

# PESQUISA QUANTITATIVA NO DIREITO: UMA REVISÃO BIBLIOGRÁFICA

Bruna Cabrera De Bonito<sup>1</sup>  
Vanessa Korasaki<sup>2</sup>  
Cristina Veloso de Castro<sup>3</sup>

## RESUMO

Tradicionalmente, há um predomínio da pesquisa qualitativa sobre a quantitativa nos estudos acadêmicos da área jurídica. Entretanto, a utilização de métodos quantitativos no Direito (análise estatísticas, modelagem, simulação, entre outros) tornou-se uma área em ascensão no ramo jurídico, buscando compreender os impactos legais de forma mais precisa e objetiva. Nesse sentido, o presente artigo tem como objetivo central a revisão da literatura disponível sobre o cenário da pesquisa quantitativa no Direito, aprofundando em alguns trabalhos empíricos realizados com a legislação, jurisprudência e patentes. Metodologicamente, trata-se de um estudo descritivo, cuja base de dados utilizada foi o Google Acadêmico. Este trabalho justifica-se na medida em que a pesquisa jurídica tomada, em sua grande maioria, por estudos qualitativos, pode perder o contato com a realidade fática, vez que as ideias não são testadas nem confrontadas com dados. Logo, com a revisão de bibliografia, busca-se demonstrar sua importância, assim como exemplificar possibilidades de aplicação a futuros trabalhos. Nesse sentido, o diálogo entre teoria e dados se somam, eis que enquanto alguns estudos priorizam o teste de hipótese, outros almejam a elucidação sobre observações específicas e desenvolvimento de novas hipóteses.

**PALAVRAS-CHAVE:** análise de dados; dados jurídicos; processamento de linguagem natural; estatística.

<sup>1</sup> Universidade do Estado de Minas Gerais.

<sup>2</sup> Universidade do Estado de Minas Gerais.

<sup>3</sup> Universidade do Estado de Minas Gerais.

# QUANTITATIVE RESEARCH IN LAW: A BIBLIOGRAPHIC REVIEW

Bruna Cabrera De Bonito  
Vanesca Korasaki  
Cristina Veloso de Castro

## ABSTRACT

Traditionally, there has been a predominance of qualitative research over quantitative research in academic studies in the legal field. However, the use of quantitative methods in Law (statistical analysis, modeling, simulation, among others) has become a growing area in the legal field, seeking to understand legal impacts in a more precise and objective way. In this sense, this article has as its main objective the review of the available literature on the scenario of quantitative research in Law, delving into some empirical work carried out with legislation, jurisprudence and patents. Methodologically, this is a descriptive study, whose database used was Google Scholar. This work is justified to the extent that legal research taken, for the most part, by qualitative studies, can lose contact with factual reality, since ideas are not tested or confronted with data. Therefore, with the review of bibliography, we seek to demonstrate its importance, as well as to exemplify possibilities of application to future works. In this sense, the dialogue between theory and data is added, since while some studies prioritize hypothesis testing, others aim to elucidate specific observations and develop new hypotheses.

**KEYWORDS:** data analysis; legal data; natural language processing; statistic.

## 1 INTRODUÇÃO

O homem, por natureza, é um ser gregário. Ou seja, ele deseja viver em comunidade. Assim, em razão de toda sociedade humana ser constituída de pessoas com diferentes padrões de comportamento, é inevitável que haja um elemento de desarmonia. Isso, por sua vez, exige a atuação da lei como um barômetro social específico (Forji, 2010).

Logo, para que qualquer sociedade perdure, ela deve inescapavelmente se ancorar na lei, que regulamenta o comportamento dos membros da sociedade em questão, bem como suas relações, seus direitos e obrigações (Forji, 2010). O Direito é, portanto, uma causa vital de conduta aceita pelos integrantes de uma determinada sociedade, conforme estabelecido pela autoridade legítima desse grupo (Tyler & Darley, 1999).

Tratando sobre o conceito de Direito, no sentido jurídico do termo, pode apresentar diversas facetas. É definido como uma ciência que estuda as normas de convivência na sociedade humana; o conjunto de leis e normas jurídicas vigentes num país; o conjunto de normas da vida em sociedade que buscam expressar e alcançar um ideal de justiça, traçando as fronteiras do ilegal e do obrigatório (Oxford Languages, s.d.).

Por outro lado, o sentido mais filosófico do termo também não se dispersa desse entendimento. Kelsen (1999) estabelece o Direito como uma ordem que normatiza a conduta humana, ou seja, um conjunto de normas que direciona o comportamento humano. No que tange ao termo “norma”, o autor define como uma conduta que deve ser ou acontecer de determinada maneira.

No mesmo sentido, para Bobbio (2003), o Direito é um conjunto de normas ou regras de comportamento nas quais a vida se desenvolve. Estas regras têm em comum um elemento característico que, segundo o autor, é o propósito de influenciar a conduta individual e de grupos no sentido de determinados objetivos em oposição a outros.

À vista disso, em que pese diversos fatores informais influenciarem a forma como as pessoas interagem, as sociedades modernas têm a lei como mecanismo primário para controlar formalmente o comportamento humano. Portanto, o

direito desempenha um papel importante na compreensão de como as sociedades se transformam (Katz *et al.*, 2020).

Neste ponto de vista, para Todorov (2005), as leis são escritas para reger o comportamento. Ou seja, a principal relevância da Lei é orientar a sociedade a se comportar de uma determinada maneira ou enfrentar as sanções correspondentes (Forji, 2010). Os legisladores formulam, alteram e removem regras legais para alcançar resultados comportamentais específicos, frequentemente esforçando-se para responder às mudanças percebidas nas necessidades sociais (Wagh, 2013). Logo, as normas jurídicas pretendem induzir o comportamento de seus destinatários, prevenindo-os de agir de maneira socialmente indesejada ou encorajando-os a agir de uma forma socialmente desejada (Nunes, 2016). Também, Hart e Green (Raz & Bulloch, 2012) dispõem que a Lei aponta um modelo geral de comportamento e se destina a uma categoria geral de pessoas, esperando que seja reconhecida e acatada, principalmente na esfera criminal. Nessa lógica, é razoável afirmar que as normas que compõem o Direito direcionam o sentido das ações da coletividade. Considerando isso, a literatura (Been *et al.*, 2015; Lodi, 2017; Malta *et al.*, 2010; Priesnitz *et al.*, 2017) já buscou avaliar o impacto da legislação empiricamente em diversos cenários, concluindo que uma mudança na legislação altera o contexto social regulado.

Todavia, as leis humanas são consideradas, em suma, textuais. Ou seja, palavras normativas escritas em documentos legais que trazem a marca das instâncias oficiais, tornando-as vinculantes e judicialmente exigíveis. Por outro lado, as leis naturais tratam de padrões e invariabilidades ocultos que podem ser descobertos no transcurso da pesquisa. Em outras palavras, o método jurídico tende a considerar a interpretação textual (Capra & Mattei, 2018), sendo atípicas pesquisas com uma abordagem quantitativa. Essa abordagem textual muitas vezes limita o fornecimento de *insights* valiosos acerca dos padrões e tendências subjacentes aos fenômenos jurídicos, que em muitos casos não pode ser detectado somente com a abordagem descritiva textual, como os padrões estatísticos de correlação que podem revelar um padrão de causa e efeito. Portanto, a incorporação de abordagens quantitativas nas pesquisas jurídicas não apenas preenche uma lacuna no conhecimento, mas também enriquece a compreensão

dos fenômenos legais ao fornecer uma perspectiva mais abrangente e fundamentada empiricamente. Dessa questão, a utilização de métodos quantitativos no Direito (análise estatísticas, modelagem, simulação, entre outros) tornou-se uma área em ascensão, buscando compreender os impactos de documentos jurídicos de forma mais precisa e objetiva.

