

INTELIGÊNCIA ARTIFICIAL E DIREITO: UMA SOLUÇÃO COMPUTACIONAL CAPAZ DE PREVER DECISÕES JUDICIAIS

ARTIFICIAL INTELLIGENCE AND LAW: A COMPUTATIONAL SOLUTION CAPABLE OF PREDICTING JUDICIAL DECISIONS

Luciana Trinkaus Menon 1
Melina Carla de Souza Britto 2
Guilherme Martelli Moreira 3
Fabrício Bittencourt da Cruz 4

Graduada em Engenharia de Computação pela UEPG e pela University of Adelaide. Doutoranda em Informática pela PUCPR com pesquisa na área de Inteligência Artificial, Processamento de Textos e Reconhecimento de Padrões. Especialista em machine learning e visão computacional na Univision LTDA, com projeto voltado ao processamento de imagens biométricas.

Lattes: <http://lattes.cnpq.br/0783750652820506>.
ORCID: <https://orcid.org/0000-0002-2674-3780>.
E-mail: trinkaus.luciana@gmail.com

Mestre em Direito pela PUCPR. Graduada em Bacharelado em Direito na UEPG. Especialista em Direito Tributário pela Uniderp e em Direito Público pela UniCesumar. Aluna especial do Doutorado em Ciências Sociais Aplicadas na UEPG. Professora na ESMAFE.

Lattes: <http://lattes.cnpq.br/1558310047406116>.
ORCID: <https://orcid.org/0000-0001-7699-6957>.
E-mail: melinabritto.adv@gmail.com

Mestrando em Direito Econômico e Desenvolvimento pela PUCPR. Integrante do Grupo de Pesquisas Tributação, Complexidade e Desenvolvimento (PPGD/PUCPR/CNPq). Especialista em Direito Tributário e Processo Tributário pela PUCPR. Graduado em Direito pela UFPR. Advogado.

Lattes: <http://lattes.cnpq.br/1820888758307120>.
ORCID: <https://orcid.org/0000-0002-9286-7706>.
E-mail: guilhermemartellimoreira@gmail.com

Doutor em Direito pela Faculdade de Direito da USP. Mestre em Direito pela PUCPR. Professor Adjunto no Departamento de Direito de Estado (Graduação) e no Programa de Pós-Graduação em Ciências Sociais Aplicadas (Mestrado e Doutorado) da UEPG. Director of the International Institute for Justice Excellence na Holanda e Líder do Projeto MindTheGap Inovação em Direito.

Lattes: <http://lattes.cnpq.br/7053459589427233>.
ORCID: <https://orcid.org/0000-0003-0538-9193>.
E-mail: fabriciobittcruz@gmail.com

Resumo: Neste trabalho, apresenta-se o resultado de uma solução computacional capaz de realizar a predição de sentenças. Classifica-se petições iniciais de ações civis públicas de improbidade administrativas e de execuções de termos de ajustamento de conduta, apresentadas pelo Ministério Público do Paraná (MPPR), nos anos de 2011 a 2018, perante as Varas Cíveis e da Fazenda Pública do Paraná. A finalidade das petições é prever sentenças judiciais (se favoráveis ou não em relação aos pedidos formulados pelos representantes do MPPR nas petições iniciais). Por meio de técnicas de processamento de linguagem natural e aprendizagem de máquina, os resultados obtidos foram promissores, atingindo 78,02% de acurácia com o uso de algoritmo indutor Logistic Regression, representação baseada em Bag of Words e uso de stemização RSLPS.

Palavras-chave: Inteligência Artificial. Aprendizagem de Máquina. Predição de Decisões Judiciais.

Abstract: In this paper, we present the result of a computational solution capable of carrying out the prediction of judicial decisions (judicial sentences). We classify plaintiffs of public civil actions of administrative improbity and enforcements of conduct adjustment terms presented by State Prosecution of Paraná (in Brazil) from 2011 to 2018, at the State Courts of Paraná. The goal of these plaintiffs is to predict judicial decisions (whether favorable or not in relation to the requests made by the State Prosecution's plaintiff). Through natural language processing techniques and machine learning, the results obtained are promising, reaching 78,02% accuracy with the use of the Logistic Regression inductive algorithm, representation based on Bag of Words and using RSLPS stemming.

Keywords: Artificial Intelligence. Machine Learning. Prediction of Judicial Decisions.

Introdução

Nos anos 1950, Alan Turing (TURING, 1995) questionou se as máquinas poderiam pensar, se poderiam ter inteligência, efeméride que marcou o início dos estudos da inteligência artificial (IA) como ramo da Ciência da Computação. As pesquisas deste ramo do conhecimento visam ao desenvolvimento de máquinas aptas a simular a capacidade humana de raciocinar e de resolver problemas.

Especificamente na área jurídica, a IA objetiva, em suma, aprimorar a pesquisa jurídica e a execução de tarefas pelos profissionais do ramo. Ao estudar o estado da arte do uso de técnicas de IA no Direito, constata-se o desenvolvimento de *softwares* especializados na identificação de assuntos, das partes e dos pedidos. Além disso, modelos mais ousados de aprendizagem de máquina (AM) têm por finalidade a classificação de decisões judiciais, visando à previsão do resultado de processos.

O objetivo principal deste trabalho foi construir e avaliar uma solução computacional capaz de realizar a tarefa de classificação de petições iniciais, tendo por finalidade a predição de sentenças judiciais. Esta solução aplicou técnicas de processamento de linguagem natural e aprendizagem supervisionada para resolver o problema de generalização e processamento assertivo de ações civis públicas de improbidade administrativa e execuções de termos de ajustamento de conduta. Essas ações foram propostas pelo Ministério Público do Paraná (MPPR), nos anos de 2011 a 2018, perante as Varas Cíveis e da Fazenda Pública do Estado do Paraná.

Especificamente, o problema que se busca resolver com o presente trabalho é a predição de sentenças proferidas em ações civis públicas de improbidade administrativa e execuções de termos de ajustamento de conduta, considerando a possibilidade ou não de êxito do MPPR. Em outras palavras, utiliza-se a AM com a finalidade de averiguar se os julgamentos de mérito tendem a ser favoráveis aos pedidos formulados na petição inicial apresentada pelo representante do *Parquet*.

A redução das taxas de congestionamento judicial e o aumento da celeridade processual são as possíveis contribuições sociais deste trabalho. Contribuições científicas evidenciam-se na definição de representação vetorial adequada para os textos processados, discussão sobre a melhor maneira de contornar a situação e similaridade entre as classes da base de dados. Além disso, considera-se como contribuição a avaliação de diferentes algoritmos indutores para a tarefa de classificação de petições.

No que diz respeito ao campo do Direito, ao realizar a classificação de decisões judiciais por seu julgamento de mérito, revelam-se questões que envolvem o manejo de ações judiciais em face dos princípios da tutela jurisdicional efetiva, da celeridade e da economia processual. Além disso, em um primeiro plano, observa-se a efetividade e o êxito na atuação do MPPR quanto à proposição de ações civis públicas de improbidade administrativa e execuções de termos de ajustamento de conduta envolvendo atos de improbidade administrativa.

Na primeira seção deste artigo, analisa-se, sucintamente, a aplicação de técnicas de AM no Direito. Na sequência, são expostos os motivos pelos quais se defende a implementação de técnicas de AM no Direito. Neste contexto, explica-se o método de aprendizagem de máquina utilizado no experimento realizado neste trabalho, expondo-se a composição da base de dados utilizada e a classificação proposta, cuja finalidade é prever sentenças, bem como os resultados obtidos.

