

Machine Learning for Súmulas Vinculantes

Law: Beatriz Sabdin Chagas, Carla Marcondes Damian, Ana Clara Macedo Jaccoud,
Pedro Burlini de Oliveira

Math: Henrique Hennes, Jorge Poco, Jean Roberto Ponciano, Lucas Resck, Raphaël Tinarrage

O IMPACTO DA SÚMULA VINCULANTE 26 NA DIMINUIÇÃO DE DEMANDA SIMILAR NO STF: UMA ANÁLISE QUANTITATIVA POR MODELOS DE *MACHINE LEARNING*

THE IMPACT OF THE SÚMULA VINCULANTE 26 ON THE DECREASE OF SIMILAR DEMANDS AT THE STF: A QUANTITATIVE ANALYSIS WITH *MACHINE LEARNING* MODELS

Beatriz Sabdin Chagas
Carla Marcondes Damian
Raphaël Tinarrage

Resumo

A reforma do judiciário de 2004 permitiu ao Supremo Tribunal Federal (STF) criação de Súmulas Vinculantes (SV), enunciados que refletem jurisprudências consolidadas de caráter normativo. Entre seus objetivos, a SV pretende aumentar a eficiência judicial, ou seja, diminuir a quantidade de processos similares que chegam à última instância. Neste artigo, estudamos a SV 26, que versa a respeito da utilização de exame criminológico para progressão de regime em crimes hediondos. A fim de avaliar sua eficácia, analisamos o número de demandas ao STF relacionadas ao assunto da SV 26, antes e depois de sua criação. Para estimar quais dos 151.765 documentos do conjunto de dados se inscrevem no assunto da SV 26, mesmo antes de sua criação, aplicamos métodos de *Machine Learning*. Comparamos diferentes modelos clássicos e identificamos os mais promissores. Nossos resultados sugerem que a SV 26 cumpriu seu papel de eficácia jurídica, até 2015, quando o número de aplicações repetidas aumentou consideravelmente. Levantamos a hipótese de que este aumento reflete discussões e debates em torno da SV 26. Este artigo é a contrapartida metodológica de (JACCOUD et al., 2022), onde analisamos, de um ponto de vista jurídico-empírico, o uso do exame criminológico na progressão de regime para tipos penais hediondos.

Palavras-chave: Processamento de Linguagem Natural (NLP); Detecção automática de documentos; Súmula Vinculante 26; Diminuição de demanda similar; Progressão de regime.

Abstract/Resumen/Résumé

The 2004 reform of the judiciary allowed the Supreme Court (STF) to create *Súmulas Vinculantes* (SV), statements that reflect consolidated jurisprudence of a normative nature. Among its objectives, the SV is intended to increase judicial efficiency, that is, to reduce the number of similar cases that reach the highest court. In this article, we study the SV 26, which deals with the use of criminological examination for regime progression in heinous crimes. In order to evaluate its effectiveness, we analyze the number of cases brought before the STF and related to the topic of SV 26, before and after its creation. To estimate which of the 151,765 documents in the dataset fall under the scope of SV 26, even before its creation, we applied Machine Learning methods. We compared different classical models and identified the most promising ones. Our results suggest that SV 26 fulfilled its role of legal effectiveness, until 2015, when the number of similar cases increased considerably. We hypothesize that this increase reflects discussions and debates surrounding SV 26. This article is the methodological counterpart of (JACCOUD et al., 2022), where we analyze, from a legal-empirical point of view, the use of the criminological examination in regime progression for heinous crimes.

Keywords/Palabras-claves/Mots-clés: Natural Language Processing (NLP); Automatic document detection; Súmula Vinculante 26; Decrease of similar demands; Regime progression.

1 INTRODUÇÃO

O tipo de precedente conhecido como Súmula Vinculante (SV), como nos lembra Alexandre de Moraes, surgiu na reforma do judiciário em 2004, por meio da Emenda Constitucional nº 45, com o intuito de uniformizar decisões, a partir da importação da lógica de jurisdição, usual no *common law* (MORAES, 2021, p. 827). A justificativa principal para a introdução destes instrumentos pode ser entendida pela tentativa de padronização decisória, a qual objetiva alcançar isonomia e segurança jurídica, i.e., evitar que casos idênticos sejam decididos de maneiras distintas de modo a violar a garantia constitucional de igualdade perante a lei (CÂMARA, 2022, p. 20). De fato, tal utilização da SV está prevista na EC 45/2004, que acrescentou à Constituição Federal o art. 103-A (BRASIL, 2004), que obriga, para a criação de uma SV, o preenchimento de três requisitos: (i) versar sobre assunto objeto de controvérsia atual entre órgãos judiciários ou entre esses e a administração pública; (ii) representar risco de grave insegurança jurídica e; (iii) ser objeto de relevante multiplicação de processos sobre questão jurídica idêntica. A partir dos critérios, extrai-se que, como instrumento jurídico, a SV objetiva não somente garantir a isonomia nas decisões judiciais, como também aumentar a eficiência do Poder Judiciário, ao evitar demora na prestação jurisdicional.

Nota-se, assim, que a observância obrigatória atribuída aos enunciados apresenta ainda grande viés de eficiência, pois uma vez publicada uma SV, o assunto deixa de ser objeto de interpretação controversa e, por conseguinte, o tema deveria também deixar de originar processos com demandas idênticas e, principalmente, evitar que estes continuem sendo levados aos Tribunais Superiores. É então de interesse investigatório estabelecer, empiricamente, o grau de eficiência deste tipo de instrumento, aqui entendido como o grau de diminuição de casos semelhantes a respeito de certo tema que chegam à última instância de apelação. Neste artigo, analisamos o impacto da criação de uma Súmula Vinculante específica – a SV 26 – no número de pedidos a respeito do tema versado que foram recebidos pelo STF desde a criação do dispositivo.

A escolha da SV 26 como estudo de caso é justificada por seu histórico peculiar de elaboração, que será discutido na Seção 2. Na Seção 3, é analisada a série temporal de citações à Súmula (Seção 3.1). Para a estimativa do número de processos semelhantes antes de sua criação, no entanto, são elaborados detectores automáticos de citações potenciais, possibilitando medir assim o impacto desta criação sobre comportamento estatístico de recursos que chegam ao STF. Como o estudo automático de documentos textuais faz parte da

teoria de Processamento de Linguagens Naturais¹ (NLP, da sigla em inglês), cujas técnicas vão desde o uso de lógica simbólica a métodos recentes de *deep learning*², são sugeridas, na Seção 3.2, técnicas de vetorização (ou *embedding*) para as decisões. Na Seção 3.3, são relatados os processos de treinamento e predição de modelos de *Machine Learning* responsáveis pela detecção das potenciais citações em si.

Na Seção 4, é proposta uma abordagem mais complexa de classificação, com o objetivo de, não só identificar citações potenciais à SV 26, mas também distinguir aquelas em que, se fosse utilizado, o verbete teria importância significativa no *mérito* da decisão. Por fim, discutimos os resultados obtidos nesta abordagem e, na Seção 5, indicamos possíveis direções futuras. Vale notar que, apesar de estudos empíricos similares sobre a efetividade de Súmulas Vinculantes já terem sido conduzidos (AMARAL, 2018), não encontramos na literatura tentativas de detecção automática de utilização de Súmulas com esse propósito ou de diferenciação entre citações de importância em mérito e puramente processuais.

