

# Text Classification in Law Area: a Systematic Review

V. S. Martins<sup>1</sup>, C. D. Silva<sup>1</sup>

Universidade Federal do Pará (UFPA)  
Programa de Pós-graduação em Computação Aplicada (PPCA)  
Campus Universitário de Tucuruí - Pará - Brazil  
{victorsm,cleison}@ufpa.br

**Abstract.** Automatic Text Classification represents a great improvement in law area workflow, mainly in the migration of physical to electronic lawsuits. A systematic review of studies on text classification in law area from January 2017 up to February 2020 was conducted. The search strategy identified 20 studies, that were analyzed and compared. The review investigates from research questions: what are the state-of-art language models, its application of text classification in English and Brazilian Portuguese datasets from legal area, if there are available language models trained on Brazilian Portuguese, and datasets in Brazilian law area. It concludes that there are applications of automatic text classification in Brazil, although there is a gap on the use of language models when compared with English language dataset studies, also the importance of language model in domain pre-training to improve results, as well as there are two studies making available Brazilian Portuguese language models, and one introducing a dataset in Brazilian law area.

CCS Concepts: • **Computing methodologies** → **Machine learning algorithms**.

Keywords: Text Classification, law, machine learning

## 1. INTRODUÇÃO

A introdução do Processamento de Linguagem Natural (PLN) na área Jurídica tem promovido diversos ganhos para os órgãos do Sistemas de Justiça, que no Brasil ganha contornos gigantescos dada a quantidade de processos atualmente em tramitação, que, segundo fontes do Conselho Nacional de Justiça - CNJ no seu relatório anual “Justiça em Números”<sup>1</sup>, referente ao ano-base de 2019, era de 77,1 milhões.

Dentre as aplicações no contexto jurídico, há a classificação de documentos para fins de migração dos processos físicos para eletrônicos [Silva and Maia 2020], ou ainda para direcionamento correto de demandas peticionadas eletronicamente, tal qual analisado em [Noguti et al. 2020] e [Mota et al. 2020]. Essas atividades requerem recursos humanos e horas de trabalho que podem ser destinadas a outras áreas com necessidades mais cognitivas, possibilitando a promoção de maior celeridade no Sistema de Justiça Brasileiro. A Figura 1 mostra um fluxo geral do processo de classificação textual.

Na Seção 2 é mostrado o protocolo de pesquisa utilizado na revisão. Nas Seções 3 a 6, as análises e comparações críticas entre os trabalhos selecionados, de acordo com as perguntas de revisão respectivas. Finalmente nas Seções 7 e 8, as principais lacunas e conclusões sobre o tema após as críticas e comparações.

---

<sup>1</sup><https://www.cnj.jus.br/wp-content/uploads/2020/08/WEB-V3-Justiça-em-Números-2020-atualizado-em-25-08-2020.pdf>

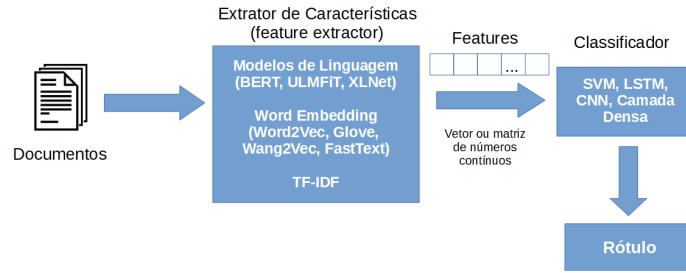


Fig. 1. Fluxo geral de um processo de classificação textual.

## 2. PROTOCOLO DE REVISÃO

O protocolo de revisão foi construído com o objetivo de responder às seguintes questões de revisão:

Q.1 Qual é o estado da arte na tarefa de classificação textual em PLN utilizando modelos pré-treinados?

Q.2 Há publicações científicas com modelos de linguagem pré-treinados utilizados no domínio jurídico?

Q.3 Quais técnicas de PLN no domínio da língua portuguesa têm sido aplicadas para classificação de textos jurídicos no Brasil?

Q.4 Há repositórios públicos de dados com textos no contexto jurídico brasileiro?

Q.5 Há repositório de modelos de linguagem pré-treinados para o português do Brasil?

