Fine-Tuning MultiFit for Enhanced Legal Sentence Basis Classification

1st David J. Barrientos R.
Escola Politécnica de Pernambuco
Recife, Brazil
djbr@ecomp.poli.br

2nd Bruno J. T. Fernandes
Escola Politécnica de Pernambuco
Recife, Brazil
bjtf@ecomp.poli.br

3rd Cleyton M. de O. Rodrigues
Escola Politécnica de Pernambuco
Recife, Brazil
cleyton.rodrigues@upe.br

4th Leandro H. S. Silva
Escola Politécnica de Pernambuco
Recife, Brazil
lhss@ecomp.poli.br

5th Allana L. S. Rocha
Escola Politécnica de Pernambuco
Recife, Brazil
alsr@ecomp.poli.br

6th Paulo Christiano Sobral
NeuroLaw
Desenvolvimento de Tecnologia LTDA
Recife, Brazil
christiano.sobral@neuro.law

2023 IEEE Latin American Conference on Computational Intelligence (LA-CCI) | 979-8-3503-4807-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/LA-CCI58595.2023.10409399

***Abstract*—Deep learning algorithms have shown promise in effectively classifying legal texts, surpassing traditional methods. However, existing approaches are primarily designed for English text and lack suitability for other languages, mainly Portuguese. This study addresses the challenge of classifying legal basis in first-degree sentences within Brazilian law by fine-tuning the multilingual MultiFit model using a novel basis dataset, comprehensively training the model for accurate legal basis classification.**

The bidirectional deep-learning MultiFit model has been subjected to rigorous fine-tuning, resulting in exceptional performance while maintaining consistently high quality. Results obtained highlight the model's remarkable proficiency in precisely categorizing legal bases in first-degree sentences, achieving an accuracy rate of 80.0%, precision of 83.3%, recall of 80.7%, and an F1 score of 82.0%. These results demonstrate the model's adaptability, versatility, and suitability for legal applications. In addition, it exhibits high precision, accuracy, and efficiency in classifying legal bases in first-degree sentences. Moreover, successfully fine-tuning pre-trained models for new tasks, leveraging extensive datasets, highlights their significant potential in enhancing performance in legal applications.

***Index Terms*—artificial intelligence, data mining, machine learning, deep learning, natural language processing**

I. Introduction

Applying deep learning (DL) techniques to classify legal texts is a well-established area of research within natural language processing [1], [13], [14], [29], [31]. Advanced Deep Learning techniques have showcased their ability to comprehensively capture language's intricate nuances and structural elements, closely mirroring human understanding [5]. These techniques have proven to be highly effective in diverse text classification tasks [33]. One notable advantage of deep models is their inherent capacity for automatic feature extraction, facilitating extracting of pertinent features for text classification. This inherent ability empowers deep learning models to adeptly comprehend the complex relationships among words and phrases, enabling them to discern contextual meanings within legal texts. Furthermore, deep learning models exhibit another advantage: their trainability on extensive and varied corpora of legal texts, facilitating robust generalization to new, previously unseen legal texts [12], [27]. However, these advances take time to transfer outside English. As a result, there has been little interest in publishing research or building datasets in other languages, even though industry applications highly require them [10].

Implementing deep learning algorithms in legal text classification demonstrates significant potential for enhancing accuracy and efficiency compared to traditional machine learning techniques [22]. Deep Neural Networks have extended their analytical and processing capacity to capture subtle language semantics and syntax; closer to human sophistication. However, several challenges must be addressed. First, legal basis classification involves dealing with diverse document types that vary in length, structure, and viewpoints [18], [19]. However, many existing approaches lack adaptations tailored to the legal domain. Furthermore, these approaches are often trained using English text, which makes them less suitable for other languages [9], particularly legal Portuguese. Furthermore, the limited availability of labeled data poses an additional obstacle in developing robust models for legal text classification. Overcoming these challenges requires advancements in domain-specific language modeling, semantic understanding, and access to more extensive, well-annotated datasets, unlocking the full potential of deep learning algorithms in legal text classification.