2 A SÚMULA VINCULANTE 26

No caso da Súmula Vinculante 26, publicada em dezembro de 2009, além destes objetivos genéricos para melhoria de eficiência, visava-se principalmente, a resolução de dois problemas que não encontravam soluções jurídico-normativas claras. O primeiro, seria quanto à retroatividade ou não da lei 11.464/07, que disciplinava como se dava a progressão de regime para crimes hediondos e equiparados, após a possibilidade desta progressão ser permitida em nosso ordenamento jurídico, e o segundo, quanto à possibilidade ou não de o juiz determinar a realização do exame criminológico na avaliação de progressão de regime prisional. Estes problemas, nota-se, se destacaram tanto politicamente, por incitarem discussões diversas, como juridicamente, por ensejarem alto número de processos a chegarem

¹ A palavra “natural” pode ser entendida tendo em vista que o objetivo deste campo de pesquisa é o estudo matemático a respeito de línguas comumente utilizadas por seres humanos, como português, inglês, etc., em contraposição às chamadas linguagens artificiais, comumente vistas como linguagens de programação.

² Para uma revisão das técnicas mais comuns de NLP, indicamos ao leitor Jones (1994), Cambria e White (2014) e Young et al. (2018). No campo jurídico, a NLP tem sido utilizada para diferentes tarefas, tais como traduzir documentos, resumi-los, estimar precedentes, ou a influência de opiniões, como em Hachey e Grover (2006), Farzindar e Lalpalme (2009), Resck et al. (2022) e Leibon et al. (2018).

ao STF, sobretudo em meados de 2006³, a partir da declaração de inconstitucionalidade do § 1º do art. 2º da lei de Crimes Hediondos, no HC 82.959/SP.

Prontamente, posterior a esta declaração de inconstitucionalidade, para assentar este entendimento no ordenamento jurídico o Congresso agiu e, por meio da lei 11.464/07, alterou a redação da norma para adequá-la a este entendimento. Contudo, além desta mudança, também alterou os lapsos temporais previstos para os crimes hediondos, tornando-os mais gravosos, de modo que era necessário assentar o entendimento acerca do uso da nova lei no tempo. A SV 26, deste modo, serviu para fortalecer na jurisprudência o princípio jurídico-penal de que a lei não pode retroagir em prejuízo do réu. Assim, a progressão seria possibilitada para todos os apenados, com a ressalva de que para os crimes praticados antes da data de publicação da nova lei, não se aplicam os novos lapsos, mas sim os previstos no artigo 112 da Lei de Execução Penal, anteriormente aplicado em ausência de lei específica.

Além disso, o STF decidiu e normatizou, através da SV 26, que o exame criminológico - não mais condicionante da progressão de regime a partir da promulgação da lei 10.792/2003, que suprimiu o parágrafo único do 112 da LEP - poderia ser realizado desde que mediante fundamentação idônea no pedido do juiz da vara de execução penal. Na concepção da Corte, a nova lei havia apenas abolido a obrigatoriedade do exame, mas não a possibilidade de uso dele pelo juiz, que deveria, em vistas do princípio da individualização da pena, manter certo grau de arbítrio para exercer seu poder geral de cautela, requisitando perícias.

A partir disto, foi consolidada seguinte redação ao verbete, que conduziria à solução de ambas problemáticas suscitadas:

“Para efeito de progressão de regime no cumprimento de pena por crime hediondo, ou equiparado, o juízo da execução observará a inconstitucionalidade do art. 2º da Lei 8.072, de 25 de julho de 1990, sem prejuízo de avaliar se o condenado preenche, ou não, os requisitos objetivos e subjetivos do benefício, podendo determinar, para tal fim, de modo fundamentado, a realização de exame criminológico.” (STF, 2009)

O que notou-se, contudo, a partir da implementação da SV 26, foi um crescimento exponencial de casos com demandas similares concernentes ao verbete. Recursos deste tipo,

³ No Anuário da Justiça de 2011, foi observado que também no STJ houve aumento de pedidos de progressão penal para crimes hediondos. Ele aponta que, entre 2004 e 2005, houve um aumento de 45%, enquanto de 2005 para 2006, de 87% no número de pedidos. Fonte: <https://www.conjur.com.br/2011-mai-29/numero-pedidos-habeas-corpus-stj-dobra-tres-anos>. Acesso em 29/08/2022.

visariam impugnar decisões de instâncias inferiores, que não haveriam acatado plenamente as novas teses jurídicas (AMARAL, 2016, p. 19). Caberia, então, comprovar, empiricamente, se os resultados da SV 26 foram tão benéficos para a jurisprudência, como o esperado.

3 ESTUDO EMPÍRICO DO IMPACTO DA SÚMULA VINCULANTE 26

Descreveremos agora a técnica utilizada no estudo empírico do grau de eficiência da SV 26. Ao invés de tentar definir uma métrica absoluta única capaz de medir, inequivocamente, por meio de um valor numérico, a capacidade da SV em diminuir os casos similares que chegam ao supremo, utilizaremos uma perspectiva mais qualitativa ao analisar as séries temporais das demandas a respeito de exame criminológico na progressão de regime para crimes hediondos. Diferentemente da abordagem por meio de métrica, acreditamos que a análise estatística apurada de toda a série temporal permite o melhor entendimento das possíveis causas de funcionamento ou falta de funcionamento da SV 26, justificando assim a escolha metodológica.

3.1 CONTABILIZAÇÃO DOS DOCUMENTOS CITANTES À SÚMULA VINCULANTE 26

O banco de dados utilizado para análise consiste em um conjunto de decisões do STF⁴, entre 1989 e 2018, mantido pelo projeto *Supremo em Números*⁵. Como o objetivo deste artigo é apenas o estudo da Súmula Vinculante 26, somente as decisões assinaladas pelo STF como de tema em “Direito Penal” ou “Direito Processual Penal” do banco de dados foram utilizadas, totalizando 151.765 documentos a respeito dos mais diversos tipos de recursos⁶.

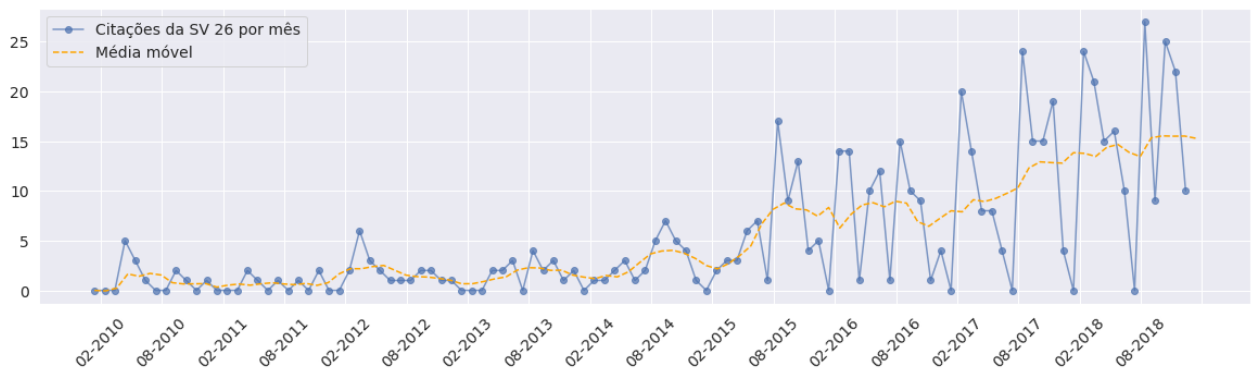
Um estudo inicial da capacidade prática da Súmula Vinculante em diminuir demandas repetitivas sobre o tema pode ser visto na Figura 1, onde o número de documentos citantes da SV 26, desde de sua criação até dezembro de 2018 é apresentado.

⁴ Vale notar que diversos esforços têm sido feitos para coletar dados de diferentes tribunais. Neste artigo, utilizaremos os dados do STF, mantidos pelo projeto *Supremo em Números*, porém, outros bancos de dados similares foram reportados, como o mantido pelo chamado Projeto Victor (PEIXTO, 2020).

⁵ Além de manter um importante banco de dados de decisões do STF, o *Supremo em Números* fez inúmeros estudos de jurimetria da corte, como pode ser visto em (PEREIRA et al., 2020) e originou outros tantos trabalhos (CORREIA et al., 2019; RESCK et al., 2022).

