Direito e Justiça*. These classes have a high frequency of non-specific words such as *altera*, *lei*, and *nacional*. Compared with *Turismo*, which only had more bills than *Direito Constitucional*, the lowest class with 268 examples, these classes had a bad result, probably from their general vocabulary, specific to the legislative domain but general concerning the other classes. The same applies to *Adminstração Pública*, even if it is the second class with more bills, it has a lower result of 69.84%, compared with *Comunicações*, *Educação*, and *Saúde* – as discussed below – once that the class has many generic terms, such as *sobre*, *lei*, and *federal*.

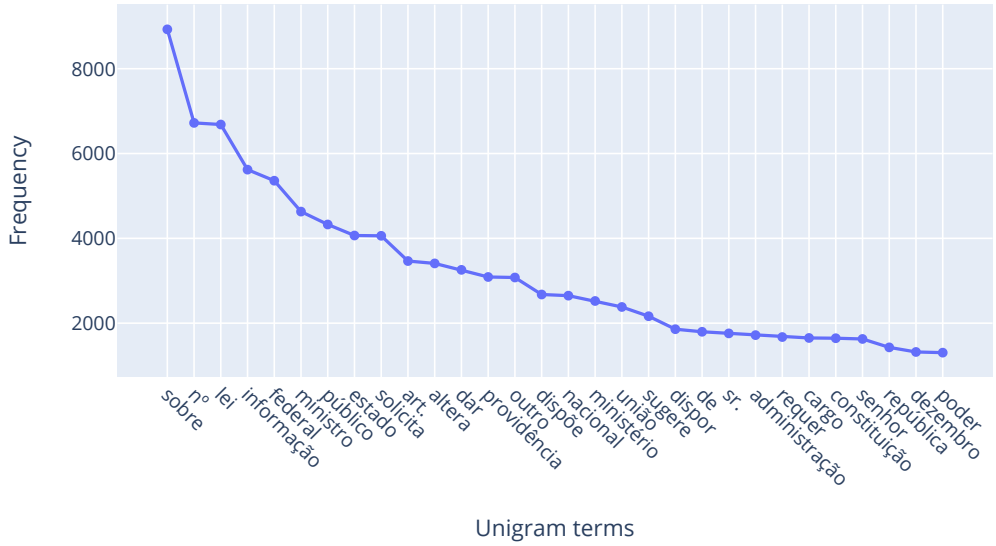
Regarding *Direito Constitucional*, the class with the worst F1 score, it is interesting to analyze its predictions. For better understanding, we selected the data test of the first fold with a set of 27 bills. Figure 7.3 shows the pipeline used to analyze bill predictions. The first column expresses our previous discussion on how classes about the law can have intersections between them with their similar vocabulary, one time that we have five predictions of *Direito e Justiça*, three of *Direitos Humanos e Minorias*, and one of *Direito Penal e Constitucional* and *Direito Civil e Processual*, even if the other classes also have content relations with politics and law. Interestingly, only five classes were in the original corpus, and none were in the law class.

A specialist in the legislative domain was also asked to curate our predictions. The specialist pointed out that nine of the fourteen incorrect predictions in the original labels were correct. Similarly, seven of the twenty five predictions pointed out as correct were bills that only had *Direito Constitucional* in their original labels. These discoveries suggest that our model can discover different subjects even when the original subject is not predicted.

Another interesting point of the analysis regards two bills labeled as *Política, Partidos e Eleições*, identified as wrong subjects by the specialist. These bills talk about plebiscites, as we can see in the follow example: "Dispõe sobre a realização de plebisc-

Figure 7.1 – Scatterplots for *Comunicações* and *Turismo* showing information about the thirty most frequency unigrams to each subject.

(a) Thirty most frequent words for the subject *Comunicações*.



(b) Thirty most frequent words for the subject *Turismo*.

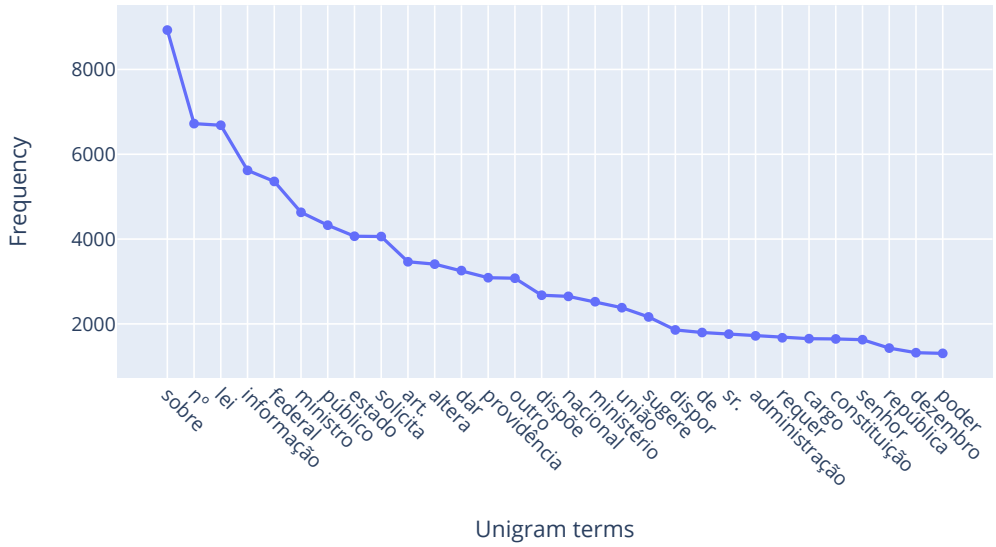
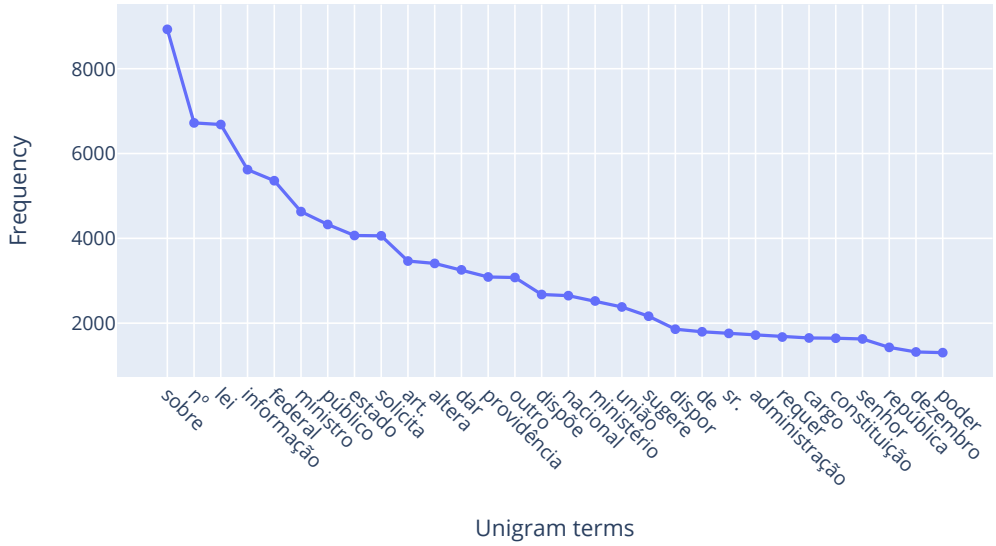


Figure 7.2 – Scatterplots for *Direito e Justiça* and *Direito Constituciona* showing information about the thirty most frequent unigrams to each subject.

(a) Thirty most frequent words for the subject *Direito e Justiça*.



(b) Thirty most frequent words for the subject *Direito Constitucional*.

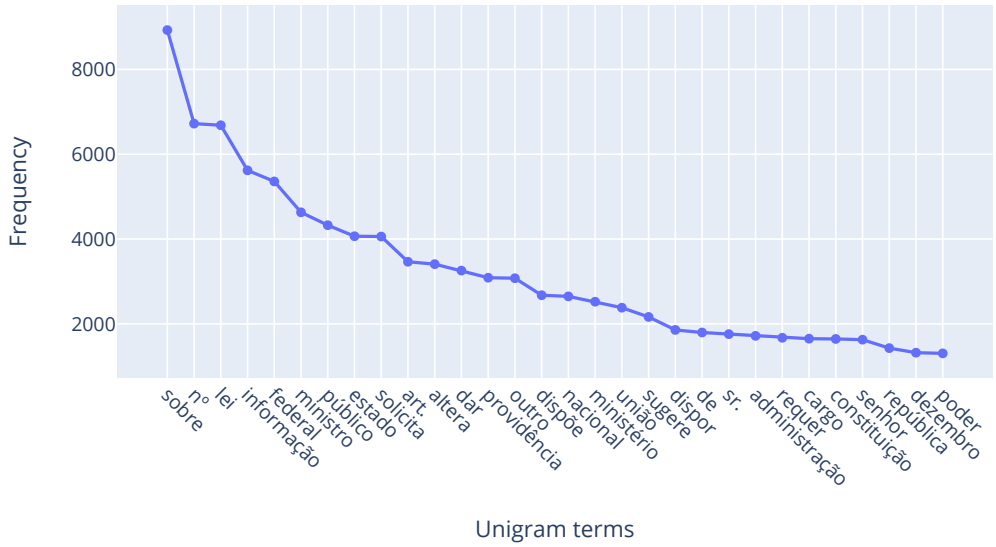
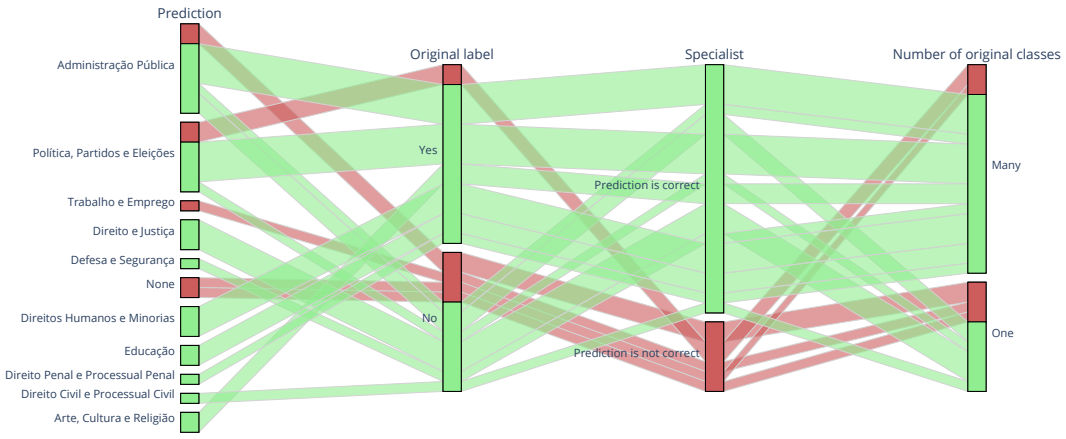