A aplicação de técnicas de aprendizagem de máquina no Direito

Dados do Relatório Justiça em Números - 2020 (CONSELHO NACIONAL DE JUSTIÇA, 2020) mostram que a taxa de congestionamento do Poder Judiciário, indicador que mede o percentual de processos em tramitação que chegam ao final de cada ano sem solução definitiva em todo país, segue em patamares insustentáveis: 68,5%. O número de processos cresceu de forma significativa, desde 2009, ano inicial da série histórica analisada pelo Conselho Nacional de Justiça (CNJ), sendo que, ao final de 2019, o Poder Judiciário contava com um acervo de 77,1 milhões de processos pendentes de baixa. Ao longo dos 10 anos de análise, a taxa de congestionamento do Poder Judiciário manteve-se acima de 70%, com pequenas variações anuais, a exemplo de 2019.

O tempo médio até a decisão de segundo grau de um processo na Justiça do Trabalho é de um ano e seis meses, em média, aumentando para quatro anos e dois meses na Justiça Estadual e seis anos e cinco meses na Justiça Federal. A demora na entrega da tutela jurisdicional viola diversos princípios constitucionais, como tutela jurisdicional efetiva, celeridade e economia processual, além de ser elemento caracterizador de insegurança jurídica.

Neste contexto, uma tarefa interessante é a automatização da classificação de decisões judiciais para auxiliar, facilitar e acelerar o trabalho dos envolvidos no processo de decisão judicial. Em 2018, por exemplo, o Supremo Tribunal Federal (STF) desenvolveu o Projeto Victor, para auxiliar no trabalho de identificação de temas de repercussão geral (FERNANDES; CARVALHO, 2018, p. 488). Dentre os artigos publicados referentes ao Projeto Victor, merece destaque o de Fabricio Atades Braz *et al.* (BRAZ *et al.*, 2018), que propõem o uso de uma arquitetura *Bidirectional Long Short-Term Memory (Bi-LSTM)* (SCHUSTER; KULDIP, 1997) de redes neurais para a classificação de peças processuais. Os autores trabalharam com uma base de 6.814 documentos do STF, divididos em 6 classes (acórdão, recurso extraordinário, agravo de recurso extraordinário, despacho administrativo, sentença e outros). A técnica consiste do uso das primeiras mil palavras das peças processuais, considerando um limite de vocabulário das cem mil palavras mais frequentes, como *input* da *Bi-LSTM*. Dado um protocolo de testes, no qual a base de dados foi dividida em 70% para treinamento, 20% para validação e 10% para teste, os resultados obtiveram uma precisão média de 85% e *F1* (VAN RIJSBERGEN, 1979) de 84%. A *F1* é a métrica utilizada na avaliação de classificadores, que consiste na média harmônica de precisão e revocação.

Ainda sobre as pesquisas relacionadas ao Projeto Victor, o trabalho de Nilton Correia da Silva *et al.* (SILVA *et al.*, 2018) propõe o uso de uma arquitetura de redes neurais convolucionais simples. Os testes foram realizados na mesma base de dados que o anterior e com a mesma distribuição estratificada de dados entre treinamento, validação e testes. A rede neural convolucional proposta pelos autores obteve resultados superiores ao anterior, com acurácia de 90,35% e *F1* de 91%.

Técnicas de classificação automática de decisões judiciais também já foram exploradas fora do Brasil. Vale destacar a previsão do resultado de casos julgados pelo Tribunal Europeu dos Direitos Humanos (*European Court of Human Rights – ECHR*), baseado apenas no conteúdo textual dos documentos. Naquela pesquisa, desenvolvida por Nikolaos Aletras *et al.* (ALETRAS *et al.*, 2016), trabalhou-se com uma base de 584 casos, com objetivo descobrir se houve ou não a violação de determinado artigo da Convenção de Direitos Humanos (CDH). Três artigos da CDH foram considerados durante os experimentos e a informação textual dos casos foi representada usando *BoW (Bag of Words)* (SALTON *et al.*, 1975) de *N-grams* (sequência contígua *N* palavras, sendo unigrama e bigrama sequências de tamanho 1 e 2, respectivamente) e tópicos (aqui, tópicos são agrupamentos de *N-grams* semanticamente semelhantes). No trabalho, um dos resultados obtidos se deu com a utilização de *Support Vector Machine (SVM)* (CORTES *et al.*, 1995), observando-se que os “tópicos” mais presentes no caso são um indicador da existência ou não de violação de um determinado artigo, com uma acurácia média de 78%. Outro resultado foi obtido com *N-grams* da seção “fatos”, com acurácia média de 76%. Por fim, os autores combinaram os dois conjuntos de características com melhor desempenho (“Fatos” e “Tópicos”) e alcançaram o melhor desempenho médio de classificação, de 79%. Segundo os autores, a principal vantagem no uso de tópicos consiste na redução da dimensionalidade do espaço de características (*features*), evitando o sobreajuste (*overfitting*) aos dados de treinamento e fornecendo uma representação semântica mais concisa.

Mais recentemente, novos testes foram realizados para a classificação de decisões judiciais da *ECHR*. No experimento desenvolvido por Zhenyu Liu e Huanhuan Chen (LIU *et al.*, 2017), conduzido com a mesma representação vetorial que o anterior, o foco foi avaliar o desempenho de cinco modelos de aprendizado de máquina quanto à predição de sentenças judiciais. Os classificadores analisados foram *SVM*, *k-Nearest Neighbors (k-NN)* (ALTMAN, 1992), Regressão Logística (HOSMER, 2000), *Bagging* (BREIMAN, 1996) e *Random Forest* (BREIMAN, 2001; HO, 1995). Os resultados mostraram que o *SVM* supera os outros modelos, seguido da Regressão Logística e *k-NN*, estes últimos apresentando resultados muito próximos.

No experimento de Octavia Maria Şulea *et al.* (ŞULEA *et al.*, 2017), investigou-se a aplicação de métodos de classificação de texto e técnicas de aprendizado de máquina, para prever tanto a área jurídica à qual um caso pertence como a decisão do tribunal com base na descrição do caso. O experimento também visou estimar o intervalo de tempo (década) em que a descrição de caso e a decisão foram emitidas, os quais possuem 8, 6 e 7 classes, respectivamente. A base de dados utilizada consistiu em 126.865 decisões judiciais do Supremo Tribunal Francês (*Cour de Cassation*). Cada instância contém a descrição de caso e quatro rótulos: área de direito, data da decisão, o caso em si e uma lista de artigos e leis citados na descrição. Como *features*, foram utilizados *unigrams* e *bigrams*. Os resultados relatados a partir do uso de um classificador *SVM* linear, testado com validação cruzada estratificada de 10 *folds* (subconjunto de amostras), são de acurácia média de 96,9% na previsão da decisão do caso, acurácia média de 90,2%, para prever a subárea do Direito de um caso e 74,3% de acurácia média na estimativa da data de uma decisão.

Noutro momento (ŞULEA *et al.*, 2017), refizeram-se os experimentos com um conjunto de classificadores *SVM* treinados individualmente, cujas decisões foram combinadas via média de probabilidade classificatória. Para o cálculo, são adicionadas estimativas de probabilidade para cada classe e atribuído o rótulo de classe com a maior probabilidade média como a previsão. Testes foram realizados a partir do mesmo protocolo experimental que o experimento anterior e resultados significativamente superiores foram obtidos. Identificou-se acurácia média de 98,6%, na previsão da decisão do caso, acurácia média de 96,8%, para prever a área do Direito de um caso e 87% de acurácia média na estimativa da data de uma decisão.