⁶ O banco de dados contém metadados dos processos que tramitam no tribunal e seus andamentos mantidos em uma base relacional. São mantidas também informações das decisões publicadas pelo tribunal e seu texto completo, o que foi utilizado no presente trabalho.

Figura 1 - Número absoluto de documentos citantes da SV 26 de 2010 a 2018 (azul escuro) e sua média móvel (laranja tracejado).



Pela imagem, fica explícito uma tendência significativa de aumento em citações após o segundo semestre de 2015. Uma análise manual dos textos indica que a partir desta data, uma grande quantidade de reclamações a respeito da fundamentação do pedido de exame criminológico, provavelmente foi impulsionada pelo aumento de reclamações sobre a fundamentação do pedido de exames por parte da Defensoria Pública de São Paulo, como abordado em artigo complementar (JACCOUD et al., 2022). Finalmente, observa-se pelo gráfico que o efeito esperado de unificar o entendimento, para que as disposições da Súmula fossem aplicadas sem discussão, de modo a diminuir o número de processos levantando a mesma demanda, não foi de todo atingido. A análise do gráfico evidencia, então, que a criação da Súmula Vinculante incitou uma nova onda de discussão sobre essa matéria, mas dessa vez, provavelmente, sobre questões de texto e aplicação ou não do enunciado.

3.2 VETORIZAÇÃO DAS DECISÕES

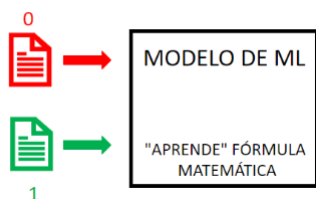
Para realizarmos a análise empírica completa da efetividade responsiva da Súmula Vinculante 26, seria necessário determinar qual o comportamento das demandas similares sobre o tema do verbete *anteriormente* à sua publicação, comparativamente ao comportamento após a publicação. Para isso, deveríamos analisar a quantidade dessas demandas repetitivas antes e após a publicação da SV, sendo possível considerar o instrumento efetivo se existirem indícios estatísticos suficientes que sugiram a queda deste volume após a aprovação da SV. Infelizmente, porém, apesar de o volume de decisões posteriores à criação da Súmula Vinculante 26 de tópico similar ao versado nesta ser acessível por meio da frequência de citações, como mostrado na Figura 1, o mesmo, claramente, não pode ser dito sobre as decisões anteriores a sua criação, dado que simplesmente, não existia

Súmula Vinculante apropriada a ser citada - assim, não podendo serem identificados casos de incidência dela de forma óbvia. Por conseguinte, fica claro que a resposta sobre o grau de diminuição de casos da SV 26 só é alcançável se identificadas as decisões que teriam suas resoluções baseadas na Súmula, caso esta existisse no momento, em nosso banco de dados. No entanto, dada a grande quantidade de documentos, trinta e oito mil novecentos e oitenta e dois (38.982), a verificação manual por especialistas da possibilidade de uma citação apropriada da SV 26, não apenas é um trabalho hercúleo, mas também fortemente propício a erros e divergência de classificação de acordo com a interpretação, por parte do leitor, a respeito de cada caso.

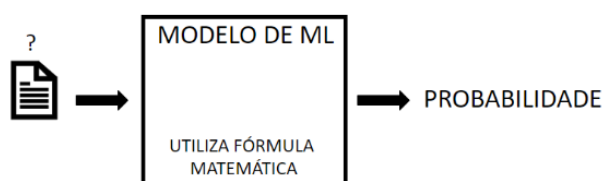
Surge então, como alternativa a elaboração de um modelo de *Machine Learning* (ML) capaz de identificar as decisões em que seria cabível aplicar a Súmula Vinculante 26, cujos *inputs* seriam decisões do STF e *outputs* a probabilidade de citação da SV 26. A Figura 2 representa uma construção genérica de tal modelo de detecção (aqui entendido simplesmente como uma fórmula matemática implícita), treinado a partir de decisões separadas em duas classes: a dos documentos que citam a SV 26, chamada, para fins de implementação, de classe “1” e a dos documentos em que uma citação da SV 26 seria inapropriada, chamada de “0”. Uma vez treinado, tal modelo é então capaz de identificar a probabilidade de um documento qualquer ser cabível de uma citação da SV 26, permitindo assim aferir quais documentos antes da criação da Súmula Vinculante provavelmente a teriam referenciado, caso esta existisse. A utilização do modelo para detecção de possíveis citações apropriadas, dentre as mais de 150.000 decisões sobre “Direito Penal” ou “Direito Processual Penal”, é chamado de estágio de previsão.

Figura 2 - Representação esquemática do processo de treinamento e generalização de um modelo de *Machine Learning* que visa prever a probabilidade de documentos pertencerem à chamada classe “1”. Note que, para o *dataset* de treino, é sabido, de antemão, a qual classe cada documento pertence, o que é indicado com a cor vermelha para os documentos em classe “0” e cor verde para os documentos na classe “1”. Entretanto, durante o processo de previsão, tal informação é, *a priori*, faltante. Em outras palavras, enquanto o *output* de todos documentos no *dataset* de treinamento é conhecido, o mesmo não segue para o *dataset* em geral ao qual se faz as previsões.

ESTÁGIO DE TREINAMENTO

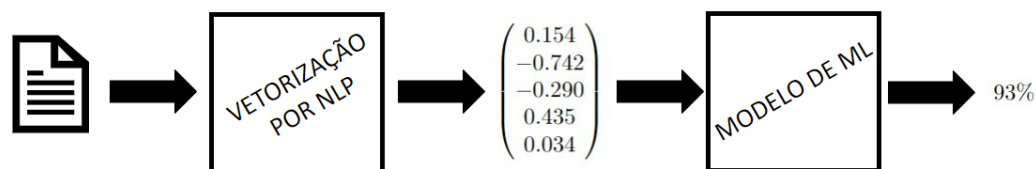


ESTÁGIO DE PREVISÃO



Antes da elaboração do modelo de *Machine Learning*, porém, é necessário transformar em vetores numéricos os documentos, permitindo assim que o modelo faça a leitura automática de cada decisão. Existem, no arcabouço da NLP inúmeras técnicas de vetorização, desde métodos puramente de contagem aos novíssimos *transformers* (VASWANI et al., 2017). Inicialmente, optamos pela utilização de técnicas de vetorização dos documentos baseadas em frequência de ocorrência de palavras no vocabulário, visto que, como indicado pela Tabela 1, a seguir, mesmo esses métodos mais simplistas foram suficientes para a obtenção de bons valores de validação cruzada. A primeira técnica de *embedding* utilizada neste estágio foi a chamada *bag-of-words* em que cada documento é representado por um vetor que conta a frequência de cada termo em sua composição. Devido à sua simplicidade, porém, o *bag-of-words* é raramente utilizado, principalmente ao notarmos que palavras de grande frequência entre os documentos, no nosso caso expressões como “ministro”, “tribunal”, etc são insignificantes para o problema de classificação em si. Apesar de o mascaramento de *stop words*⁷ aliviar a situação, em classificações temáticas, como a nossa, raramente a melhoria é significativa. Por tal motivo, compararmos também os resultados obtidos por *bag-of-words* com o *embedding* TF-IDF (*Term Frequency-Inverse Document Frequency*), em que a frequência de ocorrência dos termos (i.e a versão normalizada de *bag-of-words*) é multiplicada para cada palavra x pelo logaritmo da fração do número total de documentos e do número de documentos que utilizam x . A Figura 3 representa o caminho (ou *pipeline*, em linguagem do *Machine Learning*) pelo qual um documento de “Direito Penal” ou “Direito Processual Penal” passa até gerar o *output* que é a probabilidade de referir a tema similar ao versado na SV 26.

