

UNIVERSIDADE FEDERAL DE GOIÁS
FACULDADE DE DIREITO
PROGRAMA DE PÓS-GRADUAÇÃO EM DIREITO E POLÍTICAS PÚBLICAS

SÉRVIO TÚLIO TEIXEIRA E SILVA

Inteligência artificial na análise de patologias corruptivas: delimitação jurisprudencial nas decisões do TCU do conceito aberto de cláusula restritiva ao caráter competitivo em editais de licitação

GOIÂNIA

2020



UNIVERSIDADE FEDERAL DE GOIÁS
FACULDADE DE DIREITO

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a [Lei 9.610/98](#), o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

☒ Dissertação ☐ Tese

2. Nome completo do autor

Sérvio Túlio Teixeira e Silva

3. Título do trabalho

Inteligência artificial na análise de patologias corruptivas: delimitação jurisprudencial nas decisões do TCU do conceito aberto de cláusula restritiva ao caráter competitivo em editais de licitação

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento ☒ SIM ☐ NÃO¹

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:

- a)** consulta ao(à) autor(a) e ao(à) orientador(a);
- b)** novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.

O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;

- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.



Documento assinado eletronicamente por **SERVIO TULIO TEIXEIRA E SILVA, Discente**, em 18/02/2021, às 17:24, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Cleuler Barbosa Das Neves, Professor do Magistério Superior**, em 19/02/2021, às 01:53, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1876343** e o código CRC **310920A6**.

UNIVERSIDADE FEDERAL DE GOIÁS
FACULDADE DE DIREITO
PROGRAMA DE PÓS-GRADUAÇÃO EM DIREITO E POLÍTICAS PÚBLICAS

SÉRVIO TÚLIO TEIXEIRA E SILVA

Inteligência artificial na análise de patologias corruptivas: delimitação jurisprudencial nas decisões do TCU do conceito aberto de cláusula restritiva ao caráter competitivo em editais de licitação

Dissertação apresentada como requisito parcial para a conclusão do Mestrado Profissional em Direito e Políticas Públicas da Universidade Federal de Goiás (PPGDP-UFG), vinculado ao Projeto Institucional n. 2.2 – Patologias Corruptivas: diagnóstico e instrumentos de prevenção e combate, em atendimento ao art. 42, inciso I, do Regulamento do PPGDP.

Orientador: Prof. Dr. Cleuler Barbosa das Neves.

GOIÂNIA

2020

Ficha de identificação da obra elaborada pelo autor, através do
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Silva, Sérgio Túlio Teixeira e
Inteligência artificial na análise de patologias corruptivas:
[manuscrito] : delimitação jurisprudencial nas decisões do TCU do
conceito aberto de cláusula restritiva ao caráter competitivo em
editais de licitação / Sérgio Túlio Teixeira e Silva. - 2020.
191 f.: il.

Orientador: Prof. Dr. Cleuler Barbosa das Neves.
Dissertação (Mestrado) - Universidade Federal de Goiás,
Faculdade de Direito (FD), Programa de Pós-Graduação em Direito e
Políticas Públicas, Goiânia, 2020.

Bibliografia. Apêndice.

Inclui siglas, abreviaturas, gráfico, tabelas, lista de figuras, lista
de tabelas.

1. Jurisprudência. 2. Precedentes administrativos. 3. Inteligência
Artificial. 4. Licitação pública. 5. Cláusulas restritivas ao caráter
competitivo. I. Neves, Cleuler Barbosa das, orient. II. Título.



UNIVERSIDADE FEDERAL DE GOIÁS

FACULDADE DE DIREITO

ATA DE DEFESA DE DISSERTAÇÃO

Ata nº 14 da Turma II da sessão de Defesa de Dissertação de **Sérvio Túlio Teixeira e Silva**, que confere o título de Mestre(a) em **Direito e Políticas Públicas**, na área de concentração em **Direito da Administração e das Políticas Públicas**.

Ao/s **vinte dois dias do mês de dezembro de dois mil e vinte**, a partir da(s) **15h**, por meio de videoconferência, realizou-se a sessão pública de Defesa de Dissertação intitulada “**Inteligência artificial na análise de patologias corruptivas: delimitação jurisprudencial nas decisões do TCU do conceito aberto de cláusula restritiva ao caráter competitivo em editais de licitação**”. Os trabalhos foram instalados pelo(a) Orientador(a), Professor(a) Doutor(a) **Cleuler Barbosa das Neves (PPGDP)** com a participação dos demais membros da Banca Examinadora: Professor(a) Doutor(a) **Fabiano Hartmann Peixoto (UNB)**, membro titular externo; Professor(a) Doutor(a) **Saulo de Oliveira Pinto Coelho (PPGDP)**, membro titular interno. Durante a arguição os membros da banca **não fizeram** sugestão de alteração do título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Dissertação, tendo sido(a) o(a) candidato(a) **aprovado(a)** pelos seus membros. Proclamados os resultados pelo(a) Professor(a) Doutor(a) **Cleuler Barbosa das Neves**, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, ao(s) **vinte dois dias do mês de dezembro de dois mil e vinte**.

TÍTULO SUGERIDO PELA BANCA



Documento assinado eletronicamente por **Cleuler Barbosa Das Neves, Professor do Magistério Superior**, em 31/12/2020, às 03:22, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Saulo De Oliveira Pinto Coelho, Professor do Magistério Superior**, em 14/01/2021, às 12:32, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1766590** e o código CRC **7BEFF66D**.



UNIVERSIDADE FEDERAL DE GOIÁS
FACULDADE DE DIREITO

DESPACHO

Visando concluir a formalização da Ata de defesa (doc. 1690990), e aplicando a prerrogativa (poder-dever) prevista na Instrução Normativa n. 001/2020 da PRPG-UFG, a Coordenação do PPGDP-UFG, após confirmação de concordância do membro externo da Banca de Defesa do Trabalho Final do mestrando Sérgio Túlio Teixeira e Silva, assinou a Ata em substituição a este.

(assinado eletronicamente)

Prof. Dr. Saulo de Oliveira Pinto Coelho

Coordenador do Programa de Pós-Graduação em Direito e Políticas Públicas



Documento assinado eletronicamente por **Saulo De Oliveira Pinto Coelho, Coordenador de Pós-graduação**, em 29/01/2021, às 15:20, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1834422** e o código CRC **3F28EC8D**.



PPGDP UFG <ppgdp.ufg@gmail.com>

Ata de defesa Sérgio Túlio

Fabiano Hartmann <fabianohpeixoto@gmail.com>

28 de janeiro de 2021 10:49

Para: PPGDP UFG <ppgdp.ufg@gmail.com>

Bom dia, estou de acordo com a Ata e autorizo os procedimentos respectivos para o trâmite da aprovação da defesa e titulação.

Estou às ordens,

Prof. Fabiano Hartmann Peixoto

[Texto das mensagens anteriores oculto]

--

Fabiano Hartmann Peixoto

Professor da Faculdade de Direito da Universidade de Brasília (FD/UnB)

Professor Credenciado no Programa de Pós-Graduação em Direito da Universidade de Brasília (PPGD/UnB)

Coordenador do PPGD/UnB

Líder do Grupo de Pesquisa DR.IA (Direito, Racionalidade e Inteligência Artificial)/CNPq

Coordenador Acadêmico do Projeto Victor (UnB/STF)

PhD in Law at University of Brasília - UnB, Permanent Professor of UnB Law School and of it`s PhD program in Law. Project

Coordinator of "Development and Research project of machine learning on judicial data of general repercussion in Brazilian

Supreme Court".Leader of a research group called DRIA-UnB (stands for Law, Rationality and Artificial Intelligence at University of Brasília), certified by CNPq (Brazilian council of national scientific and technological development).

<https://unb.academia.edu/FabianoHartmann>

À minha esposa Giovanna por me apoiar nesta jornada. Ao meu filho Theo, que chegará em breve para completar as nossas vidas.

AGRADECIMENTOS

Agradeço primeiramente a Deus, que me concedeu saúde e força para superar os desafios de me aventurar num programa de pós-graduação em áreas de conhecimento diversas de minha formação.

Aos meus pais, que sempre me proporcionaram a melhor educação possível.

À minha esposa, pelo apoio incondicional neste período e às minhas cachorrinhas pela companhia nas longas horas de dedicação.

Aos meus irmãos, sogros e amigos, pelo apoio e torcida.

Ao meu orientador, Prof. Dr. Cleuler, primeiramente por ter acreditado em mim, mas sobretudo pelos ensinamentos, pela dedicação à minha pesquisa e pelo exemplo de profissional multidisciplinar que pretendo ser.

Ao meu amigo Maurício, pela inestimável ajuda nos procedimentos operacionais iniciais da pesquisa, assim como ao Wesley, pelas valiosas contribuições decorrentes de seu estágio.

À Juliana, por me tranquilizar e ajudar no planejamento das minhas atividades.

Aos meus professores do Programa de Pós-graduação em Direito e Políticas Públicas (PPGDP), pelo conhecimento repassado, e aos meus colegas, pelas discussões e cooperações resultantes de uma jornada compartilhada.

Ao Tribunal de Contas do Estado de Goiás (TCE-GO), instituição que honro diariamente com meu trabalho, pela parceria institucional com a Universidade Federal de Goiás (UFG), que possibilitou meu ingresso neste Mestrado Profissional.

Aos meus colegas de trabalho, pela compreensão e incentivos.