No estudo desenvolvido por Guido Boela *et al.* (BOELLA *et al.*, 2011), é apresentado um sistema de gerenciamento de documentos legais, que tem como uma de suas principais funcionalidades a filtragem dos documentos. Para mostrar a viabilidade do projeto, os autores utilizaram uma base de dados de 223 leis organizadas em seis classes diferentes. Usando a representação vetorial de documentos *TF-IDF* (JONES, 1972), ganho de informação para a seleção de características e *SVM* polinomial como classificador, foi relatada uma acurácia média de 97,5% para a classificação correta de um texto em um domínio específico.

Destaca-se, então, a viabilidade da classificação automática de documentos jurídicos, já que trabalhos relacionados alcançaram bons resultados neste domínio, conforme se destaca na Tabela 1.

Tabela 1. Estado da arte.

Referência	Base de dados	Contexto	Classificador	Acurácia
BRAZ <i>et al.</i> , 2018	6814 documentos no STF, 6 classes	Peças processuais	<i>Bi-LSTM</i>	85,00%
SILVA <i>et al.</i> , 2018	6814 documentos no STF, 6 classes	Peças processuais	<i>CNN</i>	90,35%
ALETRAS <i>et al.</i> , 2016	584 documentos do ECHR, 2 classes	Decisão judicial	<i>SVM</i> Linear	79,00%
ŞULEA <i>et al.</i> , 2017	126.865 documentos do Supremo Tribunal Francês, 6 classes	Decisão judicial	<i>SVM</i> Linear	96,90%
ŞULEA <i>et al.</i> , 2017	126.865 documentos do Supremo Tribunal Francês, 6 classes	Decisão judicial	<i>Ensemble de SVMs</i>	98,60%
BOELLA <i>et al.</i> , 2011	223 leis do portal <i>Regione Piemonte</i> , 6 classes	Domínio legal	<i>SVM</i> Polinomial	97,50%

Fonte: AUTORES.

Vale ressaltar também que os algoritmos de *stemming* são comumente utilizados durante a fase de pré-processamento textual para reduzir a dimensão dos dados. Este método já foi estudado em diversos domínios, incluindo o legal. O *stemming* é uma técnica que consiste em

reduzir palavras ao seu radical (*steam*). Por meio da técnica, agrupam-se conjuntos de palavras diferentes conectadas por um radical comum (por exemplo, as palavras menina, menino e meninas podem ser representadas pelo *stem* “menin”).

Os autores Robert N. de Oliveira e Methanias C. Junior (OLIVEIRA; JUNIOR, 2018) lembram que o domínio legal tem seu próprio jargão, com ênfase na não repetição de palavras e uso de vocabulário mais elaborado, características que podem influenciar diretamente a eficácia dos algoritmos de *stemming*. Como o *stemming* no contexto jurídico brasileiro ainda não havia sido explorado, os autores fizeram uma análise experimental sobre sua relação com a redução de dimensionalidade de documentos, a partir de uma base real de jurisprudência do Tribunal de Justiça do Estado de Sergipe. A base é composta por 279.448 documentos de quatro coleções de dados, julgamentos do Tribunal de Recursos, decisões monocráticas do Tribunal de Apelações, julgamentos dos Tribunais Especiais e decisões monocráticas dos Tribunais Especiais. Os resultados mostram que o algoritmo Removedor de Sufixos da Língua Portuguesa (*RSLP*) (ORENGO e HUYCK, 2001) foi o mais efetivo, com taxa de redução de termos únicos de 52%, seguido de *Porter* (PORTER, 1980), *UniNE* (SAVOY, 1999), *RSLP-S* (ORENGO et al., 2007) e não utilização de *stemming*.

Prevedo decisões judiciais em ações de improbidade administrativa e execuções de termo de ajustamento de conduta

O MPPR dispõe de considerável base de dados referentes a ações civis públicas de improbidade administrativa e de execuções de termos de ajustamento de conduta já julgadas. O órgão disponibilizou uma parte destes documentos para pesquisa junto ao Programa de Pós-Graduação em Informática Aplicada (PPGIA) da Pontifícia Universidade Católica do Paraná (PUCPR), com o intuito de treinar algoritmos de AM para realizar a predição de sentenças judiciais.

A base de dados do experimento

A base de dados é composta apenas por execuções de termos de ajustamento de conduta e por ações civis públicas de improbidade administrativa propostas pelo MPPR, entre 2011 e 2018, perante as Varas Cíveis e Varas da Fazenda Pública no Estado do Paraná. Este banco consiste em 525 petições iniciais produzidas pelos representantes do MPPR ao ingressarem com as ações civis públicas de improbidade administrativa e de execuções de termos de ajustamento de conduta.

Para a filtragem das iniciais, levou-se em conta a classificação do tipo de ação proposta e o campo conceitual de improbidade administrativa, que envolve atos praticados por agentes públicos contra princípios da Administração Pública. Tem-se, assim, como possíveis consequências, prejuízo ao patrimônio público ou enriquecimento ilícito próprio ou de terceiros.

Na formatação do modelo de AM aplicado, considerou-se a petição inicial como documento escrito pelo qual o autor da ação expõe os fatos e explica as razões pelas quais considera ser necessária a intervenção do Poder Judiciário em certa situação.

A base de dados cedida pelo MPPR conta com doze tipos de sentenças, tendo em conta características específicas do dispositivo, parte final da sentença em que se profere o julgamento após a fundamentação. Em decorrência da não existência de exemplos suficientes de alguns tipos de sentenças para treinamento dos modelos cognitivos, considerou-se o problema de categorização de sentenças judiciais de forma binária, classificando-as sob o ponto de vista do MPPR como autor da ação: favoráveis ao MPPR e desfavoráveis ao MPPR.

Destaca-se que 108 documentos da base de dados não foram utilizados neste trabalho, para efeito de aplicação de técnicas de aprendizagem de máquina, em razão da baixa resolução utilizada no momento da digitalização dificultar a extração de textos (*OCR, Optical Character Recognition* – técnica que reconhece caracteres de uma imagem ou arquivo digitalizado e transcreve o conteúdo em formato de texto). Ainda, 48 desses documentos não estavam rotulados, o que é necessário em tarefas supervisionadas de aprendizagem de máquina.

Restaram 369 documentos válidos rotulados para a realização do experimento. Para

cada uma das petições iniciais protocoladas, o MPPR cedeu a respectiva sentença de primeiro grau proferida. A Tabela 2 mostra as decisões judiciais (sentenças) presentes na base de dados e o número de exemplos de cada uma delas.

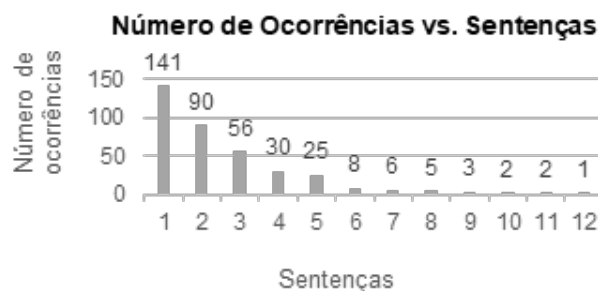
Tabela 2. Distribuição das decisões judiciais da base de dados.

ID	Decisão judicial	Número de ocorrências
1	Julgada improcedente a ação	141
2	Julgada procedente a ação	90
3	Julgada procedente em parte a ação	56
4	Extinto o processo por ausência das condições da ação	30
5	Homologada a transação	25
6	Indeferida a petição inicial	8
7	Extinto o processo por ausência de pressupostos processuais	6
8	Extinto o processo por desistência	5
9	Acordo em audiência homologado	3
10	Declarada decadência ou prescrição	2
11	Procedência do pedido – reconhecimento pelo réu	2
12	Proferida sentença condenatória	1

Fonte: AUTORES.