### 2.1 Strings de Busca

Para a coleta dos artigos analisados, foram definidas quatro *strings* de busca com as palavras-chave da Tabela I. A construção de cada *string* objetiva responder uma pergunta específica das questões de revisão. Para efetivar as buscas dos trabalhos analisados, foi considerado o período de janeiro de 2017 a fevereiro de 2020.

Agrup.	1 (2,3)	2 (3)	3 (2)	4 (2,3,4)	5 (4)	6 (3,4,5)	7 (2,3)	8 (2,3,4,5)
String	classification OR label*	“machine learning” OR “deep learning” OR “language model” OR pretrained OR “transfer learning”	“language model” OR pretrained OR “transfer learning”	denunciation, crime report, petition, legal, law	dataset OR corpus	Portuguese OR brazil	document OR text	“natural language”

Table I. Palavras-chave agrupadas. Fonte: O autor.

Como referência para a **Q.1**, foi utilizado o sítio “NLP-Progress: Tracking Progress in Natural Language Processing (Text Classification)”, especificamente na sua área de classificação Textual<sup>2</sup>, que mantém um ranking que compara os resultados nessa tarefa para diferentes modelos e arquiteturas de aprendizado de máquina, com o objetivo de reunir o estado da arte na área. Desse modo, os três trabalhos que apresentaram os melhores resultados no mês da coleta, março de 2020, foram selecionados para a revisão.

Para as demais questões de pesquisa, as buscas foram realizadas nas seguintes bases: Portal de periódicos Capes, IEEE Xplore, Science Direct – Elsevier, Scopus, Google Acadêmico e ACL-Web

<sup>2</sup>[http://nlpprogress.com/english/text\\_classification.html](http://nlpprogress.com/english/text_classification.html)

(Association for Computational Linguistics), e incluíram apenas artigos completos, escritos em inglês ou português, que utilizam *datasets* em inglês ou em línguas latinas, para as tarefas de: classificação de documentos, análise de sentimento ou reconhecimento de entidade nomeada. As *strings* de pesquisa para cada pergunta foram montadas a partir dos agrupamentos de palavras-chave apresentados na Tabela I, com o conectivo "AND" entre elas. Na coluna de agrupamentos (**Agrup.**), é indicado entre os parênteses, as questões nas quais foram utilizadas cada um deles.

## 2.2 Seleção dos artigos para revisão

Os trabalhos encontrados foram reunidos em uma lista, totalizando 1.432 artigos. Após a remoção dos artigos duplicados, restaram 1.142 trabalhos.

Após leitura dos títulos e considerando os critérios de inclusão, de exclusão e de qualidade, 20 títulos foram selecionados para a revisão, os quais estão indicados com asterisco (\*) ao final na Seção Referências Bibliográficas. O fluxo de seleção é mostrado na Figura 2

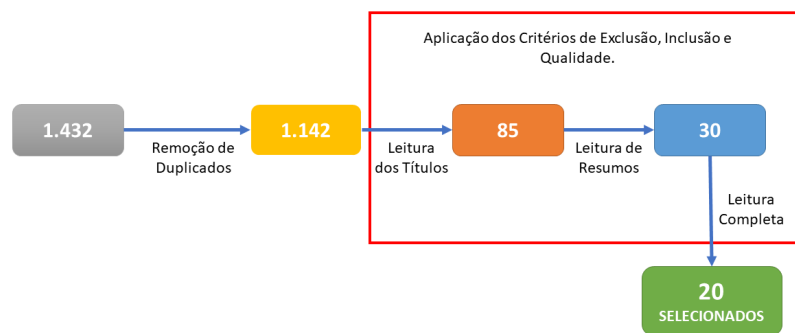


Fig. 2. Fluxo da seleção dos artigos analisados.

Os critérios utilizados garantem que a revisão atenda às questões de pesquisas, e ao mesmo tempo, restrinjam a pesquisa em modelos de PLN que envolvam tarefas de Classificação Textual que utilizam aprendizado profundo com ou sem modelos de linguagem pré-treinados para estudos com *datasets* em português, e naqueles com documentos em inglês, apenas os que propõe modelos de linguagem. Para melhor entendimento da quantidade de artigos publicados por ano, assim como o recente uso das técnicas de aprendizagem profunda no domínio jurídico, a Figura 3 mostra a quantidade de artigos publicados por ano dentre os selecionados para a revisão. Nas Seções 3 a 6 são efetuadas as análises dos trabalhos selecionados com o intuito de responder a cada uma das questões de pesquisa.