This study presents two main contributions. Firstly, it presents a specialized bidirectional deep learning model that accurately classifies the basis of first-degree legal sentences in Brazilian law. The model uses a pre-trained MultiFit model [10] and knowledge from the TCU (*Tribunal de Contas da União, Brasília*) jurisprudence corpus to capture complex language patterns and nuances, improving the accuracy and robustness of the classification task. The MultiFit model is chosen for its exceptional performance, efficient fine-tuning mechanism, and proficiency in multiple languages, making it an excellent option for legal sentence classification tasks beyond English [17]. Secondly, a labeled dataset curated explicitly for this task is introduced, which includes various foundations of varying sizes. This dataset provides essential training and evaluation resources, enabling further legal text classification developments.

II. Related Works

Text classification is essential to text analysis, especially when dealing with legal texts. Numerous well-established approaches have been widely employed in text classification, including the Naive Bayes classifier, Support Vector Machine (SVM), and Logistic Regression [24], [28]. Recently, the advent of deep learning methods has introduced promising advancements in this field, with notable techniques such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long-Short Term Memory (LSTM) networks gaining significant attention [6], [30], [32].

In order to effectively categorize legal documents, it is imperative to determine the significance of linguistic information. This information can play a crucial role in organizing legal texts into various categories, such as laws, regulations, and judgments, and can significantly aid in the research and analysis of these documents. In the work of Gao et al. [11], statistical and deep semantic features are used for legal text classification along with shallow classifiers such as Logistic Regression and Support Vector Machines.

Identifying corresponding legal provisions within written descriptions of events in English legal texts is a complex and practical issue. Li [20] presents a novel legal text classification method using feature words to tackle this challenge. Experiments demonstrate the efficacy of the proposed algorithm, which combines the Term Frequency-Inverse Document Frequency (TF-IDF) method with the chi-square statistic (CHI), in extracting feature words from a diverse range of legal texts and accurately classifying them into the appropriate legal terms. The results indicate that the proposed method is highly effective in legal text classification.

Pudaruth et al. [25] propose a deep learning-based model for classifying 490 legislations from the Republic of Mauritius into 30 categories. They show that the deep learning model was superior to traditional machine learning methods, such as Support Vector Machine and Decision Tree, by achieving an accuracy of 60.9%.

Bansal et al. [3] presented a survey about various deep learning methods applied in the legal domain. They found that models such as Convolutional Neural Networks, Long Short-Term Memory, and multi-task deep learning models achieve state-of-the-art performance in various legal tasks. Regarding legal studies in Portuguese, Silva et al. [7] implemented CNNs to examine and categorize legal papers from Brazil, yielding positive outcomes. Cruz et al. [21] used a Summarization-based model and a Support Vector Classifier to classify summons or notices in Portuguese legal documents.

However, it is important to note that most of these approaches were not specifically designed to cater to the unique characteristics of the legal domain outside the English language, which limits their suitability for effectively handling legal texts in Portuguese. Consequently, employing these methods directly in legal documents in Portuguese, leads to suboptimal results and limits the extraction of accurate insights.

III. Methodology

This study presents an approach that harnesses the robust capabilities of the Multi-lingual Fine-tuning (MultiFit) model [10]. The selection of the MultiFit model is motivated by its outstanding performance, efficient fine-tuning mechanism, transfer learning capabilities, and proficiency in multiple languages, rendering it an excellent choice for legal sentence classification tasks extending beyond the English language [17], [26]. As an extension of the ULMFit [16] model, MultiFit is specifically designed to enhance its efficiency and applicability in multi-lingual modeling tasks. Notably, it incorporates various enhancements, such as subword tokenization, as opposed to word-based tokenization, enabling a more refined representation of textual content. Moreover, MultiFit introduces a notable architectural modification by replacing the LSTM (Long Short-Term Memory) [15] architecture utilized in ULMFit with a QRNN (Quasi-Recurrent Neural Network) [4]. This architectural change offers improved computational performance while preserving the model's ability to capture long-range dependencies in sequential data, enhancing its effectiveness.

The MultiFit model undergoes extensive training on a diverse text corpus comprising multiple languages, including German, Spanish, French, Italian, Japanese, Russian, and Chinese. Utilizing a multilingual approach, the model can thoroughly understand language structures and semantics in different linguistic contexts. Moreover, training the model in multiple languages allows it to leverage the linguistic knowledge acquired from each language, thus enhancing its overall performance [10].

The results demonstrate that MultiFit surpasses the performance of both multilingual BERT [8] and LASER [2] models, despite being pretrained on a comparatively smaller dataset of 100 million tokens. These findings are particularly noteworthy, considering that multilingual BERT and LASER models have been trained using significantly more data and computational resources. The superior performance of MultiFit suggests that its training methodology is highly effective in capturing subtle language nuances and achieving state-of-the-art outcomes, even with limited resources, compared to other pretrained models.

