

PET-IA: Um Serviço de previsão de petições com Inteligência Artificial na Procuradoria - Geral da Fazenda Nacional

PET-IA: An Artificial Intelligence-based Petition Prediction Service at the National Treasury Attorney's Office

Renata de Castro Vianna
Fundação de Apoio à Pesquisa do Distrito Federal
(FAPDF)
Brasília - DF, Brasil
viannapradoadv@gmail.com

Gilmar dos Santos Marques
Fundação de Apoio à Pesquisa do Distrito Federal
(FAPDF)
Brasília - DF, Brasil
gilmar.marx@gmail.com

Geraldo Pereira Rocha Filho
Departamento de Ciências Exatas e Tecnológicas
Universidade Estadual do Sudoeste da Bahia (UESB)
Vitória da Conquista - BA, Brasil
geraldo.rocha@uesb.edu.br

Daniel de Saboia Xavier
Procuradoria Geral da Fazenda Nacional - (PGFN)
Brasília - DF, Brasil
daniel.saboia@pgfn.gov.br

Flávio Garcia Praciano
Departamento de Engenharia Elétrica - ENE
Universidade de Brasília (UnB)
Brasília - DF, Brasil
flavio.praciano@redes.unb.br

Guilherme Batista Meneses Alves
Departamento de Engenharia Elétrica - ENE
Universidade de Brasília (UnB)
Brasília - DF, Brasil
guilherme.alves@redes.unb.br

Edna Dias Canedo
Departamento de Ciência da Computação - CIC
Universidade de Brasília (UnB)
Brasília - DF, Brasil
ednacanedo@unb.br

Resumo—Assim como diversas outras áreas do conhecimento, a esfera jurídica enfrenta um cenário de saturação de informações. No entanto, a extração de dados de documentos legais e a abordagem de seus dilemas éticos constituem um dos desafios neste campo, devido à complexidade dos conceitos e terminologia jurídica. Além disso, em virtude do imenso volume de dados gerados por esses domínios, inúmeras investigações estão em curso visando à automatização da classificação de textos e à previsão de suas implicações. Frente a esse desafio, este trabalho propõe o PET-IA, um serviço de previsão de PETições com Inteligência Artificial para realizar o processo de tomada de decisões com base em dados de texto curto. Para uma análise mais precisa do texto jurídico, foi utilizado a arquitetura do

BERT e o dataset LeNER-Br no serviço proposto. Com essa estratégia, o PET-IA adquire os tokens providos do BERT como entrada no *random forest* para predizer as petições em documentos legais. Quando comparado com os algoritmos da literatura, o PET-IA demonstrou eficiência na realização de previsões de petições, alcançando um desempenho eficiente no tempo de predição que supera em seis vezes o melhor caso identificado.

Palavras-chave — *Automatização Jurídica, Inteligência Artificial, Análise de Texto Legal*

Abstract - Just like many other fields of knowledge, the legal sphere faces a scenario of information overload. However, the extraction

of data from legal documents and addressing their ethical dilemmas constitute one of the challenges in this field, due to the complexity of legal concepts and terminology. Furthermore, due to the immense volume of data generated in these domains, numerous investigations are underway aiming at the automation of text classification and the prediction of their implications. In response to this challenge, this work proposes PET-IA, a service for predicting petitions with Artificial Intelligence to carry out the decision-making process based on short text data. For a more precise analysis of legal text, the BERT architecture and the LeNER-Br dataset were used in the proposed service. With this strategy, PET-IA acquires tokens provided by BERT as input into a random forest to predict petitions in legal documents. When compared to algorithms in the literature, PET-IA demonstrated efficiency in making petition predictions, achieving a performance in prediction time that is six times better than the best case identified.

Keywords— *Legal Automation, Artificial Intelligence, Legal Text Analysis*

I. INTRODUÇÃO

Em um contexto marcado pela sobrecarga informacional, as pesquisas direcionadas à extração e categorização de documentos legais desempenham um papel crucial, com a atenção dedicada à análise de textos jurídicos aumentando significativamente [1],[2]. Esse interesse acompanha o contínuo avanço das técnicas de classificação textual ao longo do tempo, buscando enfrentar a complexidade intrínseca à linguagem humana [3],[4],[5]. Nesse panorama, os algoritmos de Inteligência Artificial (IA), em conjunto com os progressos alcançados no campo do Processamento de Linguagem Natural (PLN), emergem como soluções promissoras para abordar o desafio da extração de dados de documentos legais, visando automatizar a classificação de textos curtos e antecipar suas implicações de forma eficiente.