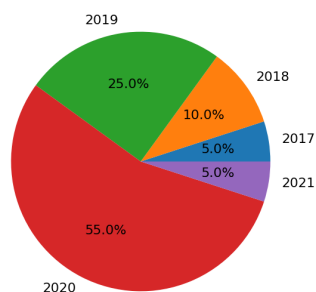


Fig. 3. Percentual de artigos selecionados para revisão por ano de referência.

## 3. MODELOS DE LINGUAGEM ESTADO DA ARTE EM CLASSIFICAÇÃO TEXTUAL (Q.1)

Os artigos [Howard and Ruder 2018], [Devlin et al. 2019] e [Yang et al. 2019], selecionados com base no ranking do NLP-Progress, apresentam os modelos de linguagem: ULMFiT (Universal Language Model Fine-tuning for Text Classification), BERT (Bidirectional Encoder Representations from Transformers) e XLNet, respectivamente.

Os trabalhos de [Howard and Ruder 2018], [Devlin et al. 2019] e [Yang et al. 2019], apresentam propostas de arquiteturas e algoritmos de treinamento para a geração de modelos de linguagem com transferência de aprendizado, representando o estado da arte na data de suas publicações, em ordem cronológica: [Howard and Ruder 2018], [Devlin et al. 2019], e [Yang et al. 2019]. O estudo realizado em [Howard and Ruder 2018] foca na classificação de textos e como executar o ajuste fino para essa tarefa no domínio do problema alvo, aplicando redes neurais recorrentes AWD-LSTM.

Nos modelos BERT e XLNet, são utilizadas técnicas embasadas em Autoencoders e mecanismos de atenção [Vaswani et al. 2017], e Transformers-XL [Dai et al. 2019], respectivamente. Esses modelos se mostraram mais precisos em tarefas de classificação textual quando comparados ao ULMFiT, como mostram os resultados apresentados na Tabela II, transcrita integralmente de [Yang et al. 2019], que compara resultados (taxa de erros) de classificação textual dos 3 modelos de linguagem em datasets de benchmark.

Modelo	IMDB	Yelp-2 <sup>3</sup>	Yelp-5	DBPedia <sup>4</sup>	AG <sup>5</sup>	Amazon-2	Amazon-5
ULMFiT	4.6	2.16	29.98	0.80	5.01	-	-
BERT	4.51	1.89	29.32	0.64	-	2.63	34.17
XLNet	<b>3.20</b>	<b>1.37</b>	<b>27.05</b>	<b>0.60</b>	<b>4.45</b>	<b>2.11</b>	<b>31.67</b>

Table II. Comparação dos resultados dos modelos estado da arte em tarefas de classificação em termos de taxa de erro de classificação. Fonte: [Yang et al. 2019].

Para os trabalhos [Howard and Ruder 2018] e [Devlin et al. 2019] além dos testes com ajuste fino do modelo, foram realizados também experimentos para avaliar a possibilidade de aplicação deles como extratores de características dos textos sem prévio ajuste para o domínio da tarefa fim. Os testes mostraram bons resultados quando comparados com os modelos que receberam o ajuste no domínio do problema, conforme dados da Tabela III.

Modelo	Dataset	Com Ajuste	Sem Ajuste
ULMFiT	IMDb	5,00 (Taxa de Erro)	6,99 (Taxa de erro)
BERT	CoNLL-2003	96,4 (F1 - Score)	95,5 (F1 - Score)

Table III. Comparativo entre a utilização ou não do ajuste fino do modelo no domínio do problema. Fonte. O autor.

Como pode ser visto na Seção 5, a maioria dos trabalhos encontrados que aplicam modelos de linguagem para Classificação Textual no domínio jurídico, obtiveram os melhores resultados com o BERT. O XLNet, provavelmente pelo fato de ser recente, foi utilizado como teste apenas em [Shaheen et al. 2020], ainda assim com restrições devido ao custo computacional do modelo.