RESUMO

No Brasil, as inovações trazidas pelo novo Código de Processo Civil (Lei nº 13.105, de 16 de março de 2015), formalizaram a *importância do uso de precedentes* em todos os níveis do ordenamento jurídico brasileiro. A problemática desta pesquisa advém da *necessidade mapear todas as decisões que versam sobre uma mesma temática*, para além de simples pesquisas por palavras-chave utilizadas atualmente, com vistas proporcionar maior segurança jurídica. No âmbito administrativo, as *decisões colegiadas do Tribunal de Contas da União (TCU) servem de orientação à atuação administrativa*, pois estabelecem critérios práticos que permitem aos gestores públicos concluir por uma solução possível diante de um caso concreto. Nesse contexto, esta pesquisa empírica teve como objetivos principais: (i) *criar um método replicável de delineamento de jurisprudência e identificação de precedentes com uso de Inteligência Artificial (IA) aplicável*, sobretudo, a decisões do TCU que versem sobre cláusulas restritivas ao caráter competitivo em editais de licitação; e (ii) *contribuir para a sedimentação da cultura de precedentes administrativos no país* e para a *delimitação jurisprudencial de conceitos jurídicos indeterminados*, mormente aqueles ligados a práticas corruptas. O recorte temático escolhido para aplicação da extração jurisprudencial, encontra-se no cerne das *fraudes em licitações, espécies do gênero corrupção*, e que apresentam como uma de suas patologias a *existência de cláusulas nos editais que restringem a competitividade do certame*. Assim, *testou-se a aplicação de técnicas de IA (text mining para classificação de documentos) nos acórdãos do TCU (entre 1992 e 2018), com intuito de sistematizar a delimitação jurisprudencial e consequente identificação inequívoca dos precedentes administrativos que dão concreção ao conceito aberto de cláusulas restritivas ao caráter competitivo em editais de licitação*. Mais de 300 acórdãos do TCU, previamente rotulados em 11 classes relacionadas à cláusulas de editais de licitação consideradas restritivas à competitividade, foram utilizados para treinar modelos de *machine learning* e *deep learning* para classificação *multi-label*, a fim de constatar se o modelo seria capaz de apontar quais outras decisões do TCU são pertencentes a alguma dessas classes. Os *resultados obtidos para as fases de treinamento e teste utilizando redes neurais convolucionais mostraram-se razoáveis*, pois apresentaram métrica de avaliação de 82,69%. Entretanto, na etapa de *avaliação supervisionada*, o *algoritmo de aprendizado profundo treinado resultou inconclusivo* para a delimitação jurisprudencial almejada. Não obstante os resultados insatisfatórios para a utilização de *IA* no delineamento da jurisprudência acerca do recorte temático escolhido, *considera-se que esta pesquisa logrou êxito, ao menos parcialmente, no atingimento de seus objetivos gerais*, visto que o *mapeamento do estado da arte computacional* aplicado ao problema da pesquisa e os *detalhes operacionais descritos* proporcionam *transferência de know-how para pesquisas que vislumbram a utilização de IA para sistematização jurisprudencial*. Demonstrou-se, também, a *importância de rastrear sistematicamente os precedentes administrativos para que haja maior segurança jurídica* não apenas aos *gestores públicos* em suas ações administrativas, mas também no *aprimoramento dos algoritmos utilizados por aplicações de IA que vasculham possíveis irregularidades*, como o robô Alice do TCU. Ainda, a fim de *contribuir para a concreção do conceito jurídico indeterminado* em estudo, apresentou-se uma *linha jurisprudencial de entendimento* com indicação dos *prováveis precedentes administrativos do TCU*, e suas respectivas *rationes decidendi*, que consideraram restritivas cláusulas em editais de licitação por exigirem capacidade técnica dos licitantes de maneira irregular.

Palavras-chave: Jurisprudência. Precedentes administrativos. Inteligência Artificial. Licitação pública. Cláusulas restritivas ao caráter competitivo.

ABSTRACT

In Brazil, the innovations brought by the new Civil Procedure Code (Law No. 13105/2015), formalized the importance of using precedents at all levels in the Brazilian legal system. This research arises from the need to congregate all decisions that deal with the same theme, in contrast to simple keyword searches used these days, in order to provide greater legal certainty. In administrative law, the Brazilian Federal Audit Court's (TCU) decisions serves as a guide for public actions, as they establish practical criteria that allow public managers to conclude for a possible solution in a specific case. In this context, this empirical research had as main objectives: (i) to create a replicable method of delineating jurisprudence and identifying precedents with the use of Artificial Intelligence (AI) applicable, above all, to TCU decisions that deal with clauses restricting the competitiveness in bidding notices; and (ii) to contribute to the administrative precedents culture sedimentation and to the jurisprudential delimitation of indeterminate legal concepts, especially those related to corrupt practices. The area chosen for jurisprudential extraction application is found at the heart of frauds in public bidding, that one of its corruption pathologies is the existence of clauses in bidding notices that restrict the competitiveness in public purchases. Thus, this research tested AI techniques (text mining for document classification) in the TCU decisions (between 1992 and 2018), with the purpose of systematize the jurisprudential delimitation and consequent unambiguous administrative precedents identification, which give concreteness to the concept of competitiveness restrictive clauses. More than 300 judgments, previously labeled in 11 different classes related to bidding clauses considered restrictive to the competitive nature of the public purchase, were used to train machine learning and deep learning models for multi-label classification, in order to verify whether the machine would be able to point out which others TCU decisions were related to any of these classes. The results using convolutional neural networks for the training and test phases proved to be reasonable, as they presented an evaluation metric of 82.69%. However, in the supervised assessment stage, the trained deep learning algorithm was inconclusive for the desired jurisprudential delimitation. Despite the unsatisfactory results this research was successful, at least partially, in reaching its general objectives, since the mapping of the computational state of art applied to the research problem and the operational details described provide know-how transferring for futures research that envisages the use of AI for jurisprudential systematization. It was also demonstrated the importance of systematically tracking administrative precedents so that there is greater legal security not only for public managers in their administrative actions, but also in improving the algorithms used by AI applications that search for possible irregularities, such as *Alice* the robot from TCU. Still, this study presents the construction of the jurisprudential line with indication of the probable administrative precedents and their *rationes decidendi* for the cases of clauses in bidding notices, considered restrictive because they demand technical capacity from bidders in an irregular manner.

Keywords: Jurisprudence. Administrative precedents. Artificial Intelligence. Public bidding. Competitiveness restrictive clauses.

LISTA DE ABREVIATURAS E SIGLAS

ABIC	Associação Brasileira da Indústria de Café
AUC	Area under the ROC curve
ARE	Agravo de Recurso Extraordinário
Atricon	Associação dos Membros dos Tribunais de Contas do Brasil
Bi-LSTM	Bidirectional Long Short-Term Memory
CAT	Certidão de Acervo Técnico
CF	Constituição Federal
CGU	Controladoria Geral da União
CNN	Convolutional Neural Network
COLIEE	Competition on Legal Information Extraction/Entailment
CPC	Código de Processo Civil
CREA	Conselho Regional de Engenharia e Agronomia
DL	Deep Learning
DNN	Deep Neural Network
DPAM	Dynamic Pairwise Attention Model
EC	Emenda Constitucional
ECHR	European Court of Human Rights
FSL	Few-Shot Learning
GMLC	Generative Multitask Learning based Classification
GPU	Unidade de Processamento Gráfico
HMN	Hierarchical Matching Network
IA	Inteligência Artificial
IDT	Iudicium Textum Dataset
IPEA	Instituto de Pesquisa Econômica Aplicada
LINDB	Lei de Introdução às Normas do Direito Brasileiro
Linear SVC	Linear Support Vector Classification
LSA	Latent Semantic Analysis
LSAN	Label Specific Attention Network
LSTM	Long Short-Term Memory
ML	Machine Learning

ML-KNN	Multi-label k Nearest Neighbors
MPF	Ministério Público Federal
NB	Naive Bayes
NLP	Processamento de Linguagem Natural
PBQPH	Programa Brasileiro da Qualidade e Produtividade do Habitat e Certificado de Boas Práticas de Fabricação e Controle
PIB	Produto Interno Bruto
POS-tagging	Part of Speech Tagging
PPGCC	Programa de Pós-graduação em Ciência da Computação
PPGDP	Programa de Pós-graduação em Direito e Políticas Públicas
RE	Recurso Extraordinário
RNN	Recurrent Neural Network
STF	Supremo Tribunal Federal
STJ	Superior Tribunal de Justiça
SVM	Support Vector Machine
TCU	Tribunal de Contas da União
TCE	Tribunal de Contas do Estado
TF-IDF	Term frequency–inverse document frequency
TPP	Three Phase Prediction
TRT	Tribunal Regional do Trabalho
UFG	Universidade Federal de Goiás
ZSL	Zero-Shot Learning

LISTA DE FIGURAS

Figura 1 – Representação gráfica da linha jurisprudencial de entendimento sobre um tema hipotético ao longo do tempo. Cada bola ou losango representa uma decisão ocorrida em um determinado tempo.	36
Figura 2 – Evolução anual da quantidade de artigos científicos publicados entre 1960 e 2018, por tópico encontrado na pesquisa de Rosca et al.	42
Figura 3 – Árvore de decisão para encontrar a área de prática de mineração de texto correta. Em destaque, as respostas escolhidas para o problema desta pesquisa.	45
Figura 4 – As sete áreas práticas de <i>text mining</i> (formas ovais) e como as tarefas específicas de <i>text mining</i> (rótulos dentro das formas ovais) existem em suas interseções. Na região destacada pela hachura em formato de losango da cor vermelha, situa-se o primeiro problema prático desta pesquisa.	47
Figura 5 – Esquema estrutural da corrupção que fundamenta o conceito de <i>corrupção</i> de Luiz Fernando Miranda.	49
Figura 6 – Tipos de corrupção enquadrados no ordenamento jurídico brasileiro.	50
Figura 7 – Compras governamentais do Brasil, em % do PIB, segundo entes da Federação (2006-2017).	53
Figura 8 – Nuvem das 30 mais frequentes palavras dos resumos dos 314 trabalhos científicos únicos extraídos das cinco bases de dados bibliográficos utilizadas.	71
Figura 9 – Análise de <i>coupling</i> dos 119 trabalhos científicos extraídos do <i>Scopus</i>	71
Figura 10 – Ilustração esquemática dos métodos de classificação de texto de 1961 a 2020. Antes de 2010, quase todos os métodos existentes eram baseados em <i>machine learning</i> (cor laranja); desde 2010, a maioria dos trabalhos nesta área tem se concentrado em esquemas de <i>deep learning</i> (cor verde).	76
Figura 11 – Fases do processo de classificação de texto.	76
Figura 12 – Detalhes do dataset <i>df_acordaos</i> , contendo 306.935 acórdãos e dois atributos para cada registro: um identificador único (ID) que remete ao tipo, ano, número e colegiado do acórdão; e uma coluna contendo o inteiro teor do relatório, voto e parte dispositiva da decisão colegiada (DECISAO).	93
Figura 13 – Quantidade de acórdãos rotulados por ano de julgamento.	95
Figura 14 – Detalhes do dataset <i>df_rotulados</i> , contendo 330 acórdãos previamente classificados em 11 classes possíveis. Às últimas colunas, foi atribuído 1 se o acórdão pertence aquela classe ou 0, se não pertence.	96
Figura 15 – Representação da transformação de leis, artigos e incisos em significantes únicos.	96
Figura 16 – Representação da transformação de citações de acórdãos anteriores do próprio TCU para significantes únicos.	97
Figura 17 – Representação em nuvem das palavras mais frequentes do <i>df_rotulados</i> , após aplicação das etapas de pré-processamento.	98