IV. Proposed Model

The MultiFit model exhibits notable fine-tuning capabilities, even when confronted with smaller task-specific datasets with limited annotated training data. This adaptability, alongside the model's bidirectional nature to augment the classifier's effectiveness and reliability, allows the creation of robust models for classification tasks.

The selected MultiFit model for this study underwent an extensive 16-hour pre-training phase on a corpus comprising 100 million tokens extracted from the Portuguese Wikipedia. Subsequently, employing transfer learning, the pre-trained MultiFit model was further trained as a legal text classifier using a comprehensive corpus of jurisprudence acquired from TCU (*Tribunal de Contas da União, Brasília*, Brazil). This corpus, abundant with diverse legal text data, proved to be an ideal training resource, enabling the pre-trained model to demonstrate exceptional proficiency in legal sentences in Portuguese. For a detailed overview of the complete training process including the fine-tuning process, refer to Figure 1.

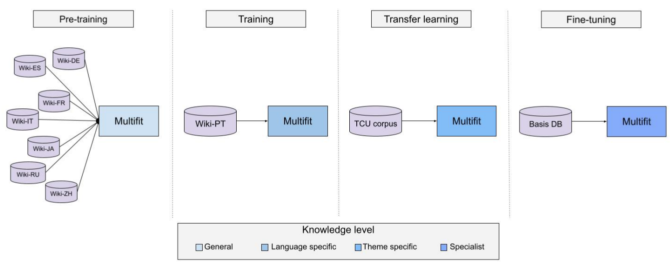

Fig. 1. MultiFit model showcasing the different training process, transfer learning and fine tuning.

This study proposes a fine-tuning approach for the selected MultiFit model, aimed at enhancing its performance in accurately classifying the basis of first-degree legal sentences. The fine-tuning process involves training the selected model on a novel basis dataset to optimize its effectiveness for this new classification task. To accomplish this, a new dataset was specifically created for this study, enabling the update of the model's parameters through a comprehensive end-to-end training process. Then, leveraging the knowledge acquired from the TCU jurisprudence corpus, the pre-trained MultiFit model was initialized, allowing it to retain its multilingual and legal text classification capabilities while improving its performance, specifically on legal documentation in Portuguese. As a result of the fine-tuning process, a specialized bidirectional deep learning model was meticulously developed, tailored to classify the legal bases in first-degree legal sentences with remarkable precision and accuracy. Refer to Figure 2 for a thorough understanding of the classification model.

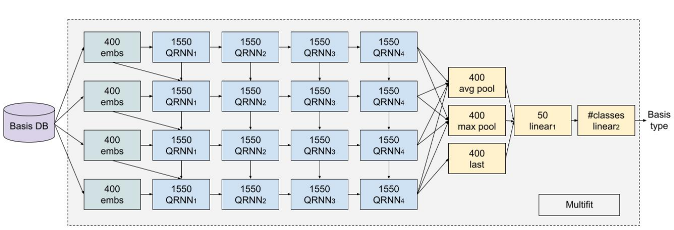

Fig. 2. Proposed model for basis classification

Following subsections provide more information on the database, hyper-parameters and training process.

A. Basis Database

The basis dataset was meticulously constructed from 612 distinct first-degree legal sentences written in Portuguese. To ensure the high quality of the data, a specialized pre-processing step was applied to each legal sentence. This step involved extracting the document's text, removing any special characters, eliminating excess whitespaces, and addressing other formatting intricacies that may arise. Next, the basis section within each legal sentence was precisely extracted and segmented using a series of regular expressions (regex). These carefully crafted expressions were specifically designed to search for targeted terms such as "Fundamentação" and "Dispensado o relatório," enabling the accurate identification and isolation of the basis section. In addition, the sequential implementation of these regular expressions further optimized the basis extraction process, ensuring comprehensive coverage even in cases where prior expressions may not have been successful.

The extracted segments underwent a meticulous manual annotation process, wherein a ground truth for the basis sections (as shown in Figure 3) was established. This process involved a collaborative effort between two researchers and two Brazilian law experts. All four individuals thoroughly reviewed each annotation to ensure the data's reliability and accuracy. Consequently, this rigorous annotation validated and improved the accuracy of the extracted basis segments, enhancing the overall reliability of the study.