Nesse sentido, o trabalho de [Sun et al. 2019] faz um estudo comparativo de como utilizar o modelo para a tarefa de classificação textual, analisando técnicas diferentes na implementação do ajuste fino e pré-treinamento nos dados do domínio do problema, concluindo que as 4 últimas das 12 camadas

<sup>3</sup><https://www.yelp.com/dataset>

<sup>4</sup><https://wiki.dbpedia.org/>

<sup>5</sup>[http://groups.di.unipi.it/~gulli/AG\\_corpus\\_of\\_news\\_articles.html](http://groups.di.unipi.it/~gulli/AG_corpus_of_news_articles.html)

escondidas do modelo BERT estudado, são mais interessantes para essa tarefa, e ainda que o pré-treinamento nos dados do problema melhora a performance do modelo.

#### 4. MODELOS PRÉ-TREINADOS EM PORTUGUÊS DO BRASIL (Q.5)

Apenas dois trabalhos, [Hartmann et al. 2017] e [Souza et al. 2020], apresentam modelos de linguagens pré-treinadas para o Português do Brasil. Em [Hartmann et al. 2017], esses modelos são técnicas de *word embeddings*, tais como Word2Vec, FastText, Glove e Wang2Vec. Em [Souza et al. 2020], o modelo pré-treinado é o BERT.

Após os treinamentos, ambos estudos disponibilizaram para a comunidade todos os modelos gerados [Hartmann et al. 2017]<sup>6</sup> e [Souza et al. 2020]<sup>7</sup>. A disponibilização desses modelos pré-treinados são importantes pois o custo computacional para o treinamento deles do zero poderia inviabilizar muitas pesquisas. O que confirma essa importância, é que trabalhos analisados na Seção 6, especialmente [Silva and Maia 2020] e [Mota et al. 2020], lançam mão dos modelos de *word embedding* gerados em [Hartmann et al. 2017] dentro da arquitetura de Classificação Textual aplicada.

#### 5. CLASSIFICAÇÃO TEXTUAL NO DOMÍNIO JURÍDICO COM MODELOS DE LINGUAGENS (Q.2)

Dos sete trabalhos estudados nessa categoria, verifica-se que a maioria ([Song et al. 2021],[Wang et al. 2020],[Chalkidis et al. 2020],[Shaheen et al. 2020] e [Chalkidis et al. 2019]) utiliza o modelo BERT, e os outros dois [Soh et al. 2019] e [Campos et al. 2020], tem como principal modelo o ULMFiT, todos alcançando F1-score acima de 0.8. Dentre os trabalhos, apenas [Wang et al. 2020] e [Campos et al. 2020] são em língua portuguesa do Brasil, os demais em língua inglesa, e um deles [Shaheen et al. 2020], apresenta um modelo multilingual que, além do inglês, inclui o francês e o alemão.

Em termos numéricos, e considerando apenas o índice de F1-score daqueles estudos que utilizaram o dataset EUROLEX57K, e o modelo BERT, o melhor resultado foi alcançado pelo modelo proposto em [Chalkidis et al. 2020], que realizou, além do treinamento no domínio global, pré-treinamento em dados do mesmo domínio do problema a partir de uma base de 12 GB de documentos legais em inglês, e ainda nos dados específicos do problema. Esse diferencial no ajuste possivelmente explica o melhor resultado obtido, conforme observa-se na Tabela IV. A coluna "Dimensão do Vetor", indica o tamanho do vetor que representa o documento em cada arquitetura, sendo esse apresentado para a camada de classificação.

Artigo	Dimensão do Vetor	Ajuste do Modelo	F1-score
[Song et al. 2021]	1024	Específico	0,745
[Chalkidis et al. 2020]	768	Geral + Domínio + Específico	<b>0,824</b>
[Shaheen et al. 2020]	1024	Geral + Específico	0,758

Table IV. Comparativo de índice F1-score dos artigos que utilizaram dataset EUROLEX57K e modelo BERT. Fonte: O autor.

##### 5.1 Testes com arquiteturas diferentes de um mesmo modelo

Dentre aqueles trabalhos que implementaram esses testes, verifica-se na Tabela V que o ULMFiT teve o pior resultado quando compara-se os modelos ajustados ou não para os dados do problema, o que indica a grande sensibilidade dele em relação à esse ajuste.