Figura 3 - *Pipeline* do processo de *Machine Learning* para detecção automática de citações potenciais



Vale notar, por fim, que nem mesmo o TF-IDF consegue resumir, de maneira totalmente eficiente, todas as informações textuais, principalmente por sua incapacidade de levar em conta a proximidade semântica entre sinônimos ou ainda marcações sintáticas

⁷ Que são palavras de maior frequência do idioma (no caso do Português, “de”, “o”, “a” são exemplos de *stop words*).

expressas pela ordem dos termos dentro das frases. De fato, esse será um problema encontrado na próxima seção, quando diferenciarmos as citações da SV 26 entre importância em mérito e puramente processuais. Felizmente, existem métodos ainda mais atuais de NLP que permitem vetorização mais abrangente, em especial, o chamado *Universal Sentence Encoder*, que faz uso de redes neurais artificiais⁸ e que será utilizado por nós como tentativa de solução ao problema.

3.3 DETECÇÃO DE POSSÍVEIS CITAÇÕES DA SÚMULA VINCULANTE 26 POR MEIO DE MODELOS DE MACHINE LEARNING

Uma vez representados como vetores, os documentos são utilizados pelo modelo de ML para aprender a função matemática capaz de prever a qual classe, não relacionada à SV 26 (“0”) ou relacionada à SV 26 (“1”) o documento pertence. Para o *dataset* de treino, utilizaremos todas as 565 decisões do STF que citam, diretamente a Súmula Vinculante 26 como classe “1”. Criar a classe de documentos não referentes à SV 26 é mais problemático. Isto pois, de maneira objetiva, o correto seria escolher, dentre todo o conjunto de texto disponíveis, aproximadamente 565 documentos que não fazem referência à tópico similar ao abordado na Súmula⁹. Apesar de tal escolha ser possível, devido à diferença de interpretação por parte dos leitores sobre a possibilidade ou não de utilização da SV 26 em cada decisão, dificilmente teríamos assim um conjunto de documentos referentes à classe “0” totalmente não-enviesados e de abrangência temática significativa. Por este motivo, optamos, ao invés de treinar apenas um modelo de *Machine Learning*, treinar vários modelos de mesmo tipo¹⁰ paralelamente, utilizando para cada um o mesmo conjunto de documentos de classe “1”, isto é, aqueles que citam diretamente a SV, e amostras aleatórias *diferentes* de documentos assinalados como “Direito Penal” ou “Direito Processual Penal” que não citam a Súmula. A probabilidade final de um documento ser apropriado para uma citação da SV 26 é, então, dada pela média aritmética dos *outputs* de cada um desses modelos, como ilustrado na Figura 4. Esse tipo de metodologia possui ampla justificativa estatística e é comumente chamado de

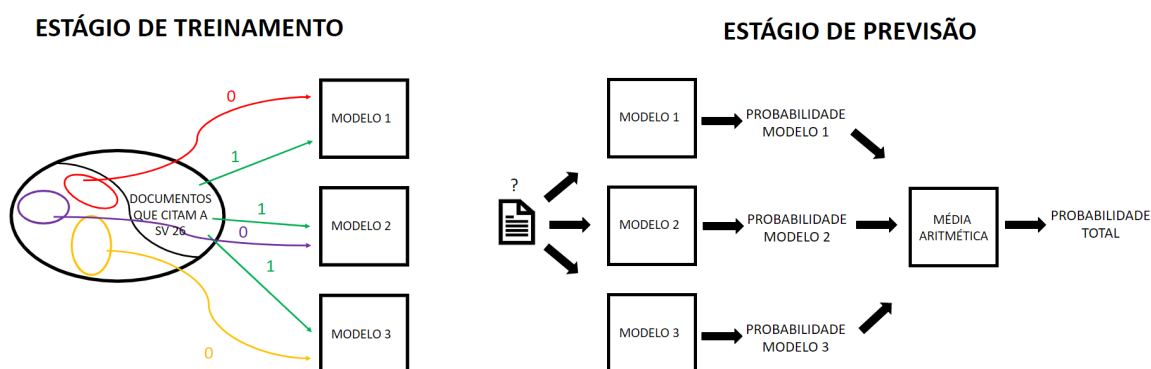
⁸ Apesar de uma explicação detalhada de seu funcionamento estar fora do escopo deste texto, recomendamos bibliografia YANG, Y. et al. **Multilingual Universal Sentence Encoder for Semantic Retrieval**. arXiv, 2019.

⁹ Como em qualquer problema de classificação e, principalmente, como um problema de detecção, é essencial o bom-balanceamento do *dataset* de treinamento.

¹⁰ Por “tipo de modelo” entende-se diferentes formas, como regressão logística, *Support Vector Machines*, *random forests*, etc.

*ensemble*¹¹. Note que a grande vantagem da utilização de *ensemble* pode ser entendida como, se por caso, por azar, alguma decisão que não cita a SV, mas que versa sobre tópico similar ao da Súmula seja incluído na classe “0”, o dano ao treinamento que tal documento pode causar ao modelo final é suavizado pela tomada da média aritmética, permitindo assim que, de fato, a verificação manual dos documentos a serem incluídos à classe “0” seja evitada.

Figura 4 - Esquemática da técnica de *ensemble* com três componentes, aplicada ao problema aqui descrito. Durante o treinamento (esquerda), se forma o *dataset* de treinamento da seguinte maneira. Primeiro, utilizamos todos os documentos que citam, diretamente, a SV 26 como sendo de classe “1” (em cor verde). Depois, se retira, aleatoriamente, três conjuntos de documentos de mesmo tamanho (565, no nosso caso) que *não* citam a Súmula e a todos se assinala a classe “0” (em cores vermelho, roxo e laranja). Então, os documentos de classe “1” e um dos conjuntos de documentos de classe “0” para cada modelo componente do *ensemble* que são assim treinados separadamente. Note que, na prática, isso significa que cada um dos componentes do *ensemble* é treinado a partir de um conjunto de dados diferente. Por fim, no momento de previsão (direita), o documento novo é introduzido a todos os três componentes e suas probabilidades finais são resumidas em média aritmética, produzindo assim o *output*.



Antes de analisarmos o resultado da aplicação de *Machine Learning* ao nosso problema específico, vamos citar, rapidamente, os modelos por nós escolhidos. Por se tratar de vetorizações de textos por contagem, é de se esperar alto grau separação linear¹², o que sugere o uso de modelos simples como regressão logística¹³. A Tabela 1 mostra os erros de

¹¹ Problemas de classificação como este geralmente utilizam-se do chamado *ensemble* por votação e não por média aritmética, como descrito aqui (GÉRON, 2019, p. 189). No entanto, tal abordagem requer a definição, *a priori*, de um limiar de aceitação, o que, como explicado a seguir no texto principal, não é de nosso interesse.

¹² Tal afirmação pode ser aferida sobre o problema específico aqui tratado pela Tabela 1 em RESCK et al. (2022).

¹³ Por brevidade, não vamos aqui discutir a construção exata desse modelo de classificação, mas indicar, ao leitor interessado, as seguintes referências (BISHOP, 2006, p. 205; GÉRON, 2019, p. 144). Notamos que (RESCK et al., 2022) indicaram uma melhora em acurácia em um problema com objetivo similar ao nosso ao utilizar *Support Vector Machines* com *kernel* linear, porém, como tais técnicas são, em essência, não probabilísticas em previsão, optamos por incluir neste texto, apenas, a regressão logística. No problema tratado por esse artigo,

validação cruzada¹⁴ para diferentes configurações do modelo de regressão logística, utilizando-se de *bag-of-word*, TF-IDF e *ensemble* com TF-IDF, respectivamente. Vale lembrar que, como explicado no parágrafo anterior, quando *ensemble* é utilizado, cada modelo é treinado separadamente do outro, juntando-os apenas ao fazer previsões, o que significa que existe mais de um erro de validação cruzada nestes casos.