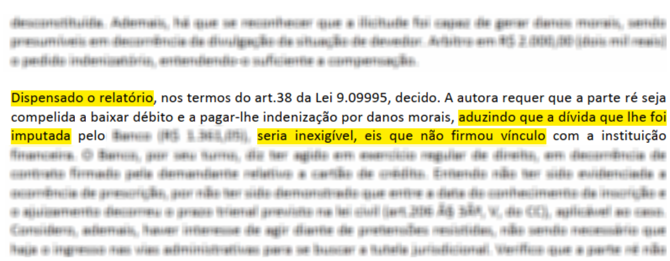
Dispensado o relatório, nos termos do art.38 da Lei 9.09995, decido. A autora requer que a parte ré seja compelida a baixar débito e a pagar-lhe indenização por danos morais, aduzindo que a dívida que lhe foi imputada pelo seria inexigível, eis que não firmou vínculo com a instituição

Fig. 3. Basis Section Example.

During the manual annotation process, it was discovered that the dataset exhibited a significant class imbalance. Specifically, 23 of the 34 basis types had only ten or fewer examples available, presenting a substantial lack of representation. In contrast, the basis type "cobrança indevida reconhecida" stood as the sole exception, boasting 230 examples, as demonstrated in Figure 4.

Given this imbalanced distribution, updating both the classifier and the dataset became necessary to accommodate a binary classification system. In the updated binary classification system, a value of 0 is assigned to indicate that the basis is not classified as "cobrança indevida reconhecida." In contrast, a value of 1 signifies that the basis indeed falls under the "cobrança indevida reconhecida." This modification was undertaken to address the data imbalance and ensure more equitable representation in the dataset, facilitating the development of a robust and accurate classifier.

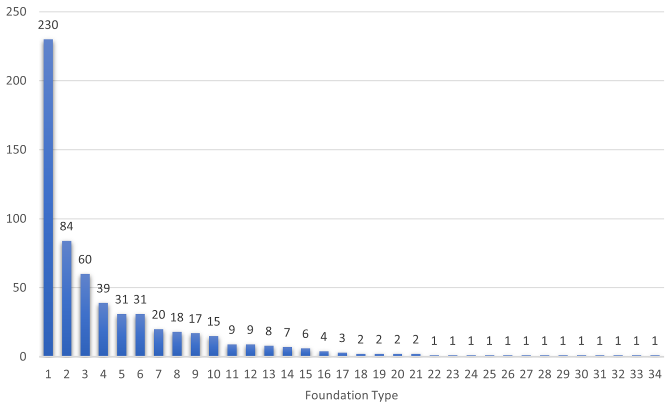

Fig. 4. Distribution of Basis Types in the Database. This histogram depicts the relative frequencies of different basis types used for data representation in the database, providing insights into its composition and informing data analysis decisions.

The newly constructed dataset comprises a total of 612 legal sentences, out of which 230 (37.58%) have been accurately annotated as "cobrança indevida reconhecida" or "Recognized Undue Charge (RUC)." In contrast, the remaining 382 (62.42%) are categorized under various other basis types. This revised distribution of examples significantly improves the balance and appropriateness of the representation across the different basis types. Moreover, the basis database encompasses a wide range of sizes, with no restrictions on its extent. This characteristic makes it versatile and adaptable to various research or application scenarios. It is important to identify instances of "cobrança indevida" due to its frequent occurrence. This highlights the pressing need to safeguard consumer rights, maintain financial transparency, and fortify trust within commercial interactions. Researchers and practitioners can leverage this extensive dataset to explore and analyze different aspects related to legal sentences and basis types, facilitating more in-depth insights and valuable discoveries.

B. Hyper-Parameters

In the fine-tuning process of the MultiFit model, the following hyperparameters were used to enhance the model's performance and convergence.The best hyperparameters were selected by testing different combinations and assessing their performance on a validation dataset, leading to the selection of those that produced the best results.