LISTA DE TABELAS

Tabela 1 – Resumo dos argumentos de busca e resultados dos trabalhos científicos encontrados para as bases de busca bibliográficas escolhidas.....	70
Tabela 2 – Critérios de qualidade aplicados aos 103 trabalhos científicos incluídos.....	72
Tabela 3 – Detalhe dos 33 trabalhos científicos qualificados após revisão sistemática.....	73
Tabela 4 – Integração dos resultados dos 26 trabalhos científicos de estudos práticos em classificação <i>multi-label</i> de textos jurídicos.....	82
Tabela 5 – Atributos utilizados dos acórdãos contidos na base de dados obtida por Fernanda Naves após requerimento formal ao TCU.	92
Tabela 6 – Quantidade de acórdãos, identificados nos trabalhos de Franklin Santos e Kleberson Souza e de Fernanda Naves. Os acórdãos foram classificados conforme as 11 classes estabelecidas por Santos e Souza e correspondem a diferentes tipos de cláusulas restritivas ao caráter competitivo em editais de licitação extraídas a partir de decisões do TCU que as consideraram restritivas à competitividade do certame.....	94
Tabela 7 – Resultados (AUC) do treinamento para modelos de ML e DL aplicados à classificação multi-label das classes 7, 2, 5, 8 e 10.	100
Tabela 8 – Quantidade de acórdãos para as subclasses da classe 7, identificados nos trabalhos de Franklin Santos e Kleberson Souza e de Fernanda Naves. Os acórdãos foram classificados conforme as subclasses estabelecidas por Santos e Souza e correspondem a decisões que versaram sobre cláusulas, em editais de licitação, que exigiram capacidade técnica dos licitantes de maneira irregular e que, por isso, foram consideradas restritivas ao caráter competitivo.	104
Tabela 9 – Precedentes administrativos (com suas respectivas rationes decidendi) do TCU identificados que versaram sobre cláusulas, em editais de licitação, que exigiram capacidade técnica dos licitantes de maneira irregular e que, por isso, foram consideradas restritivas ao caráter competitivo.	105

SUMÁRIO

1 INTRODUÇÃO	13
1.1 Delimitação epistemológica.....	13
1.2 Problematização.....	15
1.3 Hipótese	16
1.4 Justificativa	16
1.5 Análise da articulação com a prática profissional.....	19
1.6 Referenciais teóricos.....	19
1.7 Objetivos.....	22
1.7.1 Objetivos gerais	22
1.7.2 Objetivos específicos	22
1.8 Aspectos metodológicos.....	22
1.9 Organização do trabalho	23
2 LINGUAGEM, DIREITO E INTELIGÊNCIA ARTIFICIAL.....	24
2.1 A linguagem para Ludwig Wittgenstein em seus dois momentos.....	24
2.2 Da Teoria dos Modelos do Direito à construção argumentativa das decisões	27
2.3 Uso de Inteligência Artificial para identificação de precedentes administrativos.....	39
3 CLÁUSULA RESTRITIVA AO CARÁTER COMPETITIVO EM LICITAÇÕES: UMA PATOLOGIA CORRUPTIVA DE CONCEITO JURÍDICO INDETERMINADO	48
3.1 Fraude em licitações públicas: uma patologia corruptiva	48
3.2 Cláusula restritiva ao caráter competitivo das licitações: um conceito jurídico indeterminado	55
3.3 Uso de Inteligência Artificial pelo TCU para detecção de fraudes em licitações	64
4 ESTRATÉGIA EMPÍRICA: TENTATIVA DE CONCEPÇÃO DE UM MÉTODO DE EXTRAÇÃO SISTEMÁTICA DE JURISPRUDÊNCIA	68
4.1 Revisão sistemática da literatura com enfoque meta-analítico do estado da arte aplicado ao problema de pesquisa.....	68
4.1.1 Contextualização sobre a atividade de classificação de textos	75
4.1.2 Descobertas sobre o estado da arte em classificação <i>multi-label</i> de textos jurídicos ..	78
4.2 Procedimentos empíricos: aplicação das fases do processo de classificação de texto ao problema da pesquisa.....	90
4.2.1 Captura dos dados	90
4.2.2 Pré-processamento	96
4.2.3 Extração dos atributos.....	98

4.2.4 Redução da dimensionalidade	99
4.2.5 Treinamento dos classificadores	99
4.2.6 Avaliação da solução	100
4.3 Ensaio para identificação de precedentes administrativos que dão concretude ao conceito aberto de cláusulas restritivas em editais de licitação	103
5 CONCLUSÃO.....	116
REFERÊNCIAS	119
APÊNDICE A – LISTA DAS 330 DECISÕES COLEGIADAS DO TCU CLASSIFICADAS EM 11 CLASSES REFERENTES A PATOLOGIAS CORRUPATIVAS DE CLÁUSULAS RESTRITIVAS AO CARÁTER COMPETITIVO EM EDITAIS DE LICITAÇÃO.....	128
APÊNDICE B – <i>SCRIPT</i> PARA GERAR O <i>DF_ACORDAOS</i>.....	136
APÊNDICE C – <i>SCRIPT</i> PARA GERAR O <i>DF_ROTULADOS</i>	142
APÊNDICE D – <i>SCRIPT</i> DE PRÉ-PROCESSAMENTO DO <i>DF_ACORDAOS</i>	147
APÊNDICE E – <i>SCRIPT</i> DE PRÉ-PROCESSAMENTO DO <i>DF_ROTULADOS</i>.....	148
APÊNDICE F – <i>SCRIPT</i> DE TREINAMENTOS E TESTES DE MODELOS DE <i>MACHINE LEARNING</i>	152
APÊNDICE G – <i>SCRIPT</i> DE TREINAMENTOS E TESTES DE MODELOS DE <i>DEEP LEARNING</i>	164
APÊNDICE H – LINHA JURISPRUDENCIAL DE ENTENDIMENTO DA CLASSE 7	176

1 INTRODUÇÃO

1.1 Delimitação epistemológica

Em um cenário pós-positivista, a Ciência do Direito tem sofrido consequências de sua aproximação com a Racionalidade, dentre elas a necessidade de motivação (justificação racional) das decisões, bem como a constitucionalização das questões de justiça pelos princípios.¹ Tais efeitos manifestam-se com força no campo dos *precedentes judiciais*², na medida em que os princípios da *isonomia* e *universalizabilidade*³ aduzem o dever para o aplicador do Direito de observá-los. Por essas razões, a dicotomia entre um Direito inteiramente codificado (*civil law*) e outro jurisprudencial (*commom law*) tem ficado para trás.⁴

No Brasil, além das previsões constitucionais sobre a vinculatividade dos precedentes das cortes supremas (Constituição Federal – CF, art. 102, § 2º e art. 103-A)⁵, também são exemplos da aproximação entre *commom law* e *civil law* as inovações trazidas pelo novo

¹ Cf. BUSTAMANTE, Thomas da Rosa de. *Teoria do Precedente Judicial*. 1. ed. São Paulo: Noeses, 2012, p. 22-23.

² Conforme relatado no Capítulo 2 (seção 2.2), este trabalho adota o conceito de precedente judicial aduzido por Rafael Carvalho Rezende Oliveira, que assim preconiza: “Os precedentes são as decisões judiciais que enunciam regras ou princípios jurídicos que devem ser respeitados em casos futuros. Não configura precedente judicial vinculante a decisão judicial que simplesmente aplica precedente [...]”. OLIVEIRA, Rafael Carvalho Rezende. *Precedentes no Direito Administrativo*. 1. ed. Rio de Janeiro: Forense, 2018, p. 43.

³ Esse princípio, núcleo da concepção kentiana de racionalidade prática, exige dos aplicadores do Direito a utilização e extração de enunciados gerais e universais para contextos de similaridade em seus aspectos relevantes. Cf. Bustamante, *Teoria do Precedente Judicial*, op. cit., p. 113; 274; 298.

⁴ Cf. *ibid.* p. 93-96.

⁵ “Art. 102. Compete ao Supremo Tribunal Federal, precipuamente, a guarda da Constituição, cabendo-lhe: [...]

§ 2º As decisões definitivas de mérito, proferidas pelo Supremo Tribunal Federal, nas ações diretas de inconstitucionalidade e nas ações declaratórias de constitucionalidade produzirão eficácia contra todos e efeito vinculante, relativamente aos demais órgãos do Poder Judiciário e à administração pública direta e indireta, nas esferas federal, estadual e municipal.

[...]

Art. 103-A. O Supremo Tribunal Federal poderá, de ofício ou por provocação, mediante decisão de dois terços dos seus membros, após reiteradas decisões sobre matéria constitucional, aprovar súmula que, a partir de sua publicação na imprensa oficial, terá efeito vinculante em relação aos demais órgãos do Poder Judiciário e à administração pública direta e indireta, nas esferas federal, estadual e municipal, bem como proceder à sua revisão ou cancelamento, na forma estabelecida em lei.

§ 1º A súmula terá por objetivo a validade, a interpretação e a eficácia de normas determinadas, acerca das quais haja controvérsia atual entre órgãos judiciários ou entre esses e a administração pública que acarrete grave insegurança jurídica e relevante multiplicação de processos sobre questão idêntica.

§ 2º Sem prejuízo do que vier a ser estabelecido em lei, a aprovação, revisão ou cancelamento de súmula poderá ser provocada por aqueles que podem propor a ação direta de inconstitucionalidade.

§ 3º Do ato administrativo ou decisão judicial que contrariar a súmula aplicável ou que indevidamente a aplicar, caberá reclamação ao Supremo Tribunal Federal que, julgando-a procedente, anulará o ato administrativo ou cassará a decisão judicial reclamada, e determinará que outra seja proferida com ou sem a aplicação da súmula, conforme o caso.” BRASIL. *Constituição Federal de 5 de outubro de 1988*. Disponível em: http://www.planalto.gov.br/ccivil_03/Constituicao/Constituicao.htm. Acesso em: 23 out. 2020.

Código de Processo Civil – CPC (Lei nº 13.105, de 16 de março de 2015)⁶, que formalizaram a importância do uso de precedentes em todos os níveis do ordenamento jurídico brasileiro. Todavia, não existe no Brasil uma cultura difundida quanto ao uso de precedentes como ocorre na tradição anglo-saxônica, gerando riscos de resultar em simples citações de julgados.⁷

Nesse cenário, fazem-se necessários estudos aprofundados sobre o uso de precedentes na tradição jurídica da *civil law* brasileira, seja no âmbito do judiciário ou na esfera administrativa. Especialmente nessa última, a atuação administrativa por meio de precedentes, na medida em que impõe a análise de circunstâncias fáticas para aplicação a situações semelhantes, aumenta a segurança jurídica, agindo como direcionador da atuação dos agentes públicos e balizador de conceitos jurídicos indeterminados.⁸

Urge-se, nessa esteira, a *necessidade mapear as decisões que versam sobre uma mesma temática*, para além de simples pesquisas por palavras-chave, com vistas a servir de insumo tanto para juízes quanto para as partes interessadas. A *Inteligência Artificial*⁹ (IA) vem ao encontro dessa necessidade, na medida em que viabiliza a sistematização do conhecimento extraído das expressões linguísticas. Com os avanços tecnológicos das últimas décadas,¹⁰ existe grande oportunidade de utilização de técnicas avançadas de análise de expressões linguísticas pela máquina, sobretudo na esfera administrativa, podendo contribuir, inclusive, para a *prevenção e o combate a patologias corruptivas*.