Figure 7.3 – Diagram illustrating the analysis pipeline for bills categorized under the subject *Direito Constitucional* in the first fold. The first column shows the predicted class, connected to the second column indicating their presence in the original labels of each bill. Subsequently, the specialist's assessment is included to determine if the predicted class aligns with the original set, distinguishing between multi-label and single-label bill categorization.



ito para a criação do Território Federal do Marajó." ("Deals holding a plebiscite for the creation of the Federal Territory of Marajó."). This excerpt exemplifies the divergence general classes can deliver, thus providing an ambiguous understanding of their meanings and classifications. As per our specialist's insight, the main point of these bills lies not in the plebiscite itself but rather in the methods employed.

As emphasized by Caled et al. (2022), highlighting the significance of interpretability is essential, as it allows for a deeper understanding of how various factors influence the model's classification of legislative bills. Through insight into the rationale behind selecting specific subjects, interpretability empowers individuals to make informed decisions based on the model's output and domain-specific knowledge. This understanding not only aids in curating and validating the model's predictions but also ensures their alignment with legislative requirements and objectives.

To enhance interpretability, we employed the transformers-interpret library[1], following the approach of Schwarzenberg and Figueroa (2023), Ng and Carley (2022), and Khan et al. (2022), which use the Captum library (KOKHLIKYAN et al., 2020) for the library visualizations. This library employs integrated gradients to quantify the importance of each feature in the final classification.

[1]https://github.com/cdpierse/transformers-interpret

This interpretability method is effective because it extends beyond lexical analysis, such as unigram frequency assessment, and leverages the neural attention layer of BERT to identify the most pivotal tokens for a given prediction. The output visualization highlights green tokens relevant to the prediction, with darker tokens considered more crucial. Similarly, tokens are highlighted in red to indicate a decrease in the output. By adopting this approach, we provide an intuitive means of comprehending the influence of individual tokens on a model's decision.

Figure 7.4 shows the attention weights associated with three unlabeled bills. In Figure 7.4a, the classifier predicted the subject *Comunicações*, in which the word *TV* stood out significantly, as evidenced by its prominent attention weight. Terms such as *Televisão* and *rádio* are identified as highly relevant. This instance serves as a compelling demonstration of BERT's power in capturing complex relationships, including those that may not be immediately evident, such as the geographical context of a subscription television service, which has also been highlighted as influential.

Similarly, Figure 7.4b shows the weights of a bill predicted as *Saúde*. Notably, the term *seringas* is highlighted as crucial, showing a specific health-related vocabulary distinct from *segurança*, which is indicated as a decreasing term. This further underscores the relevance of a specific domain vocabulary, as previously discussed, to the subject *Defesa e Segurança*. Similarly, Figure 7.4c reveals significant terms in the economic domain, with *moeda*, *Conselho Monetário Nacional*, and *riqueza nacional* emerging as main keywords.

Figure 7.4 – The attention weights and predicted classes for three unlabeled summaries.
(a) Attention weights for the subject *Comunicações* to bill with id 13105.

Legend: Negative Neutral Positive

True Label	Prediction Score	Attribution Label	Attribution Score	Word Importance
None	(1.00)	Comunicações	5.66	[CLS] Sub ##met ##e à apre ##ciação do Congresso Nacional o ato que " outor ##ga concessão à Sociedade Rádio Alv ##ora ##da Ltda . , para explorar , pelo prazo de 15 (quinze) anos , sem direito de exclus ##ividade , serviço Especial de Televisão por Ass ##ina ##tura - TV ##A , na região Metropolitana de Belo Horizonte , Estado de Minas Gerais " [SEP]

(b) Attention weights for the subject *Saúde* to bill with id 25390.

Legend: Negative Neutral Positive

True Label	Prediction Score	Attribution Label	Attribution Score	Word Importance
None	(0.91)	Saúde	2.21	[CLS] Torn ##a obrigatória a inclusão de dispositivo de segurança que impe ##ça a re ##util ##ização nas ser ##inga ##s descar ##táveis . [SEP]

(c) Attention weights for the subject *Economia* to bill with id 24279.

Legend: Negative Neutral Positive

True Label	Prediction Score	Attribution Label	Attribution Score	Word Importance
None	(0.98)	Economia	5.96	[CLS] Sol ##ici ##ta homo ##log ##ação do Congresso Nacional para a emissão adicional de papel - moeda autor ##izada pelo Conselho Mon ##etário Nacional , no Vo ##to CM ##N [UNK] 06 ##4 / 94 , no valor de CR $ 2 , 5 tril ##hões (dois tril ##hões e quinh ##entos bilhões de cruzeiro ##s reais) , para atender às exigências das atividades de produção e da circulação de riqueza nacional , no mês de maio do corrente exercício [SEP]

8 CONCLUSIONS AND FUTURE WORK

In this study, we explored and described an approach to multilabel classification in the Brazilian Chamber of Deputies bills domain. Additionally, we produced a corpus with bills and subject classification. We explored two different BERT models in Brazilian Portuguese and two variations of our corpus, one with the whole dataset, and the other excluding the two classes with less than ten samples.

The decision to exclude these classes positively increased the macro F1 for the two models, demonstrating the importance of having sufficient data for each class in multilabel classification tasks. However, the weighted F1 differed slightly between the two corpora, which may not have practical significance.

In practical applications, the classifier can help classify new bills by suggesting subjects to be curated by a specialist. In the same way, it facilitates an automatic classification of 457,854 bills that do not have any classification.

In future work, techniques such as oversampling and undersampling can be explored to increase the results by balancing our data distribution. Also, different approaches, such as few- and zero-shot learning, can help in the identification of the classes with limited data, such as *Ciências Exatas e da Terra* and *Ciências Sociais e Humanas*. Classes with lower F1 scores require further investigation. Are there common characteristics of these classes that make the classification challenging? This analysis could lead to strategies for improving the model performance on these subjects.

REFERENCES

AGUIAR, A. et al. Text classification in legal documents extracted from lawsuits in brazilian courts. In: SPRINGER. **Intelligent Systems: 10th Brazilian Conference, BRACIS 2021, Virtual Event, November 29–December 3, 2021, Proceedings, Part II 10**. [S.l.], 2021. p. 586–600.

ALBUQUERQUE, H. O. et al. Ulyssesner-br: a corpus of brazilian legislative documents for named entity recognition. In: SPRINGER. **International Conference on Computational Processing of the Portuguese Language**. [S.l.], 2022. p. 3–14.

AMBALAVANAN, A. K.; DEVARAKONDA, M. V. Using the contextual language model bert for multi-criteria classification of scientific articles. **Journal of biomedical informatics**, Elsevier, v. 112, p. 103578, 2020.

ARAGY, R.; FERNANDES, E. R.; CACERES, E. N. Rhetorical role identification for portuguese legal documents. In: SPRINGER. **Intelligent Systems: 10th Brazilian Conference, BRACIS 2021, Virtual Event, November 29–December 3, 2021, Proceedings, Part II 10**. [S.l.], 2021. p. 557–571.

ASSOGBA, Y. et al. Many bills: engaging citizens through visualizations of congressional legislation. In: **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**. [S.l.: s.n.], 2011. p. 433–442.

AVRAM, A.-M.; PAIS, V.; TUFIS, D. Pyeurovoc: A tool for multilingual legal document classification with eurovoc descriptors. **arXiv preprint arXiv:2108.01139**, 2021.

BRISKILAL, J.; SUBALALITHA, C. An ensemble model for classifying idioms and literal texts using bert and roberta. **Information Processing & Management**, Elsevier, v. 59, n. 1, p. 102756, 2022.

CALED, D. et al. Multi-label classification of legislative contents with hierarchical label attention networks. **International Journal on Digital Libraries**, Springer, p. 1–14, 2022.

CHALKIDIS, I. et al. Large-scale multi-label text classification on eu legislation. **arXiv preprint arXiv:1906.02192**, 2019.

CHALKIDIS, I. et al. Legal-bert: The muppets straight out of law school. **arXiv preprint arXiv:2010.02559**, 2020.

CONSOLI, B. S.; VIEIRA, R. Enriching portuguese word embeddings with visual information. In: SPRINGER. **Brazilian Conference on Intelligent Systems**. [S.l.], 2021. p. 434–448.

CONSTITUIÇÃO, B. **Constituição da República Federativa do Brasil promulgada em 5 de outubro de 1988: atualizada até a Emenda Constitucional n. 48, de 10-8-2005**. 38. ed. São Paulo: Saraiva, 2006. (Coleção Saraiva de Legislação).