<sup>6</sup><http://nilc.icmc.usp.br/nilc/index.php/repositorio-de-word-embeddings-do-nilc>

<sup>7</sup><https://github.com/neuralmind-ai/portuguese-bert>

Artigo	Modelo	Com Ajuste	Sem Ajuste
[Song et al. 2021]	BERT	0,8030	0,7900
[Chalkidis et al. 2020]	BERT	0,8240	0,8320
[Campos et al. 2020]	<b>ULMFiT</b>	<b>0,8974</b>	<b>0,4724</b>

Table V. Compara F1-score dos modelos BERT e ULMFiT aplicados com e sem o ajuste do modelo de linguagem nos dados do problema.

## 6. CLASSIFICAÇÃO TEXTUAL NO DOMÍNIO JURÍDICO EM PORTUGUÊS DO BRASIL (Q.3 E Q.4)

Os estudos em [Noguti et al. 2020], [Luz de Araujo et al. 2020], [Silva et al. 2018], [Silva and Maia 2020], [Dal Pont et al. 2020], [Wang et al. 2020], [Bertalan and Ruiz 2020], e [Mota et al. 2020], mostram a aplicação de modelos de aprendizagem profunda, clássicos e estatísticos para classificação de documentos no contexto jurídico do Brasil. O artigo [Campos et al. 2020], como mostrado na Seção 5, também faz um estudo no contexto jurídico brasileiro, porém utiliza como proposta principal o ULMFiT. A Figura 4.(a) mostra a proporção de utilização de cada técnica.

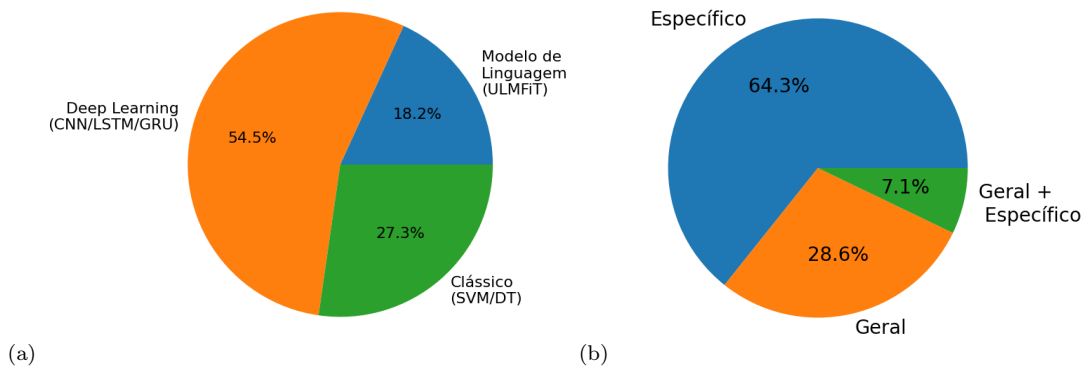


Fig. 4. (a) Proporção de cada estratégia utilizada nos classificadores dos estudos em português do Brasil. (b) Estratégia de treinamento dos modelos de linguagem.

Como regra, os trabalhos utilizaram modelos de aprendizagem profunda para a classificação, dentre eles: CNN e LSTM. Dois estudos, [Luz de Araujo et al. 2020] e [Bertalan and Ruiz 2020], obtiveram resultados de F1-score acima de 0,89 com XGBoost e Árvores de Decisão (Decision Tree - DT), respectivamente. Em relação ao [Luz de Araujo et al. 2020], os autores utilizaram o XGBoost para a tarefa de classificar processos, que possuem em média 47 páginas, pois em testes com outros classificadores obtiveram resultados inferiores, em virtude da quantidade de palavras dentro de um único processo, o que evidencia que o tamanho do documento ou texto de entrada, é fator importante no desempenho da arquitetura utilizada para a classificação textual.

### 6.1 Resultados na tarefa de classificação textual

Os trabalhos em língua portuguesa conseguiram excelentes resultados utilizando modelos clássicos e de aprendizagem profunda, normalmente acima de 0,90 em termos de F1-score. O que aponta para a eficiência dessas abordagens, principalmente quando é considerado o custo computacional comparado aos modelos de linguagem apresentados na Seção 3.

Em comparação com os trabalhos apresentados na Seção 5, que fazem uso do BERT e do ULMFiT, é preciso considerar que a maioria realiza a classificação multilabel em um dataset com pelo menos

66 classes, sendo que os trabalhos [Song et al. 2021], [Chalkidis et al. 2020] e [Shaheen et al. 2020], aplicam a classificação para 4.271 classes, e no caso dos estudos do Brasil aquele que apresenta mais classes é o [Luz de Araujo et al. 2020], com 29.