Nesse sentido, o presente artigo tem como objetivo central realizar o levantamento da literatura disponível no contexto da pesquisa quantitativa no Direito, por meio de um estudo descritivo. Para isso, fora utilizada a base de dados do Google Acadêmico. A opção da utilização do Google Acadêmico como fonte de dados foi baseada na disponibilidade de dados abertos, o que proporciona a transparência e replicabilidade do estudo de forma acessível a qualquer pesquisador. As palavras-chaves utilizadas foram “pesquisa quantitativa no Direito” e “processamento de linguagem natural no Direito”, em português e inglês, sem limitação temporal ou espacial, considerando que a temática já é restrita e utilizar esses critérios poderia limitar ainda mais o objetivo do presente trabalho.

Na primeira etapa, os artigos retornados com as palavras-chaves utilizadas foram selecionados e realizada a leitura do resumo. Na segunda etapa, os artigos que apresentavam relação com o objeto proposto passaram pela segunda seleção e estes foram em sua íntegra e realizada a análise para, posteriormente, serem selecionados para o presente trabalho. No mais, o trabalho se aprofunda em abordar alguns trabalhos empíricos realizados com algumas das fontes do direito mais utilizadas, como a legislação, jurisprudência e patentes.

Metodologicamente, utilizou-se do método indutivo, baseando-se na ideia inicial de que a pesquisa jurídica tende a ser majoritariamente qualitativa e que a possível ampliação de tal pesquisa no Direito pode possibilitar uma compreensão mais completa do objeto de estudo. Assim, ao final, foi possível concluir que tal limitação pode fazer com que os trabalhos da área jurídica percam o contato com a realidade fática, cujas ideias não são colocadas à análise ou confrontadas com dados.

## 2 MÉTODOS QUANTITATIVOS NO DIREITO

Métodos quantitativos são empregados há tempos na solução de dilemas práticos nos mais diversos setores. Brevemente, a pesquisa quantitativa pode ser caracterizada pela quantificação na coleta e tratamento de informações por meio de técnicas estatísticas (Richardson, 1999). A estatística, por sua vez, é uma disciplina definida por sua metodologia e possibilita a aplicação a qualquer objeto sujeito à experimentação e observação (Nunes, 2016). Ademais, a estatística moderna disponibiliza tecnologia quantitativa para a ciência empírica, mensurando a incerteza e explorando suas consequências no planejamento e interpretação da experimentação e observação (Stigler, 1986). Mattar (2001) afirma que a pesquisa quantitativa objetiva a validação das hipóteses por meio da utilização de dados estruturados estatísticos com apreciação de um grande número de casos representativos, recomendando um curso final da ação. Tal pesquisa quantifica os dados e estende os resultados da amostra para os interessados.

Tradicionalmente, nota-se o predomínio da abordagem qualitativa sobre a quantitativa nos estudos acadêmicos da área jurídica. De modo sucinto, enquanto a primeira se importa com abordagens pautadas no método discursivo, com aprofundamento sobre os temas estudados, a segunda se fundamenta em mensurações numéricas, dados exatos e métodos estatísticos replicáveis (Andrade, 2018).

O uso do modelo quantitativo torna exequível a iniciativa de testar hipóteses, asseverando não somente o aperfeiçoamento do conhecimento empírico quanto ao fenômeno social ou jurídico estudado, mas também a recusa de hipóteses empiricamente não confirmadas, lapidando a teoria formulada (Andrade, 2018).

Na pesquisa quantitativa, há que se considerar a possibilidade do uso da estatística descritiva e inferencial, ou combinadas. Enquanto na primeira a estatística é feita uma descrição do que foi observado, na segunda podem ser feitas considerações a respeito do que não foi observado, além de inferências sobre o comportamento futuro de variáveis ou a associação de variáveis (Nunes, 2016).

Na abordagem descritiva, descreve-se, de maneira exaustiva ou sumarizada, o conjunto de dados coletados. Um exemplo é a estimação do tempo médio de duração de uma ação de dissolução de sociedade no Brasil. Para isso, Nunes (2012)

coletou o tempo de duração, entre a distribuição e o julgamento da apelação em segunda instância, de uma amostra com 718 ações de dissolução nas 27 unidades da federação. Foi feita a soma de todos os tempos e dividido pelo número de ações da amostra, obteve-se como média 1.782 dias, ou seja, 4 anos, 10 meses e 22 dias. De tal resultado foi avaliado o desvio-padrão, que indica o grau de variação de um conjunto de elementos (Instituto de Pesquisa Econômica Aplicada, 2006). Ou seja, em que pese a média de duração das ações serem próxima a cinco anos, o desvio padrão foi de 1.063 dias, indicando que os processos observados estavam 2 anos, 11 meses e 3 dias acima ou abaixo da média. Em outras palavras, muitas ações durando mais de 7 anos e outras sendo encerradas com apenas 3 anos. Para além do desvio padrão, há diversas outras medidas descritivas que podem ser feitas nos dados, como mediana, moda, média aritmética, entre outras. Todas objetivam sumarizar informações a respeito de uma grande quantidade de dados (Nunes, 2016).

Tratando da estatística inferencial (ou indutiva), ela complementa a descritiva. Ou seja, enquanto a última resume, explora e descreve os dados, a primeira faz afirmações que vão além da mera descrição dos dados. Por exemplo, permite conclusões sobre uma população em caso de amostras, previsões sobre o comportamento futuro das variáveis e reconhecimento de tendências, associações e correlações nas variáveis. Ou seja, sempre que são feitas afirmações sobre uma amostra ou análises se a alteração no comportamento de uma variável é estatisticamente significativa, trata-se de estatística inferencial (Nunes, 2016).

Ainda assim, uma eventual mudança da dialética para uma ferramenta estatística no estudo jurídico não se propõe a substituir a análise qualitativa dos fenômenos políticos, jurídicos ou sociais. Mas sim, visa expandir a consistência das inferências extraídas das hipóteses elaboradas pelo pesquisador. Assim, propondo testes de hipóteses e análise de dados constituídos por padrões numéricos ou nominais. A tradicional abordagem qualitativa existente na área jurídica é imprescindível para a interpretação e subsequente compreensão de fenômenos políticos, jurídicos e sociais. Apesar disso, deve conciliar, quando cabível, ao empirismo, aos testes de hipóteses e, em especial, à abordagem quantitativa, com

o objetivo de permitir que experiências possam ser replicadas e assim, fornecer maior precisão às inferências extraídas de cada estudo (Andrade, 2018).

Sobretudo, cabe ressaltar que a ciência é um empreendimento coletivo. Ou seja, enquanto diversos estudos priorizam o teste de hipótese, outros focam na explicação sobre observações específicas e geração de novas hipóteses. Por fim, todos são parte de um diálogo entre teoria e dados (Andrade, 2018; Gschwend & Schimmelfennig, 2011).

Ademais, o Direito, mesmo com a incerteza no cerne de sua aplicação, não utiliza métodos quantitativos de maneira formal no cotidiano, ainda que esta associação seja de longa data (Zabala & Silveira, 2014). Outrossim, em que pese a lei dirigir todos os setores da sociedade, abrangendo um espaço bastante vasto, uma análise feita por Katz *et al.* (2023) revelou que a geração de textos jurídicos desempenhou um papel bastante limitado na última década.

Estudiosos do Direito tradicionalmente consideram a Lei como dificilmente quantificável (Katz *et al.*, 2020). Entretanto, em que pese os estudos jurídicos empíricos não serem escassos (Heise, 2011; Ho & Kramer, 2013; Epstein & Martin, 2014), há pouco tempo pesquisadores começaram a aplicar métodos de ciência de dados à Lei (Whalen, 2016; Coupette, 2019; Livermore & Rockmore, 2019; Frankenreiter & Livermore, 2020). Existem relativamente poucos trabalhos quantitativos que abordam explicitamente a mudança legal (Cross, Smith, & Tomarchio, 2007; Buchanan, Chai, & Deakin, 2014; Rockmore *et al.*, 2018; Ruhl, Katz, & Bommarito, 2017; Rutherford *et al.*, 2018; Fjelstul, 2019) e limitado estudo que avalie os resultados em evolução no tempo dos poderes legislativo e executivo dos governos nacionais em escala (Katz *et al.*, 2020).