Nesse contexto, a inserção da IA na Procuradoria Geral da Fazenda Nacional (PGFN) é uma mudança significativa no cenário fiscal brasileiro. Diante da crescente complexidade das questões tributárias e da necessidade de proteger os interesses financeiros da União, a busca por inovações visa aprimorar a eficiência da PGFN. Neste contexto, a IA emerge como ferramenta indispensável, capacitando o órgão a enfrentar desafios fiscais com agilidade e sofisticação tecnológica. Essa integração não só revoluciona as operações diárias, mas promete uma gestão mais eficaz e transparente dos recursos públicos, gerando benefícios diretos para a sociedade brasileira.

Extrair dados de documentos legais é um desafio multifacetado, uma vez que a diversidade da linguagem jurídica e a variação no conteúdo de cada caso complicam a padronização documental [2],[6]. Além disso, a evolução constante dos conceitos jurídicos exige uma manutenção onerosa das bases de conhecimento, e o uso de termos específicos pode dificultar o reconhecimento pelos métodos tradicionais [7],[8], [2]. Adicionalmente, o volume significativo de documentos judiciais ultrapassa a capacidade humana para coletar e processar informações manualmente [9], [8], [2].

Com o objetivo de ir além do estado da arte, este trabalho propõe o PET-IA, um serviço de previsão de PETições com Inteligência Artificial, destinado a conduzir o processo de tomada de decisões com base em dados de texto curto. Para alcançar esse objetivo, o PET-IA é modelado a partir da integração entre a arquitetura BERT e o *dataset* LeNER-Br para

aprimorar a confiabilidade na classificação de textos curtos através da saída dos seus tokens. Com essa estratégia, o PET-IA adquire os tokens do BERT que são utilizados como entrada no *random forest*, possibilitando a predição de petições em documentos legais. Os resultados destacam a eficácia do PET-IA na execução de previsões de petições, assegurando uma resposta rápida no tempo de predição quando comparado com outros algoritmos da literatura.

O restante deste trabalho está estruturado da seguinte forma. A Seção II apresenta os trabalhos relacionados, discutindo os desafios e lacunas para esta pesquisa. A Seção III apresenta como o PET-IA foi modelado e desenvolvido. A Seção IV apresenta os resultados obtidos para validar o PET-IA. Por fim, a Seção V apresenta as conclusões e trabalhos futuros.

II. TRABALHOS RELACIONADOS

Várias soluções foram propostas para resolver o problema de extração de dados de documentos legais [2], [10], [11], [12], [13], [14]. A maioria destas soluções foram desenvolvidas considerando padrões pré-definidos e sem considerar a evolução constante e dinâmica dos conceitos jurídicos. Stephen Few [15] destaca a importância crucial da compreensão profunda dos dados quantitativos, enfatizando habilidades interdisciplinares essenciais. Esse entendimento demanda sólido conhecimento na área relevante, junto ao pensamento crítico, científico e domínio estatístico. Essas competências formam a base para análise e interpretação de dados quantitativos [16], características desta pesquisa. No entanto, na área de documentos legais, os trabalhos não abordam diretamente o papel do pensamento ético na tomada de decisões.

Barros em seu estudo [10], realizou uma pesquisa focada na análise de documentos judiciais na jurisprudência trabalhista. Para tanto, utilizou-se da mineração de dados para avaliar o posicionamento da turma de julgamento, categorizando as decisões quanto à parte favorecida (*i.e.*, empregador ou empregado). Inicialmente, documentos com recursos de ambas as partes foram excluídos por inconclusividade. Os acórdãos judiciais escolhidos passaram por anotações detalhadas e alimentaram um algoritmo de aprendizado de máquina baseado em redes Bayesianas. Além disso, as decisões foram processadas, extraindo características relevantes com o índice TF-IDF e as Redes Bayesianas. No entanto, as redes Bayesianas dependem da disponibilidade e qualidade das informações prévias, ao contrário de nossa abordagem, que consegue lidar com conjuntos de dados grandes e complexos, mantendo ao tempo alta eficiência na predição da petição.