Tabela 1 - Erros de validação cruzada calculados para cada configuração de vetorização e modelo de *Machine Learning*. Em todos os casos em que *ensemble* foi utilizado, foram treinados um total de três modelos do mesmo tipo (i.e regressão logística). Quando *ensemble* foi utilizado, três modelos foram treinados e o valor indicado é do menor erro em validação cruzada.

Técnica de vetorização e modelo de <i>Machine Learning</i>	Erro de validação cruzada
Vetorização: <i>bag-of-words</i> Modelo: regressão logística, sem <i>ensemble</i>	0.97
Vetorização: <i>TF IDF</i> Modelo: regressão logística, sem <i>ensemble</i>	0,98
Vetorização: <i>TF IDF</i> Modelo: regressão logística, com <i>ensemble</i>	0,97

A Figura 5 apresenta os coeficientes relacionados a cada palavra do vocabulário, em grau de importância, para determinar se um documento pertence à classe “0” ou “1” calculados a partir do aprendizado de um modelo em regressão logística em que os documentos foram vetorizados utilizando *bag-of-words*. Para fins de interpretação do gráfico, quanto maior a frequência no texto de uma palavra representada pelas barras positivas (em azul), maior a chance do texto específico pertencer a classe “1”, isto é, ser de tema similar à Súmula, enquanto que quanto maior a frequência de uma palavra de barra negativa (vermelho), menor esta probabilidade. Assim, como era de se esperar, um texto contendo

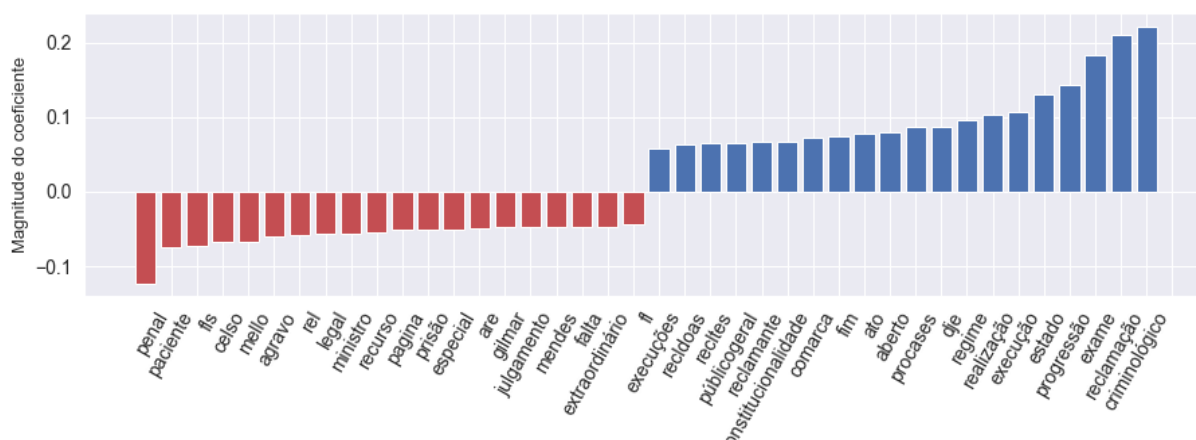
também foram utilizadas técnicas de ML para a detecção de potenciais citações de SVs, mas ao invés de utilizar apenas uma Súmula, foi-se criado um algoritmo para todas as 10 Súmulas mais citadas.

¹⁴ Validação cruzada (*cross validation*) é um estimador de acurácia de modelos de ML, onde uma validação cruzada igual a 0 indica um modelo que sempre erra e a 1 um modelo que sempre acerta. Note que, diferentemente de métricas mais tradicionais de desempenho de modelos, a validação cruzada não requer estabelecer, *a priori*, uma porcentagem fixa do *dataset* para avaliação do modelo – algo de grande importância para nós, dado o número pequeno de dados disponíveis para treinamento, porém, permite que todos os dados disponíveis sejam utilizados para o treinamento, o que, em geral, significa maior acurácia, principalmente em casos, como o nosso, em que o volume de dados é relativamente pequeno. No entanto, validação cruzada não permite uma interpretação de generalização, isto é, se estimada em um valor 0.8, por exemplo, não temos motivos para acreditar que o modelo acertará 80% das vezes quando deparado com dados novos, algo que é verdade para algumas outras métricas de desempenho.

palavras como “reclamação”, “criminológico” e “hediondo” é bem mais provável de versar sobre tema similar ao da SV 26 do que um texto que possui, apenas, “penal”, “extraordinário” e “república”. Note que isso não significa que uma decisão da classe “1” não pode conter esse último conjunto de palavras em sua redação, apenas que, sozinhas, elas não são suficientes para indicar alta probabilidade do documento pertencer a tal classe.

Além disso, quanto maior a barra positiva relacionada a uma palavra, maior o impacto que ela possui em aumentar a probabilidade de uma decisão ser de classe “1”. Por exemplo, a presença de “progressão” possui maior efeito de definir o texto como relacionado à classe similar ao versado na Súmula que “fim”. Note que as palavras “Súmula”, “Vinculante”, “SV” e “26” não estão entre as mais importantes na determinação de probabilidade de citação da Súmula, o contrário que se esperaria, a princípio, de um modelo que foi treinado como classe “1” sendo exatamente os textos que *citam* a Súmula. Isto, porém, pode ser entendido por meio de uma escolha metodológica. Porque o objetivo da criação do modelo é identificar documentos relacionados com a Súmula *antes* da criação desta, optamos, durante o processo de aprendizado do modelo, por “mascarar” essas palavras. Também foram mascarados caracteres numéricos (evitando o aprendizado de datas por parte do modelo) e a palavra “São Paulo” e suas derivadas, uma vez que mais de 70% dos documentos citantes à SV 26 são de origem deste estado¹⁵. Assim, esperamos que os algoritmos sejam capazes de identificar possíveis documentos semelhantes pelo contexto, e não pelas marcações de citação em si, o que de fato parece o caso se considerarmos a Figura 5.

Figura 5 - Coeficientes do modelo de regressão logística das variáveis (i.e entradas do vetor) correspondentes a cada uma das palavras apresentadas por vetorização *bag-of-words*.