De qualquer modo, tanto em português do Brasil quanto em inglês, é notório o ganho de performance quando se ajusta o modelo, pré-treinado ou não, com os dados do problema ou outros do mesmo domínio de conhecimento. A Figura 4.(b) reforça essa afirmação, em que observa-se mais de 70% dos artigos analisados utilizam apenas os dados do domínio do problema, ou uma combinação desse com corpus do domínio geral.

## 6.2 Repositórios públicos de dados no domínio jurídico no Brasil

Como resposta a quinta questão de revisão, o trabalho [Luz de Araujo et al. 2020] é o único que disponibiliza Dataset do contexto jurídico para o público. É um conjunto de dados bem amplo, com 45.532 Recursos extraordinários do Supremo Tribunal Federal (STF), sendo que dentro de cada um desses processos há diversos documentos, totalizando 692.966. Desses processos, 44.855 estão rotulados, bem como seus 628.820 documentos. Essa disponibilização de datasets no domínio jurídico em português do Brasil é fundamental para as pesquisas na área.

## 7. LACUNAS

Após a análise dos trabalhos, especificamente em relação ao idioma português do Brasil, ficou claro que ainda falta um maior estudo da aplicação de modelos de linguagem estado da arte no contexto jurídico. Ainda que dois trabalhos, [Wang et al. 2020] e [Campos et al. 2020], lancem mão dessa técnica, apenas o último trata de classificação textual. Além disso, nenhum dos trabalhos selecionados que utilizaram *datasets* em português do Brasil, apresentou estudos de classificação *multilabel*, situação bem comum em documentos jurídicos.

## 8. CONCLUSÃO

As comparações mostram que há estudos no Brasil e na Europa visando a aplicação de PLN no contexto jurídico. Além disso, evidenciam a viabilidade técnica da aplicação, com resultados de F1-score não inferiores a 0.80 nos estudos comparados.

Em idioma inglês, notadamente em *datasets* do domínio jurídico Europeu, foi possível observar que há aplicações dos modelos de linguagens propostos em [Howard and Ruder 2018] e [Devlin et al. 2019], enquanto nos trabalhos com *datasets* em idioma português do Brasil isso ainda está restrito a apenas dois artigos dos nove selecionados nessa categoria.

Portanto, é possível concluir a partir do estudo e comparação dos artigos que: há a possibilidade de maior aplicação de modelos de linguagem em língua portuguesa e no domínio jurídico do Brasil; o ajuste dos modelos de linguagem em documentos do mesmo domínio do problema melhoram a performance do classificador; e por fim que os modelos de linguagem pré-treinados no domínio geral do idioma podem ser usados sem ajustes em tarefas de classificação.

## REFERENCES

- BERTALAN, V. G. F. AND RUIZ, E. Predicting judicial outcomes in the brazilian legal system using textual features. In *DHandNLP@PROPOR*, 2020\*.
- CAMPOS, T., LUZ DE ARAUJO, P. H., AND SOUSA, M. pp. 76–86. In , *Inferring the Source of Official Texts: Can SVM Beat ULMFiT?* pp. 76–86, 2020\*.
- CHALKIDIS, I., ANDROUTSOPOULOS, I., AND ALETRAS, N. Neural legal judgment prediction in English. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, pp. 4317–4323, 2019\*.