Por exemplo, o Tribunal de Contas da União (TCU) divulga, quinzenalmente, em seu sítio eletrônico, o Informativo de Licitações e Contratos,¹¹ em que se apresentam selecionadas decisões das Câmaras e do Plenário do Tribunal, nas áreas de *licitação* e *contratos*. No entanto, como a própria instituição afirma, não se trata de repositório oficial de jurisprudência¹² do TCU, o que pode gerar incertezas para caracterização de conceitos abertos, a exemplo das cláusulas

⁶ BRASIL. Lei nº 13.105, de 16 de março de 2015. Código de Processo Civil. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2015/Lei/L13105.htm. Acesso em: 23 out. 2020.

⁷ Cf. LUVIZOTTO, Juliana Cristina. *Precedentes administrativos & a vinculação da atividade administrativa*. 2. ed. Curitiba: Juruá, 2017, p. 56. Na mesma linha, cf. Oliveira, *Precedentes no Direito Administrativo*, op. cit., p. 65.

⁸ Cf. Oliveira, *Precedentes no Direito Administrativo*, op. cit., p. 112-116.

⁹ A Inteligência Artificial é uma subárea da ciência da computação que faz modelagens computacionais do comportamento humano e pode ser vista como potencial solucionadora de problemas humanos dada sua capacidade de operacionalizar tarefas humanamente hercúleas com maior acurácia e velocidade, como a consolidação e organização de precedentes. PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblik Martins da. *Inteligência Artificial e Direito*. 1. ed. Curitiba: Alteridade, 2019, p. 22; 30; 44.

¹⁰ Cf. *ibid.*, p. 26-31.

¹¹ Disponível em: <https://portal.tcu.gov.br/jurisprudencia/boletins-e-informativos/informativo-de-licitacoes-e-contratos.htm>. Acesso em 26 out. 2019.

¹² Conforme relatado no Capítulo 2 (seção 2.2), este trabalho adota o conceito de jurisprudência trazido por Rafael Carvalho Rezende Oliveira, para quem, em sentido amplo, o termo refere-se ao conjunto de decisões uniformes ou divergentes sobre determinada temática. Cf. Oliveira, *Precedentes no Direito Administrativo*, op. cit., p. 45.

restritivas ao caráter competitivo em licitações (art. 3º, § 1º, inciso I, da Lei nº 8.666, de 21 de junho de 1993 – Lei Geral de Licitações)¹³.

Assim, *aplicar técnicas de IA nas decisões colegiadas (acórdãos) do TCU (entre 1992 e 2018), de modo a sistematizar a delimitação jurisprudencial e consequente identificação dos precedentes administrativos* para os casos de patologias corruptivas relacionadas à existência de *cláusulas restritivas ao caráter competitivo em editais de licitação*, constitui-se o desafio em que se insere este trabalho.

1.2 Problemática

Em linhas gerais, a *problemática* do tema da pesquisa se resume em: *como sistematizar, com uso de Inteligência Artificial, a extração da jurisprudência e dos precedentes administrativos do Tribunal de Contas da União, referentes a patologias corruptivas de modo a dar concretude jurisprudencial ao conceito jurídico indeterminado¹⁴ de cláusulas restritivas ao caráter competitivo em editais de licitação?*

Essa problemática pode ser dividida em *dois problemas de ordem prática*:

- a) Como delimitar, via Inteligência Artificial, todas as decisões colegiadas (jurisprudência) do TCU que versam sobre patologias corruptivas relacionadas a cláusulas restritivas à competitividade em editais de licitação?
- b) Uma vez delimitada a jurisprudência, como identificar os precedentes administrativos que dão concretude ao conceito aberto de cláusulas restritivas ao caráter competitivo em editais de licitação?

¹³ “Art. 3º A licitação destina-se a garantir a observância do princípio constitucional da isonomia, a seleção da proposta mais vantajosa para a administração e a promoção do desenvolvimento nacional sustentável e será processada e julgada em estrita conformidade com os princípios básicos da legalidade, da impessoalidade, da moralidade, da igualdade, da publicidade, da probidade administrativa, da vinculação ao instrumento convocatório, do julgamento objetivo e dos que lhes são correlatos.

§ 1º É vedado aos agentes públicos:

I - admitir, prever, incluir ou tolerar, nos atos de convocação, **cláusulas ou condições que comprometam, restrinjam ou frustrem o seu caráter competitivo**, inclusive nos casos de sociedades cooperativas, e estabeleçam preferências ou distinções em razão da naturalidade, da sede ou domicílio dos licitantes ou de qualquer outra circunstância impertinente ou irrelevante para o específico objeto do contrato, ressalvado o disposto nos §§ 5º a 12 deste artigo e no art. 3º da Lei nº 8.248, de 23 de outubro de 1991” (grifos nossos). BRASIL. *Lei nº 8.666, de 21 de junho de 1993 (Lei Geral de Licitações)*. Regulamenta o art. 37, inciso XXI, da Constituição Federal, institui normas para licitações e contratos da Administração Pública e dá outras providências. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/L8666cons.htm. Acesso em: 23 out. 2020.

¹⁴ Esta pesquisa adota como sinônimo de *conceitos jurídicos indeterminados* a expressão *conceitos abertos*. Maria Sylvia Zanella Di Pietro define *conceito jurídico indeterminado* como aquelas expressões linguísticas que são “empregadas para designar vocábulos ou expressões que não têm um sentido preciso, objetivo, determinado, mas que são encontrados com grande frequência nas regras jurídicas dos vários ramos do direito”. DI PIETRO, Maria Sylvia Zanella. *Discrecionalidade administrativa na Constituição de 1988*. 3. ed. São Paulo: Atlas, 2012, p. 90.

1.3 Hipótese

De modo análogo, apresenta-se as *hipóteses separadas por cada problema prático*, seguindo padrão sugerido por Miracy Barbosa de Sousa Gustin e Maria Teresa Fonseca Dias¹⁵:

- a) Considerando que toda expressão linguística possui uma forma lógica que remete ao seu significado, ideia traçada na obra *Tractatus Logico-Philosophicus* de Ludwig Wittgenstein, e o conceito de *ratio decidendi* como sendo o conjunto de normas adscritas identificáveis a partir do modelo silogístico utilizado na justificação das decisões, aduzido na Teoria dos Precedentes de Thomas da Rosa de Bustamante, afirma-se ser possível, utilizando IA e com acurácia razoável (ao menos 80%), *delimitar a jurisprudência (conjunto de decisões colegiadas) do TCU (entre 1992 e 2018) que trate de patologias corruptivas relacionadas a cláusulas restritivas ao caráter competitivo em editais de licitação*.
- b) Considerando a relação entre o uso das expressões linguísticas em inúmeros contextos e seus diferentes significados, traçada na obra *Investigações Filosóficas* de Ludwig Wittgenstein, e os *modelos jurídicos jurisdicionais*, responsáveis pela adequação das normas legais aos casos concretos e conceituados pela Teoria dos Modelos do Direito de Miguel Reale, afirma-se ser possível *identificar os precedentes administrativos que dão concretude ao conceito aberto de cláusulas restritivas ao caráter competitivo em editais de licitação, a partir da jurisprudência do TCU (entre 1992 e 2018) previamente delimitada*.

1.4 Justificativa

Justifica-se esta pesquisa dada a imperiosa necessidade de *rastrear sistematicamente*, dentre as decisões da Corte de Contas da União, os *precedentes administrativos* que dão concretude a tipologias corruptivas ligadas às licitações, como é o caso do conceito aberto de *cláusulas restritivas ao caráter competitivo em editais de licitação* (art. 3º, § 1º, inciso I, da Lei Geral de Licitações)¹⁶.

¹⁵ Cf. GUSTIN, Miracy Barbosa de Sousa; DIAS, Maria Teresa Fonseca. *(Re)Pensando a pesquisa jurídica*. 4. ed. rev. e atual. Belo Horizonte: Del Rey, 2013, p. 57.

¹⁶ Brasil, *Lei nº 8.666, de 21 de junho de 1993 (Lei Geral de Licitações)*, op. cit.

Essa urgência situa-se dentro de uma perspectiva na qual o conhecimento humano perpassa, necessariamente, pela linguagem, tal como aludido por João Maurício Adeodato no conceito de “*abismo gnoseológico*”. Para o jurista brasileiro, esse abismo constitui-se o centro das relações humanas com o mundo circundante, resultado de uma incompatibilidade entre o evento real, a ideia e a expressão linguística.¹⁷

Nesse contexto de “correlação sincrônica entre significante e significado”¹⁸, essa pesquisa guarda estrita afinidade com os modelos do direito de Miguel Reale, mais especificamente, com os *modelos jurídicos jurisdicionais*¹⁹, sua relação com a eficácia jurídica²⁰ e sua capacidade de suprir lacunas legais ou determinação hermenêutica dos modelos em vigor. Detalha-se, por esse prisma, a problemática da corrupção.

Abordada como patologia social, econômica e política, a corrupção tem se tornado uma preocupação em nível mundial pois gera efeitos indesejados na sociedade, impedindo que o interesse público seja atingido em sua totalidade.²¹ No cotidiano, a expressão tem absorvido característica “*guarda-chuva*”, abrigando diversas condutas.²²

Considerando a definição de *fraude*, estabelecida no artigo 90 da Lei Geral de Licitações²³, esta pesquisa entende que as *fraudes em licitações* são *espécies* do *gênero* *corrupção* e que uma tipologia de fraude é a presença de *cláusulas no edital de licitação que comprometem a livre concorrência*, induzindo o resultado para licitantes favorecidos.²⁴ Ademais, afirma-se que a expressão *cláusula restritiva ao caráter competitivo*, trazida pelo art. 3º, § 1º, inciso I, da Lei Geral de Licitações,²⁵ é um conceito jurídico indeterminado de

¹⁷ Cf. ADEODATO, João Maurício. *Ética & Retórica: Para uma teoria da dogmática jurídica*. 4. ed. São Paulo: Saraiva, 2012, p. 216; Cf. ADEODATO, João Maurício. *Uma teoria retórica da norma jurídica e do direito subjetivo*. 2. ed. São Paulo: Saraiva, 2014, p. 39-42.

¹⁸ Cf. REALE, Miguel. *Fontes e modelos do direito: para um novo paradigma hermenêutico*. 1. ed. São Paulo: Saraiva, 1994, p. 44.

¹⁹ Cf. *ibid.*, p. 69-72.

²⁰ Cf. *ibid.*, p. 4.

²¹ Cf. LEAL, Rogério Gesta; SCHNEIDER, Yuri. Os efeitos deletérios da corrupção em face dos Direitos Humanos e Fundamentais. *Revista da AJURIS*, v. 41, n. 136, dez. 2014, p. 431.