Trabalhar com bases de dados desbalanceadas, geralmente, é um problema, dado que os algoritmos de indução convencionais tendem a aprender a classe majoritária e errar com frequência as amostras oriundas das classes minoritárias. A Figura 1 apresenta a distribuição das amostras nas 12 classes (categorias) de sentenças presentes na base de documentos disponível.

Figura 1. Distribuição das sentenças na base de dados.



Fonte: AUTORES.

Para contornar a dificuldade gerada pelo desbalanceamento de classes, geralmente o conjunto de dados é equilibrado artificialmente, a exemplos da subamostragem ou da sobre-amostragem das classes. Neste contexto, foi idealizado um reagrupamento das classes de sentenças. Inicialmente, a base de dados foi agrupada em dois grandes segmentos: ações extintas sem resolução de mérito e ações extintas com resolução de mérito, segundo os artigos 487 e 485 do Código de Processo Civil (ver Tabela 3).

Tabela 3. Reagrupamento inicial das sentenças da base de dados.

Extinção sem resolução de mérito	Extinção com resolução de mérito
Extinto por ausência das condições da ação Indeferida a petição inicial Extinto por ausência de pressupostos processuais Extinto por desistência	Julgada procedente a ação Julgada procedente em parte a ação Homologada a transação Acordo em audiência homologado Procedência do pedido – reconhecimento pelo réu Proferida sentença condenatória Julgada improcedente a ação Declarada decadência ou prescrição

Fonte: AUTORES.

No entanto, como pode ser observado na Figura 2, nesta classificação, os dados continuaram bastante desbalanceados.

A partir de uma análise mais detalhada da distribuição dos dados, foi proposta a classificação apenas das sentenças extintas com resolução de mérito, dado que existem poucos exemplos de sentenças extintas sem essa resolução, as quais podem vir a atrapalhar o classificador proposto. As ações extintas com resolução de mérito foram então reagrupadas sob o ponto de vista do MPPR como autor da ação (e, portanto, levando-se em conta os pedidos contidos na petição inicial), como favoráveis ao MPPR e desfavoráveis ao MPPR. As ações julgadas procedentes, em parte, poderiam ser consideradas tanto como favoráveis quanto como desfavoráveis ao MPPR. Portanto, foram excluídas da classificação em razão de sua ambiguidade intrínseca. A Tabela 4 apresenta o agrupamento em duas classes.

Tabela 4. Reagrupamento das sentenças sob o ponto de vista do MPPR como autor da ação.

Extinção com resolução de mérito	
Favorável ao MPPR	Desfavorável ao MPPR
Julgada procedente a ação	Julgada improcedente a ação
Homologada a transação	Declarada decadência ou prescrição
Acordo em audiência homologado	
Procedência do pedido – reconhecimento pelo réu	
Proferida sentença condenatória	

Fonte: AUTORES.

Após este reagrupamento dos dados, as classes favorável ao MPPR e desfavorável ao MPPR são relativamente balanceadas, possuindo 121 e 143 documentos, respectivamente, conforme é possível constatar na Figura 3.

Figura 2. Distribuição dos dados quanto à resolução de mérito.

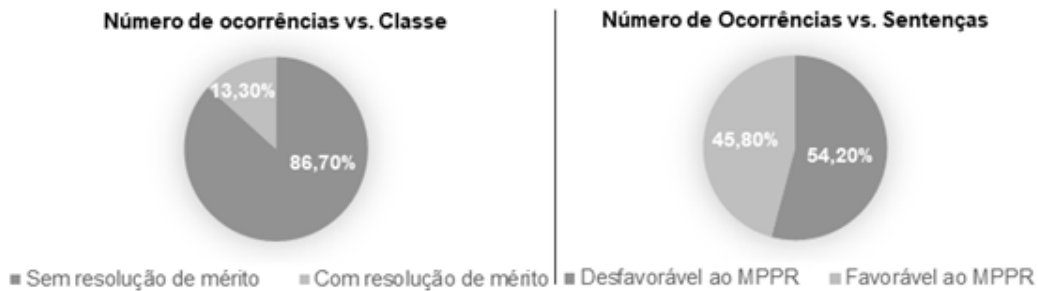
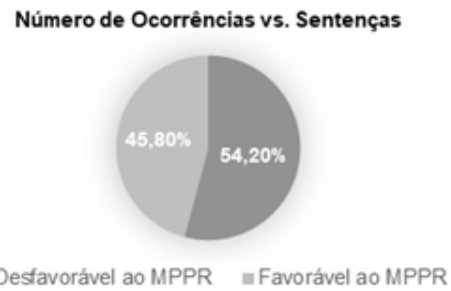


Figura 3. Distribuição dos dados quanto à resolução de mérito favoráveis e desfavoráveis ao MPPR.



Fonte: AUTORES.

Método de aprendizagem de máquina utilizado para a predição

O método aplicado é constituído de três estágios: (i) o primeiro é a etapa de pré-processamento dos dados. Esta etapa é de suma importância e consiste na limpeza de dados e detecção de partes relevantes dos documentos; (ii) o segundo, consiste na representação vetorial dos textos. Nesta etapa são exploradas diferentes técnicas de *word-embedding* (formas de representação da palavra) para construir uma representação robusta do vocabulário encontrado nas sentenças judiciais; (iii) o terceiro e último estágio consiste na exploração de aprendizagem supervisionada, considerando classificadores monolíticos. Métricas de acurácia, *F1*, precisão e revocação (*recall*) são utilizadas para avaliar a combinação de diferentes técnicas de pré-processamento e modelos de classificação utilizados.

Processamento da base de dados

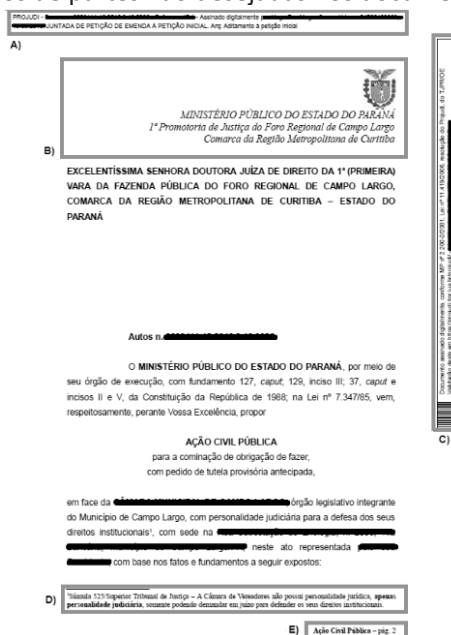
As petições iniciais foram redigidas por diversos promotores do MPPR e não possuem padrão de formatação. Assim, a primeira dificuldade relacionada à formatação dos arquivos foi a delimitação dos documentos, já que, em alguns deles, havia compilados de vários documentos relacionados à petição ou que continham anexos. Ressalta-se que a atividade de filtragem destes documentos, com a separação exata do que consistia a petição inicial, foi realizada por servidores do MPPR.

Com os documentos já delimitados apenas à petição inicial, o passo seguinte foi a remoção de cabeçalhos e rodapés, assim como a exclusão de texto de protocolo digital junto ao PROJUDI (2020) (Processo Judicial Digital). Os rodapés de identificação foram excluídos (aqueles que se reperem inteiros ou parcialmente em todas as páginas da petição) e as informações contidas nas notas de rodapé foram mantidas (BIRD, 2006). Exemplos destes problemas podem ser observados na Figura 4.

Os *PDFs* foram convertidos em documentos de texto com o auxílio da ferramenta denominada *PDFMiner* (PDFMINER, 2021).