Abordando a crescente problemática da ampla disponibilidade de textos jurídicos não estruturados, [11] propõem uma metodologia de mineração de texto baseada em diversos mecanismos de classificação. O objetivo central consiste na avaliação do método mais adequado para lidar com um extenso conjunto de dados diversificados de jurisprudências. A hipótese central desses autores fundamenta-se na capacidade das técnicas de aprendizado de máquina supervisionado em apresentar informações relevantes sem a necessidade de extensa leitura manual. No entanto, além do desafio de lidar com a grande quantidade de dados, é essencial considerar o tempo computacional necessário para treinar cada modelo, visto que um treinamento excessivamente longo pode inviabilizar sua aplicação prática. Os resultados indicam um

desempenho sólido, porém a aplicação em cenários diferentes ou com conjuntos de dados distintos pode exigir ajustes nos processos de execução. Vale ressaltar que o aumento do tamanho da amostra nem sempre resulta em melhorias nos resultados, pois pode afetar negativamente a eficácia dos algoritmos, demandando uma nova parametrização.

No estudo de Dantas et al. [12], a classificação de textos em português sobre dúvidas tributárias no Estado do Rio de Janeiro foi abordada por meio de duas estratégias: uma direta e outra hierárquica, utilizando os algoritmos SVM e LSTM. Devido à escassez de dados em determinadas categorias, foi necessária a redução do número delas, impulsionando esforços para a expansão do conjunto de documentos. A ausência de dúvidas em categorias específicas demandou a redução total dessas categorias, estimulando a busca por métodos de ampliação do corpus documental disponível. Os resultados evidenciaram que o SVM se destacou na generalização e no tratamento de bases desequilibradas, enquanto o LSTM, apesar de apresentar boa acurácia global, enfrentou desafios com categorias semelhantes e baixa quantidade de dados. Isso enfatiza a necessidade de revisão das categorias de assuntos, característica presente nesta pesquisa. Ainda, tanto a LSTM quanto a SVM são sensíveis a conjuntos de dados desbalanceados, diferentemente desta pesquisa que pode lidar melhor com desequilíbrios, mantendo um desempenho mais estável em classes minoritárias.

III. PET-IA: UM SERVIÇO DE PREVISÃO DE PETIÇÕES VIA IA

Esta seção descreve como o PET-IA foi modelado para solucionar o problema de extração de dados em documentos legais, como apresentado na Figura 1. Utilizando o BERT e o LeNER-Br para aprimorar a compreensão e o processamento de texto no domínio jurídico em português, o PET-IA faz uso da representação contextual dos dados fornecida pelo BERT, que é então processada pelo *random forest* para conduzir o processo de predição. O PET-IA tem como objetivo central não apenas otimizar a eficiência na previsão durante a categorização de textos curtos, mas também facilitar sua aplicação na extração de dados em documentos legais de maneira eficaz. Para uma melhor compreensão do serviço proposto, a seguir será o ambiente de armazenamento, seguido pelo ambiente de decisão e do mecanismo de predição.

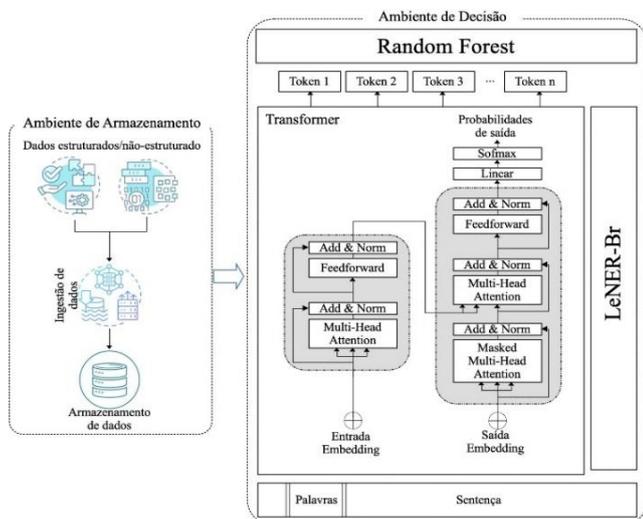


Figura 1: Cenário de funcionamento do PET-IA.