²² Cf. TRIBUNAL DE CONTAS DA UNIÃO. *Referencial de combate à fraude e corrupção: aplicável a órgãos e entidades da Administração Pública*. 2. ed. 2018, p. 13-14. Disponível em: <https://portal.tcu.gov.br/referencial-de-combate-a-fraude-e-corruptao.htm>. Acesso em: 17 jul. 2019.

²³ “Art. 90. Frustrar ou fraudar, mediante ajuste, combinação ou qualquer outro expediente, o caráter competitivo do procedimento licitatório, com o intuito de obter, para si ou para outrem, vantagem decorrente da adjudicação do objeto da licitação:

Pena - detenção, de 2 (dois) a 4 (quatro) anos, e multa”. Brasil, *Lei nº 8.666, de 21 de junho de 1993 (Lei Geral de Licitações)*, op. cit.

²⁴ Cf. SANTOS, Franklin Brasil; SOUZA, Kleber Roberto de. *Como combater a corrupção em licitações: detecção e prevenção de fraudes*. 3. ed. Belo Horizonte: Fórum, 2020, p. 75-110.

²⁵ Brasil, *Lei nº 8.666, de 21 de junho de 1993 (Lei Geral de Licitações)*, op. cit.

*experiência*²⁶ e, por isso, possui um *núcleo de incerteza* que deve ser delimitado a partir da *praxe administrativa*²⁷, que estabelece critérios práticos que permitem concluir por uma solução possível diante de um caso concreto.

As *fraudes em licitações* merecem atenção por parte dos gestores e controladores públicos, visto que as compras governamentais envolvem grande volume de recursos humanos e financeiros (média de 12,5% do Produto Interno Bruto anual)²⁸.

Nessa conjuntura, considerando a função constitucional de controle dos atos administrativos dos Tribunais de Contas, condutas que se encaixam como patologias corruptivas são identificadas no decorrer de processos de controle externo e emergem nas decisões dessas Cortes. Quando o assunto é licitações, a Corte de Contas da União possui força institucional de direcionar a atuação dos administradores públicos em todas as esferas, agindo como indutor de boas práticas e evitando a ocorrência de irregularidades.²⁹

Dessarte, à luz dos modelos jurisdicionais de Miguel Reale, os *precedentes administrativos da Corte de Contas da União dão concretude ao conceito jurídico indeterminado de cláusulas restritivas de editais de licitação*. Reitera-se assim, a importância

²⁶ Para Maria Sylvia Zanella Di Pietro nos casos de *conceitos jurídicos indeterminados de experiência*, “a discricionariedade [administrativa] fica afastada porque existem critérios objetivos, práticos, extraídos da experiência comum, que permitem concluir qual a única solução possível”. Di Pietro, *Discricionariedade administrativa na Constituição de 1988*, op. cit., p. 118.

²⁷ “A praxe administrativa é a atividade interna, reiterada e uniforme da Administração Pública na aplicação das normas e dos atos jurídicos”. Oliveira, *Precedentes no Direito Administrativo*, op. cit., p. 118.

²⁸ RIBEIRO, Cássio Garcia; INÁCIO JÚNIOR, Edmundo. O mercado de compras governamentais brasileiro (2006-2017): mensuração e análise. *Texto para Discussão*, n. 2476, 2019, p. 18-19. Disponível em: http://repositorio.ipea.gov.br/bitstream/11058/9315/1/td_2476.pdf. Acesso em 23 out. 2020.

²⁹ “SÚMULA Nº 222. As Decisões do Tribunal de Contas da União, relativas à aplicação de normas gerais de licitação, sobre as quais cabe privativamente à União legislar, devem ser acatadas pelos administradores dos Poderes da União, dos Estados, do Distrito Federal e dos Municípios.” TRIBUNAL DE CONTAS DA UNIÃO. *Súmulas* Nº 001 a 289. Disponível em: <https://portal.tcu.gov.br/lumis/portal/file/fileDownload.jsp?fileId=8A8182A25753C20F0157679AA5617071&inLine=1>. Acesso em: 28 out. 2020.

Cf. também o Art. 22, XXVII c/c art. 37, XXI, CF/88:

“Art. 22. Compete privativamente à União legislar sobre:

[...]

XXVII – normas gerais de licitação e contratação, em todas as modalidades, para as administrações públicas diretas, autárquicas e fundacionais da União, Estados, Distrito Federal e Municípios, obedecido o disposto no art. 37, XXI, e para as empresas públicas e sociedades de economia mista, nos termos do art. 173, § 1º, III;

[...]

Art. 37. A administração pública direta e indireta de qualquer dos Poderes da União, dos Estados, do Distrito Federal e dos Municípios obedecerá aos princípios de legalidade, impessoalidade, moralidade, publicidade e eficiência e, também, ao seguinte:

[...]

XXI - ressalvados os casos especificados na legislação, as obras, serviços, compras e alienações serão contratados mediante processo de licitação pública que assegure igualdade de condições a todos os concorrentes, com cláusulas que estabeleçam obrigações de pagamento, mantidas as condições efetivas da proposta, nos termos da lei, o qual somente permitirá as exigências de qualificação técnica e econômica indispensáveis à garantia do cumprimento das obrigações.” Brasil, *Constituição Federal de 5 de outubro de 1988*, op. cit.

de delimitar, de maneira sistemática, a jurisprudência correlata de modo a possibilitar a identificação inequívoca dos precedentes administrativos relacionados a esse conceito aberto.

1.5 Análise da articulação com a prática profissional

A pesquisa que por ora se apresenta relaciona-se com a prática profissional, pois visa trazer um aspecto *inovador ao diagnóstico de patologias corruptivas por meio do uso de Inteligência Artificial na determinação de um conceito jurídico indeterminado*. O sítio eletrônico do TCU é referência para os demais tribunais de contas quanto a pesquisa nas bases de dados jurisprudenciais. Entretanto, ele se limita às buscas por palavras-chave em qualquer parte do texto das decisões colegiadas³⁰ ou à visualização de acórdãos previamente selecionados pela própria Corte (aproximadamente 14 mil acórdãos) e que são classificados em 10 áreas, 393 temas e 817 subtemas³¹.

Por isso almeja-se, ao final, apresentar um *método de extração jurisprudencial* que se aplique universalmente às Cortes de Contas, independentemente do enfoque específico, contribuindo para a delimitação de conceitos abertos, mormente aqueles ligados a práticas corruptas. Tal método poderia até ser estendido para a pesquisa da concreção de conceitos a partir das decisões dos Tribunais de Justiça, estaduais ou federais, inferiores ou superiores.

Assim, busca-se *colaborar para sedimentação da cultura dos precedentes administrativos no país* ao desenvolver um procedimento que visa aumentar a segurança jurídica e ajudar os operadores do Direito em pesquisas jurisprudenciais.³² Ademais, o desafio proposto nesta pesquisa pode contribuir para que exista *maior segurança jurídica* também aos *auditores de controle externo no aprimoramento dos algoritmos de IA utilizados para busca por irregularidades em licitações*, em especial, aquelas relacionadas a cláusulas restritivas à competitividade.

1.6 Referenciais teóricos

O marco teórico em que esta pesquisa se assenta está na relação entre fato, significante e significado, abordada de diferentes formas nas obras de do filósofo austríaco Ludwig

³⁰ O Portal de Jurisprudência de acórdãos do TCU está disponível em: <https://pesquisa.apps.tcu.gov.br/#/pesquisa/acordao-completo>. Acesso em 18 out. 2019.

³¹ O Portal de Jurisprudência Selecionada do TCU está disponível em: <https://pesquisa.apps.tcu.gov.br/#/pesquisa/jurisprudencia-selecionada>. Acesso em 18 out. 2019.

³² Cf. Oliveira, *Precedentes no Direito Administrativo*, op. cit., p. 165.

Wittgenstein. Com a linguagem como tema principal de sua reflexão, pode-se separar as contribuições filosóficas de Wittgenstein em dois períodos: o primeiro, que corresponde ao seu *Tractatus Logico-Philosophicus*³³, publicado em 1921, e o segundo, cuja obra principal é *Investigações Filosóficas*³⁴, publicada postumamente, em 1952.

No *Tractatus* o autor dá ênfase à função designativa – ou representativa – da linguagem, a partir da construção da relação entre uma figuração (seja ela espacial, por meio de um diagrama, ou expressão linguística, por meio de uma proposição) e o fato afigurado, isto é, aquilo que se quer representar. Para o filósofo, o que a figuração tem em comum com o fato diagramado é a “forma de afiguração”. Advém daí, a ideia de isomorfismo entre a linguagem e o real: a expressão linguística possui uma forma lógica (léxico, sintaxe)³⁵ que remete a seu significado.³⁶

Posteriormente, Wittgenstein amplia seu entendimento sobre linguagem em *Investigações Filosóficas*. A partir de então, o filósofo passa a defender que não há propriamente nenhuma ideia única atrás do significante: há apenas ele, e a função por ele desempenhada nos jogos de linguagem dos homens.³⁷ Nessa acepção, o significado de uma expressão linguística não seria mais estabelecido unicamente pela forma da sentença, mas por sua função exercida no contexto utilizado.³⁸

Dado o caráter interdisciplinar desta pesquisa, que relaciona Linguística, Direito e Ciência de Dados (*Data Science*), faz-se mister destacar o marco teórico jurídico pelos quais os problemas da pesquisa são observados.

Apoia-se na *Teoria dos Modelos do Direito* de Miguel Reale³⁹, no sentido de que *modelos* são o conteúdo das fontes do Direito, sendo “estruturas que surgem e se elaboram no contexto [...] da experiência”⁴⁰. Logo, a concreção de um conceito jurídico indeterminado⁴¹ –

³³ WITTGENSTEIN, Ludwig. *Tractatus Logico-Philosophicus*. Trad. Luiz Henrique Borges dos Santos (edição bilíngue). 3. ed. São Paulo: Edusp, 2017.

³⁴ WITTGENSTEIN, Ludwig. *Investigações filosóficas*. Trad. Emmanuel Carneiro Leão. 1. ed. São Paulo: Universitária São Francisco, 2014.

³⁵ Na mesma linha, cf. FERRAZ JR., Tercio Sampaio. *Introdução ao Estudo do Direito: técnica, decisão, dominação*. 10. ed. São Paulo: Atlas, 2018, p. 240-241; e cf. BENGFORT, Benjamin; BILBRO, Rebecca; OJEDA, Tony. *Applied Text Analysis with Python*. 1. ed. Sebastopol: O'Reilly Media, 2018, p. 15.

³⁶ Cf. Wittgenstein, *Tractatus Logico-Philosophicus*, op. cit., p. 51-60; 137-145.

³⁷ Cf. Wittgenstein, *Investigações filosóficas*, op. cit., p. 18-19.