Visando à normalização dos textos, foram aplicadas etapas básicas do pré-processamento de textos. A normalização inclui converter todas as letras para minúsculas, remover números, caracteres especiais e alfanuméricos, espaços em branco excessivos, acentuação, pontuação e termos muito frequentes (*stopwords*) ou muito esparsos.

Figura 4: Exemplos de partes não desejadas nos documentos.



Nota: As partes não desejadas são: a) PROJUDI; b) cabeçalho; c) mais informações PROJUDI; d) nota de rodapé com informações importantes; e) rodapé a ser descartado.

Fonte: AUTORES.

A última técnica aplicada foi a *stemização* dos textos, na qual foram utilizados o *RSLP stemmer* e o *Snowball stemmer* (PORTER, 2001). Para fins de comparação, foram mantidos o texto sem *stemização* e os textos processados com os dois *stemmers*. Todo o pré-processamento dos textos foi realizado com a ferramenta *NLTK*. A remoção de *stopwords* e a *stemização* com ambos os *stemizadores* foram executados a partir das implementações padrão disponíveis na ferramenta.

Um grande desafio para análise desta base de dados decorre da inconstância dos tamanhos dos documentos e do detalhamento das informações contidas neles (que variam de cinco a 134 páginas). Para normalizá-los, foram utilizadas técnicas de sumarização para padronizar o tamanho dos documentos. A partir da distribuição de frequência das palavras do texto, foi criado um *score* para cada sentença, baseado na soma da frequência de cada palavra presente. Com base nesse *score*, é possível selecionar as “n” sentenças mais importantes e criar um texto sumarizado com essas. Para fins de comparação, foram considerados o texto inteiro e o texto sumarizado.

Representação vetorial

Para a representação vetorial dos dados foram consideradas duas técnicas baseadas na frequência das palavras nos textos, *BoW* e *TF-IDF*. Para a vetorização dos documentos com ambas as técnicas, foi utilizada a ferramenta *Scikit-Learn* (PEDREGOSA *et al.*, 2011).

Ambos os métodos foram utilizados com *unigrams* e *bigrams*. Para impedir a criação de vetores de características muito extensas e evitar o processamento desnecessário de palavras irrelevantes para a classificação dos documentos, os termos que apresentaram uma frequência inferior a 5% ou superior a 70% foram ignorados. No caso específico do *TF-IDF*, foi aplicado o escalonamento *sublinear* de *tf*, no qual *tf* é substituído por $1 + \log(tf)$.

Classificação

Seguindo o que é usualmente utilizado na literatura, os classificadores empregados foram SVM, *Multilayer Perceptron (MLP)* (HORNIK, 1989), *Naive Bayes (NB)* (RISH, 2001), *Random Forest* e Regressão Logística.

Devido à dificuldade de trabalhar com *ensembles* e métodos de aprendizagem profunda com poucos dados, os experimentos foram realizados com os cinco classificadores monolíticos previamente descritos. Todos os classificadores utilizados nos experimentos preliminares foram executados a partir da ferramenta *Scikit-Learn*. O *NB* foi utilizado na sua forma multinomial com os parâmetros padrão, implementado na biblioteca como *MultinomialNB*. A regressão logística também foi utilizada em seu formato padrão, implementada como *Logistic Regression*.

O *SVM* foi utilizado com *kernel* linear, implementado no *Scikit-Learn* como *SVC*. A *Random Forest* foi implementada por meio da classe *RandomForestClassifier*, com 200 árvores e profundidade máxima de 3 níveis. O *MLP* foi utilizado via *MLPClassifier*, com o método de ajuste de pesos *lbfgs*, 70 neurônios na camada oculta e 0,00001 como parâmetro de regularização (*alpha*).

Resultados

Todos os resultados relatados nesta seção são referentes a testes realizados na base de dados acima relatada, a qual contém 264 sentenças (julgamentos de mérito proferidos no primeiro grau por Varas Cíveis e Varas da Fazenda Pública do Estado do Paraná), sendo 121 rotuladas como favoráveis ao MPPR e 143 rotuladas como desfavoráveis ao MPPR. Todos os testes foram realizados sob um protocolo experimental de validação cruzada estratificada de 10 *folds*.

O primeiro experimento realizado evidenciou a importância do pré-processamento dos textos. A intenção foi comparar resultados de textos puros e de textos pré-processados quanto à dimensionalidade dos vetores de características e quanto à sua performance na classificação. Em ambos os testes, a limpeza inicial dos documentos foi realizada considerando a delimitação dos documentos apenas à petição inicial e remoção de cabeçalhos, rodapés e textos do PROJUDI.

Sem nenhum processo de pré-processamento adicional, os vetores de características retornados pelo *BoW* e *TF-IDF* de *unigrams* e *bigrams* dos textos possuem 425.533 características. Já o vetor de característica dos textos pré-processados a partir da transformação de letras para minúsculo e remoção de acentuação, *stopwords*, numerais, caracteres especiais, alfanuméricos, termos muito frequentes (mais de 70% dos documentos) ou muito esparsos (menos de 5% dos documentos), diminuiu para 15.962 características.

Quanto à eficácia dos classificadores, os resultados obtidos, a partir do texto pré-processado, podem ser observados nas Tabelas 5 e 6. No entanto, ressalta-se que, devido ao tamanho dos vetores gerados pelos textos sem o pré-processamento, estes sobrepujaram a memória disponível no computador durante sua execução e não puderam ser finalizados.

Tabela 5: Resultados da classificação dos documentos de textos pré-processados *BoW*.

Classificador	Precisão	Recall	F1	Acurácia
SVM	0,7668	0,7611	0,7590	0,7611 ± 0,2208
MLP	0,7350	0,7308	0,7251	0,7308 ± 0,2179
NB	0,7625	0,7157	0,6882	0,7157 ± 0,1523
<i>Random Forest (RF)</i>	0,7451	0,6931	0,6592	0,6931 ± 0,1698
Regressão Logística (LR)	0,7520	0,7496	0,7456	0,7496 ± 0,2364

Fonte: AUTORES.

Tabela 6: Resultados da classificação dos documentos de textos pré-processados *TF-IDF*.

Classificador	Precisão	Recall	F1	Acurácia
SVM	0,7723	0,7389	0,7210	0,7389 ± 0,2121
MLP	0,7472	0,7386	0,7324	0,7386 ± 0,1967
NB	0,7521	0,6705	0,6193	0,6705 ± 0,1426
<i>Random Forest (RF)</i>	0,7667	0,7114	0,6809	0,7114 ± 0,1648
Regressão Logística (LR)	0,7652	0,7198	0,6962	0,7198 ± 0,1992

Fonte: AUTORES.

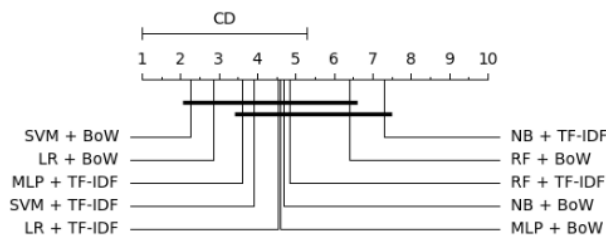
Nas Tabelas 5 e 6, é possível observar que as duas técnicas de representação de textos analisadas obtiveram resultados de acurácia bastante próximos e desvios padrão altos em todos os classificadores.

A fim de verificar se existe diferença significativa na acurácia média dos métodos analisados, foi utilizado o teste estatístico de Milton Friedman (FRIEDMAN, 1937), conforme proposto por Janez Demšar (DEMŠAR, 2006), aplicado com nível de significância de 0,05. Concluiu-se que há diferença significativa entre, pelo menos, uma das técnicas examinadas.