A. Ambiente de Armazenamento

Para extrair os dados da plataforma de Processo Judicial Eletrônico (PJE) (Rótulo A, Figura 2) no PET-IA e, posteriormente, armazená-los em um ambiente seguro, foi necessário realizar um *scraping*. Diante disso, foi utilizado o Selenium, uma ferramenta de automação, para automatizar o acesso à plataforma e realizar o download das petições judiciais. As petições judiciais obtidas são arquivadas no sistema de arquivos do servidor Selenium (Rótulo B, Figura 2), garantindo organização e segurança para processamento subsequente. O servidor Selenium realiza o processamento das petições armazenadas, iniciando a extração de texto. As petições são submetidas a um processo de extração de texto da plataforma eletrônica de maneira direta, isto é, são identificados e recuperados os conteúdos textuais. Dada a natureza sensível das informações jurídicas, a confidencialidade é prioritária. O acesso aos dados é estritamente controlado, restrito a pessoal autorizado submetido a rigorosas verificações de identidade e treinamento em protocolos de segurança.

O servidor de armazenamento (Rótulo C, Figura 2) dedicado desempenha um papel crucial, fornecendo um ambiente altamente seguro para a preservação das petições e documentos, assegurando integridade e confidencialidade. Além das restrições de acesso, são implementadas medidas de segurança proativas, incluindo *backups* regulares e prevenção contra ameaças cibernéticas, visando mitigar riscos. A equipe responsável pela gestão dessas informações passa por treinamento contínuo para aderir às melhores práticas de segurança de dados.

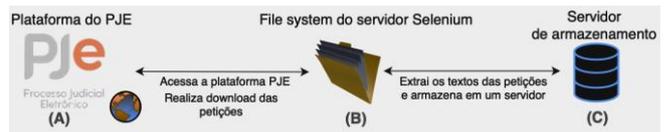


Figura 2: Extração e armazenamento automático de dados do PJE.

Essa estratégia integrada busca estabelecer um ambiente confiável e seguro para o armazenamento e gerenciamento de documentos provenientes de petições, mantendo padrões elevados de segurança no contexto jurídico.

B. Ambiente de Decisão

Contribuir para uma análise mais precisa do texto jurídico, utilizamos o BERT [17] e o LeNER-Br [18] no PET-IA. É importante destacar que o LeNER-Br foi empregado para lidar com a natureza singular dos textos jurídicos em língua portuguesa. O LeNER-Br é um conjunto de dados composto por documentos legais de diversos tribunais brasileiros, incluindo cortes superiores e estaduais, como o Supremo Tribunal Federal, o Superior Tribunal de Justiça, o Tribunal de Justiça de Minas Gerais e o Tribunal de Contas da União.

Neste estudo, o BERT e seu pré-treinamento foram baseados em [17]. Portanto, nosso BERT é um modelo de rede neural com múltiplas camadas de *transformers* bidirecionais, permitindo a captura de informações contextuais jurídicas em ambos os sentidos do texto. Nossa arquitetura é composta por camadas de codificadores que processam *tokens* de entrada. Durante o pré-treinamento, o BERT é exposto aos textos,

aprendendo representações ricas e contextuais de palavras e frases por meio de tarefas de previsão mascarada (*Masked Language Modeling* - MLM) e predição de sentenças seguintes (*Next Sentence Prediction* - NSP).

Na entrada, o BERT recebe o texto tokenizado, dividido em segmentos de tokens separados por um token especial [SEP] e precedido por um *token* [CLS], usado para representar toda a sequência. Cada *token* é codificado em uma representação vetorial que incorpora informações contextuais das camadas anteriores, permitindo ao modelo compreender e representar o contexto das palavras no texto. Nesse processo, marcadores especiais, como [CLS] e [SEP], são adicionados no início e no final do texto, respectivamente, antes de sua divisão em tokens de subpalavras. A saída do BERT consiste em representações contextualizadas de cada token de entrada. Para tarefas de predição em textos jurídicos, essas representações serão empregadas no mecanismo de predição que será apresentado a seguir.

C. Mecanismo de Predição

O objetivo do mecanismo de predição é analisar as petições dentro do texto jurídico do PJE para realizar previsões. Optou-se por utilizar o *random forest* devido à sua eficácia em lidar com conjuntos de dados complexos e variados, reconhecido por sua capacidade de produzir resultados precisos em tarefas de predição. Além disso, a randomização desempenha um papel fundamental no *random forest*, tanto na seleção dos subconjuntos de dados de treinamento quanto na escolha das características (palavras ou atributos) usadas para construir cada árvore de decisão. Essa abordagem de randomização desempenha um papel crucial na mitigação do *overfitting* e na melhoria da capacidade de generalização do modelo.