³⁸ MARCONDES, Danilo. *Textos básicos de linguagem: de Platão a Foucault*. 1 ed. Rio de Janeiro: Zahar, 2010, p. 105-106.

³⁹ REALE, Miguel. *Fontes e modelos do Direito: para um novo paradigma hermenêutico*. 1 ed. São Paulo: Saraiva, 1994.

⁴⁰ Ibid., p. 49-50.

⁴¹ “[...] o conceito indeterminado sempre admite uma determinação, isto é, conceitos indeterminados são, presumidamente, determináveis, o que acontece por um processo de refinamento progressivo de seu sentido.” Ferraz Jr., *Introdução ao Estudo do Direito: técnica, decisão, dominação*, op. cit. p. 249.

como é o caso de cláusulas restritivas ao caráter competitivo em editais de licitação – perpassa pela identificação de *modelos* utilizados no plano da eficácia, dentre eles, os *precedentes administrativos* do TCU.

Assim, transporta-se da tese de doutorado de Thomas da Rosa de Bustamante⁴², importantes conceitos de sua *Teoria dos Precedentes* para presente o estudo, mormente quanto ao de *ratio decidendi* como sendo o conjunto de normas adscritas identificáveis a partir do modelo silogístico utilizado na justificação das decisões.

Dessa forma, pelos dois prismas linguísticos de Wittgenstein e à luz dos marcos teóricos jurídicos da Teoria dos Modelos do Direito de Reale e da Teoria dos Precedentes de Bustamante, pretende-se nortear a escolha das técnicas de IA mais adequadas para o problema da pesquisa. Ou seja, com o uso de IA e a partir do significado das *rationes decidendi* que congregam decisões correlatas, delimitar a jurisprudência (conjunto de decisões colegiadas) que versa sobre cláusulas restritivas ao caráter competitivo em editais de licitações. Em seguida, diferenciar a função exercida por essas expressões linguísticas dentro do contexto do inteiro teor de cada acórdão e, conseqüentemente, caracterizar os *modelos jurídicos* por meio da identificação dos *precedentes administrativos* que dão concretude ao conceito jurídico indeterminado em estudo.

Por fim, ressalta-se que este trabalho parte do pressuposto conceitual⁴³ do gênero *corrupção* na visão de Luiz Fernando Miranda⁴⁴ dentro do qual a *fraude em licitações* constitui-se uma *espécie*. Ainda, esta pesquisa *parte dos estudos empíricos* de Franklin Brasil Santos e Kleberson Roberto de Souza⁴⁵, bem como de Fernanda de Moura Ribeiro Naves⁴⁶, que *juntos identificaram mais de 300 acórdãos do TCU que versaram sobre a ilegalidade de cláusulas de editais de licitação consideradas restritivas ao caráter competitivo* em casos concretos, e que foram *classificados em 11 tipologias diferentes* para essa patologia corruptiva.

⁴² Bustamante, *Teoria do Precedente Judicial*, op. cit., 2012.

⁴³ Os pressupostos conceituais de uma pesquisa “são os conceitos ou proposições genericamente aceitos em determinado campo ou área do conhecimento [...]. Esses pressupostos, entretanto, nem sempre são igualmente conceituados entre doutrinadores ou teóricos de campos de conhecimento diferenciados. Quando isso ocorre, torna-se indispensável que o problema [de pesquisa] seja seguido do conceito exato de cada elemento para não pairar dúvida sobre a conceituação que se está usando”. Gustin; Dias, *(Re)Pensando a pesquisa jurídica*, op. cit., p. 51.

⁴⁴ MIRANDA, Luiz Fernando. Unificando os conceitos de corrupção: uma abordagem através da nova metodologia dos conceitos. *Revista Brasileira de Ciência Política*, n. 25, 2018.

⁴⁵ Cf. Santos; Souza, *Como combater a corrupção em licitações: detecção e prevenção de fraudes*, op. cit.

⁴⁶ NAVES, Fernanda de Moura Ribeiro. *TCU e stare decisis administrativo: o caso das cláusulas restritivas em editais de licitação de obras de engenharia*. Dissertação (mestrado em Direito e Políticas Públicas) – Universidade Federal de Goiás, Goiânia, 2019.

1.7 Objetivos

1.7.1 Objetivos gerais

a) *Criar um método replicável de delineamento de jurisprudência e identificação de precedentes com uso de IA* aplicável, sobretudo, a decisões do TCU que versem sobre cláusulas restritivas ao caráter competitivo em editais de licitação.

b) *Contribuir para a sedimentação da cultura de precedentes administrativos no país*, bem como para a *delimitação jurisprudencial de conceitos jurídicos indeterminados*, mormente aqueles ligados a práticas corruptas.

1.7.2 Objetivos específicos

a) *Aprofundar os marcos teóricos linguísticos e jurídicos* pelos quais o problema de pesquisa será observado;

b) *Escolher as técnicas computacionais de IA* mais adequadas ao desenvolvimento empírico deste trabalho;

c) *Apresentar o recorte temático* para aplicar a extração jurisprudencial com uso de IA por meio da *delimitação dos pressupostos conceituais: corrupção, fraude em licitações e cláusula restritiva ao caráter competitivo*;

d) *Traçar um panorama do uso de IA pelo TCU* na busca por *fraudes em licitações*;

e) *Apresentar o estado da arte computacional aplicado à extração jurisprudencial*, a partir das técnicas de IA previamente escolhidas (b);

f) *Delinear sistematicamente*, a partir das decisões colegiadas do TCU (entre 1992 e 2018), utilizando IA e com acurácia razoável (ao menos 80%), a *jurisprudência* (conjunto de acórdãos) que versa sobre patologias corruptivas relacionadas a cláusulas restritivas ao caráter competitivo em editais de licitações.

g) *Identificar os precedentes administrativos* que dão concretude ao conceito aberto de cláusulas restritivas ao caráter competitivo em editais de licitação, a partir da jurisprudência do TCU (entre 1992 e 2018) previamente delineada por meio de IA.

1.8 Aspectos metodológicos

A pesquisa pertence à *vertente jurídico-dogmática*⁴⁷ por centrar-se no estudo da jurisprudência do Tribunal de Contas da União para análise da eficácia do Direito na concreção de conceitos jurídicos indeterminados. Por conseguinte, encaixa-se no *gênero da pesquisa empírica*, uma vez que se pretende chegar a conclusões baseadas em observação e experimentação.⁴⁸

O trabalho segue o tipo de *investigação histórico-jurídica*⁴⁹, de modo a analisar a evolução dos acórdãos do TCU através do espaço/tempo, utilizando-se do *raciocínio investigativo hipotético-dedutivo*, por meio de testes de falseamento de hipóteses.⁵⁰

Quanto à natureza (tipologia) e limite temporal dos dados, utiliza-se de *dados primários*,⁵¹ quais sejam, todas as decisões colegiadas daquele Tribunal entre 1992 a 2018. Pelo *aspecto operacional*, esses dados são manipulados de acordo com as *melhores práticas computacionais* aplicadas a problemas análogos solucionados com uso de Inteligência Artificial.

1.9 Organização do trabalho

O *Capítulo 2* consiste no *aprofundamento dos marcos teóricos (linguístico e jurídico)*, de modo a relacioná-los com a problemática da pesquisa, bem como explicitar de que forma eles norteiam as abordagens empíricas escolhidas. Apresenta-se, também, a *conformação teórica do problema de pesquisa* no âmbito nas áreas práticas de Inteligência Artificial.

O *Capítulo 3*, destina-se à *apresentação do recorte temático escolhido* para aplicação das técnicas de IA para extração jurisprudencial. Detalha-se a *patologia corruptiva de cláusulas restritivas ao caráter competitivo em editais de licitação*, destaca-se o *papel dos precedentes administrativos* do TCU na *concreção desse conceito jurídico indeterminado* e apresenta-se um *panorama do uso de IA pelo TCU na busca por fraudes em licitações*.

O *Capítulo 4*, destina-se à *apresentação dos resultados empíricos* deste estudo. Inicialmente detalha-se o *método da pesquisa* a partir de uma revisão sistemática da literatura, com enfoque meta-analítico. Em seguida, apresentam-se os *modelos implementados* e os *resultados obtidos*. Por fim, o *Capítulo 5* contém as conclusões e considerações finais.

⁴⁷ Cf. Gustin; Dias, (Re)Pensando a pesquisa jurídica, op. cit., p. 21.

⁴⁸ Cf. EPSTEIN, Lee; KING, Gary. *Pesquisa empírica em Direito*. São Paulo: Direito GV, 2013, p. 24. Disponível em: <http://bibliotecadigital.fgv.br/dspace/handle/10438/11444>. Acesso em: 28 out. 2019.

⁴⁹ Cf. Gustin; Dias, (Re)Pensando a pesquisa jurídica, op. cit., p. 25.

⁵⁰ POPPER, Karl Raimund. *A lógica da pesquisa científica*. Tradução Leônidas Hegenberg; Octanny Silveira da Mota. 2. ed. São Paulo: Cultrix, 2013 [1972].

⁵¹ Gustin; Dias, (Re)Pensando a pesquisa jurídica, op. cit., p. 29-30 e 74-46.

2 LINGUAGEM, DIREITO E INTELIGÊNCIA ARTIFICIAL

Neste capítulo, inicialmente são detalhados os *marcos teóricos linguístico e jurídico* pelos quais o problema de pesquisa é observado. Na última seção, é apresentado um histórico da Inteligência Artificial aplicada a problemas jurídicos e, à luz dos marcos teóricos descritos, são escolhidas as *técnicas computacionais de Inteligência Artificial mais adequadas* ao desenvolvimento empírico deste trabalho.

2.1 A linguagem para Ludwig Wittgenstein em seus dois momentos

Com a linguagem como tema principal de sua reflexão, pode-se separar as contribuições filosóficas de Ludwig Wittgenstein em dois períodos: o primeiro, que corresponderia ao seu *Tractatus Logico-Philosophicus*, publicado em 1921, e o segundo, cuja obra principal é *Investigações Filosóficas*, publicada postumamente, em 1952.

No *Tractatus* o autor dá ênfase à *função designativa ou representativa da linguagem*, a partir da construção da relação entre uma figuração – seja ela espacial, por meio de um diagrama, ou expressão linguística, por meio de uma proposição – e o fato afigurado, isto é, aquilo que se quer representar. Para o filósofo, o que a figuração tem em comum com o fato afigurado é a “forma afiguração”.⁵² Veja-se:

2.151 A forma de afiguração é a possibilidade de que as coisas estejam umas para as outras tal como os elementos da figuração.

2.1511 É assim que a figuração se enlaça com a realidade; ela vai até a realidade.

2.1512 Ela é como uma régua aposta à realidade.

2.15121 Apenas os pontos mais externos das marcas da régua tocam o objeto a ser medido.

[...]