O primeiro trabalho conhecido na literatura data de 1709 e foi apresentado por Nicolau I Bernoulli, que aborda temas como probabilidade de sobrevivência de pessoas, precificação de seguros, preços de loterias, questões de herança, confiança em testemunhas e probabilidade de inocência de um acusado (Bernoulli, 1709; Hald, 1998). Em 1897, Holmes apontava a necessidade do desenvolvimento de um pensamento quantitativo, afirmando que o homem do futuro seria o homem da estatística e o mestre da economia (Holmes, 2009; Zabala & Silveira, 2014).

Nos últimos anos, com o aumento exponencial do uso de tecnologia, tem-se observado uma grande quantidade de dados sendo produzidos, armazenados e pesquisados (Wagh, 2013). Considerando que praticamente todos os processos jurídicos envolvem a produção ou o consumo de documentos (Katz *et al.*, 2023) e que o processo de raciocínio jurídico e tomada de decisão são fortemente dependentes de informações armazenadas em arquivos de texto (Wagh, 2013), é inegável que a pesquisa jurídica e as metodologias de pesquisa jurídica foram impactadas (Goanta, 2017). Em especial, a disponibilidade de documentos judiciais em formato digital oferece inúmeras oportunidades para extração e emprego de informações (Anand & Wagh, 2022), essencial para compreender como o Direito e a sociedade interagem (Katz *et al.*, 2020). Portanto, considerando que a base dos dados utilizada em pesquisas no Direito são arquivos de palavras, transformá-las em uma abordagem quantitativa é a questão a seguir.

Embora haja muitas abordagens que tentam lidar com o desafio de transformar textos legais em um conjunto de condições legíveis por máquina, os resultados ainda são insatisfatórios e esse continua sendo um grande desafio em aberto (Dragoni *et al.*, 2016). A pesquisa nas áreas de extração de informações (processamento de linguagem natural, inteligência artificial, entre outros) aumentou o processo de mineração de texto para aprimorar o processo de descoberta de conhecimento neste domínio (Wagh, 2013).

Conceitualmente, entende-se por mineração de texto o processo de buscar ou extrair informações úteis de dados textuais (Vijayarani, Ilamathi, & Nithya, 2015). Diversas pesquisas mostraram o potencial da mineração de texto em encarregar-se das limitações humanas de tempo e cognição, dada a taxa de crescimento das publicações científicas (Juventia *et al.*, 2020; Aureli, 2017; Nunez-Mir *et al.*, 2016; Tamames & Lorenzo, 2010; Tobback *et al.*, 2018; Westergaard *et al.*, 2018), possibilitando uma exploração sistemática e abrangente rápida de uma grande quantidade de fontes baseadas em texto não estruturadas<sup>4</sup> (Bayrak, 2020; Juventia *et al.*, 2020).

---

<sup>4</sup> Dados não estruturados geralmente se referem a informações que não residem em um banco de dados linha-coluna tradicional (Vijayarani, Ilamathi, & Nithya, 2015).

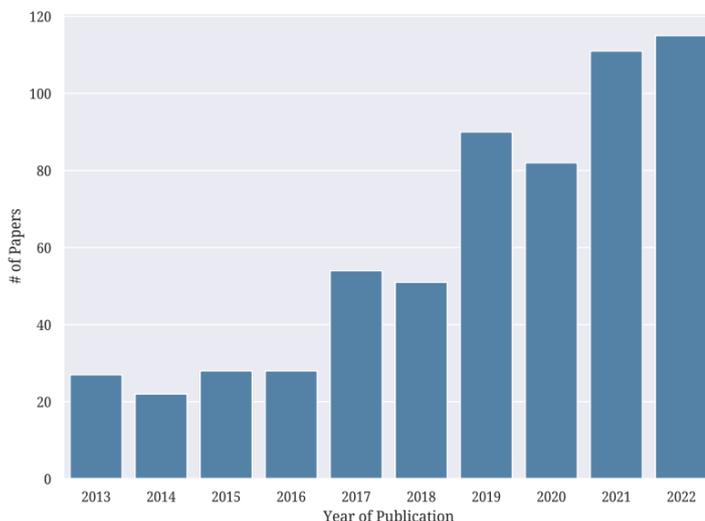
Adentrando no tema, mineração de texto é um termo amplo que retrata uma variedade de técnicas que visam extrair informações úteis de coleções de documentos em dados textuais não estruturados de vários tipos de documentos. Uma definição mais formal restringe a mineração de texto à criação de novas informações não explícitas (como padrões, tendências ou relacionamentos) a partir de uma coleção de documentos textuais (Truyens & Van Eecke, 2014). Trata-se de uma coleção de métodos utilizados para “navegar, organizar, achar e descobrir informação em bases textuais” (Aranha & Passos, 2006, p. 2). Dentre esses métodos têm-se técnicas já consolidadas, como indexação, processamento de linguagem natural, mineração de dados, entre outros (Aranha & Passos, 2006).

Atualmente, o processamento da linguagem natural (PLN) é o método de inteligência artificial mais habitualmente usado na prática do direito (Haney, 2020). Sobretudo, ao considerar que o Direito é fortemente baseado na palavra escrita, é um dos campos que podem se beneficiar dessa abordagem (Aletas *et al.*, 2016; Ikram & Chakir, 2019; Katz, Bommarito & Blackman, 2017; Kowsrihawat, Vateekul, & Boonkwan, 2018; Long *et al.*, 2019; Şulea *et al.*, 2017a; Şulea *et al.*, 2017b; Virtucio *et al.*, 2018; Medvedeva, Wieling, & Vols, 2023). Trata-se de uma técnica chave para mineração de textos, normalmente, participando do pré-processamento dos dados, transformando os textos em números (Aranha & Passos, 2006).

O processamento de linguagem natural tem sido usado com êxito em muitas aplicações de ciência da informação relacionadas às ciências sociais (Ji *et al.*, 2020; De Fortuny *et al.*, 2014; Li Wu, & Wang, 2020; Qian *et al.*, 2019; Schumaker & Chen, 2009; Tuke *et al.*, 2020). Sobretudo, segundo Medvedeva, Wieling e Vols (2023), o atual acesso digital a uma grande quantidade de casos legais disponibilizada promove uma oportunidade única de processar esses dados automaticamente em grande escala usando técnicas de processamento de linguagem natural. Inclusive, o número de artigos tratando da temática mais que dobrou nos últimos 5 anos (Figura 1).

## Figura 1

### Número de artigos de processamento de linguagem natural na área jurídica de 2013 a 2022



Fonte: Katz *et al.* (2023, p. 3).

Considerando que as informações legais geralmente são representadas em formato de texto (casos legais, contratos, patentes), o processamento de textos jurídicos é uma área em crescimento no PLN com diversas aplicações, como classificação de tópicos jurídicos (Nallapati & Manning, 2008; Chalkidis, Androutsopoulos & Aletras, 2019), geração de opinião judicial (Ye *et al.*, 2018) e análise (Wang *et al.*, 2012), extração de informações legais (Chalkidis, Androutsopoulos & Michos, 2018) e reconhecimento de entidade (Cardellino *et al.*, 2017; Chalkidis, Androutsopoulos & Michos, 2017; Chalkidis, Androutsopoulos & Aletras, 2019).

A análise de documentos jurídicos utilizando técnicas de PLN envolve aplicar tais métodos para o domínio jurídico e, possivelmente, personalizá-los e especializá-los para melhor atender às diferentes tarefas em questão (Mumcuoğlu *et al.*, 2021). Também, para os mais diversos documentos jurídicos, como tratados internacionais, constituições, legislações, regulamentos, políticas, jurisprudência, patentes, entre outros. Os mais frequentemente abordados pela literatura são as legislações e a jurisprudência.

## JURISPRUDÊNCIA

Uma das aplicações de métodos quantitativos ao Direito, em especial, à jurisprudência, é cunhada como jurimetria. O uso dessa ferramenta não é contemporâneo, mas pouco explorada em países de tradição jurídica romano-germânica, como o Brasil. Isso porque há um distanciamento entre pesquisadores de ciências de campos científicos diferentes, aplicando-se aos voltados à análise das ciências jurídicas, cuja abordagem quantitativa dos estudos não é frequente (Andrade, 2018).