O Algoritmo 1 apresenta o mecanismo de predição aplicado à análise de textos jurídicos, combinando o poder do modelo BERT e LeNER-Br. Inicialmente, o processo inicia-se dividindo o conjunto de dados rotulados X e y em conjuntos de treino (X_{train} , y_{train}) e teste (X_{test} , y_{test}). Durante a fase de treinamento, a função *TrainTree* é utilizada para construir o modelo de *random forest*. Para cada árvore i no modelo, é realizada uma amostragem aleatória (X_{sample} , y_{sample}) a partir do conjunto de treino. Cada árvore é então treinada com as características extraídas (isto é, os tokens) da saída do BERT, representadas por X_{sample} . Esses tokens provenientes do BERT são incorporadas no treinamento de cada árvore individualmente, proporcionando um contexto detalhado para a tomada de decisão.

Algoritmo 1 Mecanismo de predição das petições dos textos jurídicos

Inicialização:

```
1: Dividir  $X$  e  $y$  em conjuntos de treino ( $X_{train}$ ,  $y_{train}$ ) e teste ( $X_{test}$ ,  $y_{test}$ )
2:  $n \leftarrow$  número de árvores no random forest
3: forest  $\leftarrow$  []
4: function TrainTree( $X_{train}$ ,  $y_{train}$ )
5:   Para cada árvore  $i$  no modelo:
6:     tree  $\leftarrow$  Árvore de Decisão
7:     Amostrar aleatoriamente ( $X_{sample}$ ,  $y_{sample}$ ) de ( $X_{train}$ ,  $y_{train}$ )
8:     Treinar tree com as características  $X_{sample}$  da saída do BERT
```

```
9:   Adicionar tree a forest
10: Retornar forest
11: end function
Treinamento do random forest para predição:
1: forest  $\leftarrow$  TrainTree( $X_{train}$ ,  $y_{train}$ )
1: function Predict( $X_{test}$ )
2:   predictions  $\leftarrow$  []
3:   Para cada texto  $x$  em  $X_{test}$ :
4:     tree_predictions  $\leftarrow$  []
5:     Para cada árvore tree em forest:
6:       tree_predictions.append(tree.predict( $x$ ))
7:     final_prediction  $\leftarrow$  MajorityVote(tree_predictions)
8:     Adicionar final_prediction a predictions
9:   Retornar predictions
10: end function
```

Posteriormente, durante a etapa de predição, a função *Predict* é empregada para prever os rótulos dos textos de teste. Para cada texto x em X_{test} , as previsões de cada árvore do modelo são agregadas utilizando um esquema de votação majoritária. Esse processo garante uma predição final baseada nas decisões coletivas das árvores do *random forest*, resultando em previsões mais robustas e contextualmente informadas para os textos jurídicos.

IV. AVALIAÇÃO DE DESEMPENHO

Esta seção valida o desempenho do PET-IA comparando-o com três algoritmos da literatura, os quais são: (i) *Random forest* - RF, (ii) Rede neural artificial do tipo multicamadas - RNA-MLP; e (iii) Gradiente descendente estocástico - SGD. O cenário de avaliação, as métricas utilizadas e os resultados alcançados serão apresentados nas próximas subseções.

A. Cenário de Avaliação

Para validar o serviço proposto, modelamos um *dataset* desenvolvido a partir das petições eletrônicas arquivadas nos sistemas da Justiça Federal, abrangendo seis tribunais regionais no Brasil. O *dataset* inclui informações essenciais de cada petição, como classe processual (Mandado de Segurança ou Procedimento Comum) e região de origem (1 a 6). Ademais, o *dataset* incorpora dados relativos à classificação dos processos efetuada na PGFN, em que procuradores, atuando como servidores do órgão, são responsáveis pela categorização. O *dataset* possui aproximadamente 112 mil petições, distribuídas em 202 matérias distintas, organizadas hierarquicamente para facilitar a análise. Utilizou-se a técnica de *hold-out* no *dataset* alocando 10% para validação, 30% para teste e 60% para treinamento. É válido salientar que os algoritmos avaliados foram executados no mesmo cenário, com os melhores parâmetros de cada um, para obter uma comparação justa com a solução proposta.