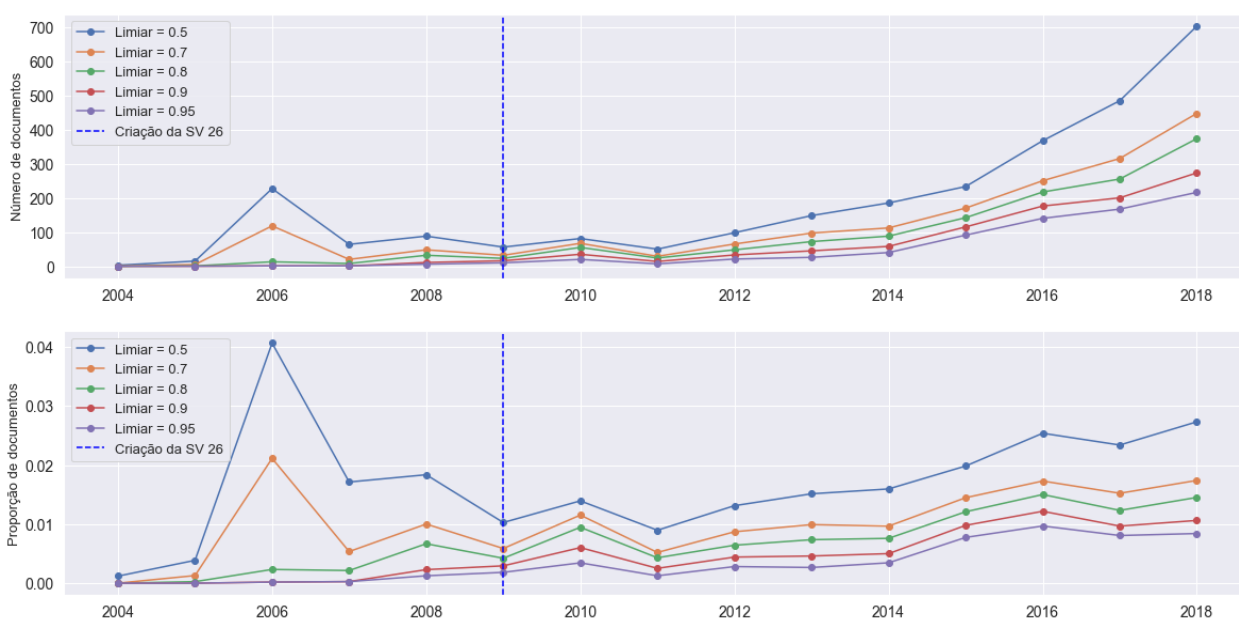


Finalmente, é possível fazer uso dos modelos treinados para a elaboração das séries temporais de frequência de decisões sobre casos repetitivos no STF, incluindo também

¹⁵ Em (JACCOUD et al., 2022) fazemos uma ampla investigação objetivando o entendimento dessa quantidade significativa de processos oriundos de São Paulo.

aquelas anteriores à criação da SV 26, como pode ser constatado na Figura 6. As diferentes curvas para um mesmo modelo indicam o chamado limiar de aceitação, isto é, a probabilidade mínima de *output* em que consideramos um documento como de tema provavelmente similar ao versado na Súmula. Note que valores de limiar muito pequenos, como de 0.5, fazem com que a frequência de documentos similares detectados seja superestimada, visto que muitos falsos positivos (i.e. documentos que possuem alguma similaridade com os documentos que propriamente citam a SV, mas não seriam cabíveis de citação apropriada) são escolhidos como pertencentes à classe semelhante. Por outro lado, valores de limiar muito altos, como 0.95 fazem com que muitos casos em que citações da SV são perfeitamente cabíveis acabem passando despercebidos e, assim, subestimamos a frequência de demandas repetitivas sobre o tema. Desta maneira, a escolha de um limiar ótimo se torna essencial para que todos os casos parecidos com o tema do verbete - e apenas esses - sejam contabilizados no estudo da série de frequências. Existe, na literatura de *Machine Learning*, vasta exploração sobre boas técnicas de escolha de limiar (BISHOP, 2006; HAND, 2009; BROWNLEE, 2022; FANOUS, 2021) contudo, como nosso objetivo aqui é desenvolver apenas um estudo quantitativo inicial do impacto empírico da SV, optamos por não fazer uma escolha única de limiar de probabilidade, mas sim analisar as séries temporais em diferentes configurações.

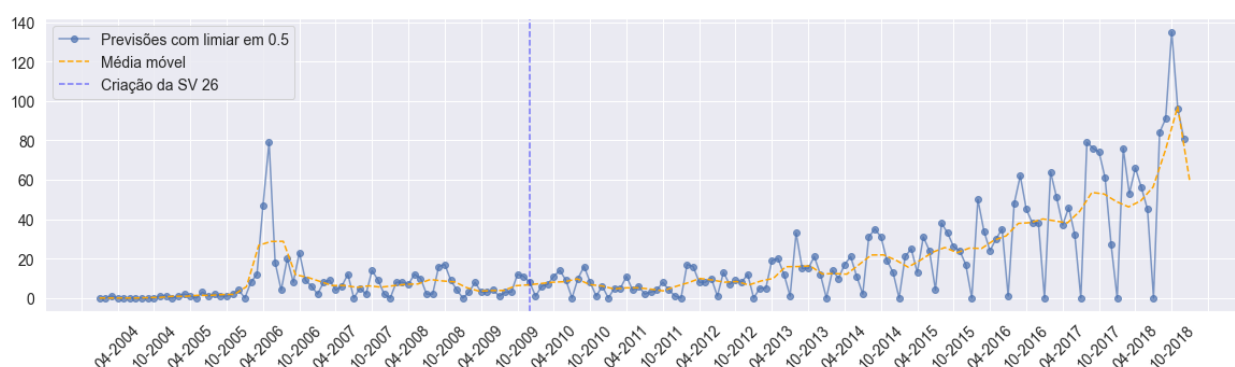
Figura 6 - Número absoluto (acima) e proporção (abaixo) de documentos detectados com probabilidade acima do limiar de decisão por ano, de 2004 a 2018. O ano de criação da SV 26 (2009) é destacado pela linha vertical azul tracejada.



A Figura 6 parece indicar que a quantidade de casos similares que chegam à última instância, para os valores de limiar inferiores a 0,7, teve um grande pico em 2006, e se manteve constante após a criação da Súmula Vinculante 26, com a criação de uma tendência de subida após 2014, algo já sabido a partir da contabilização do número de citações representada na Figura 1 da seção anterior. Tal comportamento fica ainda mais aparente ao considerarmos a série temporal mensal do previsto pelo modelo, na Figura 7 abaixo, onde apenas as previsões em limiar 0.5 são mostradas. Não apenas se pode perceber a partir do gráfico que o pico observado em 2006 ultrapassa em muito os anos adjacentes, mas também que tal valor, exacerbadamente acima do esperado, é concentrado apenas em poucos meses do ano, a dizer março e abril.

Temos algumas hipóteses, que merecem mais atenção em um momento futuro, para explicar a curva encontrada. Acreditamos que a justificativa do pico em 2006 é justamente a declaração de inconstitucionalidade do §1º, art. 2º da lei de Crimes Hediondos, com mais demandas entrando no sistema brasileiro sobre *como fazer progredir* condenados por crimes previstos nesta lei. A reação quase instantânea do Congresso Brasileiro em discutir uma lei que remediasse a inconstitucionalidade decretada pelo STF, com a apresentação de um projeto de lei já em 23/03/2006¹⁶, explicaria a queda do pico de 2006, visto na Figura 7 abaixo.

Figura 7 - Número absoluto de documentos detectados com probabilidade acima de 0.5 por mês, de 2004 a 2018 (azul escuro) e sua média móvel (laranja tracejado). O ano de criação da SV 26 (2009) é, novamente, destacado pela linha vertical azul tracejada.



¹⁶ Ver o histórico em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=318362>.

4 CONSIDERAÇÕES A RESPEITO DO MÉRITO DE USO DA SÚMULA VINCULANTE 26

A metodologia supra descrita pode ser criticada com base na formação dos dados da classe “1”, isto é, a classe de documentos relacionados à SV 26. Isso pois, ao formarmos a lista dessas decisões, consideramos todas que citam a Súmula Vinculante, ou seja, mesmo documentos que a utilizam de forma puramente processual, tendo pouca ou nenhuma importância no mérito da decisão.

Para clarificar a definição, pode-se entender uma decisão em mérito como aquela que explora os efeitos da súmula no caso do processo específico, ou seja, a decisão argui se no caso incide ou não a Súmula Vinculante, e o porquê dessa aplicação ou não. Por outro lado, as decisões puramente processuais foram entendidas como todas que não são de mérito. Estas, de modo geral, simplesmente enunciam que o processo gerador foi feito de modo procedimentalmente equivocado (regras do Regimento Interno do STF ou do CPC, como perda superveniente de objeto, litispendência, etc) ou que faltam informações importantes ao julgamento, subsequentemente requisitadas ao órgão público responsável (MP, Presídio, etc). Vale notar que existem decisões que, por mais que sejam viciadas processualmente, contém julgamento de mérito. Da mesma forma, existem casos de mérito com não aplicação da súmula assinalados como puramente processuais, quando o relator não justifica o motivo da não aplicabilidade do precedente.