O termo *jurimetrics* foi criado por Loevinger (1948) que conciliou pela primeira vez a teoria jurídica, métodos computacionais e estatística. O intuito foi analisar a jurisprudência e tornar o uso do Direito mais previsível (Zabala & Silveira, 2014). Loevinger (1963) afirma que a jurimetria se preocupa com a análise quantitativa da conduta judicial, a utilização da lógica matemática na lei, a recuperação de informações jurídicas por meio de meios eletrônicos e mecânicos, a elaboração de cálculos para previsibilidade jurídica.

Em razão dos precedentes serem a base do Direito dos Estados Unidos (EUA), Loevinger se dispôs a criar um mecanismo capaz de transferir esses precedentes para o meio eletrônico, a fim facilitar o armazenamento e a localização das decisões por meio de mecanismos de busca. Na época vivenciada pelo autor, os precedentes jurisprudenciais americanos se amontoavam e iniciavam o arquivo nos primeiros sistemas computacionais das cortes. Para além da facilitação nas buscas, tal ferramenta seria útil também para mensurar a frequência das decisões através de uma metodologia objetiva, sujeita à testes de veracidade. Para o autor, a real ciência deve ser falseável e essa abordagem poderia descrever o fenômeno jurídico de forma imparcial e ampla, aferindo o Direito dentro de padrões de falseabilidade (Nunes, 2016).

Para Andrade (2018), a jurimetria fornece um ponto de vista sistemático dos fatores que influenciam ou agem na tomada de decisões pelo magistrado, dado que contribui para estabelecer padrões de comportamento legal com suporte em elementos quantitativos. Além disso, prever decisões legais é fundamental para

entender as consequências do comportamento e para melhorar a qualidade do produto do trabalho jurídico (Lage-Freitas *et al.*, 2022).

A metodologia da jurimetria, de acordo com Nunes (2016), é um conjunto de métodos estatísticos capaz de obter informações objetivas sobre o desempenho de uma ordem jurídica, bem como realizar previsões a respeito de seu comportamento futuro. Por meio de tais métodos, o pesquisador tem condições de coletar, descrever, sumarizar e analisar criticamente a produção de normas, bem como antecipar suas consequências. Logo, o objetivo da jurimetria é oferecer uma compreensão dos fatores que influenciam a tomada de decisões pelo magistrado, contribuindo para definir padrões de comportamento legal com suporte em elementos quantitativos (Andrade, 2018). Tal investigação pode também ampliar a legitimidade do sistema jurídico, desempenhando um papel de supervisão e fiscalização sobre os atores e mecanismos envolvidos (Visser, 2006; Ramírez, Díaz, & Fernández, 2016; Andrade, 2018).

Couto e Oliveira (2016) apresentam aspectos positivos e negativos quanto à utilização de ferramentas matemáticas no Direito. Como aspecto positivo, têm-se a contribuição para identificar problemas e direcionar planos para equacionamento das dificuldades da gestão do conhecimento. Como aspecto negativo, têm-se a dificuldade na obtenção de dados, o conservadorismo e o engessamento do Poder Judiciário. No mesmo sentido, porém por outro ângulo, Menezes e Barros (2017) afirmam que há a dificuldade do jurista em entender a linguagem estatística, a interdisciplinaridade, a mudança da pesquisa puramente bibliográfica pela prática e, por fim, o manejo das incertezas que faz parte da estatística, mas não do Direito.

Considerando tais aspectos, tal ferramenta foi aplicada por estudos em diversos países e tribunais, nas mais diversas áreas. Inclusive, em tribunais internacionais, como no Tribunal Europeu de Direitos Humanos. Aletras *et al.* (2016) previram as decisões de casos julgados pelo Tribunal Europeu de Direitos Humanos com uma média de 79% de precisão. Para isso, foram utilizados 584 casos do Tribunal Europeu que estariam violando os artigos 3, 6 e 8 da Convenção Europeia de Direitos Humanos. Medvedeva, Vols e Wieling (2020) aplicaram a abordagem de Aletras *et al.* (2016) a 14 artigos da Convenção para prever decisões

de violação, cujos resultados mais altos em termos de métricas de pontuação são 77% para prever se os artigos serão violados e com uma pontuação de 79% em um experimento que usou apenas os nomes dos juízes.

Há ainda trabalhos que avaliam as decisões de tribunais de modo mais específico (Lage-Freitas *et al.*, 2022). Inicialmente, será considerado o continente asiático. Yang *et al.* (2019) previram decisões legais do Supremo Tribunal Popular da China com uma pontuação de 86,7%. Li *et al.* (2019) usaram mais de um milhão de casos do Supremo Tribunal Popular da China, extraindo características fisiológicas e descrições dos fatos para prever artigos de lei específicos usados em decisões legais, bem como acusações e penas de prisão, atingindo pontuações de 41%-96%. Previsões de cobranças tiveram a pontuação mais alta; as previsões de prisão são as mais baixas. Ao abordar a previsão da Suprema Corte da Tailândia, Kowsrihawat, Vateekul e Boonkwan (2018) usam *Bi-GRU* e modelo de atenção e têm um F1-score de 66,67%. Para prever os resultados do caso da Suprema Corte das Filipinas, Virtucio *et al.* (2018) propõem uma abordagem que atinge uma precisão de 59% usando o classificador Random Forest e n-gramas para extração de características.

Em países europeus, como Reino Unido, Strickson e de la Iglesia (2020) abordaram a previsão de decisões judiciais e seu melhor modelo alcançou uma pontuação de 69,02%. Niklaus, Chalkidis e Stürmer (2021) avaliam métodos baseados em *BERT* de última geração usando um corpus multilíngue em alemão, francês e italiano com aproximadamente 85 mil registros. Os melhores resultados atingiram uma pontuação F1-macro de 70% para os idiomas alemão e francês.

No caso do Brasil, há diversas pesquisas que buscaram prever decisões dos tribunais. Bertalan e Ruiz (2020) propõem uma abordagem para prever decisões judiciais brasileiras para crimes de homicídio em segundo grau e corrupção ativa, com 782 casos do Tribunal de Justiça do Estado de São Paulo. O score foi de 98% para o conjunto de dados de corrupção ativa, que continha 158 registros “inocentes” e 31 “culpados”. Barros, Lorenzi e Wives (2018) usaram redes bayesianas para classificar decisões judiciais de um tribunal trabalhista brasileiro, concluindo que as taxas de sucesso de litígios de empregados e empregadores são aproximadamente iguais. Menezes-Neto e Clementino (2022) compararam o desempenho de três

modelos de aprendizado profundo (ULMFiT, BERT e Big Bird) e 22 especialistas altamente qualificados em prever o resultado de 612.961 recursos de Juizados Especiais Federais no 5º Tribunal Regional Federal do Brasil. Os autores concluíram que todos os modelos superaram os especialistas humanos. Moraes e Demercian (2017) discutiram o uso da jurimetria como mecanismo automático para subsidiar a forma de atuação repressiva do Ministério Público brasileiro.

Na América do Norte, Almuslim e Inkpen (2022) utilizaram decisões do tribunal de apelação do Canadá para prever o resultado de casos (apenas usando a descrição dos casos escritos pelo Tribunal. Os resultados tiveram precisão de F1-scores de 0,92, assim como os melhores resultados da literatura. Katz, Bommarito e Blackman (2017) previram decisões da Suprema Corte dos EUA classificando as decisões em três categorias e apresentando perfis com 70,2% de precisão, cujo conjunto de dados contou com 28.000 casos. Ruger *et al.* (2004) compararam o desempenho preditivo de especialistas jurídicos com um modelo estatístico treinado, usando menos de 200 casos de dados da Suprema Corte dos Estados Unidos.