- CHALKIDIS, I., FERGADIOTIS, M., MALAKASIOTIS, P., ALETRAS, N., AND ANDROUTSOPOULOS, I. LEGAL-BERT: The muppets straight out of law school. In *Findings of the Association for Computational Linguistics: EMNLP 2020*. Association for Computational Linguistics, Online, pp. 2898–2904, 2020\*.
- DAI, Z., YANG, Z., YANG, Y., CARBONELL, J., LE, Q., AND SALAKHUTDINOV, R. Transformer-XL: Attentive language models beyond a fixed-length context. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, pp. 2978–2988, 2019.
- DAL PONT, T. R., SABO, I. C., HÜBNER, J. F., AND ROVER, A. J. Impact of text specificity and size on word embeddings performance: An empirical evaluation in brazilian legal domain. In *Intelligent Systems*, R. Cerri and R. C. Prati (Eds.). Springer International Publishing, Cham, pp. 521–535, 2020\*.
- DEVLIN, J., CHANG, M.-W., LEE, K., AND TOUTANOVA, K. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, Minneapolis, Minnesota, pp. 4171–4186, 2019\*.
- HARTMANN, N. S., FONSECA, E. R., SHULBY, C. D., TREVISI, M. V., RODRIGUES, J. S., AND ALUÍSIO, S. M. Portuguese word embeddings evaluating on word analogies and natural language tasks. In *Anais do XI Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*. SBC, Porto Alegre, RS, Brasil, pp. 122–131, 2017\*.
- HOWARD, J. AND RUDER, S. Universal language model fine-tuning for text classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Melbourne, Australia, pp. 328–339, 2018\*.
- LUZ DE ARAUJO, P. H., DE CAMPOS, T. E., ATAÍDES BRAZ, F., AND CORREIA DA SILVA, N. VICTOR: a dataset for Brazilian legal documents classification. In *Proceedings of the 12th Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, pp. 1449–1458, 2020\*.
- MOTA, C., LIMA, A., NASCIMENTO, A., MIRANDA, P., AND DE MELLO, R. Classificação de páginas de petições iniciais utilizando redes neurais convolucionais multimodais. In *Anais do XVII Encontro Nacional de Inteligência Artificial e Computacional*. SBC, Porto Alegre, RS, Brasil, pp. 318–329, 2020\*.
- NOGUTI, M. Y., VELLASQUES, E., AND OLIVEIRA, L. S. Legal document classification: An application to law area prediction of petitions to public prosecution service. In *2020 International Joint Conference on Neural Networks (IJCNN)*. pp. 1–8, 2020\*.
- SHAHEEN, Z., WOHLGENANT, G., AND FILTZ, E. Large scale legal text classification using transformer models. *Computer Science ArXiv* vol. abs/2010.12871, 2020\*.
- SILVA, A. C. AND MAIA, L. C. G. The use of machine learning in the classification of electronic lawsuits: An application in the court of justice of minas gerais. In *Intelligent Systems*, R. Cerri and R. C. Prati (Eds.). Springer International Publishing, Cham, pp. 606–620, 2020\*.
- SILVA, N., BRAZ, F., AND DE CAMPOS, T. Document type classification for brazil’s supreme court using a convolutional neural network. pp. 7–11, 2018\*.
- SOH, J., LIM, H. K., AND CHAI, I. E. Legal area classification: A comparative study of text classifiers on Singapore Supreme Court judgments. In *Proceedings of the Natural Legal Language Processing Workshop 2019*. Association for Computational Linguistics, Minneapolis, Minnesota, pp. 67–77, 2019\*.
- SONG, D., VOLD, A., MADAN, K., AND SCHILDER, F. Multi-label legal document classification: A deep learning-based approach with label-attention and domain-specific pre-training. *Information Systems*, 2021\*.
- SOUZA, F., NOGUEIRA, R., AND LOTUFO, R. Bertimbau: Pretrained bert models for brazilian portuguese. In *Intelligent Systems*, R. Cerri and R. C. Prati (Eds.). Springer International Publishing, Cham, pp. 403–417, 2020\*.
- SUN, C., QIU, X., XU, Y., AND HUANG, X. How to fine-tune bert for text classification? In *Chinese Computational Linguistics*, M. Sun, X. Huang, H. Ji, Z. Liu, and Y. Liu (Eds.). Springer International Publishing, Cham, pp. 194–206, 2019\*.
- VASWANI, A., SHAZEER, N., PARMAR, N., USZKOREIT, J., JONES, L., GOMEZ, A. N., KAISER, L. U., AND POLOSUKHIN, I. Attention is all you need. In *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Vol. 30. Curran Associates, Inc., 2017.
- WANG, Z., WU, Y., LEI, P., AND PENG, C. Named entity recognition method of brazilian legal text based on pre-training model. *Journal of Physics: Conference Series* vol. 1550, pp. 032149, 05, 2020\*.
- YANG, Z., DAI, Z., YANG, Y., CARBONELL, J., SALAKHUTDINOV, R. R., AND LE, Q. V. Xlnet: Generalized autoregressive pretraining for language understanding. In *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox, and R. Garnett (Eds.). Vol. 32. Curran Associates, Inc., 2019\*.