Um modelo treinado a identificar como classe “1” também decisões puramente processuais tenderá a entender certos tópicos, claramente não relacionados ao uso de exame criminológico para progressão de regime em crimes hediondos como da classe de possíveis citações à Súmula, causando assim uma inflação da quantidade de citações potenciais. Para resolver essa questão, treinamos modelos em que, não só documentos que não citam a SV foram chamados de classe “0”, mas também citações puramente processuais nessa mesma classe. Para a separação entre documentos que citam a SV avaliando-a em mérito e aqueles que a citam em casos apenas processuais, uma verificação manual¹⁷ de cada decisão por meio de anotações coletadas pelo *software doccano* (NAKAYAMA et al., 2018) foi utilizada, classificando-se um total de 485 documentos citantes da SV 26 entre processual e mérito. *Datasets* de treino com 251 documentos em classe “1” e 252 documentos de classe “0” (i.e

¹⁷Em primeiro momento, a classificação foi feita apenas por um pesquisador. Para as demais etapas de verificação manual da pesquisa, contudo, será adotada uma verificação e cruzamento de classificações de pesquisadores distintos, permitindo assim a utilização de métodos de votação e estatística *fuzzy*.

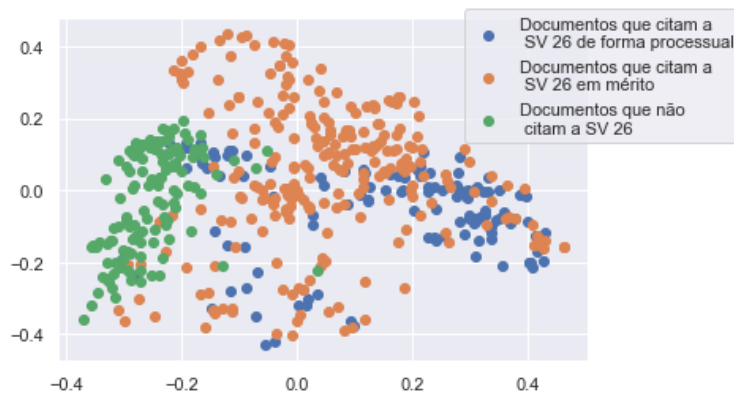
documentos com citações processuais ou sem citação alguma) foram criados e a validação cruzada por regressão logística é indicada na segunda linha da Tabela 2, abaixo.

Tabela 2 - Erros de validação cruzada calculados para cada configuração de vetorização por TF-IDF e modelo de *Machine Learning*, com cada modelo treinado três vezes por *ensemble*, como descrito no texto principal. O do erro em valor validação cruzada mostrado representa a menor destas quantias.

Modelo	Validação cruzada
Regressão logística	0,80
<i>Support Vector Machine</i> com kernel RBF	0,69
<i>Random forest</i>	0,78
<i>Naive Bayes</i> com distribuição multinomial	0,68

Uma primeira justificativa para a observada queda de acurácia do modelo é que, nele, os dados utilizados para treinamento são menos da metade do *dataset* de aprendizado utilizado na Tabela XX. Entretanto, mesmo com um *dataset* reduzido, um modelo em regressão logística que não diferencia citações processuais de mérito apresenta erro de validação cruzada semelhante (terceira linha da Tabela XX), indicando que o problema não é a falta de dados propriamente dita. Concluiu-se, então, que a queda de capacidade de aprendizado surge da maior dificuldade do problema em si. De fato, como demonstrado na Figura 8, em que uma redução de dimensionalidade por PCA é utilizado no *dataset* de treinamento vetorizado por TF-IDF, apesar dos documentos que citam a SV 26 e dos que falham em citar serem aproximadamente linearmente separáveis, mesmo em dimensão tão baixa, o mesmo não pode se dizer sobre os documentos que citam a SV 26 de forma puramente processual e os que citam em mérito.

Figura 8 - Projeção em duas dimensões, por PCA, do *dataset* de treino com as três classes de documentos distintas destacadas: “Documentos que citam a SV 26 de forma processual” (azul), “Documentos que citam a SV 26 em mérito” (laranja) e “Documentos que não citam a SV 26” (verde).



Se apresenta, então, como possibilidade, a utilização de modelos não-lineares, por exemplo *Support Vector Machine* (SVM) com kernel RBF, *Random Forests* e *Naive Bayes* (NB) com distribuição multinomial - que possui histórico de boa aplicação em problemas linguísticos de classificação¹⁸ (RASCHKA, 2014; SANG-BUM KIM et al., 2006; FRANK, 2006). A Tabela 2 traz também os erros de validação cruzada para esses modelos alternativos. É interessante notar que os modelos SVM e NB possuem métricas de treinamento significativamente piores que as dos outros dois.

As ainda baixas validações cruzadas parecem sugerir que o uso de *embeddings* por contagem, como o TF-IDF, são insuficientes para diferenciar decisões na utilização por mérito ou processual. Na verdade, esta é uma conclusão esperada, visto que muitas vezes documentos que utilizam processualmente da SV 26 citam, *ipsum litteris*, outras decisões que, por sua vez, utilizam da Súmula em mérito. Assim, independente do modelo de *Machine Learning* utilizado, é extremamente difícil uma classificação apropriada. Surge, então, a possibilidade do uso de técnicas *embeddings* mais avançadas, baseados em redes neurais profundas, capazes de “entender” as mesmas nuances textuais que diferenciam os dois tipos de citação e que foram observadas pelos leitores. Dentre essas vetorizações, destaca-se o chamado *Universal Sentence Encoder* (USE) (YANG et al., 2019), devido ao seu cuidadoso processo de aprendizado multi-linguístico, em especial, em português. Justificado por tal aprendizado, optamos pelo não retreinamento do modelo com os textos de decisões do STF, utilizando apenas o USE para predizer o *embedding*. A Tabela 3 mostra os erros de validação cruzada para alguns dos modelos antes discutidos¹⁹, utilizando-se, porém, da vetorização USE. Vale notar que a regressão logística possui queda significativa com esse *embedding* em comparação ao TF-IDF, o que parece indicar que os modelos de rede neurais tendem a criar vetorizações com menor separação linear, algo que será abordado em trabalhos futuros.

Tabela 3 - Erros de validação cruzada calculados para cada configuração de vetorização por USE e modelo de *Machine Learning*, com cada modelo treinado três vezes por *ensemble*, como descrito no texto principal. O do erro em valor validação cruzada mostrado representa a menor destas quantias..

Modelo	Validação cruzada
Regressão logística	0,72
<i>Support Vector Machine</i> com kernel RBF	0,80
<i>Random forest</i>	0,74

¹⁸ Apesar de alternativas, a princípio, possíveis, redes neurais profundas não são indicadas para esse problema tendo em vista a baixíssima quantidade de documentos disponíveis para treinamento.

¹⁹ Como a versão de Naive Bayes antes utilizada assumia distribuições multimodais (de suporte positivo) e USE possivelmente gera vetores com entradas negativas, esse modelo não pôde ser aqui testado.

É curioso que mesmo para o modelo não linear *random forest*, o USE representa uma queda considerável em acurácia de validação cruzada do que para o *embedding* mais simples por TF-IDF. Pode-se pensar que a causa dessa piora de generalização está na falta de retreinamento do modelo: como o USE inicialmente aprendeu a partir de textos de tópicos variados, a alta proporção de documentos sobre Direito, em especial Direito Penal, pode fazer com que os vetores estimados estejam muito mais próximos do que se esperaria no caso de retreinamento a partir dos dados.

5 CONSIDERAÇÕES FINAIS

Neste artigo, analisamos a eficiência da Súmula Vinculante 26 de maneira empírica, estimando o impacto de sua criação no número de recursos similares no STF. Para a análise de documentos após o estabelecimento da SV, uma contagem simples das citações mensais foi realizada, indicando um aumento significativo de decisões referenciando o verbete após 2015.