Em seguida, o teste *post-hoc* Nemenyi (NEMENYI, 1962, p. 263) foi utilizado para identificar as diferenças internas entre cada um dos experimentos realizados. O teste Nemenyi compara todas as variáveis observadas, uma contra a outra, identificando se uma variável é significativamente diferente da outra, quando a média de classificação observada é maior que a distância crítica calculada (CD). Neste caso, CD foi igual a 4,28. A Figura 5 mostra que quase todas as combinações de classificadores e vetorizadores de textos obtiveram acurácias significativamente não distintas, com exceção do SVM e da Regressão Linear, com vetor de características extraído a partir do *BoW*, que tem nível de acurácia mais elevado que o *NB*, com vetor de características extraído a partir do *TF-IDF*.

Como os resultados estão muito próximos e não é possível distinguir quais classificadores e quais técnicas de vetorização estão obtendo melhores resultados, os demais experimentos foram realizados com todos os classificadores e ambas as técnicas de vetorização.

Figura 5: Teste Nemenyi dos rankings médios obtidos pelo teste de Friedman.



Fonte: AUTORES.

A técnica de sumarização de documentos descrita na seção sobre métodos deste trabalho foi aplicada, com o número de sentenças variando entre 5 e 100. No entanto, em todos os experimentos realizados, os resultados foram inferiores aos apresentados quando trabalhado o texto inteiro pré-processado. Portanto, esta técnica não foi utilizada na comparação dos resultados.

Com o próximo experimento, objetivou-se verificar o uso dos stemizadores *RSLP* e *Snowball*. Novamente, a intenção era comparar resultados de textos stemizados e de textos apenas pré-processados quanto à dimensionalidade dos vetores de características e quanto à performance na classificação. A Tabela 7 mostra a dimensão dos vetores de características da base de dados pré-processada, seguida ou não de stemização com *RSLP* e *Snowball*.

Tabela 7: Dimensões dos vetores de características de textos pré-processados, com e sem stemização – *RSLP* e *Snowball*.

Stemizador	Dimensão Vetor de Características
Nenhum	15962
<i>Snowball</i>	14687
<i>RSLP</i>	14369

Fonte: AUTORES.

Como os textos já estavam pré-processados, a redução causada pelos stemizadores não foi tão significativa. No entanto, o *RSLP* mostrou uma capacidade de redução do vetor de características um pouco superior que o *Snowball*. Os resultados de classificação obtidos a partir da utilização de cada um dos *stemmers*, assim como o resultado sem a stemização, podem

ser observados nas Tabelas 8 e 9.

Tabela 8: Resultados da classificação dos documentos de textos pré-processados – Vetorização com BoW, seguido de Stemização com *RSLP* ou *Snowball*.

Classificador	Precisão	Recall	F1	Acurácia
SVM	0,7668	0,7611	0,7590	0,7611 ± 0,2208
MLP	0,7350	0,7308	0,7251	0,7308 ± 0,2179
NB	0,7625	0,7157	0,6882	0,7157 ± 0,1523
<i>Random Forest (RF)</i>	0,7451	0,6931	0,6592	0,6931 ± 0,1698
Regressão Logística (<i>LR</i>)	0,7520	0,7496	0,7456	0,7496 ± 0,2364
SVM + RSLP	0,7686	0,7610	0,7586	0,7610 ± 0,2041
MPL + RSLP	0,7698	0,7644	0,7612	0,7644 ± 0,2260
NB + RSLP	0,7705	0,7269	0,7018	0,7269 ± 0,1563
RF + RSLP	0,7460	0,7007	0,6692	0,7007 ± 0,1860
LR + RSLP	0,7902	0,7802	0,7766	0,7802 ± 0,1920
SVM + SNOWBALL	0,7637	0,7574	0,7547	0,7574 ± 0,2137
MLP + SNOWBALL	0,4589	0,4961	0,4430	0,4961 ± 0,0953
NB + SNOWBALL	0,7672	0,7234	0,6974	0,7234 ± 0,1621
RF + SNOWBALL	0,7424	0,6931	0,6595	0,6931 ± 0,1767
LR + SNOWBALL	0,7758	0,7687	0,7654	0,7697 ± 0,2036

Fonte: AUTORES.

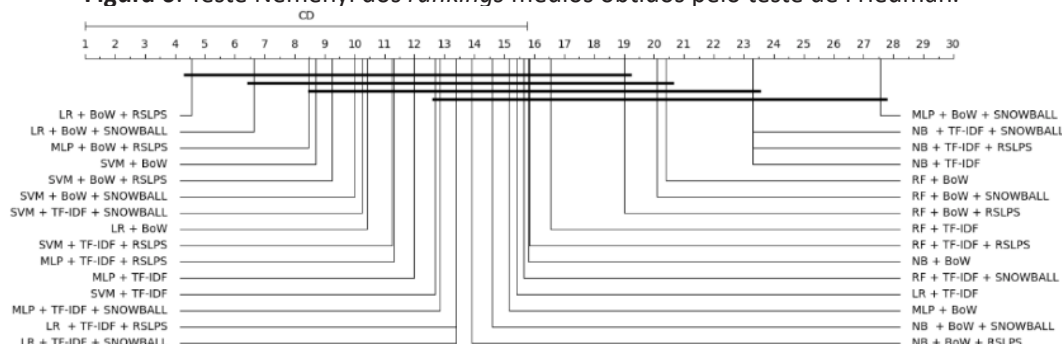
Tabela 9: Resultados da classificação dos documentos de textos pré-processados – Vetorização com *TF-IDF*, seguido de stemização com *RSLP* ou *Snowball*.

Classificador	Precisão	Recall	F1	Acurácia
SVM	0,7723	0,7389	0,7210	0,7389 ± 0,2121
MLP	0,7472	0,7386	0,7324	0,7386 ± 0,1967
NB	0,7521	0,6705	0,6193	0,6705 ± 0,1426
<i>Random Forest (RF)</i>	0,7667	0,7114	0,6809	0,7114 ± 0,1648
Regressão Logística (<i>LR</i>)	0,7652	0,7198	0,6962	0,7198 ± 0,1992
SVM + RSLP	0,7804	0,7459	0,7280	0,7459 ± 0,2008
MPL + RSLP	0,7529	0,7426	0,7342	0,7426 ± 0,1935
NB + RSLP	0,7521	0,6705	0,6193	0,6705 ± 0,1426
RF + RSLP	0,7749	0,7121	0,6803	0,7121 ± 0,1351
LR + RSLP	0,7740	0,7310	0,7094	0,7310 ± 0,1957
SVM + SNOWBALL	0,7857	0,7496	0,7318	0,7496 ± 0,2041
MLP + SNOWBALL	0,7432	0,7355	0,7271	0,7355 ± 0,2016
NB + SNOWBALL	0,7521	0,6705	0,6193	0,6705 ± 0,1426
RF + SNOWBALL	0,7807	0,7157	0,6844	0,7157 ± 0,1338
LR + SNOWBALL	0,7740	0,7310	0,7094	0,7310 ± 0,1957

Fonte: AUTORES.

Para melhor visualizar os resultados e verificar se há diferença significativa entre os métodos abordados, o teste de Friedman, seguido do teste *post-hoc* Nemenyi, foi aplicado (DEMŠAR, 2006). Com um nível de significância de 0,05, concluiu-se que há diferença significativa entre, pelo menos, uma das técnicas examinadas. O teste de Nemenyi resultou em $CD = 14,75$ e os *rankings* obtidos pelo teste podem ser observados na Figura 6.

Figura 6: Teste Nemenyi dos rankings médios obtidos pelo teste de Friedman.



Fonte: AUTORES.