- Loss Function: The FlattenedLoss is a simple loss function that flattens the prediction tensor and computes the cross-entropy between the flattened prediction and target. The LabelSmoothing CrossEntropy [23] loss is a regularization technique that modifies the cross-entropy loss to account for label smoothing. Label smoothing is a technique that alters the one-hot encoded target labels to have a small amount of uncertainty. This combination loss function provides a regularization effect through label smoothing and serves as a performance evaluator for the model during optimization.
- Number of Layers: The number of layers in the model affects its capacity to learn complex relationships. Four layers were used to balance complexity and efficiency in this fine-tuning scenario.
- Hidden Parameters by Layer: The number of neurons in each model layer is determined by the hidden parameters. In this case, each of the four layers had 1550 neurons.
- Batch Size: The batch size determines the number of examples processed in each training iteration. A batch size of 48 was used to balance computational efficiency and optimization stability.
- Epochs (Forward and Backward): The model underwent 20 forward and 20 backward epochs, respectively, allowing it to learn and refine its predictions fully.
- Training and Testing Set Sizes: The model was trained on 90% of the data, or 556 examples, and tested on 10% of the data, or 55 examples. The dataset was split using Random Splitting.

It is important to note that these hyperparameters can significantly impact the performance and convergence of the MultiFit model. Table I summarizes all the hyperparameters used in the fine-tuning process.

TABLE I
HYPER-PARAMETERS SELECTED FOR THE TRAINING PROCESS

Hyper Parameter	Selected Value
Loss Function	Label Smoothing Cross Entropy Flattened Loss
Number of layers	4
Hidden parameters by layer	1550
Batch Size	48
Epochs (Forward)	20
Epochs (Backward)	20
Training Set Size	556 (90%)
Testing Set Size	55 (10%)

V. RESULTS

This section presents the performance results of the binary classifier. The classifier was trained and evaluated on a carefully curated dataset consisting of 556 instances for training and 55 instances for testing.

The confusion matrix, illustrating the classifier's performance, is presented in Table II, offering valuable insights into the classification outcomes. Additionally, Table III provides a concise summary of the performance metrics attained by the binary classifier when evaluated on the test dataset. Notably, the classifier demonstrated an accuracy of 0.800, indicating the overall correctness of its predictions. Furthermore, a precision of 0.833 showcases the classifier's ability to accurately identify instances belonging to the positive class, while a recall of 0.807 reflects its capability to capture relevant instances effectively. Finally, the F1 score of 0.82 further signifies a harmonious balance between precision and recall, underlining the classifier's overall balanced performance.

These performance metrics highlight the classifier's effectiveness in accurately classifying the basis types, demonstrating its robustness and potential for real-world applications.

TABLE II
BASIS CLASSIFIER CONFUSION MATRIX

		Predicted	
		RUC	Others
Real	RUC	25	6
	Others	5	19

TABLE III
PERFORMANCE METRICS FOR THE BASIS CLASSIFIER

Measure	Value
Accuracy	0.800
Precision	0.833
Recall	0.807
F1 Score	0.820

Regarding complexity, the binary classifier boasted an impressive 46,020,150 parameters, making it a pretty sophisticated model. However, despite its intricacy, the training process proved remarkably efficient, averaging 52 seconds per epoch. Overall, the model completed its training in a relatively short period of 34.7 minutes, spanning 40 epochs (20 forward and 20 backward).

This study's findings clearly show how well the binary classifier performs on the given dataset. With its high accuracy, well-balanced precision and recall, and efficient training time, the classifier showcases its potential for practical applications. It is worth noting, though, that these findings are based on a limited dataset. Further evaluations using more extensive and diverse datasets are essential to gain a comprehensive understanding of the classifier's performance.

VI. CONCLUSIONS

In this study, a pre-trained MultiFit model was fine-tuned to classify the basis of first-degree legal sentences effectively. Utilizing a specialized legal basis dataset optimized the model's performance for this task while preserving its exceptional multilingual and legal text classification capabilities. The result is a sophisticated bidirectional deep-learning model demonstrating outstanding precision and accuracy. The classifier achieved impressive performance metrics, including an 80.0% accuracy, 83.3% precision, 80.7% recall, and an F1 score of 82.0%. Despite its high complexity, with 46,020,150 parameters, the model proved highly efficient during training. The model delivers accurate results without compromising efficiency, with an average training time of only 52 seconds per epoch and a total training duration of 34.7 minutes. This specialized bidirectional model is a valuable tool for accurately classifying legal bases in first-degree sentences. It can be combined with other classifiers to create a comprehensive solution that covers all legal bases, leveraging the unique strengths of each classifier. Moreover, the fine-tuned MultiFit model can be adapted to other legal classifications in Portuguese, offering flexibility and adaptability within the legal field. In conclusion, our proposed model is valuable for classifying the basis of first-degree legal sentences. It provides high precision and accuracy while demonstrating efficiency during training. Furthermore, this study highlights the potential of pre-trained models to be successfully adapted to new tasks, harnessing the knowledge gained from extensive datasets and improving performance through fine-tuning.