2.16 O fato, para ser uma figuração, deve ter algo em comum com o afigurado.

2.161 Na figuração e no afigurado deve haver algo de idêntico, a fim de que um possa ser, de modo geral, uma figuração do outro.

2.17 O que a figuração deve ter em comum com a realidade para poder afigurá-la à sua maneira – correta ou falsamente – é sua forma de afiguração.

[...]

2.18 O que toda figuração, qualquer que seja sua forma, deve ter em comum com a realidade para poder de algum modo – correta ou falsamente – afigurá-la é a forma lógica, isto é, a forma da realidade.⁵³

⁵² Cf. Wittgenstein, *Tractatus Logico-Philosophicus*, op. cit., p. 51-60.

⁵³ Ibid., p. 137.

De toda sorte, o filósofo austríaco pondera haver algo que distinga os dois (figuração e fato afigurado), sem o que eles seriam um só. Decorre daí a “*tese da bipolaridade*”⁵⁴, na qual se exprime no princípio da independência do sentido da figuração com respeito à sua verdade ou falsidade, ou seja, a verdade de uma figuração não pode estar entre as condições de sua significatividade. Em suma, toda proposição é uma figuração lógica, correta ou incorreta da realidade.⁵⁵

1.2 O mundo resolve-se em fatos.

[...]

2.063 A realidade total é o mundo.

[...]

2.1 Figuramos os fatos.

[...]

2.12 A figuração é um modelo da realidade.

[...]

2.21 A figuração concorda ou não com a realidade; é correta ou incorreta, verdadeira ou falsa.

[...]

2.221 O que a figuração representa é seu sentido.

2.222 Na concordância ou discordância de seu sentido com a realidade consiste sua verdade ou falsidade.⁵⁶

Nessa esteira, Wittgenstein desenvolve a ideia de *isomorfismo entre a linguagem e o real*,⁵⁷ adotada aqui como o ponto de partida para o primeiro problema prático da pesquisa: *a expressão linguística (proposição, para o filósofo) possui uma forma lógica (léxico, sintaxe)*⁵⁸ *que remete a seu significado*.

3.1 Na proposição o pensamento exprime-se de maneira sensivelmente perceptível.

3.11 Utilizamos o sinal sensivelmente perceptível (sinal escrito ou sonoro, etc. (*sic*)) da proposição como projeção da situação possível. [...]

[...]

3.12 O sinal por meio do que exprimimos o pensamento chamo de sinal proposicional. E a proposição é o sinal proposicional em sua relação projetiva com o mundo.

3.13 [...] Na proposição, portanto, ainda não está contido seu sentido, mas sim a possibilidade de exprimi-lo.

[...]

3.14 O sinal proposicional consiste em que seus elementos, as palavras, nele estão, uns para os outros, de uma determinada maneira. [...]

[...]

3.141 A proposição não é uma mistura de palavras. [...]

[...]

⁵⁴ Cf. *ibid.*, p. 51.

⁵⁵ Cf. *ibid.*, p. 62.

⁵⁶ *Ibid.*, p. 129-139.

⁵⁷ Cf. Marcondes, *Textos básicos de linguagem: de Platão a Foucault*, op. cit., p. 104.

⁵⁸ Na mesma linha, cf. Ferraz Jr., *Introdução ao Estudo do Direito: técnica, decisão, dominação*, op. cit., p. 240-241; e cf. Bengfort et al., *Applied Text Analysis with Python*, op. cit., p. 15.

3.3 Só a proposição tem sentido; é só no contexto da proposição que um nome tem significado.⁵⁹

Ao final de sua carreira filosófica, Ludwig Wittgenstein *amplia seu entendimento* sobre linguagem na obra *Investigações Filosóficas*. A partir de então, o filósofo passa a defender que não há propriamente nenhuma ideia única atrás do significante: há apenas ele, e a *função por ele desempenhada* nos “jogos de linguagem” dos homens.⁶⁰

13. Quando dizemos: “cada palavra da linguagem designa alguma coisa”, com isso ainda não se disse por enquanto *absolutamente* nada; a não ser que explicássemos, exatamente, *que* distinção desejamos fazer. [...]

[...]

15. A palavra “designar” é empregada de modo mais direto talvez lá onde o signo repousa sobre o objeto que designa. [...]

[...]

40. Falemos, primeiramente, sobre o ponto central desta argumentação: a palavra não tem significado algum quando nada lhe corresponde. – É importante constatar que a palavra “significado” é usada de um modo que vai contra a linguagem quando com ela se designa a coisa que “corresponde” à palavra. Isto significa: confundir o significado de um nome com o *portador* do nome. [...]

[...]

43. Para uma *grande* classe de casos – mesmo que não para *todos* – de utilização da palavra “significado”, pode se explicar esta palavra do seguinte modo: O significado de uma palavra é seu uso na linguagem.⁶¹

Nesse sentido, *o significado de uma expressão linguística não seria mais estabelecido unicamente pela sua forma lógica, mas pelo seu uso ou função exercida no contexto utilizado*: “O significado é assim indeterminado, só podendo ser compreendido através da consideração do jogo de linguagem, o que envolve mais do que a simples análise da sentença”⁶². Assim, Wittgenstein concebeu o que ficou consagrada como “*the use theory of meaning*”⁶³, em que o *significado de uma palavra é definido pelas circunstâncias de seu uso*.

Portanto, à luz do *primeiro enfoque linguístico de Wittgenstein*, almeja-se *delimitar a jurisprudência* (conjunto de decisões colegiadas) do TCU que verse sobre cláusulas restritivas ao caráter competitivo em editais de licitações, a partir do *significado advindo da forma lógica das expressões linguísticas*.

Por outro lado, a partir da *concepção de “jogos de linguagem”*, operacionaliza-se o segundo problema prático deste trabalho, uma vez que se pretende *identificar os precedentes administrativos que dão concretude ao conceito aberto de cláusulas restritivas ao caráter*

⁵⁹ Wittgenstein, *Tractatus Logico-Philosophicus*, op. cit., p. 141-145.

⁶⁰ Cf. Wittgenstein, *Investigações filosóficas*, op. cit., p. 18-19.

⁶¹ Wittgenstein, *Investigações filosóficas*, op. cit., p. 21;37-38.

⁶² Marcondes, *Textos básicos de linguagem: de Platão a Foucault*, op. cit., p. 106.

⁶³ MANNING, Christopher D.; SCHUTZE, Hinrich. *Foundations of statistical natural language processing*. 6. ed. Cambridge: MIT Press, 2003, p. 17.

competitivo em editais de licitação, com base na função exercida pelas expressões linguísticas no contexto de cada decisão pertencente à jurisprudência delimitada.

2.2 Da Teoria dos Modelos do Direito à construção argumentativa das decisões

Para melhor compreensão da *Teoria dos Modelos do Direito* de Miguel Reale, faz-se mister relembrar a “fórmula realiana”: “o Direito é uma integração normativa de fatos segundo valores”⁶⁴, advinda da *Teoria Tridimensional do Direito*, do mesmo autor. Por essa perspectiva, os três elementos (fato, valor e norma) se dialetizam consoante a elasticidade da norma⁶⁵ em suas variações interpretativas.⁶⁶

Respalado nesse raciocínio, Reale introduz o conceito de *modelos* como conteúdo das fontes do Direito, cabendo ao intérprete se atentar tanto à fonte originária, cuja intenção não é desimportante, quanto aos fatos e valores de cada caso concreto. Assim, os modelos materializam a elasticidade das normas, promovendo a efetivação das regras em todo o leque de possibilidades hermenêuticas, até que surja a necessidade de revogação de uma norma vigente.⁶⁷

Miguel Reale os divide em *modelos jurídicos* e *modelos hermenêuticos*, cuja diferença se resume na natureza prescritiva do primeiro, ou seja, na capacidade de reger atos futuros, enquanto os modelos hermenêuticos, por seu viés teórico, não obrigam alguém a agir em conformidade com suas conclusões.⁶⁸ Dessa divisão, o autor *define modelos jurídicos* como “resultado da ordenação racional do conteúdo das normas reveladas ou formalizadas pelas fontes de direito, para atender aos característicos de *validade objetiva autônoma e de atualização prospectiva* dessas mesmas normas”⁶⁹ ou ainda, “estruturas que surgem e se elaboram no contexto [...] da experiência [jurídica]”⁷⁰. Finalmente:

[...] *modelo jurídico*, apresentando-o como ‘*estrutura normativa de atos e fatos pertinentes unitariamente a dado campo da experiência social, prescrevendo a atualização racional e garantida dos valores que lhes são próprios*’. Nos termos da teoria tridimensional do Direito, pode-se, analiticamente, esclarecer que a estrutura de um modelo jurídico pressupõe:

a) dado campo de atos ou fatos da experiência social;

⁶⁴ REALE, Miguel. *Teoria Tridimensional do Direito*. 5. ed. São Paulo: Saraiva, 1994, p. 119.

⁶⁵ Para Reale, norma jurídica é “uma estrutura integrante de fatos segundo valores”. Reale, *Fontes e modelos do Direito: para um novo paradigma hermenêutico*, op. cit., p. 35.

⁶⁶ Cf. Reale, *Teoria Tridimensional do Direito*, op. cit., p. 127.

⁶⁷ Cf. Reale, *Fontes e modelos do Direito: para um novo paradigma hermenêutico*, op. cit., p. 29-32.

⁶⁸ Cf. *ibid.*, p. 37.

⁶⁹ *Ibid.*, p. 40.

⁷⁰ *Ibid.*, p. 49-50.

- b) uma ordenação normativa racionalmente garantida;
- c) o propósito de realizar valores ou impedir desvalores, de conformidade com a natureza de cada porção de realidade objeto da investigação científica.⁷¹

Superado o conceito de modelo jurídico, Reale os classifica em quatro diferentes espécies, em função de diferentes fontes possíveis: *legais*, que resultam de leis e decretos; *costumeiro*, originados dos usos e costumes; *jurisdicionais*, oriundos de decisões jurisdicionais; e *negociais*, fruto de acordo das vontades de partes individuais.⁷²

Para esta pesquisa, é importante detalhar, brevemente, os *modelos jurídicos jurisdicionais*. Derivados do poder decisório, normalmente surgem da adequação das normas legais aos casos concretos e têm sua importância no suprimento de lacunas legais ou determinação hermenêutica dos modelos legais em vigor. Em síntese, os *modelos jurídicos jurisdicionais dão caráter mais operacional aos modelos legais*, desde que sejam sempre revistos, em razão de alterações na correlação entre fatos e valores.⁷³

Faz-se relevante uma pequena digressão para mencionar o *conceito de ordenamento jurídico e suas características*.