DEPUTADOS, C. dos. Regimento interno da câmara dos deputados. **Resolução nº 17**, 1989.

DEVLIN, J. et al. Bert: Pre-training of deep bidirectional transformers for language understanding. **arXiv preprint arXiv:1810.04805**, 2018.

FACELI, K. et al. Inteligência artificial: uma abordagem de aprendizado de máquina. 2021.

FILHO, J. A. W. et al. The brwac corpus: a new open resource for brazilian portuguese. In: **Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)**. [S.l.: s.n.], 2018.

INFORMAÇÃO, D. de Inovação e Tecnologia da. **Dados Abertos da Câmara dos Deputados**. 2022. <https://dadosabertos.camara.leg.br/swagger/api.html>. Accessed: 2022-12-03.

JURAFSKY, D.; MARTIN, J. H. **Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition**. 3rd draft. ed. [S.l.: s.n.], 2023. <https://web.stanford.edu/~jurafsky/slp3/>(Accessed: 2023-08-07).

KADHIM, A. I. Survey on supervised machine learning techniques for automatic text classification. **Artificial Intelligence Review**, Springer, v. 52, n. 1, p. 273–292, 2019.

KHAN, P. I. et al. Performance comparison of transformer-based models on twitter health mention classification. **IEEE Transactions on Computational Social Systems**, IEEE, 2022.

KOKHLIKYAN, N. et al. Captum: A unified and generic model interpretability library for pytorch. **arXiv preprint arXiv:2009.07896**, 2020.

LILLIS, D.; NULTY, P.; ZHANG, G. Enhancing legal argument mining with domain pre-training and neural networks. **Journal of Data Mining & Digital Humanities**, Episciences. org, 2022.

LIMSOPATHAM, N. Effectively leveraging bert for legal document classification. In: **Proceedings of the Natural Legal Language Processing Workshop 2021**. [S.l.: s.n.], 2021. p. 210–216.

LOSHCHILOV, I.; HUTTER, F. Decoupled weight decay regularization. **arXiv preprint arXiv:1711.05101**, 2017.

MANNING, C. D. **An introduction to information retrieval**. [S.l.]: Cambridge university press, 2009.

MÉNDEZ, G. G.; MORENO, O.; MENDOZA, P. Legislatio: A visualization tool for legislative roll-call vote data. In: **Proceedings of the 15th International Symposium on Visual Information Communication and Interaction**. [S.l.: s.n.], 2022. p. 1–8.

MONTESQUIEU, C. D. **Montesquieu: The spirit of the laws**. [S.l.]: Cambridge University Press, 1989.

NG, L. H. X.; CARLEY, K. M. Is my stance the same as your stance? a cross validation study of stance detection datasets. **Information Processing & Management**, Elsevier, v. 59, n. 6, p. 103070, 2022.

POLO, F. M. et al. Legalnlp–natural language processing methods for the brazilian legal language. **arXiv preprint arXiv:2110.15709**, 2021.

RUSSELL, S. J. **Artificial intelligence a modern approach**. [S.l.]: Pearson Education, Inc., 2010.

SCHWARZENBERG, P.; FIGUEROA, A. Textual pre-trained models for gender identification across community question-answering members. **IEEE Access**, IEEE, v. 11, p. 3983–3995, 2023.

SECHIDIS, K.; TSOUMAKAS, G.; VLAHAVAS, I. On the stratification of multi-label data. In: SPRINGER. **Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2011, Athens, Greece, September 5-9, 2011, Proceedings, Part III 22**. [S.l.], 2011. p. 145–158.

SERRAS, F. R.; FINGER, M. verbert: Automating brazilian case law document multi-label categorization using bert. **arXiv preprint arXiv:2203.06224**, 2022.

SILVA, N. F. d. et al. Evaluating topic models in portuguese political comments about bills from brazil's chamber of deputies. In: SPRINGER. **Intelligent Systems: 10th Brazilian Conference, BRACIS 2021, Virtual Event, November 29–December 3, 2021, Proceedings, Part II 10**. [S.l.], 2021. p. 104–120.

SILVA, R. N. M. da; SPRITZER, A.; FREITAS, C. D. S. Visualization of roll call data for supporting analyses of political profiles. In: IEEE. **2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)**. [S.l.], 2018. p. 150–157.

SOUZA, F.; NOGUEIRA, R.; LOTUFO, R. Bertimbau: pretrained bert models for brazilian portuguese. In: SPRINGER. **Intelligent Systems: 9th Brazilian Conference, BRACIS 2020, Rio Grande, Brazil, October 20–23, 2020, Proceedings, Part I 9**. [S.l.], 2020. p. 403–417.

SUN, C. et al. How to fine-tune bert for text classification? In: SPRINGER. **Chinese Computational Linguistics: 18th China National Conference, CCL 2019, Kunming, China, October 18–20, 2019, Proceedings 18**. [S.l.], 2019. p. 194–206.

TANG, T.; TANG, X.; YUAN, T. Fine-tuning bert for multi-label sentiment analysis in unbalanced code-switching text. **IEEE Access**, IEEE, v. 8, p. 193248–193256, 2020.

VASWANI, A. et al. Attention is all you need. **Advances in neural information processing systems**, v. 30, 2017.

VIANNA, D.; MOURA, E. Silva de. Organizing portuguese legal documents through topic discovery. In: **Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval**. New York, NY, USA: Association for Computing Machinery, 2022. (SIGIR '22), p. 3388–3392. ISBN 9781450387323. Available from Internet: <https://doi.org/10.1145/3477495.3536329>.

WOLF, T. et al. Huggingface's transformers: State-of-the-art natural language processing. **arXiv preprint arXiv:1910.03771**, 2019.

ZANUZ, L.; RIGO, S. J. Fostering judiciary applications with new fine-tuned models for legal named entity recognition in portuguese. In: SPRINGER. **International Conference on Computational Processing of the Portuguese Language**. [S.l.], 2022. p. 219–229.

APPENDIX A — EXTENDED RESULTS

In this appendix, we expand the findings presented in Chapter 7 by providing additional results that encompass both BERTimbau and BERTikal. These extended results include outcomes obtained using the standard corpus and its variations, wherein the classes *Ciências Exatas e da Terra* and *Ciências Sociais e Humanas* were excluded for fine-tuning purposes, and also include the variation of the maximum sentence length in the model tokenizer, varying between 75 and 512. Our objective is to offer a more comprehensive overview of the results, encompassing a diverse set of metrics across all tests conducted. Furthermore, we introduced metrics, such as the area under the ROC curve and the area under the precision-recall curve. Although these metrics were not employed in the main study, they could be significant in a more nuanced analysis. Through these extended results, we aim to enhance transparency and provide more comprehensive insight into the models' performance across various scenarios.

In Tables A.1, A.2, A.3, A.4, A.5, A.6, A.8 and A.7 we present the global results of our experiments. Additionally, Tables A.9, A.10, A.11, A.12, A.13, A.14, A.15 and A.16 showcase the results for each individual class. Because the discussion has already been covered in Chapter 7, we refrained from providing a detailed analysis in this appendix.

Table A.1 – BERTimbau fine-tuned with the standard corpus metrics to each fold, using 512 tokens as the maximum sentence length. In the last two rows, we have the average and standard deviation for each metric in the 10-fold.

Folds	$F1_{macro}$	$F1_{micro}$	$F1_{weighted}$	*PR AUC*	$Recall_{macro}$	$Precision_{macro}$	*ROC AUC*	*Accuracy*
Fold 0	68.52	79.23	78.86	86.90	65.44	72.34	87.47	65.06
Fold 1	68.81	79.42	79.09	86.92	66.39	72.07	87.77	65.10
Fold 2	68.65	79.63	79.25	87.35	65.94	72.06	87.83	65.14
Fold 3	67.78	79.43	78.98	86.90	65.17	71.80	87.71	65.04
Fold 4	67.95	79.70	79.32	87.12	65.10	71.70	87.80	65.26
Fold 5	68.64	79.48	79.07	87.24	65.85	72.72	87.73	64.84
Fold 6	67.90	79.22	78.84	86.72	65.18	71.61	87.55	64.76
Fold 7	68.06	79.37	78.99	86.98	65.50	71.65	87.81	64.97
Fold 8	68.31	79.31	78.98	87.08	65.46	75.10	87.65	64.56
Fold 9	67.88	78.98	78.55	86.52	65.39	71.06	87.47	64.53
Average	68.25	79.38	78.99	86.97	65.54	72.21	87.68	64.92
Std. Dev.	0.38	0.21	0.22	0.24	0.41	1.11	0.14	0.25

Table A.2 – BERTimbau fine-tuned with the standard corpus without the classes *Ciências Exatas e da Terra* and *Ciências Sociais e Humanas* metrics to each fold, using 512 tokens as the maximum sentence length. In the last two rows, we have the average and standard deviation for each metric in the 10-fold.

Folds	$F1_{macro}$	$F1_{micro}$	$F1_{weighted}$	*PR AUC*	$Recall_{macro}$	$Precision_{macro}$	*ROC AUC*	*Accuracy*
Fold 0	72.57	79.38	78.99	87.09	69.99	75.95	87.74	65.14
Fold 1	72.89	79.78	79.32	87.08	69.74	77.27	87.75	65.22
Fold 2	73.34	79.74	79.31	87.24	70.28	77.39	87.75	65.38
Fold 3	72.42	79.00	78.58	86.68	69.46	76.56	87.35	64.45
Fold 4	72.53	79.12	78.76	86.82	69.78	76.10	87.65	64.79
Fold 5	72.96	79.32	78.94	87.04	70.26	76.71	87.65	64.89
Fold 6	72.70	79.23	78.84	86.92	69.44	77.01	87.47	64.96
Fold 7	73.05	79.26	78.87	86.90	70.31	76.60	87.62	64.76
Fold 8	72.68	79.35	78.99	87.09	69.92	76.18	87.67	64.83
Fold 9	72.66	79.22	78.81	87.02	70.04	76.39	87.54	65.23
Average	72.78	79.34	78.94	86.99	69.92	76.62	87.62	64.96
Std. Dev.	0.28	0.25	0.23	0.16	0.32	0.49	0.13	0.28

Table A.3 – BERTikal fine-tuned with the standard corpus metrics to each fold, using 512 tokens as the maximum sentence length. In the last two rows, we have the average and standard deviation for each metric in the 10-fold.