VII. LIMITATIONS AND FUTURE DIRECTIONS

Our model has demonstrated great alignment with our objectives for the task at hand, delivering impressive results. Its primary focus was to serve as a binary classifier for the "Recognized Undue Charge" (RUC) basis, addressing the inherent imbalance observed within the database. This deliberate design choice effectively achieved our legal task, showcasing the model's strength. However, the model's ability to handle other basis types in the dataset is limited due to this emphasis on binary classification. Therefore, while the potential exists for training the model as a multiclass classifier, it was not prioritized in our current objective. Nevertheless, it is important to note that the model exhibits proficiency in accurately classifying sentences into multiple basis types, offering potential avenues for exploration and analysis. To further enhance our work, several promising future directions can be pursued. Firstly, we can explore the possibility of integrating our existing approach as a hybrid model, combining it with another model specifically designed to identify text similarity. This hybrid approach can significantly improve performance in multiclass classification, enabling our model to effectively handle a broader range of basis types present in the dataset. Furthermore, by leveraging the strengths of both models, it can achieve more accurate classifications.

ACKNOWLEDGMENT

We would like to express our gratitude to Neurolaw Desenvolvimento de Tecnologia LTDA for their support and collaboration in the development of this article. This research was partially financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001, Fundação de Amparo a Ciência e Tecnologia do Estado de Pernambuco (FACEPE), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)—Brazilian research agencies.

REFERENCES

[1] Onur Akça, Giyaseddin Bayrak, Abdul Majeed Issifu, and Murat Can Ganz. Traditional machine learning and deep learning-based text classification for turkish law documents using transformers and domain adaptation. *2022 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, pages 1–6, 2022.

[2] Mikel Artetxe and Holger Schwenk. Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. *Transactions of the Association for Computational Linguistics*, 7:597–610, nov 2019.

[3] Neha Bansal, Arun Sharma, and R. K. Singh. A review on the application of deep learning in legal domain. In John MacIntyre, Ilias Maglogiannis, Lazaros Iliadis, and Elias Pimenidis, editors, *Artificial Intelligence Applications and Innovations*, pages 374–381, Cham, 2019. Springer International Publishing.

[4] James Bradbury, Stephen Merity, Caiming Xiong, and Richard Socher. Quasi-recurrent neural networks, 2016.

[5] Ilias Chalkidis and Dimitrios Kampas. Deep learning in law: early adaptation and legal word embeddings trained on large corpora. *Artificial Intelligence and Law*, 27:171 – 198, 2018.