Além disso, utilizamos também métodos de *Machine Learning* para a detecção automática de decisões sobre o tema antes da criação da SV 26. Do ponto de vista metodológico, concluímos que a utilização de TF-IDF e regressão logística mostraram-se suficientes para a solução do problema proposto. Ficou comprovado, assim, que os casos de incidência temática do verbete aumentaram ao longo do tempo. Atribuímos à primeira declaração de inconstitucionalidade do § 1º do art. 2º da lei de Crimes Hediondos o primeiro grande aumento do número de casos com demandas similares às concernentes ao verbete, em 2006. Este debate, e consequentemente maior possibilidade de abertura de incidência da inconstitucionalidade, logo se cessaram dada pacificação do entendimento pelo Congresso, ao editar lei 11.464/07. Em seguida, verificamos grande diminuição do número de casos sobre a temática, até a criação da Súmula Vinculante 26. A partir de então, identificamos uma ascendência do debate e aumento do número de casos por pequeno período de tempo, atribuídos à repercussão inicial do verbete. Logo, uma tendência de constância foi notada ao longo dos anos, contudo, com tendência de subida após 2014, mantida até dezembro de 2018, último período analisado. Sobre este aumento, teoricamente inesperado ao funcionamento geral das súmulas, supomos que decorre de pressões para mudança de entendimento vinda de instâncias inferiores, sobretudo do Estado de São Paulo, como exploramos com mais detalhes em outro artigo. (JACCOUD et al., 2022)

Infelizmente, porém, tais técnicas foram incapazes de distinguir citações que chamamos de "puramente processuais", daquelas em que o impacto de mérito é significativo,

mesmo ao recorrermos a técnicas de *embedding* mais avançadas como *transformers*. Será objeto de trabalhos futuros propor novos algoritmos que permitam tal separação e verificar, manualmente, a qualidade da detecção dos modelos, para que possamos, ao fim, analisar apenas a utilização prática, de mérito, da Súmula Vinculante 26 no Supremo Tribunal Federal, extraíndo, enfim, dados concretos sobre sua utilização ao longo do tempo.

6 REFERÊNCIAS

AMARAL, Thiago Bottino Do. **A súmula vinculante vincula? Um estudo da eficiência da súmula vinculante nº 24.** REVISTA BRASILEIRA DE CIÊNCIAS CRIMINAIS, v. 143, p. 177-219, 2018.

AMARAL, Thiago Bottino Do. **Habeas corpus nos Tribunais Superiores: uma análise e proposta de reflexão.** 2016, p. 19

BISHOP, C. **Pattern recognition and machine learning.** Springer Verlag, 2006.

BRASIL, Constituição (1988). **Emenda constitucional n.º 45, de 30 de dezembro de 2004.** Altera dispositivos dos arts. 5º, 36, 52, 92, 93, 95, 98, 99, 102, 103, 104, 105, 107, 109, 111, 112, 114, 115, 125, 126, 127, 128, 129, 134 e 168 da Constituição Federal, e acrescenta os arts. 103-A, 103B, 111-A e 130-A, e dá outras providências. Disponível em: <http://www.planalto.gov.br/ccivil_03/Constituicao/Emendas/Emc/emc45.htm>. Acesso em: 24 ago. 2022.

BROWNLEE, J. **A Gentle Introduction to Threshold-Moving for Imbalanced Classification.** Disponível em: <<https://machinelearningmastery.com/threshold-moving-for-imbalanced-classification/>>. Acesso em: 25 ago. 2022.

C MARA, Alexandre F. **Manual de Direito Processual Civil.** 1ª ed. Rio de Janeiro: Grupo GEN, 2022. E-book. 9786559771776. Disponível em: <<https://integrada.minhabiblioteca.com.br/#/books/9786559771776/>>. Acesso em: 24 ago. 2022.

CAMBRIA, E.; WHITE, B. Jumping NLP Curves: A Review of Natural Language Processing Research. IEEE Computational Intelligence Magazine, v. 9, n. 2, p. 48–57, maio 2014.

CORREIA, F. A. et al. **An Exploratory Analysis of Precedent Relevance in the Brazilian Supreme Court Rulings.** Proceedings of the ACM Symposium on Document Engineering 2019, 23 set. 2019.

FANOUS, K. **Classification Models and Thresholds.** Disponível em: <<https://towardsdatascience.com/classification-models-and-thresholds-97821aa5760f>>. Acesso em: 29 ago. 2022.

FARZINDAR, A.; LAPALME, G. **Machine Translation of Legal Information and Its Evaluation**. *Advances in Artificial Intelligence*, v. 5549, p. 64–73, 2009.

FRANK, E.; BOUCKAERT, R. R. **Naive Bayes for Text Classification with Unbalanced Classes**. Disponível em: <https://link.springer.com/chapter/10.1007%2F11871637_49>. Acesso em: 29 ago. 2022.

GÉRON, Aurélien. **Hands-on machine learning with Scikit-Learn and TensorFlow concepts, tools, and techniques to build intelligent systems**. O'Reilly Media, Inc., 2019.

HACHEY, B.; GROVER, C. **Extractive summarisation of legal texts**. *Artificial Intelligence and Law*, v. 14, n. 4, p. 305–345, 8 mar. 2007.

HAND, D. J. **Measuring classifier performance: a coherent alternative to the area under the ROC curve**. *Machine Learning*, v. 77, n. 1, p. 103–123, 16 jun. 2009.

JACCOUD, Ana Clara; OLIVEIRA, Pedro Burlini De; TINARRAGE, Raphaël. **Progressão de regime em crimes hediondos no Supremo Tribunal Federal: uma análise empírica pela Súmula Vinculante 26**. Submetido à publicação. 2022.

JONES, K. S. **Natural Language Processing: A Historical Review**. *Current Issues in Computational Linguistics: In Honour of Don Walker*, v. 9, p. 3–16, 1994.

LEIBON, G. et al. **Bending the law: geometric tools for quantifying influence in the multinetwork of legal opinions**. *Artificial Intelligence and Law*, v. 26, n. 2, p. 145–167, 9 mar. 2018.

MORAES, Alexandre de. **Direito Constitucional**. São Paulo: Atlas, 2021. p. 827.

NAKAYAMA, Hiroki et al., **Doccano: Text Annotation Tool for Human**. GitHub, 2018. Software available from <https://github.com/doccano/doccano>.

PEIXTO, F. H. **Projeto Victor: relato do desenvolvimento da inteligência artificial na repercussão geral do Supremo Tribunal Federal**. *Revista Brasileira de Inteligência Artificial e Direito-RBIAD*, v. 1, n. 1, 2020.

PEREIRA, T.; WERNECK ARGUELHES, D.; DA FRANCA COUTO FERNANDES DE ALMEIDA, G. **VIII Relatório Supremo em Números : Quem decide no Supremo? : tipos de decisão colegiada no tribunal**. Rio de Janeiro: FGV Direito Rio, 2020.

RASCHKA, S. **Naive Bayes and Text Classification I - Introduction and Theory**. arXiv, 2014.

RESCK, L. E. et al. **LegalVis: Exploring and Inferring Precedent Citations in Legal Documents**. *IEEE Transactions on Visualization and Computer Graphics*, p. 1–1, 2022.

SANG-BUM, Kim et al. **Some Effective Techniques for Naive Bayes Text Classification.** IEEE Transactions on Knowledge and Data Engineering, v. 18, n. 11, p. 1457–1466, nov. 2006.

VASWANI, A. et al. **Attention Is All You Need.** arXiv, 2017.

YOUNG, T. et al. **Recent Trends in Deep Learning Based Natural Language Processing.** IEEE Computational Intelligence Magazine, v. 13, n. 3, p. 55–75, ago. 2018.