Folds	$F1_{macro}$	$F1_{micro}$	$F1_{weighted}$	*PR AUC*	$Recall_{macro}$	$Precision_{macro}$	*ROC AUC*	*Accuracy*
Fold 0	66.86	78.17	77.71	85.86	63.94	70.76	86.93	63.47
Fold 1	67.25	78.27	77.86	85.84	64.59	71.26	86.97	63.64
Fold 2	67.34	78.70	78.23	86.26	64.73	71.14	87.22	64.39
Fold 3	66.78	78.35	77.83	85.84	63.81	71.59	86.93	63.71
Fold 4	67.14	78.55	78.09	86.17	64.21	74.10	87.06	64.04
Fold 5	66.84	78.16	77.67	85.94	63.83	71.80	86.91	63.41
Fold 6	66.54	78.22	77.74	85.93	63.53	70.68	86.77	63.70
Fold 7	66.31	77.98	77.46	85.64	63.20	71.55	86.73	63.49
Fold 8	66.87	78.61	78.14	86.08	63.74	71.65	87.00	64.13
Fold 9	66.71	77.99	77.47	85.62	63.84	71.15	86.77	63.36
Average	66.86	78.30	77.82	85.92	63.94	71.57	86.93	63.74
Std. Dev.	0.31	0.25	0.27	0.21	0.46	0.96	0.15	0.34

Table A.4 – BERTikal fine-tuned with the standard corpus without the classes *Ciências Exatas e da Terra* and *Ciências Sociais e Humanas* metrics to each fold, using 512 tokens as the maximum sentence length. In the last two rows, we have the average and standard deviation for each metric in the 10-fold.

Folds	$F1_{macro}$	$F1_{micro}$	$F1_{weighted}$	*PR AUC*	$Recall_{macro}$	$Precision_{macro}$	*ROC AUC*	*Accuracy*
Fold 0	71.87	78.58	78.13	86.12	68.81	76.26	87.17	64.01
Fold 1	71.18	78.46	77.92	85.83	67.73	76.49	86.93	63.80
Fold 2	71.52	78.60	78.07	86.07	68.45	76.07	86.99	64.06
Fold 3	70.73	77.90	77.35	85.38	67.64	75.84	86.61	63.01
Fold 4	71.16	78.13	77.68	85.72	68.08	75.67	86.77	63.63
Fold 5	71.59	78.47	78.01	86.10	68.55	75.84	87.07	63.69
Fold 6	71.74	78.50	78.03	86.08	68.51	76.13	87.00	64.52
Fold 7	72.24	78.45	78.00	85.91	69.51	76.48	87.06	63.52
Fold 8	71.27	78.29	77.82	85.79	67.94	75.98	86.82	63.93
Fold 9	71.53	78.49	78.02	86.08	68.62	75.75	87.10	64.28
Average	71.48	78.39	77.90	85.91	68.38	76.05	86.95	63.85
Std. Dev.	0.42	0.22	0.23	0.24	0.56	0.29	0.17	0.42

Table A.5 – BERTimbau fine=tuned with the standard corpus metrics to each fold, using 75 tokens as the maximum sentence length. In the last two rows, we have the average and standard deviation for each metric in the 10-fold.

Folds	$F1_{macro}$	$F1_{micro}$	$F1_{weighted}$	*PR AUC*	*Recall* $_{macro}$	*Precision* $_{macro}$	*ROC AUC*	*Accuracy*
Fold 0	68.17	78.99	78.60	86.63	64.95	72.22	87.36	64.80
Fold 1	68.56	79.18	78.83	86.69	65.85	72.19	87.55	64.65
Fold 2	68.28	79.41	79.03	87.11	65.60	71.73	87.65	64.97
Fold 3	67.91	79.31	78.91	86.76	65.28	71.72	87.70	64.93
Fold 4	68.14	79.41	79.06	86.88	65.59	71.50	87.80	64.94
Fold 5	68.71	79.36	78.94	87.02	65.78	72.99	87.61	64.80
Fold 6	67.87	78.99	78.61	86.59	64.83	71.76	87.25	64.81
Fold 7	67.42	79.02	78.62	86.75	64.58	71.10	87.44	64.83
Fold 8	68.07	79.12	78.75	86.93	65.38	71.64	87.52	64.45
Fold 9	67.50	78.64	78.22	86.28	64.70	70.99	87.21	64.25
Average	68.06	79.14	78.76	86.77	65.25	71.78	87.51	64.74
Std. Dev.	0.41	0.24	0.25	0.24	0.46	0.58	0.19	0.23

Table A.6 – BERTimbau fine-tuned with the standard corpus without the classes *Ciências Exatas e da Terra* and *Ciências Sociais e Humanas* metrics to each fold, using 75 tokens as the maximum sentence length. In the last two rows, we have the average and standard deviation for each metric in the 10-fold.

Folds	$F1_{macro}$	$F1_{micro}$	$F1_{weighted}$	*PR AUC*	*Recall* $_{macro}$	*Precision* $_{macro}$	*ROC AUC*	*Accuracy*
Fold 0	72.45	79.06	78.69	86.89	69.44	76.31	87.46	64.67
Fold 1	72.45	79.29	78.83	86.76	68.88	77.30	87.45	64.81
Fold 2	72.83	79.42	79.00	87.12	69.66	77.04	87.60	65.00
Fold 3	72.38	78.76	78.35	86.54	69.44	77.58	87.27	64.59
Fold 4	72.49	78.97	78.57	86.56	69.45	76.74	87.42	64.62
Fold 5	72.83	79.33	78.97	86.95	69.87	77.04	87.57	64.91
Fold 6	72.61	79.12	78.71	87.00	69.29	77.35	87.32	64.73
Fold 7	72.74	79.16	78.74	86.71	70.08	76.38	87.51	64.42
Fold 8	72.09	79.04	78.63	86.81	68.89	76.14	87.29	65.02
Fold 9	72.12	79.11	78.69	86.82	69.94	75.35	87.62	64.81
Average	72.50	79.13	78.72	86.82	69.49	76.72	87.45	64.76
Std. Dev.	0.26	0.19	0.19	0.18	0.41	0.68	0.13	0.19

Table A.7 – BERTikal fine-tuned with the standard corpus without the classes *Ciências Exatas e da Terra* and *Ciências Sociais e Humanas* metrics to each fold, using 75 tokens as the maximum sentence length. In the last two rows, we have the average and standard deviation for each metric in the 10-fold.

Folds	$F1_{macro}$	$F1_{micro}$	$F1_{weighted}$	*PR AUC*	*Recall* $_{macro}$	*Precision* $_{macro}$	*ROC AUC*	*Accuracy*
Fold 0	66.86	78.00	77.47	85.43	63.35	71.59	86.64	63.56
Fold 1	66.76	78.00	77.56	85.57	63.30	71.58	86.64	63.74
Fold 2	67.45	78.67	78.20	86.13	64.28	72.10	86.99	64.26
Fold 3	66.88	78.34	77.85	85.61	63.78	71.72	86.84	64.05
Fold 4	66.34	78.20	77.71	85.77	63.44	70.45	86.83	63.63
Fold 5	66.46	78.18	77.63	85.71	63.21	71.07	86.70	63.35
Fold 6	66.49	78.24	77.76	85.80	63.34	70.88	86.70	63.77
Fold 7	66.88	78.18	77.66	85.55	63.67	71.46	86.77	63.83
Fold 8	66.10	77.93	77.42	85.47	62.39	72.20	86.45	63.38
Fold 9	66.40	77.79	77.24	85.26	63.34	70.69	86.52	63.21
Average	66.66	78.15	77.65	85.63	63.41	71.38	86.71	63.68
Std. Dev.	0.38	0.25	0.26	0.24	0.48	0.59	0.16	0.32

Table A.8 – BERTikal fine-tuned with the standard corpus metrics to each fold, using 75 tokens as the maximum sentence length. In the last two rows, we have the average and standard deviation for each metric in the 10-fold.