Há que se fazer uma ressalva de que os trabalhos desenvolvidos na jurimetria são de natureza probabilística, não determinística. Ou seja, uma interpretação legal majoritária e não determina necessariamente o sentido de uma decisão pelo tribunal, mas aumenta a probabilidade de ser decidida naquele sentido (Nunes, 2016). Além disso, as pesquisas apontadas acima divergem no número de decisões (amostra) analisadas (variando de menos de mil a mais de um milhão) e em quanto foi capaz de prever o resultado (60% a 100%).

Isto posto, as pesquisas jurimétricas podem ser vistas como uma forma de comprimir tempo e espaço com o objetivo de viabilizar a análise de casos ocorridos em locais e momentos passados. Ela aborda a ordem jurídica como um objeto concreto, com um passado e um presente, sendo, portanto, possível prever seu comportamento futuro (Nunes, 2016).

Desta forma, a jurimetria pode contribuir com estadistas, juristas, juízes e formuladores de políticas públicas a antecipar, ainda que não de modo absoluto, as consequências de suas decisões em tribunais.

## LEGISLAÇÃO

Outra fonte de pesquisa estatística no Direito são as leis. Segundo Katz *et al.* (2020), as legislações são criadas, modificadas e revogadas para alcançar resultados comportamentais específicos, muitas vezes em um esforço para responder às mudanças percebidas nas necessidades sociais. Assim, as sociedades contemporâneas utilizam a lei como principal instrumento para regular formalmente o comportamento humano.

A aplicação da estatística no estudo dos comportamentos coletivos em função das normas jurídicas viabiliza a criação de modelos capazes de aproximar os resultados produzidos pela ordem jurídica das expectativas e aspirações da sociedade. Quer seja para entender como elas são produzidas ou como se conformam à reação dos destinatários quando são aplicadas. Além disso, quanto mais próximo o comportamento da sociedade ficar do comportamento esperado – o que podemos chamar de aderência das leis à realidade –, mais organizada será a sociedade e mais bem sucedida será a lei (Nunes, 2016).

Ocorre que um estudo quantitativo das legislações não se trata de uma atividade trivial. São as mais diversas possibilidades de se analisar o complexo legislativo. Como para medir se e quanto uma lei foi capaz de alterar o comportamento da sociedade no cenário que regula, se e quanto as variáveis (sociais, econômicas, ambientais etc.) do contexto se relacionam com a respectiva legislação, seja considerando a legislação e sua mudança ao longo do tempo, se de um país ou uma grande amostra, não se limitando a isso.

Nesse sentido, ainda que não seja, via de regra, um papel desempenhado por pesquisadores, reformas legislativas podem possibilitar o teste de hipóteses por meio do monitoramento do efeito em diferentes grupos. Logo, tal análise pode ser feita antes da implementação da legislação, ou ainda após sua promulgação, objetivando aprender com os seus resultados (Nunes, 2016).

Katz *et al.* (2020) investigaram o crescimento da legislação federal nos Estados Unidos da América e Alemanha pelo período de 25 anos (de 1994 a 2018). Na estatística descritiva, o volume das legislações de 2018 são 1,5 vezes maiores do que comparado à 1994. Os autores finalizam recomendando para trabalhos futuros

que os dados legislativos sejam combinados com outros dados da ciência social quantitativa para melhorar a compreensão da interação entre as legislações que afetam o comportamento de indivíduos e sociedades. Por fim, ainda, levanta o questionamento: que relações causais podemos estabelecer entre mudança legal e mudança social?

Empiricamente, diversos estudos abordaram tal temática. Lodi (2017), no cenário econômico, concluiu que a legislação e as instituições judiciárias desempenham considerável influência no setor econômico do país, sendo capaz de conter ou incentivar o desenvolvimento econômico-social. Priesnitz *et al.* (2017) avaliaram o impacto da Lei n. 10.973 de 2004 (Lei de Inovação) nas universidades brasileiras utilizando o número de patentes realizadas pelas universidades federais brasileiras antes e após a entrada em vigor da lei. Os resultados apontaram um total de 4.626 pedidos de depósitos registrados até o ano de 2015, sendo que 604 ocorreram até o ano de 2004 e 4.022 entre os anos de 2005 e 2015, relatando um grande aumento no número de pedidos de depósitos após a Lei de Inovação. Com relação à legislação de trânsito, Malta *et al.* (2010) avaliaram o impacto das restrições trazidas pela Lei n. 11.705 de 2008, popularmente conhecida como Lei Seca. Ao comparar o mesmo período de 2007 e 2008 (após a entrada em vigor da lei) nas capitais, apontaram uma redução de cerca de 23% nas internações por acidente de transporte terrestre e uma redução de cerca de 22% na mortalidade.

Considerando o cenário de saúde pública, Been *et al.* (2015) avaliaram o impacto da legislação antifumo na mortalidade perinatal e infantil. Os resultados mostraram que tal legislação foi associada a uma redução imediata de 7,8% ( $p < 0,001$ ) em natimortos, uma redução de 3,9% ( $p < 0,001$ ) no baixo peso ao nascer e 7,6 % ( $p = 0,001$ ) redução na mortalidade neonatal. Especialmente neste último campo, Gostin, Burris e Lazzarini (1999) dão particular ênfase, pois afirmam que a legislação é uma disciplina fundamental dentro da saúde pública.

De acordo com Burris *et al.* (2010), os indícios revelados pela pesquisa empírica são significativos para a prática e para os estudos acadêmicos em direito de saúde pública, pois trata-se de episódios que fundamentam o sentido da atividade normativa para quais políticas devem ser privilegiadas. Os autores trazem uma base substancial de evidências nas quais se investigou a eficácia da legislação

intervencionista, englobando revisões de literatura, meta-análises e revisões sistemáticas conduzidas por entidades como a Campbell Collaboration e a Força-Tarefa dos EUA sobre Prevenção Comunitária Serviços. Esses “estudos de intervenção” objetivam avaliar o efeito de uma intervenção legal sobre os resultados de saúde ou fatores mediadores que influenciam os resultados de saúde.

Tal tópico ganhou notabilidade nas últimas duas décadas uma vez que doenças infecciosas ressurgiram, trazendo atenção e uma crescente consciência da complexidade da regulamentação da saúde nos níveis local, nacional e global e restauraram a importância do direito na saúde pública e no direito acadêmico (Burriss *et al.*, 2010). Em particular, a legislação atua como mecanismo de intervenção proeminente para alcançar objetivos próprios de saúde pública (Burriss *et al.*, 2010). Ainda, devido às tendências sociais e ambientais que geram uma crescente pressão sobre os recursos hídricos, há razões para acreditar que a legislação será empregada cada vez mais para aprimorar a eficiência e a segurança dos serviços relacionados à água (Judd & Judd, 2010).

## PATENTES

Jun e Sung Park (2013) realizaram uma pesquisa quantitativa utilizando as patentes da Apple, cujo objetivo foi analisar todas as patentes para prever as futuras tecnologias da empresa. Os autores pontuaram que a maioria dos trabalhos anteriores sobre a inovação da Apple foram baseados em abordagens qualitativas, como opiniões de especialistas, oferecendo um ponto de vista subjetivo. Nesse trabalho, os autores estudaram a inovação tecnológica e as previsões da empresa, analisando seus pedidos de patente, sendo uma abordagem objetiva para examinar a inovação da Apple de uma perspectiva tecnológica.

## 3 CONSIDERAÇÕES FINAIS

O objetivo do presente trabalho foi apresentar um levantamento bibliográfico das pesquisas quantitativas no Direito e suas possibilidades e aplicações futuras. A pesquisa quantitativa no cenário jurídico, ainda que iniciada há tempos, como foi o caso do surgimento da jurimetria, não é, ainda, amplamente aplicada. Inclusive, considerando que os documentos jurídicos tiveram uma utilização restrita nos últimos anos. Além disso, em especial, com a recente disponibilização digital de documentos judiciais, têm-se novas oportunidades de pesquisas.