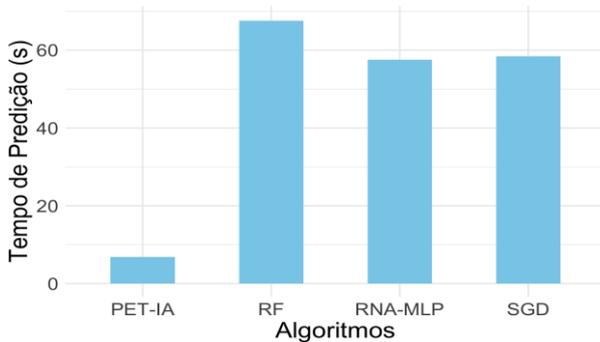
A eficácia dos algoritmos foi avaliada utilizando as seguintes métricas: (i) tempo de predição, (ii) tempo de treinamento, (iii) precisão, (iv) *hamming loss*, e (v) f1-score. Os experimentos foram realizados em um servidor com um processador Intel® Core™ i7 de 16 núcleos a 2.9 GHz, 32 GB de memória RAM, 81 GB de espaço livre em disco e 4 placas de vídeo NVIDIA Corporation GA104 [GeForce RTX 3060 Ti Lite Hash Rate]. O ambiente operacional utilizado foi o Ubuntu 23.04. A seguir será apresentado os impactos dos resultados obtidos.

B. Impacto dos Resultados Obtidos

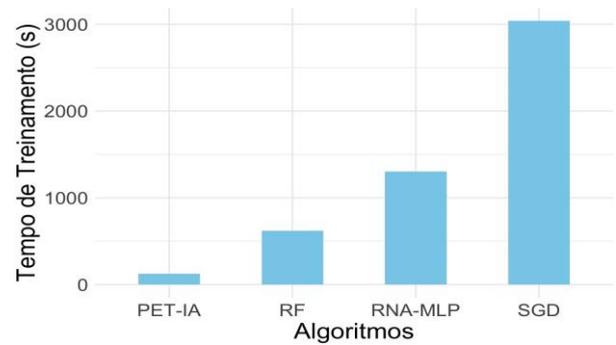
Na Figura 3, são apresentados os impactos no desempenho do PET-IA realizando-se uma análise comparativa entre o PET-IA, o RF, a RNA-MLP e o SGD. Destaca-se que o tempo de predição (Figura 3(a)) do PET-IA é em média oito vezes mais rápido em comparação com a RNA-MLP, que se configura como o segundo algoritmo de melhor desempenho. Isto faz sentido, pois o PET-IA explora as vantagens complementares do BERT para representação contextual e do RF para modelagem preditiva, resultando em uma execução mais eficiente e ágil em comparação com os outros algoritmos. Isso pode ser ratificado por meio da Figura 3(b), que apresenta o tempo de treinamento do PET-IA tendo uma eficiência de até seis vezes superior em comparação com o segundo melhor desempenho, representado pelo RF.

Após a análise do desempenho temporal, avaliou-se a efetividade do PET-IA no contexto da predição de petições. Observa-se que o PET-IA obteve uma precisão 54% superior ao SGD, enquanto apresentou desempenho semelhante ao RF e à RNA-MLP, como apresentado na Figura 3(c). Isso ocorre porque o PET-IA, ao combinar o BERT com o RF, consegue aprimorar a extração de características contextuais e lidar eficientemente com grandes conjuntos de dados. Como resultado, obtém-se uma modelagem mais precisa e com menor suscetibilidade ao *overfitting* em comparação com métodos convencionais, como o SGD.

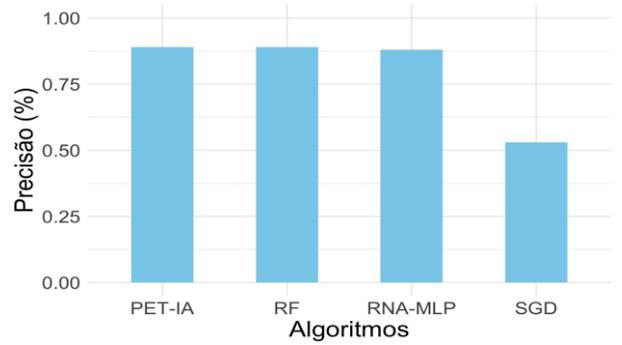
Ao analisar a métrica *Hamming Loss* (Figura 3(d)), que quantifica a fração de petições erroneamente previstas pelo modelo em relação ao total de previsões, observa-se a eficiência do PET-IA em comparação aos demais algoritmos. Saliencia-se que valores mais baixos do *Hamming Loss* indicam um desempenho mais preciso na classificação multirrotulo. Além disso, observa-se que o PET-IA mantém sua eficiência na predição, independentemente da disparidade entre as classes, como apresentado na Figura 3(e). Assim, o PET-IA demonstra habilidade em lidar com as complexidades inerentes à predição em textos legais, mantendo uma precisão considerável e proporcionando, adicionalmente, tempos de predição mais rápidos.