Folds	*F1 macro*	*F1 micro*	*F1 weighted*	*PR AUC*	*Recall macro*	*Precision macro*	*ROC AUC*	*Accuracy*
Fold 0	71.49	78.22	77.73	85.81	68.07	76.45	86.77	63.55
Fold 1	71.55	78.40	77.92	85.61	67.97	76.80	86.90	64.17
Fold 2	71.70	78.40	77.95	85.99	68.24	76.65	86.84	63.81
Fold 3	70.92	77.81	77.31	85.24	67.91	74.99	86.58	63.28
Fold 4	70.95	77.97	77.55	85.45	67.80	75.44	86.73	63.46
Fold 5	71.28	78.19	77.74	85.77	68.04	75.66	86.80	63.62
Fold 6	71.45	78.38	77.92	85.77	67.91	76.87	86.88	64.33
Fold 7	71.63	78.31	77.86	85.61	68.50	76.16	86.81	63.48
Fold 8	71.07	78.09	77.62	85.47	67.56	75.97	86.69	63.78
Fold 9	71.27	78.10	77.67	85.58	68.63	75.20	86.91	63.68
Average	71.33	78.19	77.73	85.63	68.06	76.02	86.79	63.72
Std. Dev.	0.28	0.20	0.20	0.21	0.32	0.68	0.10	0.32
0.32								

Table A.9 – BERTimbau fine-tuned with the standard corpus local metrics to each class, using 512 tokens as the maximum sentence length. Each value is in the format "xx ± yy", which is the average and standard deviation of the stratified 10-fold cross-validation

Order	*Target*	*F1*	*PR AUC*	*Precision*	*Recall*
1	Comunicações	95.92 ± 0.26	98.96 ± 0.15	96.28 ± 0.39	95.56 ± 0.44
4	Educação	92.60 ± 0.45	96.46 ± 0.54	93.25 ± 0.59	91.97 ± 0.48
3	Saúde	86.80 ± 0.49	93.53 ± 0.53	88.49 ± 0.65	85.17 ± 0.60
20	Esporte e Lazer	86.42 ± 1.24	92.22 ± 1.36	87.57 ± 1.79	85.36 ± 2.17
15	Energia, Recursos Hídricos e Minerais	84.87 ± 1.44	90.62 ± 1.04	85.50 ± 1.47	84.29 ± 2.27
21	Homenagens e Datas Comemorativas	84.54 ± 1.07	87.71 ± 1.68	81.04 ± 1.87	88.40 ± 1.31
12	Direito Penal e Processual Penal	83.94 ± 1.49	90.41 ± 1.20	87.37 ± 1.28	80.79 ± 2.06
19	Relações Internacionais e Comércio Exterior	83.80 ± 1.12	88.52 ± 1.21	84.86 ± 1.55	82.81 ± 2.07
24	Política, Partidos e Eleições	83.49 ± 1.08	86.89 ± 1.90	82.06 ± 1.09	85.01 ± 2.01
14	Meio Ambiente e Desenvolvimento Sustentável	81.01 ± 1.23	86.75 ± 1.41	83.81 ± 1.50	78.46 ± 2.47
7	Trabalho e Emprego	80.48 ± 0.86	87.53 ± 0.80	84.78 ± 1.07	76.61 ± 1.38
29	Turismo	79.21 ± 3.74	81.42 ± 4.10	80.69 ± 4.68	77.89 ± 3.98
9	Viação, Transporte e Mobilidade	79.10 ± 1.16	85.79 ± 1.19	79.48 ± 1.17	78.75 ± 1.85
5	Finanças Públicas e Orçamento	77.39 ± 0.99	85.04 ± 0.93	83.38 ± 1.51	72.23 ± 1.24
16	Previdência e Assistência Social	76.43 ± 0.86	81.74 ± 1.12	81.44 ± 1.32	72.06 ± 2.04
26	Processo Legislativo e Atuação Parlamentar	75.69 ± 2.46	81.41 ± 2.85	78.83 ± 3.25	72.90 ± 3.27
13	Cidades e Desenvolvimento Urbano	75.54 ± 0.85	83.10 ± 0.79	81.12 ± 1.62	70.72 ± 1.43
18	Agricultura, Pecuária, Pesca e Extrativismo	75.29 ± 1.42	82.92 ± 1.39	81.05 ± 1.84	70.32 ± 1.82
17	Direito Civil e Processual Civil	74.78 ± 2.25	80.34 ± 2.24	80.37 ± 1.18	70.04 ± 3.90
6	Direitos Humanos e Minorias	73.23 ± 1.21	79.88 ± 1.26	78.76 ± 1.38	68.44 ± 1.42
10	Defesa e Segurança	71.84 ± 1.09	78.29 ± 0.91	75.72 ± 1.50	68.36 ± 1.03
22	Estrutura Fundiária	71.11 ± 1.78	76.64 ± 2.09	73.54 ± 1.22	68.88 ± 2.80
25	Arte, Cultura e Religião	70.58 ± 2.82	76.09 ± 2.89	75.97 ± 3.27	66.04 ± 3.92
2	Administração Pública	69.99 ± 0.79	77.73 ± 0.83	75.53 ± 0.91	65.22 ± 1.08
8	Economia	68.35 ± 1.62	75.10 ± 1.88	75.17 ± 1.50	62.71 ± 2.29
27	Ciência, Tecnologia e Inovação	62.68 ± 2.57	68.34 ± 2.18	68.26 ± 2.46	57.99 ± 3.12
11	Indústria, Comércio e Serviços	57.57 ± 1.38	61.30 ± 1.93	63.58 ± 2.05	52.61 ± 1.07
23	Direito e Defesa do Consumidor	56.88 ± 3.30	59.62 ± 2.66	63.85 ± 2.40	51.47 ± 4.75
28	Direito e Justiça	23.74 ± 5.88	28.83 ± 4.09	49.03 ± 10.25	15.96 ± 4.56
30	Direito Constitucional	0.71 ± 2.14	17.36 ± 7.16	10.00 ± 30.00	0.37 ± 1.11
32	Ciências Exatas e da Terra	0.00 ± 0.00	0.13 ± 0.24	0.00 ± 0.00	0.00 ± 0.00
31	Ciências Sociais e Humanas	0.00 ± 0.00	0.08 ± 0.13	0.00 ± 0.00	0.00 ± 0.00

Table A.10 – BERTimbau fine-tuned with the standard corpus without the classes *Ciências Exatas e da Terra* and *Ciências Sociais e Humanas* local metrics to each class, using 512 tokens as the maximum sentence length. Each value is in the format "xx ± yy", which is the average and standard deviation of the stratified 10-fold cross-validation.

Order	*Target*	*F1*	*PR AUC*	*Precision*	*Recall*
1	Comunicações	95.96 ± 0.33	98.89 ± 0.14	96.36 ± 0.49	95.57 ± 0.38
4	Educação	92.63 ± 0.50	96.55 ± 0.53	93.25 ± 0.78	92.03 ± 0.66
3	Saúde	86.83 ± 0.52	93.66 ± 0.33	88.20 ± 0.49	85.51 ± 1.11
20	Esporte e Lazer	86.80 ± 1.46	92.11 ± 0.97	88.30 ± 2.14	85.42 ± 2.35
15	Energia, Recursos Hídricos e Minerais	85.26 ± 0.70	90.88 ± 1.20	85.44 ± 1.00	85.11 ± 1.75
12	Direito Penal e Processual Penal	84.09 ± 0.58	90.47 ± 0.50	88.05 ± 1.79	80.53 ± 1.58
21	Homenagens e Datas Comemorativas	84.06 ± 1.50	86.90 ± 1.46	80.54 ± 1.83	87.94 ± 2.27
19	Relações Internacionais e Comércio Exterior	83.61 ± 1.46	88.00 ± 1.66	84.65 ± 1.44	82.63 ± 2.12
24	Política, Partidos e Eleições	83.59 ± 1.62	87.68 ± 1.71	82.13 ± 1.49	85.17 ± 2.86
14	Meio Ambiente e Desenvolvimento Sustentável	80.88 ± 1.26	86.64 ± 1.21	83.24 ± 1.64	78.72 ± 2.57
7	Trabalho e Emprego	80.20 ± 1.03	87.57 ± 0.70	84.36 ± 1.33	76.44 ± 1.27
29	Turismo	79.68 ± 3.47	80.84 ± 2.35	79.95 ± 4.30	79.46 ± 3.31
9	Viação, Transporte e Mobilidade	78.82 ± 0.73	85.63 ± 1.02	79.38 ± 1.25	78.28 ± 0.96
5	Finanças Públicas e Orçamento	77.29 ± 0.98	85.08 ± 0.72	82.88 ± 1.27	72.41 ± 1.22
26	Processo Legislativo e Atuação Parlamentar	76.45 ± 1.09	80.63 ± 1.97	79.40 ± 2.94	73.80 ± 1.38
16	Previdência e Assistência Social	75.97 ± 1.74	81.31 ± 1.72	81.21 ± 1.86	71.42 ± 2.49
18	Agricultura, Pecuária, Pesca e Extrativismo	75.47 ± 1.76	83.07 ± 1.74	80.92 ± 1.56	70.82 ± 3.24
13	Cidades e Desenvolvimento Urbano	75.12 ± 0.95	82.99 ± 1.10	81.48 ± 1.51	69.76 ± 2.14
17	Direito Civil e Processual Civil	74.51 ± 1.35	80.35 ± 1.14	79.98 ± 2.16	69.78 ± 1.71
6	Direitos Humanos e Minorias	73.33 ± 0.74	80.02 ± 0.97	78.64 ± 1.01	68.72 ± 1.14
10	Defesa e Segurança	71.40 ± 1.49	78.45 ± 1.31	75.17 ± 2.18	68.01 ± 1.39
22	Estrutura Fundiária	70.50 ± 1.86	76.02 ± 1.17	73.78 ± 1.69	67.60 ± 3.30
25	Arte, Cultura e Religião	70.07 ± 2.38	76.20 ± 2.85	75.34 ± 3.47	65.56 ± 2.43
2	Administração Pública	69.84 ± 0.86	77.84 ± 0.61	76.08 ± 0.86	64.56 ± 1.23
8	Economia	68.43 ± 1.39	75.19 ± 1.41	75.27 ± 1.19	62.77 ± 2.08
27	Ciência, Tecnologia e Inovação	64.68 ± 3.06	69.12 ± 4.53	69.45 ± 5.74	60.75 ± 2.80
11	Indústria, Comércio e Serviços	57.46 ± 1.28	61.65 ± 1.19	64.59 ± 1.99	51.81 ± 1.95
23	Direito e Defesa do Consumidor	56.73 ± 2.23	59.68 ± 2.02	64.14 ± 2.05	51.03 ± 3.68
28	Direito e Justiça	23.75 ± 4.09	28.81 ± 2.08	46.30 ± 5.18	16.08 ± 3.34
30	Direito Constitucional	0.00 ± 0.00	17.07 ± 5.58	0.00 ± 0.00	0.00 ± 0.00

Table A.11 – BERTikal fine-tuned with the standard corpus local metrics to each class, using 512 tokens as the maximum sentence length. Each value is in the format "xx ± yy", which is the average and standard deviation of the stratified 10-fold cross-validation.