Portanto, este estudo destaca a crescente importância das análises quantitativas no campo do Direito, especialmente em um contexto no qual as bases jurídicas estão cada vez mais disponíveis online e o avanço computacional permite a análise automatizada de grandes conjuntos de dados. Por meio dessas técnicas, torna-se possível captar informações e identificar padrões que não seriam facilmente acessíveis apenas por meio da análise textual tradicional. A integração de abordagens quantitativas no estudo do Direito não apenas enriquece a compreensão dos fenômenos legais, mas também promove uma análise mais objetiva e fundamentada empiricamente. Ao reconhecer e incorporar o potencial das análises quantitativas, os pesquisadores e profissionais do Direito podem ampliar suas capacidades de pesquisa, tomar decisões mais informadas e contribuir para o avanço contínuo do campo jurídico em um mundo cada vez mais digitalizado e complexo.

Ainda, há que se pontuar que a ciência é um empreendimento coletivo, tratando-se de um diálogo entre teoria e dados. O presente trabalho se limitou a apresentar uma fonte promissora para pesquisas futuras, que é a utilização de ferramentas estatísticas no contexto jurídico. Tais trabalhos podem contribuir com advogados, juízes, legisladores e gestores públicos a direcionarem suas ações no sentido pretendido pela e para a sociedade.

## REFERÊNCIAS

Aletras, N. *et al.* (2016). Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective. *PeerJ Computer Science*, 2, e93. <https://peerj.com/articles/cs-93.pdf>

Almuslim, I., & Inkpen, D. (2022). Legal judgment prediction for Canadian appeal cases. In *2022 7th International Conference on Data Science and Machine Learning Applications (CDMA)* (pp. 163-168), Riyadh, Arábia Saudita. <https://doi.org/10.1109/CDMA54072.2022.00032>

Anand, D., & Wagh, R. (2022). Effective deep learning approaches for summarization of legal texts. *Journal of King Saud University-Computer and Information Sciences*, 34(5), 2141-2150. <https://doi.org/10.1016/j.jksuci.2019.11.015>

Andrade, M. D. (2018). A utilização do sistema R-Studio e da jurimetria como ferramentas complementares à pesquisa jurídica. *Revista Quaestio Iuris*, 11(2), 680-692. <https://doi.org/10.12957/rqi.2018.29221>

Aranha, C., & Passos, E. (2006). A tecnologia de mineração de textos. *Revista Eletrônica de Sistemas de Informação*, 5(2). <https://doi.org/10.21529/RESI.2006.0502001>

Aureli, S. (2017). A comparison of content analysis usage and text mining in CSR corporate disclosure. *International Journal of Digital Accounting Research*, 17, 1-32. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=2c4e6aff81a835779415af37551b12c3b7a50f19>

Barros, R., Lorenzi, F., & Wives, L. K. (2018). *Recent trends and future technology in applied intelligence, lecture notes in computer science*. Cham: Springer Cham.

Bayrak, T. (2020). A content analysis of top-ranked universities' mission statements from five global regions. *International Journal of Educational Development*, 72, 102130. <https://doi.org/10.1016/j.ijedudev.2019.102130>

Been, J. V. et al. (2015). Impact of smoke-free legislation on perinatal and infant mortality: a national quasi-experimental study. *Scientific Reports*, 5(1), 1-12. <https://link.springer.com/content/pdf/10.1038/srep13020.pdf>

Bernoulli, N. (1709). *Dissertatio inauguralis mathematico-juridica de usu artis conjectandi in jure* [Dissertação]. Typis Johannis Conradi à Mechel.

Bertalan, V. G., & Ruiz, E. E. (2020). Predicting judicial outcomes in the Brazilian legal system using textual features. In *DHandNLP@ PROPOR* (pp. 22-32). <https://ceur-ws.org/Vol-2607/paper3.pdf>

Bobbio, N. (2003). *Teoria da norma jurídica*. Bauru: Edipro.

Buchanan, J., Chai, D. H., & Deakin, S. (2014). Empirical analysis of legal institutions and institutional change: multiple-methods approaches and their application to corporate governance research. *Journal of Institutional Economics*, 10(1), 1-20. <https://doi.org/10.1017/S1744137413000349>

Burris, S. et al. (2010). Making the case for laws that improve health: a framework for public health law research. *The Milbank Quarterly*, 88(2), 169-210. <https://doi.org/10.1111/j.1468-0009.2010.00595.x>

Capra, F., & Mattei, U. (2018). *A revolução ecojurídica: o direito sistêmico em sintonia com a natureza e a comunidade*. São Paulo: Curtix.

Cardellino, C. et al. (2017). Legal NERC with ontologies, Wikipedia and curriculum learning. In M. Lapata, P. Blunsom, & A. Koller (Eds.), *15th European Chapter of the*

*Association for Computational Linguistics* (Volume 2, Short Papers) (pp. 254-259), Valência, Espanha. <https://dx.doi.org/10.18653/v1/E17-2041>

Chalkidis, I., Androutsopoulos, I., & Aletras, N. (2019). Neural legal judgment prediction in English. In Association for Computational Linguistics (Org.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 4317-4323), Florença, Itália. <https://core.ac.uk/download/pdf/223214543.pdf>

Chalkidis, I., Androutsopoulos, I., & Michos, A. (2017). Extracting contract elements. In Association for Computing Machinery (Org.), *Proceedings of the 16th Edition of the International Conference on Artificial Intelligence and Law* (pp. 19-28), Nova York, Estados Unidos da América. <https://doi.org/10.1145/3086512.3086515>

Chalkidis, I., Androutsopoulos, I., & Michos, A. (2018). Obligation and prohibition extraction using hierarchical RNNs. In Association for Computational Linguistics (Org.), *Proceedings of the Annual Meeting of the Association for Computational Linguistics* (pp. 254-259), Melbourne, Austrália. <https://doi.org/10.48550/arXiv.1805.03871>

Coupette, C. (2019). *Juristische netzwerkforschung: modellierung, quantifizierung und visualisierung relationaler daten im recht*. Tübingen: Mohr Siebeck. <https://doi.org/10.1628/978-3-16-157012-4>

Couto, M. B., & Oliveira, S. P. (2016). Gestão da justiça e do conhecimento: a contribuição da jurimetria para a administração da justiça. *Revista Jurídica*, 2(43), 771-801. <http://dx.doi.org/10.26668/revistajur.2316-753X.v2i43.1869>

Cross, F. B., Smith, T. A., & Tomarchio, A. (2007). The Reagan revolution in the network of Law. *Emory Law Journal*, 57, 1227. <https://dx.doi.org/10.2139/ssrn.909217>

De Fortuny, E. J. *et al.* (2014). Evaluating and understanding text-based stock price prediction models. *Information Processing & Management*, 50(2), 426-441.

<http://dx.doi.org/10.1016/j.ipm.2013.12.002>

Dragoni, M. *et al.* (2016). Combining NLP approaches for rule extraction from legal documents. In *1st Workshop on Mining and REasoning with Legal Texts*, Sophia Antipolis, França. <https://hal.science/hal-01572443>

Epstein, L., & Martin, A. D. (2014). *An introduction to empirical legal research*. Oxford: Oxford University Press.

Fjelstul, J. C. (2019). The evolution of European Union law: a new data set on the Acquis Communautaire. *European Union Politics*, 20(4), 670-691.

<https://doi.org/10.1177/1465116519842947>

Forji, A. G. (2010). The correlation between law and behaviour as pillars of human society. *International Journal of Punishment and Sentencing*, 6(3), 85-98.

<https://core.ac.uk/download/33737735.pdf>

Frankenreiter, J., & Livermore, M. A. (2020). Computational methods in legal analysis. *Annual Review of Law and Social Science*, 16, 39-57.

<https://doi.org/10.1146/annurev-lawsocsci-052720-121843>

Goanta, C. (2017). Big law, big data. *Law and Method*, Special Issue Comparative Law. <https://doi.org/10.5553/REM/000029>

Gostin, L. O., Burris, S., & Lazzarini, Z. (1999). The Law and the public's health: a study of infectious disease law in the United States. *Columbia Law Review*, 99, 59-128. <https://pubmed.ncbi.nlm.nih.gov/10558403/>

Gschwend, T., & Schimmelfennig, F. (2011). *Research design in political science: how to practice what they preach*. Londres: Palgrave Macmillan.