[6] Yahui Chen. Convolutional neural network for sentence classification. Master's thesis, University of Waterloo, 2015.
[7] N Correia Da Silva, FA Braz, TE De Campos, D Gusmao, F Chaves, D Mendes, D Bezerra, G Ziegler, L Horinouchi, M Ferreira, et al. Document type classification for brazil's supreme court using a convolutional neural network. In *10th International Conference on Forensic Computer Science and Cyber Law (ICoFCS), Sao Paulo, Brazil*, pages 29–30, 2018.
[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.
[9] Stella Douka, Hadi Abdine, Michalis Vazirgiannis, Rajaa El Hamdani, and David Restrepo Amariles. Juribert: A masked-language model adaptation for french legal text, 2022.
[10] Julian Martin Eisenschlos, Sebastian Ruder, Piotr Czapla, Marcin Kardas, Sylvain Gugger, and Jeremy Howard. Multifit: Efficient multilingual language model fine-tuning. *arXiv preprint arXiv:1909.04761*, 2019.
[11] Jiaming Gao, Hui Ning, Zhong yuan Han, Lei lei Kong, and Haoliang Qi. Legal text classification model based on text statistical features and deep semantic features. In *Fire*, 2020.
[12] Biyang Guo, Xin Zhang, Ziyuan Wang, Minqi Jiang, Jinran Nie, Yuxuan Ding, Jianwei Yue, and Yupeng Wu. How close is chatgpt to human experts? comparison corpus, evaluation, and detection, 2023.
[13] Eya Hammami, Imen Akermi, Rim Faiz, and Mohand Boughanem. Deep learning for french legal data categorization. In *International Conference on Model and Data Engineering*, 2019.
[14] Qian Han and Derek Snaidauf. Comparison of deep learning technologies in legal document classification. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 2701–2704, 2021.
[15] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
[16] Jeremy Howard and Sebastian Ruder. Universal language model fine-tuning for text classification. *arXiv preprint arXiv:1801.06146*, 2018.
[17] Junjie Hu, Sebastian Ruder, Aditya Siddhant, Graham Neubig, Orhan Firat, and Melvin Johnson. Xtreme: A massively multilingual multi-task benchmark for evaluating cross-lingual generalisation. In *International Conference on Machine Learning*, pages 4411–4421. PMLR, 2020.
[18] Zihan Huang, Charles Low, Mengqiu Teng, Hongyi Zhang, Daniel E. Ho, Mark S. Krass, and Matthias Grabmair. Context-aware legal citation recommendation using deep learning. *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law*, 2021.
[19] Penghua Li, Fen Zhao, Yuanyuan Li, and Ziqin Zhu. Law text classification using semi-supervised convolutional neural networks. *2018 Chinese Control And Decision Conference (CCDC)*, pages 309–313, 2018.
[20] Zhonghao Li. A classification retrieval approach for english legal texts. *2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*, pages 220–223, 2019.
[21] Marie Chantelle Cruz Medina, Lucas Matheus Da Silva Oliveira, Jean Felipe Coelho Ferreira, Leandro Honorato De S. Silva, Cleyton Mario O. Rodrigues, João Fausto L. De Oliveira, Paulo Christiano Sobral, Bruno Souza, Dionizio Feitosa, and Bruno J. T. Fernandes. Classification of legal documents in portuguese language based on summarization. In *2022 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, pages 1–6, 2022.
[22] Shervin Minaee, Nal Kalchbrenner, Erik Cambria, Narjes Nikzad, Meysam Chenaghlu, and Jianfeng Gao. Deep learning–based text classification: A comprehensive review. *ACM Comput. Surv.*, 54(3), apr 2021.
[23] Rafael Müller, Simon Kornblith, and Geoffrey Hinton. When does label smoothing help?, 2020.
[24] Ramesh Nallapati and Christopher D Manning. Legal docket classification: Where machine learning stumbles. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 438–446, 2008.
[25] Sameerchand Pudaruth, K.M.s Soyjaudah, and Gunputh Rajendra Parsad. Classification of legislations using deep learning. *The International Arab Journal of Information Technology*, 18, 01 2021.
[26] Xipeng Qiu, Tianxiang Sun, Yige Xu, Yunfan Shao, Ning Dai, and Xuanjing Huang. Pre-trained models for natural language processing: A survey. *Science China Technological Sciences*, 63(10):1872–1897, 2020.
[27] Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Kurt Shuster, Eric Michael Smith, Y.-Lan Boureau, and Jason Weston. Recipes for building an open-domain chatbot. In *Conference of the European Chapter of the Association for Computational Linguistics*, 2020.
[28] Octavia-Maria Sulea, Marcos Zampieri, Shervin Malmasi, Mihaela Vela, Liviu P Dinu, and Josef Van Genabith. Exploring the use of text classification in the legal domain. *arXiv preprint arXiv:1710.09306*, 2017.
[29] Samir Undavia, Adam Meyers, and J. Ortega. A comparative study of classifying legal documents with neural networks. *2018 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pages 515–522, 2018.
[30] Xin Wang, Yuanchao Liu, Cheng-Jie Sun, Baoxun Wang, and Xiaolong Wang. Predicting polarities of tweets by composing word embeddings with long short-term memory. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1343–1353, 2015.
[31] Fusheng Wei, Han Qin, Shi Ye, and Haozhen Zhao. Empirical study of deep learning for text classification in legal document review. *2018 IEEE International Conference on Big Data (Big Data)*, pages 3317–3320, 2018.
[32] Ye Zhang and Byron Wallace. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. *arXiv preprint arXiv:1510.03820*, 2015.
[33] Muhammad Zulqarnain, Rozaida Ghazali, Yana Mazwin Mohmad Hassim, and Muhammad Rehan. A comparative review on deep learning models for text classification. *Indonesian Journal of Electrical Engineering and Computer Science*, 19:325–335, 2020.