Order	*Target*	*F1*	*PR AUC*	*Precision*	*Recall*
1	Comunicações	95.80 ± 0.23	98.82 ± 0.15	96.51 ± 0.37	95.11 ± 0.45
4	Educação	92.17 ± 0.50	96.15 ± 0.41	92.80 ± 0.79	91.55 ± 0.36
3	Saúde	86.07 ± 0.80	92.98 ± 0.45	87.97 ± 0.86	84.26 ± 0.79
20	Esporte e Lazer	85.21 ± 1.09	90.56 ± 1.26	85.71 ± 2.02	84.77 ± 1.85
15	Energia, Recursos Hídricos e Minerais	84.29 ± 1.16	89.75 ± 1.12	84.68 ± 1.67	83.95 ± 1.81
12	Direito Penal e Processual Penal	83.20 ± 1.19	89.58 ± 1.21	86.95 ± 2.04	79.79 ± 1.11
21	Homenagens e Datas Comemorativas	83.04 ± 1.39	85.66 ± 1.98	79.68 ± 1.62	86.75 ± 2.29
24	Política, Partidos e Eleições	82.86 ± 1.30	85.71 ± 1.58	81.72 ± 1.68	84.14 ± 3.23
19	Relações Internacionais e Comércio Exterior	81.73 ± 1.37	86.54 ± 1.53	84.43 ± 1.66	79.23 ± 1.98
14	Meio Ambiente e Desenvolvimento Sustentável	79.77 ± 1.24	85.53 ± 1.26	83.43 ± 1.60	76.45 ± 1.96
7	Trabalho e Emprego	79.74 ± 1.03	86.74 ± 0.95	83.32 ± 1.35	76.47 ± 1.18
9	Viação, Transporte e Mobilidade	78.52 ± 0.93	85.19 ± 1.09	77.16 ± 1.01	79.96 ± 1.68
29	Turismo	77.22 ± 2.65	79.29 ± 4.70	77.19 ± 2.12	77.42 ± 4.54
5	Finanças Públicas e Orçamento	76.69 ± 0.69	84.06 ± 0.66	83.34 ± 1.28	71.05 ± 1.04
16	Previdência e Assistência Social	75.82 ± 1.31	80.97 ± 0.97	79.78 ± 2.11	72.31 ± 2.08
17	Direito Civil e Processual Civil	74.41 ± 1.87	79.36 ± 1.81	79.06 ± 1.43	70.35 ± 2.97
18	Agricultura, Pecuária, Pesca e Extrativismo	73.91 ± 1.35	81.49 ± 0.77	77.53 ± 0.96	70.69 ± 2.84
13	Cidades e Desenvolvimento Urbano	73.60 ± 1.07	82.14 ± 1.03	82.13 ± 2.08	66.72 ± 1.44
26	Processo Legislativo e Atuação Parlamentar	72.84 ± 2.66	76.51 ± 3.01	78.03 ± 3.92	68.37 ± 2.56
6	Direitos Humanos e Minorias	71.62 ± 1.11	78.28 ± 1.41	76.37 ± 1.74	67.45 ± 1.09
10	Defesa e Segurança	69.86 ± 0.77	77.11 ± 1.05	75.36 ± 1.98	65.16 ± 1.21
22	Estrutura Fundiária	69.16 ± 1.82	75.15 ± 2.10	73.42 ± 1.63	65.43 ± 2.75
25	Arte, Cultura e Religião	67.82 ± 3.30	73.38 ± 2.65	74.60 ± 3.03	62.33 ± 4.72
2	Administração Pública	67.81 ± 0.82	76.14 ± 0.86	75.98 ± 0.90	61.25 ± 1.26
8	Economia	65.97 ± 1.40	73.33 ± 1.47	74.71 ± 1.31	59.08 ± 1.69
27	Ciência, Tecnologia e Inovação	60.76 ± 2.22	66.01 ± 2.44	70.26 ± 2.73	53.57 ± 2.39
11	Indústria, Comércio e Serviços	56.02 ± 1.38	60.24 ± 1.87	63.50 ± 1.51	50.16 ± 1.83
23	Direito e Defesa do Consumidor	55.98 ± 2.16	57.69 ± 3.36	61.07 ± 2.61	51.75 ± 2.69
28	Direito e Justiça	16.96 ± 3.24	27.09 ± 3.90	53.46 ± 11.72	10.21 ± 2.26
30	Direito Constitucional	0.71 ± 2.14	11.20 ± 6.27	10.00 ± 30.00	0.37 ± 1.11
32	Ciências Exatas e da Terra	0.00 ± 0.00	0.80 ± 2.30	0.00 ± 0.00	0.00 ± 0.00
31	Ciências Sociais e Humanas	0.00 ± 0.00	0.11 ± 0.18	0.00 ± 0.00	0.00 ± 0.00

Table A.12 – BERTikal fine-tuned with the standard corpus without the classes *Ciências Exatas e da Terra* and *Ciências Sociais e Humanas* local metrics to each class, using 512 tokens as the maximum sentence length. Each value is in the format "xx ± yy", which is the average and standard deviation of the stratified 10-fold cross-validation.

Order	*Target*	*F1*	*PR AUC*	*Precision*	*Recall*
1	Comunicações	95.75 ± 0.37	98.76 ± 0.19	96.33 ± 0.60	95.18 ± 0.33
4	Educação	92.30 ± 0.52	96.14 ± 0.52	93.04 ± 0.73	91.58 ± 0.56
3	Saúde	86.23 ± 0.54	92.97 ± 0.47	87.54 ± 0.61	84.97 ± 1.04
20	Esporte e Lazer	85.14 ± 1.21	89.87 ± 2.10	85.30 ± 1.79	85.02 ± 1.92
15	Energia, Recursos Hídricos e Minerais	84.35 ± 0.57	89.88 ± 1.12	84.44 ± 1.18	84.29 ± 1.48
12	Direito Penal e Processual Penal	83.40 ± 0.65	89.64 ± 0.58	88.01 ± 1.34	79.27 ± 0.95
21	Homenagens e Datas Comemorativas	82.98 ± 1.08	85.96 ± 1.43	79.26 ± 1.15	87.10 ± 1.93
24	Política, Partidos e Eleições	82.42 ± 2.00	85.47 ± 2.50	80.71 ± 2.61	84.34 ± 3.63
19	Relações Internacionais e Comércio Exterior	82.27 ± 1.41	86.59 ± 1.04	84.58 ± 1.70	80.12 ± 2.05
14	Meio Ambiente e Desenvolvimento Sustentável	79.69 ± 1.44	85.18 ± 1.59	82.90 ± 1.38	76.73 ± 2.02
7	Trabalho e Emprego	79.39 ± 0.95	86.55 ± 0.62	83.16 ± 1.27	75.97 ± 1.59
9	Viação, Transporte e Mobilidade	78.40 ± 0.59	84.93 ± 0.90	77.36 ± 1.15	79.49 ± 0.88
29	Turismo	77.90 ± 2.94	79.67 ± 3.31	78.79 ± 3.24	77.17 ± 4.13
5	Finanças Públicas e Orçamento	76.89 ± 0.76	84.11 ± 0.78	83.37 ± 0.85	71.36 ± 0.90
16	Previdência e Assistência Social	76.00 ± 1.56	80.54 ± 2.09	80.08 ± 1.68	72.34 ± 1.99
17	Direito Civil e Processual Civil	74.32 ± 1.21	79.23 ± 1.33	79.43 ± 1.81	69.89 ± 1.94
18	Agricultura, Pecuária, Pesca e Extrativismo	74.08 ± 1.97	81.26 ± 2.54	77.08 ± 2.46	71.42 ± 3.10
13	Cidades e Desenvolvimento Urbano	73.91 ± 1.34	82.39 ± 1.28	82.04 ± 1.71	67.29 ± 1.88
26	Processo Legislativo e Atuação Parlamentar	72.39 ± 2.18	76.89 ± 2.44	77.36 ± 3.15	68.08 ± 2.31
6	Direitos Humanos e Minorias	71.63 ± 1.07	78.21 ± 1.13	77.00 ± 1.63	66.99 ± 1.42
10	Defesa e Segurança	69.93 ± 1.21	76.91 ± 1.42	75.60 ± 2.12	65.07 ± 0.94
22	Estrutura Fundiária	69.23 ± 1.69	74.88 ± 1.78	74.91 ± 2.95	64.49 ± 2.82
25	Arte, Cultura e Religião	68.45 ± 2.69	74.10 ± 3.14	74.64 ± 3.37	63.32 ± 3.28
2	Administração Pública	68.00 ± 0.85	76.22 ± 0.66	76.13 ± 0.94	61.45 ± 1.27
8	Economia	66.15 ± 1.14	73.42 ± 1.52	74.85 ± 1.40	59.31 ± 1.93
27	Ciência, Tecnologia e Inovação	62.40 ± 3.00	66.17 ± 4.02	69.74 ± 3.81	56.55 ± 3.31
11	Indústria, Comércio e Serviços	55.87 ± 1.16	59.99 ± 1.18	64.31 ± 1.61	49.43 ± 1.50
23	Direito e Defesa do Consumidor	55.85 ± 1.86	56.97 ± 2.10	61.38 ± 1.49	51.30 ± 2.74
28	Direito e Justiça	19.19 ± 7.38	26.76 ± 3.20	52.25 ± 8.83	12.02 ± 5.10
30	Direito Constitucional	0.00 ± 0.00	14.88 ± 6.93	0.00 ± 0.00	0.00 ± 0.00

Table A.13 – BERTimbau fine-tuned with the standard corpus local metrics to each class, using 75 tokens as the maximum sentence length. Each value is in the format "xx ± yy", which is the average and standard deviation of the stratified 10-fold cross-validation.