Hald, A. (1998). *A history of mathematical statistics from 1750 to 1930*. Nova Jersey: Wiley-Interscience.

Haney, B. (2020). Applied natural language processing for Law practice. *The Intellectual Property and Technology Forum Journal*.

<https://dx.doi.org/10.2139/ssrn.3476351>

Heise, M. (2011). An empirical analysis of empirical legal scholarship production, 1990-2009. *University of Illinois Law Review*, 2011(5), 1739-1752.

<https://scholarship.law.cornell.edu/facpub/700>

Ho, D. E., & Kramer, L. (2013). Introduction: the empirical revolution in Law.

*Stanford Law Review*, 65(6), 1195-1202. <https://www.jstor.org/stable/23528249>

Holmes Jr, O. W. (2009). *The path of the law*. *The Floating Press*.

Ikram, A. Y., & Chakir, L. (2019). Arabic text classification in the legal domain. In *2019 3rd International Conference on Intelligent Computing in Data Sciences* (pp. 1-6), Marrakesh, Marrocos.

<http://dx.doi.org/10.1109/ICDS47004.2019.8942343>

Instituto de Pesquisa Econômica Aplicada. (2006, 6 de junho). O que é? Desvio padrão.

[https://www.ipea.gov.br/desafios/index.php?option=com\\_content&view=article&id=2104:catid=28&Itemid=2](https://www.ipea.gov.br/desafios/index.php?option=com_content&view=article&id=2104:catid=28&Itemid=2)

Ji, D. et al. (2020). An end-to-end joint model for evidence information extraction from court record document. *Information Processing & Management*, 57(6),

102305. <http://dx.doi.org/10.1016/j.ipm.2020.102305>

„ S., & Judd, C. (Eds.). (2010). *The MBR book: principles and applications of membrane bioreactors for water and wastewater treatment*. Amserdam: Elsevier.

Jun, S., & Sung Park, S. (2013). Examining technological innovation of Apple using patent analysis. *Industrial Management & Data Systems*, 113(6), 890-907. <https://doi.org/10.1108/IMDS-01-2013-0032>

Juventia, S. D. *et al.* (2020). Text mining national commitments towards agrobiodiversity conservation and use. *Sustainability*, 12(2), 715. <https://doi.org/10.3390/su12020715>

Katz, D. M., Bommarito, M. J., & Blackman, J. (2017). A general approach for predicting the behavior of the Supreme Court of the United States. *PLoS ONE*, 12(4), e0174698. <http://dx.doi.org/10.1371/journal.pone.0174698>

Katz, D. M. *et al.* (2020). Complex societies and the growth of the Law. *Scientific Reports*, 10, 18737. <https://doi.org/10.1038/s41598-020-73623-x>

Katz, D. M. *et al.* (2023). *Natural language processing in the legal domain*. <https://doi.org/10.48550/arXiv.2302.12039>

Kelsen, H. (1999). *Teoria pura do Direito*. São Paulo: Livraria Martins Fontes.

Kowsrihawatt, K., Vateekul, P., & Boonkwan, P. (2018). Predicting judicial decisions of criminal cases from Thai Supreme Court using bi-directional GRU with attention mechanism. In *2018 5th Asian Conference on Defense Technology* (pp. 50-55), Hanói, Vietnã. <http://dx.doi.org/10.1109/ACDT.2018.8592948>

Lage-Freitas, A. *et al.* (2022). Predicting Brazilian court decisions. *PeerJ Computer Science*, 8, e904. <https://doi.org/10.7717/peerj-cs.904>

Li, Shang *et al.* (2019). Mann: A multichannel attentive neural network for legal judgment prediction. *IEEE Access*, 7, 151144-151155.

<https://doi.org/10.1109/ACCESS.2019.2945771>

Li, X., Wu, P., & Wang, W. (2020). Incorporating stock prices and news sentiments for stock market prediction: a case of Hong Kong. *Information Processing & Management*, 57(5), 102212. <http://dx.doi.org/10.1016/j.ipm.2020.102212>

Livermore, M., & Rockmore, D. (2019). *Law as data: computation, text, and the future of legal analysis*. Santa Fe: Santa Fe Institute Press.

Lodi, G. (2017). O impacto da legislação e das instituições judiciárias no desenvolvimento econômico e social do Brasil. *A Economia em Revista - AERE*, 25(1), 109-129. <https://doi.org/10.4025/aere.v25i1.36947>

Loevinger, L. (1948). Jurimetrics: the next step forward. *Minnesota Law Review*, 33, 455. <https://scholarship.law.umn.edu/mlr/1796>

Loevinger, L. (1963). Jurimetrics: the methodology of legal inquiry. *Law and Contemporary Problems*, 28(1), 5-35. <https://doi.org/10.2307/1190721>

Long, S., *et al.* (2019). Automatic judgment prediction via legal reading comprehension. In M. Sun *et al.* (Eds.), *Chinese computational linguistics* (pp. 558-572). Cham: Springer International Publishing. [http://dx.doi.org/10.1007/978-3-030-32381-3\\_45](http://dx.doi.org/10.1007/978-3-030-32381-3_45)

Malta, D. C. *et al.* (2010). Impacto da legislação restritiva do álcool na morbimortalidade por acidentes de transporte terrestre – Brasil, 2008. *Epidemiologia e Serviços de Saúde*, 19(1), 78. <http://dx.doi.org/10.5123/S1679-49742010000100009>

Mattar, F. N. (2001). *Pesquisa de marketing*. Barueri: Atlas.

Medvedeva, M., Vols, M., & Wieling, M. (2020). Using machine learning to predict decisions of the European Court of Human Rights. *Artificial Intelligence and Law*, 28, 237-266. <https://doi.org/10.1007/s10506-019-09255-y>

Medvedeva, M., Wieling, M., & Vols, M. (2023). Rethinking the field of automatic prediction of court decisions. *Artificial Intelligence and Law*, 31(1), 195-212. <https://doi.org/10.1007/s10506-021-09306-3>

Menezes, D., & Barros, G. P. (2017). Breve análise sobre a jurimetria, os desafios para a sua implementação e as vantagens correspondentes. *Duc In Altum - Cadernos de Direito*, 9(19). <https://doi.org/10.22293/2179-507x.v9i19.667>

Menezes-Neto, E. J., & Clementino, M. B. (2022). Using deep learning to predict outcomes of legal appeals better than human experts: a study with data from Brazilian federal courts. *PLoS ONE*, 17(7), e0272287. <https://doi.org/10.1371/journal.pone.0272287>

Moraes, A. R., & Demercian, P. (2017). Um novo modelo de atuação criminal para o Ministério Público brasileiro: agências e laboratório de jurimetria. *Revista Jurídica da Escola Superior do Ministério Público de São Paulo*, 11(1), 14-40.

Mumcuoğlu, E. et al. (2021). Natural language processing in Law: prediction of outcomes in the higher courts of Turkey. *Information Processing & Management*, 58(5), 102684. <https://doi.org/10.1016/j.ipm.2021.102684>

Nallapati, R., & Manning, C. D. (2008). Legal docket classification: where machine learning stumbles. In Association for Computational Linguistics (Org.), *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing* (pp. 438-446), Honolulu, Estados Unidos da América. <https://aclanthology.org/D08-1046.pdf>

Niklaus, J., Chalkidis, I., & Stürmer, M. (2021). Swiss-judgment-prediction: a multilingual legal judgment prediction benchmark. In Association for Computational Linguistics (Org.), *Natural Legal Language Processing Workshop 2021* (pp. 19-35), Punta Cana, República Dominicana.  
<https://doi.org/10.48550/arXiv.2110.00806>

Nunes, M. G. (2012). *Jurimetria aplicada ao Direito societário: um estudo estatístico da dissolução de sociedade no Brasil* [Tese de Doutorado, Programa de Pós-graduação em Direito, Pontifícia Universidade Católica de São Paulo].

Nunes, M. G. (2016). *Jurimetria: como a estatística pode reinventar o direito*. São Paulo: Revista dos Tribunais.