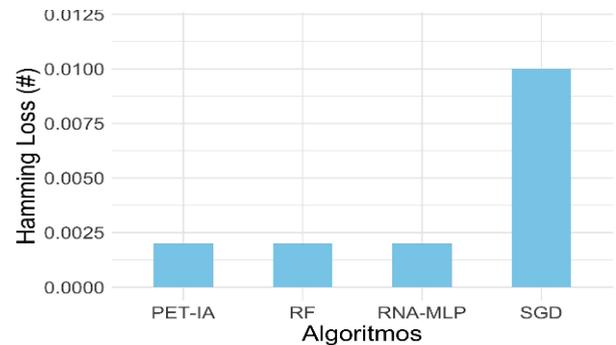
(a) Tempo de Predição.



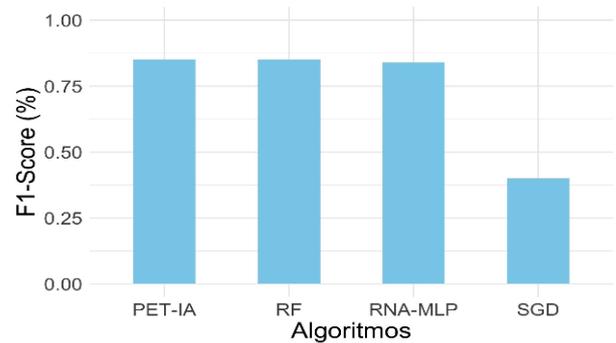
(b) Tempo de treinamento.



(c) Precisão.



(d) *Hamming Loss*.



(e) F1-Score.

Figura 3: Impacto de desempenho do PET-IA em relação às métricas tempos de predição, treinamento, precisão, *Hamming Loss* e F1-Score.

V. CONCLUSÃO E TRABALHOS FUTUROS

Este artigo apresentou o PET-IA, um serviço de previsão de petições por meio de inteligência artificial, projetado para conduzir o processo de tomada de decisões com base no problema de extração de dados de documentos legais. O

objetivo fundamental do PET-IA foi além de simplesmente oferecer eficácia na previsão durante a categorização de textos breves; ele visou habilitar sua aplicação efetiva na extração de dados contidos em documentos legais. Os resultados destacaram a efetividade do PET-IA ao realizar previsões de petições, proporcionando uma resposta rápida no tempo de predição em comparação com outros algoritmos mencionados na literatura.

Como perspectiva para trabalhos futuros, planeja-se ampliar este estudo, aplicando o PET-IA em dados locais e desenvolvendo um mecanismo para detecção de anomalias em documentos legais jurídicos.

VI. ACKNOWLEDGEMENT

Os autores agradecem o apoio técnico e computacional do Laboratório LATITUDE, da Universidade de Brasília ao TED (nº PGFN 23106.148934/2019- 67), apoiado pela Procuradoria Geral da Fazenda Nacional, um agradecimento especial para Thiago Oliveira Hoerlle - Auditor Federal de Finanças e Controle – Coordenação Geral de TI. Ao TED da Advocacia Geral da União (nº AGU 697.935/2019), ao Projeto SISTER City –Sistemas Inteligentes Seguros e em Tempo Efetivo Real para Cidades Inteligentes (Outorga 625/2022), ao Projeto “Sistema de Controle e Unificação de Projetos para o Governo Distrito Federal – Sispro-DF” (Outorga 497/2023), ao Decanato de Pesquisa e Inovação – DPI/UnB e a Fundação de Apoio a Pesquisa do DF - FAP/DF.