É possível identificar, na Figura 6, que as diversas combinações de classificadores e vetORIZADORES de textos obtiveram acurácias similares, sendo que muitos deles não têm diferenças significativas entre seus resultados. No entanto, é possível observar que, de maneira geral, a utilização de técnicas de *stemming* resultaram num aumento de acurácia. Além disso, aponta-se que os classificadores regressão logística, *MLP* e *SVM* estão mais bem ranqueados que o *NB* e *Random Forest (RF)*. Sobre as técnicas de vetorização de documentos, *BoW* está nas seis primeiras posições do ranking, no entanto, sem diferença significativa das abordagens que utilizam *TF-IDF*.

O melhor resultado observado atingiu 78,02% de acurácia e consiste na combinação (*LR + BoW + RSLPS*), ou seja, algoritmo indutor *Logistic Regression*, representação dos textos com *Bag of Words* e uso de *stemização RSLPS*. Vale destacar que a técnica *LR* é caixa-branca, o que a torna adequada para a área do direito na qual a IA explicável é preferível.

Conclusão

Estratégias de processamento de textos para a classificação e predição de decisões judiciais ainda são relativamente sub-exploradas na língua portuguesa, na qual sua aplicação tem sido restrita à classificação de peças processuais. Os trabalhos relacionados mostram a utilização de diferentes tipos de classificadores, como monolíticos, *ensembles* e de aprendizagem profunda, o que reforça o método proposto em relação à necessidade de análise de diferentes tipos de classificadores para encontrar aquele que melhor se adapta ao contexto específico do conjunto de dados.

Além disso, o uso de técnicas de *stemização* ainda não foi abordado no contexto de classificação e predição de decisões judiciais. Pesquisas indicam sobre sua capacidade de redução de dimensão dos dados na área jurídica, mas ainda não existem trabalhos relacionados à qualidade dos *stemizadores* no contexto de classificação de documentos quanto à decisão judicial.

Desta forma, utilizaram-se técnicas de processamento da linguagem natural e AM, para investigar diferentes formas de representação e limpeza de dados, bem como técnicas de aprendizagem supervisionada para predição de sentenças judiciais. Apesar do grande número de documentos disponíveis no MPPR, poucos estavam rotulados, o que dificultou a utilização de técnicas puramente supervisionadas de AM. Como a linguagem jurídica tem um vocabulário específico e a massa de documentos vem de origem muito variada e sem padrão pré-definido, a definição de uma representação adequada para generalização e o processamento assertivo de linguagem natural foram fatores de complexidade.

Outro problema a ser destacado, na condução desta pesquisa, relaciona-se à qualidade dos dados para processamento de AM, pois os resultados dos algoritmos podem ser sensíveis a ruídos. Muitas vezes, textos jurídicos contam com cabeçalhos, rodapés, anexos, referências a outros documentos legais (leis, artigos etc.) e, em alguns casos, tabelas e imagens, características que, frequentemente, não são as mais relevantes para a predição da sentença judicial. Nesse sentido, técnicas de limpeza e seleção de dados também precisam ser exploradas.

Não obstante as dificuldades apontadas neste trabalho, obteve-se um resultado signifi-

ficavam relevante, tanto para a Computação quanto para o Direito. No primeiro campo, evidenciou-se a possibilidade de aplicação de variadas técnicas de processamento de dados e de AM na área do Direito, sendo possível a predição de sentenças. Do experimento realizado, o melhor resultado atingiu 78,02% de acurácia, consistindo na combinação (*LR + BoW + RSLPS*, ou seja, algoritmo indutor *Logistic Regression*), com representação dos textos com *Bag of Words* e uso de *stemização RSLPS*. No segundo campo, do Direito, como resultado desta, observou-se que 264 ações civis públicas de improbidade administrativa e execuções de termos de conduta envolvendo atos de improbidade administrativa, protocoladas pelo MPPR, no período entre 2011 e 2018, obtiveram julgamento de mérito. Destas, 121 sentenças foram favoráveis ao MPPR, visto que foram julgadas procedentes ou homologada transação e acordo, e 143 foram desfavoráveis ao MPPR, pois foram julgadas improcedentes ou declarada a prescrição da pretensão ou decadência do direito.

Por fim, vale ressaltar que o experimento desenvolvido foi preliminar. A coleta e a incrementação da base de dados do Ministério Público foi suspensa em razão da pandemia e do consequente teletrabalho da grande maioria dos seus servidores. Não obstante, pretende-se dar continuidade à pesquisa assim que as atividades voltem à forma presencial.

Nota. Os autores agradecem o Ministério Público do Paraná pela disponibilização da base de dados e pela cooperação de seus servidores e promotores na realização da presente pesquisa.

Referências

ALETRAS, N.; TSARAPATSANIS, D.; PREOȚIUC-PIETRO, D.; LAMPOS, V. Predicting Judicial Decisions of the European Court of Human Rights: A natural Language Processing Perspective. **PeerJ Computer Science**, PeerJ Inc., v. 2, 2016. Disponível em: <https://peerj.com/articles/cs-93/>. Acesso em: 27 mai. 2021.

ALTMAN, N. S. An Introduction to Kernel and Nearest-neighbor Nonparametric Regression. **The American Statistician**, v. 46, n. 3, pp. 175-185, 1992. Disponível em: www.jstor.org/stable/2685209. Acesso em: 27 mai. 2021.

BIRD, S. **NLTK: The Natural Language Toolkit. Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions**, Association for Computational Linguistics, Sydney, Australia, 2006. Disponível em: <https://doi.org/10.3115/1225403.1225421>. Acesso em: 27 mai. 2021.

BOELLA, G.; CARO, L. Di; HUMPHREYS, L. Using Classification to Support Legal Knowledge Engineers in the Eunomos Legal Document Management System. **Fifth International Workshop on Juris-informatics**, JURISIN, 2011.

BRAZ, F. A.; SILVA, N. C. da; CAMPOS, T. E. de; CHAVES, F. B. de S.; FERREIRA, M. H. P.; INAZAWA, P. H. G.; COELHO, V. H. D.; SUKIENNIK, B. P.; ALMEIDA, A. P. G. S. de; VIDAL, F. B.; BEZERRA, D. A.; GUSMÃO, D. B.; ZIEGLER, G. G.; FERNANDES, R. V. de C.; ZUMBLICK, R.; PEIXOTO, F. H. **Document Classification Using Abi-lstm to Unclog Brazil's Supreme Court**. Brasília, Brasil, 2018. Disponível em: https://cic.unb.br/~teodecampos/ViP/braz_etal_nips_ml4d_2018.pdf. Acesso em: 27 mai. 2021.

BREIMAN, L. Bagging Predictors. **Machine Learning**, v. 24, n. 2, pp. 123-140, 1996. Disponível em: <https://doi.org/10.1007/BF00058655>. Acesso em: 27 mai. 2021.

_____. Random Forests. **Machine Learning**, v. 45, pp. 5-32, 2001. Disponível em: <https://doi.org/10.1023/A:1010933404324>. Acesso em: 27 mai. 2021.

CONSELHO NACIONAL DE JUSTIÇA. **Relatório Justiça em Números 2020**. Brasília, 2020. Disponível: <https://www.cnj.jus.br/wp-content/uploads/2020/08/WEB-V3-Justi%C3%A7a-em-N%C3%BAmeros-2020-atualizado-em-25-08-2020.pdf>. Acesso em: 27 mai. 2021.

CORTES, C.; VAPNIK, V. Support-vector Networks. **Machine Learning**, v. 20, pp. 273-297, 1995. Disponível: <https://doi.org/10.1007/BF00994018>. Acesso em: 27 mai. 2021.