Order	*Target*	*F1*	*PR AUC*	*Precision*	*Recall*
1	Comunicações	95.97 ± 0.25	98.84 ± 0.20	96.39 ± 0.43	95.57 ± 0.35
4	Educação	92.39 ± 0.57	96.44 ± 0.35	93.00 ± 0.78	91.79 ± 0.63
3	Saúde	86.70 ± 0.57	93.45 ± 0.48	88.48 ± 0.58	84.99 ± 0.77
20	Esporte e Lazer	86.29 ± 1.15	91.94 ± 0.89	87.73 ± 1.19	84.95 ± 2.64
15	Energia, Recursos Hídricos e Minerais	84.87 ± 1.25	90.31 ± 0.87	86.00 ± 1.46	83.78 ± 1.66
19	Relações Internacionais e Comércio Exterior	83.97 ± 1.34	88.25 ± 1.37	85.53 ± 1.21	82.49 ± 2.01
12	Direito Penal e Processual Penal	83.92 ± 1.34	90.41 ± 1.19	87.30 ± 1.01	80.79 ± 1.77
21	Homenagens e Datas Comemorativas	83.84 ± 1.10	87.61 ± 1.76	80.67 ± 1.84	87.33 ± 2.09
24	Política, Partidos e Eleições	83.04 ± 1.18	87.03 ± 1.50	81.83 ± 1.33	84.34 ± 2.33
14	Meio Ambiente e Desenvolvimento Sustentável	80.60 ± 1.59	86.53 ± 1.50	83.30 ± 1.27	78.10 ± 2.43
7	Trabalho e Emprego	80.23 ± 0.79	87.18 ± 0.80	84.84 ± 0.95	76.11 ± 1.48
29	Turismo	79.44 ± 4.08	81.23 ± 5.44	80.66 ± 5.51	78.38 ± 3.69
9	Viação, Transporte e Mobilidade	78.62 ± 1.04	85.18 ± 1.13	79.21 ± 0.86	78.08 ± 2.05
5	Finanças Públicas e Orçamento	77.20 ± 1.08	84.95 ± 0.99	83.36 ± 1.40	71.89 ± 1.33
16	Previdência e Assistência Social	76.38 ± 0.89	81.58 ± 1.00	80.69 ± 1.51	72.54 ± 1.32
26	Processo Legislativo e Atuação Parlamentar	76.19 ± 2.95	81.98 ± 3.02	79.12 ± 3.59	73.59 ± 3.93
13	Cidades e Desenvolvimento Urbano	74.74 ± 0.72	82.90 ± 0.79	81.01 ± 0.91	69.39 ± 1.26
17	Direito Civil e Processual Civil	74.55 ± 2.02	80.05 ± 2.14	79.95 ± 1.33	69.96 ± 3.74
18	Agricultura, Pecuária, Pesca e Extrativismo	74.42 ± 1.46	82.59 ± 1.22	80.51 ± 1.82	69.22 ± 1.91
6	Direitos Humanos e Minorias	72.68 ± 1.09	79.43 ± 1.45	78.50 ± 1.30	67.68 ± 1.19
10	Defesa e Segurança	70.98 ± 1.17	77.88 ± 0.77	75.23 ± 1.74	67.20 ± 1.38
22	Estrutura Fundiária	70.66 ± 2.21	76.32 ± 2.46	72.88 ± 2.21	68.64 ± 3.21
25	Arte, Cultura e Religião	70.53 ± 3.26	75.87 ± 2.53	75.80 ± 2.61	66.04 ± 4.41
2	Administração Pública	69.83 ± 0.82	77.46 ± 0.95	75.13 ± 0.66	65.23 ± 1.09
8	Economia	68.29 ± 1.38	75.15 ± 1.90	75.41 ± 1.56	62.43 ± 2.07
27	Ciência, Tecnologia e Inovação	62.74 ± 2.35	68.74 ± 1.65	69.66 ± 2.08	57.16 ± 3.32
11	Indústria, Comércio e Serviços	57.16 ± 2.08	61.24 ± 2.01	63.69 ± 2.18	51.87 ± 2.20
23	Direito e Defesa do Consumidor	56.49 ± 2.81	59.76 ± 3.34	63.24 ± 2.56	51.17 ± 3.86
28	Direito e Justiça	25.35 ± 4.04	28.71 ± 4.98	47.94 ± 8.49	17.40 ± 3.14
32	Ciências Exatas e da Terra	0.00 ± 0.00	0.27 ± 0.73	0.00 ± 0.00	0.00 ± 0.00
31	Ciências Sociais e Humanas	0.00 ± 0.00	0.05 ± 0.06	0.00 ± 0.00	0.00 ± 0.00
30	Direito Constitucional	0.00 ± 0.00	16.25 ± 6.78	0.00 ± 0.00	0.00 ± 0.00

Table A.14 – BERTimbau fine-tuned with the standard corpus without the classes *Ciências Exatas e da Terra* and *Ciências Sociais e Humanas* local metrics to each class, using 75 tokens as the maximum sentence length. Each value is in the format "xx ± yy", which is the average and standard deviation of the stratified 10-fold cross-validation.

Order	*Target*	*F1*	*PR AUC*	*Precision*	*Recall*
1	Comunicações	95.87 ± 0.28	98.86 ± 0.16	96.36 ± 0.45	95.39 ± 0.59
4	Educação	92.35 ± 0.41	96.32 ± 0.61	93.14 ± 1.01	91.58 ± 0.81
3	Saúde	86.86 ± 0.57	93.55 ± 0.40	88.40 ± 0.61	85.38 ± 1.22
20	Esporte e Lazer	86.59 ± 1.82	91.97 ± 1.00	88.18 ± 2.12	85.14 ± 2.94
15	Energia, Recursos Hídricos e Minerais	84.75 ± 1.13	90.27 ± 1.33	85.07 ± 1.03	84.48 ± 2.31
19	Relações Internacionais e Comércio Exterior	84.06 ± 1.49	88.29 ± 1.29	85.15 ± 1.87	83.05 ± 2.19
12	Direito Penal e Processual Penal	84.05 ± 0.69	90.40 ± 0.50	87.80 ± 1.70	80.64 ± 1.23
21	Homenagens e Datas Comemorativas	83.86 ± 1.25	87.71 ± 1.56	80.17 ± 1.63	88.01 ± 2.98
24	Política, Partidos e Eleições	83.22 ± 1.95	87.24 ± 2.29	81.76 ± 2.23	84.81 ± 3.04
14	Meio Ambiente e Desenvolvimento Sustentável	80.69 ± 1.39	86.73 ± 1.33	83.59 ± 1.40	78.03 ± 2.44
7	Trabalho e Emprego	80.11 ± 0.85	87.29 ± 0.71	84.75 ± 1.16	75.99 ± 1.64
29	Turismo	79.37 ± 2.40	80.20 ± 3.39	81.20 ± 2.40	77.65 ± 2.94
9	Viação, Transporte e Mobilidade	78.60 ± 0.97	85.57 ± 1.00	79.03 ± 1.54	78.19 ± 1.14
5	Finanças Públicas e Orçamento	77.14 ± 0.75	84.90 ± 0.79	83.10 ± 1.66	72.00 ± 0.99
26	Processo Legislativo e Atuação Parlamentar	76.21 ± 1.59	81.10 ± 1.40	78.56 ± 2.47	74.08 ± 2.47
16	Previdência e Assistência Social	75.78 ± 1.54	81.27 ± 1.30	81.69 ± 1.49	70.70 ± 2.20
13	Cidades e Desenvolvimento Urbano	74.92 ± 1.10	82.76 ± 1.00	81.06 ± 1.54	69.68 ± 1.53
17	Direito Civil e Processual Civil	74.74 ± 1.71	80.02 ± 1.57	79.81 ± 1.42	70.30 ± 2.16
18	Agricultura, Pecuária, Pesca e Extrativismo	74.45 ± 1.67	82.34 ± 1.93	80.80 ± 1.96	69.12 ± 2.87
6	Direitos Humanos e Minorias	73.12 ± 0.96	79.73 ± 1.01	77.94 ± 1.02	68.88 ± 1.62
10	Defesa e Segurança	70.90 ± 1.40	77.84 ± 1.31	74.77 ± 2.36	67.45 ± 1.48
22	Estrutura Fundiária	70.66 ± 1.93	76.26 ± 1.08	74.10 ± 2.43	67.64 ± 3.15
25	Arte, Cultura e Religião	69.76 ± 1.90	76.08 ± 2.28	75.03 ± 2.95	65.29 ± 2.64
2	Administração Pública	69.54 ± 0.85	77.80 ± 0.44	76.19 ± 1.05	63.98 ± 1.31
8	Economia	68.11 ± 1.38	75.31 ± 1.40	75.36 ± 1.99	62.21 ± 2.31
27	Ciência, Tecnologia e Inovação	62.79 ± 3.40	68.01 ± 4.54	68.93 ± 6.56	58.04 ± 3.99
11	Indústria, Comércio e Serviços	56.81 ± 1.35	61.54 ± 0.75	63.72 ± 1.67	51.33 ± 2.30
23	Direito e Defesa do Consumidor	56.51 ± 1.63	59.94 ± 2.32	64.47 ± 3.47	50.49 ± 2.71
28	Direito e Justiça	22.44 ± 4.17	28.06 ± 3.30	46.55 ± 6.53	14.88 ± 3.16
30	Direito Constitucional	0.69 ± 2.07	15.62 ± 5.32	5.00 ± 15.00	0.37 ± 1.11

Table A.15 – BERTikal fine-tuned with the standard corpus local metrics to each class, using 75 tokens as the maximum sentence length. Each value is in the format "xx ± yy", which is the average and standard deviation of the stratified 10-fold cross-validation.