REFERÊNCIAS BIBLIOGRÁFICA

- [1] K. van Noordwijk and K. van Noordwijk, “Automatic document classification in integrated legal content collections,” 2017.
- [2] L. Enamoto, A. R. Santos, R. Maia, L. Weigang, and G. P. R. Filho, “Multi-label legal text classification with bilstm and attention,” *International Journal of Computer Applications in Technology*, vol. 68, no. 4, pp. 369–378, 2022.
- [3] L. Enamoto, L. Weigang, and G. P. R. Filho, “Generic framework for multilingual short text categorization using convolutional neural network,” *Multimedia Tools and Applications*, vol. 80, pp. 13475–13490, 2021.
- [4] A. C. Godinho, C. M. Nunes, R. I. Meneguette, V. P. Gonçalves, M. A. M. de Oliveira, J. C. Oliveira, and G. P. Rocha Filho, “Stalla: Um framework para análise de fontes abertas durante a pandemia do covid-19,” in *Anais do VII Workshop de Computação Urbana*, pp. 54–67, SBC, 2023.
- [5] M. Schmitz, R. Immich, G. Pessin, and G. P. Rocha Filho, “Towards the categorization of brazilian financial market headlines,” *IEEE Latin America Transactions*, vol. 20, no. 2, pp. 344–351, 2021.
- [6] E. Mumcuoglu, C. E. Öztürk, H. M. Ozaktas, and A. Koç, “Natural language processing in law: Prediction of outcomes in the higher courts of turkey,” *Information Processing & Management*, vol. 58, no. 5, p. 102684, 2021.
- [7] X. Rao and Z. Ke, “Hierarchical rnn for information extraction from lawsuit documents,” *arXiv preprint arXiv:1804.09321*, 2018.
- [8] J. Deriu, A. Rodrigo, A. Otegi, G. Echegoyen, S. Rosset, E. Agirre, and M. Cieliebak, “Survey on evaluation methods for dialogue systems,” *Artificial Intelligence Review*, vol. 54, pp. 755–810, 2021.
- [9] H. Zhong, C. Xiao, C. Tu, T. Zhang, Z. Liu, and M. Sun, “How does nlp benefit legal system: A summary of legal artificial intelligence,” *arXiv preprint arXiv:2004.12158*, 2020.
- [10] R. P. L. Barros, “Análise jurisprudencial com técnica de aprendizado de máquina qual a tendência de opinião de cada turma do trt da 3ª região,” in *Congresso*, 2019. Acesso em: 04 de outubro de 2023.
- [11] E. C. M. d. Silva and B. A. Medeiros, “Comparação de métodos de mineração de texto para classificação de documentos jurídicos,” *Ciência da Computação-Tubarão*, 2020.
- [12] R. Dantas, K. Figueiredo, and L. Andrade, “Estudo de abordagens para classificação de textos sobre dúvidas tributárias utilizando mineração de texto,” *sbic.org.br* 2021, paper 155, 2021.
- [13] D. S. Kieckbusch, P. Geraldo Filho, V. Di Oliveira, and L. Wei-gang, “Scan-nf: A cnn-based system for the classification of electronic invoices through short-text product description,” in *WEBIST*, pp. 501–508, 2021.
- [14] J. P. Schulte, F. T. Giuntini, R. A. Nobre, K. C. d. Nascimento, R. I. Meneguette, W. Li, V. P. Gonçalves, and G. P. Rocha Filho, “Elinac: autoencoder approach for electronic invoices data clustering,” *Applied Sciences*, vol. 12, no. 6, p. 3008, 2022.
- [15] S. Few, “Show me the numbers,” *Analytics Pres*, 2004.
- [16] J. Lowgren, J. M. Carroll, M. Hassenzahl, T. Erickson, and A. Blackwell, “The encyclopedia of human-computer interaction,” *Interaction design foundation*, 2019.
- [17] J. D. M.-W. C. Kenton and L. K. Toutanova, “Bert: Pretraining of deep bidirectional transformers for language understanding,” in *Proceedings of naacL-HLT*, vol. 1, p. 2, 2019.
- [18] P. H. Luz de Araujo, T. E. de Campos, R. R. de Oliveira, M. Stauffer, S. Couto, and P. Bermejo, “Lener-br: a dataset for named entity recognition in brazilian legal text,” in *Computational Processing of the Portuguese Language: 13th International Conference, PROPOR 2018, Canela, Brazil, September 24–26, 2018, Proceedings 13*, pp. 313–323, Springer, 2018.