DEMŠAR, J. Statistical Comparisons of Classifiers Over Multiple Data Sets. **Journal of Machine Learning Research**, v. 7, pp. 1-30, 2006. Disponível: <https://www.jmlr.org/papers/volume7/demsar06a/demsar06a.pdf>. Acesso em: 27 mai. 2021.

FERNANDES, R. V. de C.; CARVALHO, A. G. P. de. **Tecnologia jurídica Direito Digital: II Congresso Internacional de Direito, Governo e Tecnologia – 2018**. Belo Horizonte, Brasil: Fórum, 2018.

FRIEDMAN, M. The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. **Journal of the American Statistical Association**, Taylor & Francis, v. 32, n. 200, p. 675-701, 1937. Disponível: <http://www.jstor.org/stable/2279372?origin=JSTOR-pdf>. Acesso em: 27 mai. 2021.

HO, T. K. Random Decision Forests. **Proceedings of the 3rd International Conference on Document Analysis and Recognition**, Montreal, QC, pp. 278-282, 1995. Disponível: <https://doi.ieee-computersociety.org/10.1109/ICDAR.1995.598994>. Acesso em: 27 mai. 2021.

HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Multilayer Feedforward Networks are Universal Approximators. **Neural Networks**, v. 2, n. 5, pp. 359-366, 1989. Disponível: <https://www.sciencedirect.com/science/article/pii/0893608089900208>. Acesso em: 27 mai. 2021.

HOSMER, D. W.; LEMESHOW, S. **Applied Logistic Regression**. 2 ed. New York: Wiley, 2000. Disponível: http://resource.heartonline.cn/20150528/1_3kOQSTg.pdf. Acesso em: 27 mai. 2021.

JONES, K. S. A Statistical Interpretation of Term Specificity and its Application in Retrieval. **Journal of Documentation**, v. 28 n. 1, pp. 11-21, jan. 1972. Disponível: <https://doi.org/10.1108/eb026526>. Acesso em: 27 mai. 2021.

LIU, Z.; CHEN, H. A Predictive Performance Comparison of Machine Learning Models for Judicial Cases. **2017 IEEE Symposium Series on Computational Intelligence**, SSCI, 2017. Disponível: <https://ieeexplore.ieee.org/document/8285436>. Acesso em: 27 mai. 2021.

NEMENYI, P. Distribution-free Multiple Comparisons. **International Biometrical Soc, DC 20005-2210**, v. 18, n. 2, p. 263, 1962.

OLIVEIRA, R. N. de; JUNIOR, M. C. Experimental Analysis of Stemming on Jurisprudential Documents Retrieval. **Information**, MDPI AG, v. 9, n. 2, 28, 2018. Disponível: <https://doi.org/10.3390/info9020028>. Acesso em: 27 mai. 2021.

ORENGO, V. M.; BURIOL, L. S.; COELHO, A. R. A Study on the Use of Stemming for Monolingual Ad-Hoc Portuguese Information Retrieval. In: **Evaluation of Multilingual and Multi-Modal Information Retrieval**. Springer Science & Business Media: Berlin, Germany, 2007. Disponível: https://doi.org/10.1007/978-3-540-74999-8_12. Acesso em: 27 mai. 2021.

ORENGO, V. M.; HUYCK, C. A Stemming Algorithm for the Portuguese Language. **Proceedings of the Eighth International Symposium on String Processing and Information Retrieval**, SPIRE 2001, Laguna de San Rafael, Chile, pp. 13–15, 2001; **IEEE**: Piscataway, NJ, USA, 2001, pp. 186-193. Disponível: <https://doi.ieee-computersociety.org/10.1109/SPIRE.2001.10024>. Acesso em:

27 mai. 2021.

PDFMiner. Disponível em: <https://pypi.org/project/pdfminer/>. Acesso em: 20 mai. 2021.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, v. 12, pp. 2825-2830, 2011. Disponível: <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>. Acesso em: 27 mai. 2021.

PORTER, M. F. An Algorithm for Suffix Stripping. **Program**, 1980.

PORTER, M. F. **Snowball**: A Language for Stemming Algorithms. 2001. Disponível: <http://snowball.tartarus.org/texts/introduction.html>. Acesso em: 27 mai. 2021.

PROJUDI – **Processo Judicial Digital**. Disponível em: <https://projudi.tjpr.jus.br/projudi/>. Acesso em: 28 jun. 2020.

RISH, I. An Empirical Study of the Naive Bayes Classifier. **IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence**, v. 3. n. 22, 2001. Disponível: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.330.2788&rep=rep1&type=pdf>. Acesso em: 27 mai. 2021.

SAVOY, J. A Stemming Procedure and Stopword List for General French Corpora. **J. Am. Soc. Inf. Sci.**, v. 50, pp. 944-952, 1999. Disponível: <http://members.unine.ch/jacques.savoy/Papers/FRJasis.pdf>. Acesso em: 27 mai. 2021.

SCHUSTER, M.; KULDIP K. P. Bidirectional Recurrent Neural Networks. **IEEE Transactions on Signal Processing**, v. 45, n. 11 pp. 2673-2681, 1997. Disponível: <https://ieeexplore.ieee.org/document/650093>. Acesso em: 27 mai. 2021.

SALTON, G.; WONG, A.; YANG, C. A Vector Space Model for Automatic Indexing. **Communications of the ACM**, v. 18, n. 11, pp. 613-620, 1975. Disponível: <https://doi.org/10.1145/361219.361220>. Acesso em: 27 mai. 2021.

SILVA, N. C. da; BRAZ, F. A.; CAMPOS, T. E. de; GUEDES, A. B. S.; MENDES, D. B.; BEZERRA, D. A.; GUSMÃO, D. B.; CHAVES, F. B. de S.; ZIEGLER, G. G.; HORINOUCI, L. H.; FERREIRA, M. H. P.; INAZAWA, P. H. G.; COELHO, V. H. D.; FERNANDES, R. V. de C.; PEIXOTO, F. H.; FILHO, M. S. M.; SUKIENNIK, B. P.; ROSA, L. da S.; SILVA, R. Z. M. da; JUNQUILHO, T. A.; CARVALHO, G. H. T. A. Document Type Classification for Brazil's Supreme Court Using a Convolutional Neural Network. **The Tenth International Conference on Forensic Computer Science and Cyber Law, ICoFCS**, São Paulo, Brasil, p. 7-11, 2018. Disponível: https://cic.unb.br/~teodecampos/ViP/correiaDaSilva_etal_icofcs2018.pdf. Acesso em: 27 mai. 2021.

ŞULEA, O. M.; ZAMPIERI, M.; MALMASI, S.; VELA, M.; DINU, L. P.; GENABITH, J. van. Exploring the Use of Text Classification in the Legal Domain. **2nd Workshop on Automated Semantic Analysis of Information in Legal Texts**, London United Kingdom, 2017.

ŞULEA, O. M.; ZAMPIERI, M.; VELA, M.; GENABITH, J. van. Predicting the Law Area and Decisions of French Supreme Court Cases. **Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP**, 2017. Disponível: <https://acl-bg.org/proceedings/2017/RANLP%202017/pdf/RANLP092.pdf>. Acesso em: 27 mai. 2021.

TURING, A. M. **Computing Machinery and Intelligence**. Computers & Thought. Cambridge: MIT Press, pp. 11-35, 1995.

VAN RIJSBERGEN, C. J. **Information Retrieval**. 2 ed. Butterworth-Heinemann, 1979. Disponível: <http://www.dcs.gla.ac.uk/Keith/Preface.html>. Acesso em: 27 mai. 2021.

Recebido em 27 de maio de 2021
Aceito em 14 de junho de 2021