Order	*Target*	*F1*	*PR AUC*	*Precision*	*Recall*
1	Comunicações	95.78 ± 0.27	98.78 ± 0.17	96.47 ± 0.36	95.10 ± 0.63
4	Educação	92.06 ± 0.46	95.91 ± 0.44	92.58 ± 0.67	91.56 ± 0.64
3	Saúde	86.10 ± 0.53	92.73 ± 0.47	88.00 ± 0.59	84.30 ± 0.82
20	Esporte e Lazer	85.26 ± 1.26	89.87 ± 1.74	86.35 ± 1.88	84.27 ± 2.54
15	Energia, Recursos Hídricos e Minerais	83.85 ± 1.33	88.93 ± 1.12	85.13 ± 1.33	82.64 ± 1.93
12	Direito Penal e Processual Penal	83.20 ± 1.34	89.46 ± 1.05	87.31 ± 1.92	79.50 ± 1.74
21	Homenagens e Datas Comemorativas	83.20 ± 1.30	85.55 ± 1.54	79.55 ± 2.10	87.23 ± 1.18
24	Política, Partidos e Eleições	82.76 ± 1.08	85.44 ± 1.88	80.72 ± 1.98	85.01 ± 2.58
19	Relações Internacionais e Comércio Exterior	81.83 ± 1.61	85.96 ± 1.46	85.24 ± 1.06	78.70 ± 2.15
14	Meio Ambiente e Desenvolvimento Sustentável	79.63 ± 1.31	84.81 ± 1.43	83.86 ± 1.27	75.84 ± 2.05
7	Trabalho e Emprego	79.47 ± 0.85	86.31 ± 0.65	83.70 ± 1.06	75.68 ± 1.52
9	Viação, Transporte e Mobilidade	78.38 ± 1.01	84.92 ± 1.09	77.68 ± 1.43	79.13 ± 1.44
29	Turismo	77.09 ± 2.95	78.04 ± 4.88	79.45 ± 3.78	75.12 ± 4.66
5	Finanças Públicas e Orçamento	76.31 ± 1.06	83.94 ± 0.92	83.33 ± 1.71	70.40 ± 1.06
16	Previdência e Assistência Social	75.97 ± 1.19	80.66 ± 1.33	80.59 ± 1.64	71.91 ± 2.32
17	Direito Civil e Processual Civil	74.24 ± 1.88	79.46 ± 1.96	79.39 ± 0.83	69.81 ± 3.42
13	Cidades e Desenvolvimento Urbano	73.79 ± 0.96	82.09 ± 1.19	82.08 ± 2.30	67.08 ± 1.54
18	Agricultura, Pecuária, Pesca e Extrativismo	73.06 ± 1.46	80.59 ± 1.22	77.80 ± 2.11	68.93 ± 2.30
26	Processo Legislativo e Atuação Parlamentar	71.94 ± 2.48	76.03 ± 3.56	77.68 ± 3.05	67.10 ± 3.27
6	Direitos Humanos e Minorias	71.26 ± 0.99	77.79 ± 1.56	77.46 ± 1.59	66.00 ± 1.14
10	Defesa e Segurança	69.58 ± 0.82	76.34 ± 0.79	74.96 ± 1.11	64.93 ± 1.12
22	Estrutura Fundiária	68.70 ± 1.66	74.53 ± 2.16	73.67 ± 1.29	64.42 ± 2.67
25	Arte, Cultura e Religião	67.84 ± 2.61	72.69 ± 2.38	76.26 ± 2.73	61.19 ± 3.47
2	Administração Pública	67.71 ± 0.65	75.88 ± 0.88	76.03 ± 0.85	61.03 ± 0.82
8	Economia	66.04 ± 1.22	73.38 ± 1.65	75.37 ± 1.02	58.81 ± 2.06
27	Ciência, Tecnologia e Inovação	60.50 ± 3.80	65.23 ± 3.24	71.25 ± 2.74	52.73 ± 4.96
11	Indústria, Comércio e Serviços	55.28 ± 1.53	59.67 ± 1.74	63.14 ± 1.77	49.19 ± 1.72
23	Direito e Defesa do Consumidor	55.15 ± 2.05	56.64 ± 3.28	60.27 ± 2.01	50.93 ± 2.98
28	Direito e Justiça	17.19 ± 5.83	24.94 ± 4.15	48.65 ± 13.47	10.57 ± 3.93
32	Ciências Exatas e da Terra	0.00 ± 0.00	0.03 ± 0.05	0.00 ± 0.00	0.00 ± 0.00
31	Ciências Sociais e Humanas	0.00 ± 0.00	0.21 ± 0.29	0.00 ± 0.00	0.00 ± 0.00
30	Direito Constitucional	0.00 ± 0.00	12.48 ± 6.14	0.00 ± 0.00	0.00 ± 0.00

Table A.16 – BERTikal fine-tuned with the standard corpus without the classes *Ciências Exatas e da Terra* and *Ciências Sociais e Humanas* local metrics to each class, using 75 tokens as the maximum sentence length. Each value is in the format "xx ± yy", which is the average and standard deviation of the stratified 10-fold cross-validation.

Order	*Target*	*F1*	*PR AUC*	*Precision*	*Recall*
1	Comunicações	95.76 ± 0.40	98.70 ± 0.13	96.66 ± 0.55	94.88 ± 0.52
4	Educação	92.13 ± 0.55	95.89 ± 0.58	93.09 ± 0.88	91.20 ± 0.72
3	Saúde	86.10 ± 0.39	92.70 ± 0.50	87.93 ± 0.60	84.36 ± 0.88
20	Esporte e Lazer	85.28 ± 0.89	90.32 ± 1.27	85.67 ± 1.17	84.92 ± 1.54
15	Energia, Recursos Hídricos e Minerais	84.49 ± 0.97	89.30 ± 1.43	84.79 ± 0.74	84.22 ± 1.96
12	Direito Penal e Processual Penal	83.24 ± 0.71	89.57 ± 0.58	87.69 ± 1.55	79.25 ± 1.12
21	Homenagens e Datas Comemorativas	82.89 ± 0.78	86.08 ± 1.43	79.44 ± 1.10	86.68 ± 1.64
24	Política, Partidos e Eleições	82.70 ± 1.92	85.22 ± 2.45	81.25 ± 2.86	84.34 ± 3.30
19	Relações Internacionais e Comércio Exterior	82.08 ± 1.50	86.53 ± 1.34	85.05 ± 1.09	79.35 ± 2.58
14	Meio Ambiente e Desenvolvimento Sustentável	79.60 ± 1.64	84.86 ± 1.40	83.21 ± 1.66	76.31 ± 2.09
7	Trabalho e Emprego	79.12 ± 0.82	86.25 ± 0.70	83.24 ± 0.85	75.42 ± 1.73
9	Viação, Transporte e Mobilidade	78.42 ± 0.70	84.91 ± 0.96	77.23 ± 1.00	79.66 ± 1.40
29	Turismo	77.23 ± 2.98	78.88 ± 2.31	79.65 ± 4.52	75.12 ± 3.37
5	Finanças Públicas e Orçamento	76.39 ± 0.85	83.92 ± 0.78	83.57 ± 1.32	70.36 ± 0.91
16	Previdência e Assistência Social	75.57 ± 1.17	80.29 ± 1.47	79.93 ± 1.27	71.68 ± 1.76
18	Agricultura, Pecuária, Pesca e Extrativismo	74.09 ± 1.31	80.80 ± 2.25	76.97 ± 2.44	71.52 ± 2.36
13	Cidades e Desenvolvimento Urbano	73.93 ± 1.32	82.00 ± 1.34	83.21 ± 0.92	66.55 ± 2.13
17	Direito Civil e Processual Civil	73.86 ± 1.51	78.94 ± 1.55	77.73 ± 1.91	70.40 ± 2.15
26	Processo Legislativo e Atuação Parlamentar	72.54 ± 1.16	76.47 ± 1.42	78.10 ± 2.71	67.80 ± 1.72
6	Direitos Humanos e Minorias	71.51 ± 1.14	77.88 ± 1.02	77.08 ± 1.22	66.71 ± 1.69
10	Defesa e Segurança	69.76 ± 1.24	76.42 ± 1.37	75.19 ± 1.94	65.09 ± 1.26
22	Estrutura Fundiária	68.09 ± 2.89	74.39 ± 2.33	73.23 ± 2.34	63.81 ± 4.52
25	Arte, Cultura e Religião	68.07 ± 2.48	73.08 ± 2.89	75.53 ± 2.57	62.02 ± 3.18
2	Administração Pública	67.78 ± 0.76	75.79 ± 0.84	75.27 ± 1.19	61.65 ± 0.95
8	Economia	66.19 ± 1.50	73.38 ± 1.33	74.92 ± 1.46	59.34 ± 2.30
27	Ciência, Tecnologia e Inovação	61.30 ± 3.83	66.13 ± 4.56	69.50 ± 5.67	55.06 ± 4.28
23	Direito e Defesa do Consumidor	55.38 ± 1.41	56.26 ± 2.31	59.07 ± 3.06	52.32 ± 2.81
11	Indústria, Comércio e Serviços	55.15 ± 1.47	59.64 ± 1.63	64.21 ± 1.92	48.39 ± 1.93
28	Direito e Justiça	21.29 ± 4.31	27.31 ± 2.80	52.16 ± 9.35	13.44 ± 2.98
30	Direito Constitucional	0.00 ± 0.00	13.74 ± 6.97	0.00 ± 0.00	0.00 ± 0.00