Nunez-Mir, G. C. et al. (2016). Automated content analysis: addressing the big literature challenge in ecology and evolution. *Methods in Ecology and Evolution*, 7(11), 1262-1272. <https://doi.org/10.1111/2041-210X.12602>

Oxford Languages. (s.d.). Conceito de Direito. <https://languages.oup.com/>

Priesnitz, M. C. et al. (2017). Depósitos de patentes pelas universidades federais brasileiras: distribuição regional e o impacto da Lei da Inovação. In S. L. Russo, & M. R. Santos (Orgs.), *Propriedade intelectual, tecnologias e empreendedorismo* (pp. 15-21). Aracaju: Associação Acadêmica de Propriedade Intelectual.

Qian, Y. et al. (2019). On detecting business event from the headlines and leads of massive online news articles. *Information Processing & Management*, 56(6), 102086. <http://dx.doi.org/10.1016/j.ipm.2019.102086>

Ramírez, R. A., Díaz, Y. V., & Fernández, Y. A. (2016). Jurimetría: una opción para la sociedad. *Serie Científica de la Universidad de las Ciencias Informáticas*, 9(4), 1-10.

Raz, J., & Bulloch, P. A. (Eds.). (2012). *The concept of Law*. Oxford: Oxford University Press.

Richardson, R. J. (1999). *Pesquisa social: métodos e técnicas*. São Paulo: Editora Atlas.

Rockmore, D. N. et al. (2018). The cultural evolution of national constitutions. *Journal of the Association for Information Science and Technology*, 69(3), 483-494. <https://doi.org/10.1002/asi.23971>

Ruger, T. W. et al. (2004). The Supreme Court forecasting project: legal and political science approaches to predicting Supreme Court decisionmaking. *Columbia Law Review*, 104(4), 1150-1210. <https://doi.org/10.2307/4099370>

Ruhl, J. B., Katz, D. M., & Bommarito, M. J. (2017). Harnessing legal complexity. *Science*, 355(6332), 1377-1378.

Rutherford, A. et al. (2018). Inferring mechanisms for global constitutional progress. *Nature Human Behaviour*, 2, 592-599. <https://doi.org/10.1038/s41562-018-0382-8>

Schumaker, R. P., & Chen, H. (2009). A quantitative stock prediction system based on financial news. *Information Processing & Management*, 45(5), 571-583. <http://dx.doi.org/10.1016/j.ipm.2009.05.001>

Stigler, S. M. (1986). *The history of statistics: the measurement of uncertainty before 1900*. Cambridge: Harvard University Press.

Strickson, B., & de la Iglesia, B. (2020). Legal judgement prediction for uk courts. In Association for Computing Machinery (Org.), *Proceedings of the 3rd International Conference on Information Science and Systems* (pp. 204-209), Cambridge, Inglaterra. <https://doi.org/10.1145/3388176.3388183>

Șulea, O. M. *et al.* (2017a). Exploring the use of text classification in the legal domain. In *Proceedings of the 2nd Workshop on Automated Semantic Analysis of Information in Legal Texts (ASAIL)*, Londres, Inglaterra.  
<https://doi.org/10.48550/arXiv.1710.09306>

Șulea, O. M., *et al.* (2017b). Predicting the Law area and decisions of French Supreme Court cases. In *Proceedings of the International Conference Recent Advances in Natural Language Processing* (pp. 716-722), Varna, Bulgária.  
[http://dx.doi.org/10.26615/978-954-452-049-6\\_092](http://dx.doi.org/10.26615/978-954-452-049-6_092)

Tamames, J., & de Lorenzo, V. (2010). EnvMine: A text-mining system for the automatic extraction of contextual information. *BMC Bioinformatics*, 11, 294.  
<https://doi.org/10.1186/1471-2105-11-294>

Tobback, E. *et al.* (2018). Belgian economic policy uncertainty index: improvement through text mining. *International Journal of Forecasting*, 34(2), 355-365.  
<https://doi.org/10.1016/j.ijforecast.2016.08.006>

Todorov, J. C. (2005). Laws and the complex control of behavior. *Behavior and Social Issues*, 14(2), 86-91. <http://dx.doi.org/10.5210/bsi.v14i2.360>

Truyens, M., & Van Eecke, P. (2014). Legal aspects of text mining. *Computer Law & Security Review*, 30(2), 153-170. <https://doi.org/10.1016/j.clsr.2014.01.009>

Tuke, J. *et al.* (2020). Pachinko prediction: a Bayesian method for event prediction from social media data. *Information Processing & Management*, 57(2), 102147.  
<http://dx.doi.org/10.1016/j.ipm.2019.102147>

Tyler, T. R., & Darley, J. M. (1999). Building a law-abiding society: taking public views about morality and the legitimacy of legal authorities into account when formulating substantive law. *Hofstra Law Review*, 28, 707.

Vijayarani, S., Ilamathi, M. J., & Nithya, M. (2015). Preprocessing techniques for text mining-an overview. *International Journal of Computer Science & Communication Networks*, 5(1), 7-16.

Virtucio, M. B. et al. (2018). Predicting decisions of the Philippine Supreme Court using natural language processing and machine learning. In IEEE Computer Society (Org.), *2018 IEEE 42nd Annual Computer Software and Applications Conference* (pp. 130-135), Tóquio, Japão.  
<http://dx.doi.org/10.1109/COMPSAC.2018.10348>

Visser, J. (2006). Jurimetrics, safety and security. *International Review of Law, Computers & Technology*, 20(1&2), 123-145.

Wagh, R. S. (2013). Knowledge discovery from legal documents dataset using text mining techniques. *International Journal of Computer Applications*, 66(23), 32-34.  
<http://dx.doi.org/10.5120/11258-6501>

Wang, W. Y. et al. (2012). Historical analysis of legal opinions with a sparse mixed-effects latent variable model. In Association for Computational Linguistics (Org.), *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers) (pp. 740-749), Jeju, Coréia do Sul.

Westergaard, D. et al. (2018). A comprehensive and quantitative comparison of text-mining in 15 million full-text articles versus their corresponding abstracts. *PLoS Computational Biology*, 14(2), e1005962.  
<https://doi.org/10.1371/journal.pcbi.1005962>

Whalen, R. (2016). Legal networks: the promises and challenges of legal network analysis. *Michigan State Law Review*, 2016(539), 539-565.

Yang, W. *et al.* (2019). Legal judgment prediction via multi-perspective bi-feedback network. In AAAI Press (Org.), *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence* (4085-4091), Macau, China. <https://doi.org/10.24963/ijcai.2019/567>

Ye, H. *et al.* (2018). Interpretable charge predictions for criminal cases: learning to generate court views from fact descriptions. <https://doi.org/10.48550/arXiv.1802.08504>

Zabala, F. J., & Silveira, F. F. (2014). Jurimetria: estatística aplicada ao direito. *Revista Direito e Liberdade*, 16(1), 87-103.

**Bruna Cabrera De Bonito:** Graduada em Direito na Universidade do Estado de Minas Gerais (UEMG). Mestranda em Ciências Ambientais na Universidade do Estado de Minas Gerais (UEMG) e Bolsista da CAPES.

**Vanesca Korasaki:** Engenheira Agrônoma e Mestre em Agronomia pela Universidade Estadual de Londrina (UEL) e doutora em Agronomia/Entomologia pela Universidade Federal de Lavras (UFLA). Professora da Universidade do Estado de Minas Gerais (UEMG), sub-coordenadora do PPG Mestrado em Ciências Ambientais (Ciamb) e professora permanente do PPG Mestrado em Propriedade Intelectual e Transferência de Tecnologia para Inovação.

**Cristina Veloso de Castro:** Docente do Programa de Pós-graduação em Ciências Ambientais da Universidade do Estado de Minas Gerais. Professora de Direito Constitucional na Universidade do Estado de Minas Gerais - UEMG. Mestrado em Direito pela Universidade de Ribeirão Preto e doutorado em Derecho Empresarial pela Universidade de Extremadura.

Data de submissão: 27/04/2023

Data de aprovação: 24/06/2024