Norberto Bobbio, em sua obra “Teoria do ordenamento jurídico”, inicia com uma definição genérica de ordenamento jurídico: “o ordenamento jurídico (como todo sistema normativo) é um conjunto de normas”⁷⁴. Ainda, destaca que a *unidade*, *coerência* (sistema) e *completude* são aspectos de um ordenamento jurídico,⁷⁵ sobre os quais discorrer-se-á a seguir.

Para o autor italiano, a unidade do ordenamento jurídico pressupõe uma norma fundamental com a qual se possa relacionar todas as normas.⁷⁶ Quanto à segunda característica, o ordenamento jurídico constitui um sistema pois não admite a coexistência de normas incompatíveis (antinomias), exigindo uma solução quando isso acontece.⁷⁷ Nessa lógica, afirma que duas normas incompatíveis e válidas não podem ser, para o mesmo caso concreto, eficazes, sob pena de comprometimento da exigência de coerência do ordenamento jurídico: “A coerência não é condição de validade, mas é sempre condição de justiça do ordenamento”⁷⁸.

⁷¹ Ibid., p. 48.

⁷² Cf. ibid., p. 63-75.

⁷³ Cf. ibid., p. 69-72.

⁷⁴ BOBBIO, Norberto. *Teoria do ordenamento jurídico*. Trad. Maria Celeste Cordeiro Leite Santos. 10. ed. Brasília: Editora UnB, 1999, p. 31.

⁷⁵ Cf. ibid., p. 34-35.

⁷⁶ Cf. ibid., p. 71.

⁷⁷ Cf. ibid., p. 80.

⁷⁸ Ibid., p. 113.

Por fim, como completude, entende-se como a necessidade de, independentemente do caso concreto, o ordenamento possuir uma norma para regulá-lo, ou seja, não haver lacunas.⁷⁹

Miguel Reale faz alusão a um horizonte normativo em comum – característica de unidade do ordenamento para Bobbio – quando define o ordenamento jurídico como um “ecossistema complexo e variegado que abrange uma multiplicidade de sistemas e subsistemas normativos que se escalonam uns distintos dos outros [...], todos eles [...] no âmbito do comum horizonte de validade da Constituição de cada povo”⁸⁰. Na sequência, o autor classifica o ordenamento jurídico como um macromodelo dinâmico, complexo e assimétrico, mas com sentido global de ordem e equilíbrio dinâmico frente aos valores sociais, sendo composto por princípios, códigos, leis, regulamentos, sentenças e contratos.⁸¹

Nesse diapasão, a *Teoria dos Modelos realiana vem ao encontro da exigência de coerência e necessidade de completude do ordenamento jurídico*, na medida em que *promove soluções para eficácia quando da aplicação das diferentes regras jurídicas que compõem o ordenamento e que se relacionam com o caso concreto*.

Voltando aos modelos jurídicos jurisdicionais, Reale cita como exemplo as *súmulas*, “ordenadoras e sistematizadoras das decisões de nossos Tribunais”⁸². A adoção de súmulas vinculantes pelo Supremo Tribunal Federal (STF), a partir da Emenda Constitucional (EC) nº 45, de 30 de dezembro de 2004, ficou expressamente reconhecida no ordenamento jurídico brasileiro, nos art. 102, §2º e art. 103-A da Constituição Federal (CF), de 5 de outubro de 1988.⁸³

⁷⁹ Cf. *ibid.*, p. 113.

⁸⁰ Reale, *Fontes e modelos do Direito: para um novo paradigma hermenêutico*, op. cit., p. 95.

⁸¹ Cf. *ibid.*, p. 95-97;102.

⁸² *Ibid.*, p. 72.

⁸³ “Art. 102. Compete ao Supremo Tribunal Federal, precipuamente, a guarda da Constituição, cabendo-lhe: [...]

§ 2º As decisões definitivas de mérito, proferidas pelo Supremo Tribunal Federal, nas ações diretas de inconstitucionalidade e nas ações declaratórias de constitucionalidade produzirão eficácia contra todos e efeito vinculante, relativamente aos demais órgãos do Poder Judiciário e à administração pública direta e indireta, nas esferas federal, estadual e municipal.

[...]

Art. 103-A. O Supremo Tribunal Federal poderá, de ofício ou por provocação, mediante decisão de dois terços dos seus membros, após reiteradas decisões sobre matéria constitucional, aprovar súmula que, a partir de sua publicação na imprensa oficial, terá efeito vinculante em relação aos demais órgãos do Poder Judiciário e à administração pública direta e indireta, nas esferas federal, estadual e municipal, bem como proceder à sua revisão ou cancelamento, na forma estabelecida em lei.

§ 1º A súmula terá por objetivo a validade, a interpretação e a eficácia de normas determinadas, acerca das quais haja controvérsia atual entre órgãos judiciários ou entre esses e a administração pública que acarrete grave insegurança jurídica e relevante multiplicação de processos sobre questão idêntica.

§ 2º Sem prejuízo do que vier a ser estabelecido em lei, a aprovação, revisão ou cancelamento de súmula poderá ser provocada por aqueles que podem propor a ação direta de inconstitucionalidade.

§ 3º Do ato administrativo ou decisão judicial que contrariar a súmula aplicável ou que indevidamente a aplicar, caberá reclamação ao Supremo Tribunal Federal que, julgando-a procedente, anulará o ato administrativo ou cassará a decisão judicial reclamada, e determinará que outra seja proferida com ou sem a aplicação da súmula, conforme o caso.” Brasil, *Constituição Federal de 5 de outubro de 1988*, op. cit.

Mais recentemente, as *inovações* trazidas pelo *novo Código de Processo Civil – CPC* (Lei nº 13.105, de 16 de março de 2015), vieram no sentido de consagrar a força de decisões anteriores sobre novas decisões e, conseqüentemente, dão concretude à teoria dos modelos jurisdicionais ao ordenamento jurídico brasileiro. Destacam-se.

O *art. 489* delimita os elementos essenciais de uma decisão, exigindo, em linhas gerais, fundamentação sobre como o caso se relaciona com o elemento normativo, inclusive quando se trata de conceitos jurídicos indeterminados.⁸⁴

O *art. 926* obriga que os tribunais mantenham uma jurisprudência uniformizada, estável, íntegra e coerente.⁸⁵ Sobre essas qualidades, Lênio Streck diz: “a coerência assegura a igualdade, isto é, que os diversos casos terão a igual consideração por parte do Poder Judiciário”⁸⁶. Ainda, para o jurista brasileiro, “a integridade exige que os juízes construam seus argumentos de forma integrada ao conjunto do direito, constituindo uma garantia contra arbitrariedades interpretativas”⁸⁷.

Já o *art. 927*, consagra a vinculação horizontal (do tribunal às suas próprias decisões) e vertical (dos tribunais inferiores às decisões das cortes hierarquicamente superiores):⁸⁸

Art. 927. Os juízes e os tribunais observarão:

I - as decisões do Supremo Tribunal Federal em controle concentrado de constitucionalidade;

II - os enunciados de súmula vinculante;

III - os acórdãos em incidente de assunção de competência ou de resolução de demandas repetitivas e em julgamento de recursos extraordinário e especial repetitivos;

IV - os enunciados das súmulas do Supremo Tribunal Federal em matéria constitucional e do Superior Tribunal de Justiça em matéria infraconstitucional;

V - a orientação do plenário ou do órgão especial aos quais estiverem vinculados.

[...]

§ 4º A modificação de enunciado de súmula, de jurisprudência pacificada ou de tese adotada em julgamento de casos repetitivos observará a necessidade de fundamentação adequada e específica, considerando os princípios da segurança jurídica, da proteção da confiança e da isonomia.

⁸⁴ Lei 13.105/2015, art. 489: “§ 1º Não se considera fundamentada qualquer decisão judicial, seja ela interlocutória, sentença ou acórdão, que:

[...]

II - empregar conceitos jurídicos indeterminados, sem explicar o motivo concreto de sua incidência no caso”. Brasil, *Lei nº 13.105, de 16 de março de 2015*, op. cit.

⁸⁵ Lei 13.105/2015: “Art. 926. Os tribunais devem uniformizar sua jurisprudência e mantê-la estável, íntegra e coerente”. Ibid.

Ressalta-se que, considerando os conceitos de *precedente* e *jurisprudência* adotados nesta pesquisa, é possível dizer que o CPC/2015, por vezes, mistura esses dois conceitos.

⁸⁶ STRECK, Lênio. Novo CPC terá mecanismos para combater decisionismos e arbitrariedades? Disponível em: <https://www.conjur.com.br/2014-dez-18/senso-incomum-cpc-mecanismoscombater-decisionismos-arbitrariedades>. Acesso em 10 ago. 2019.

⁸⁷ Ibid.

⁸⁸ Cf. Oliveira, *Precedentes no Direito Administrativo*, op. cit., p. 43/75.

§ 5º Os tribunais darão publicidade a seus precedentes, organizando-os por questão jurídica decidida e divulgando-os, preferencialmente, na rede mundial de computadores (Lei nº 13.105, de 16 de março de 2015).⁸⁹

Desta feita, não se pode negar o *fator impulsionador da consolidação jurisprudencial da Lei 13.105/2015*, que, inclusive, é perceptível no *Projeto Corpus 927*⁹⁰, desenvolvido pela Escola Nacional de Formação e Aperfeiçoamento de Magistrados em parceria com o Superior Tribunal de Justiça (STJ). Esse projeto tem como objetivo consolidar em um só local as decisões vinculantes do STF e do STJ, bem como a jurisprudência deste último.

Nesse momento, importante trazer à baila, as distinções conceituais feitas por Rafael Carvalho Rezende Oliveira – e adotadas nesta pesquisa – entre “decisão”, “precedente”, “jurisprudência” e “súmula”. Uma *decisão* é um comando único que visa solucionar um determinado conflito. *Precedente* é uma decisão que enuncie uma norma que deva ser respeitada em um caso semelhante futuro, não configurando como precedente a decisão que meramente aplica essa norma já emanada anteriormente. *Jurisprudência*, em seu sentido amplo,⁹¹ designa o conjunto de decisões, normalmente de apenas um tribunal, divergentes ou uniformes, sobre uma determinada temática. Por fim, a *súmula* representa a expressão da tese jurídica consolidada de um tribunal sobre um tema que foi objeto de várias decisões anteriores em casos semelhantes e no mesmo sentido.⁹²

Considerando esses conceitos e os exemplos aludidos de modelos jurídicos jurisdicionais, é possível afirmar que a tradição jurídica brasileira, da *família de civil law*, vem incorporando técnicas que caminham em direção à institucionalização dos precedentes e seus efeitos vinculantes (mormente para decisões da corte constitucional) em similaridade ao que ocorre em países da *common law*.⁹³ Antes de adentrar nesse mérito, tecer-se-á considerações sobre essas duas tradições jurídicas importantes, também chamadas de romano-germânica e a anglo-saxônica, respectivamente.

⁸⁹ Brasil, *Lei nº 13.105, de 16 de março de 2015*, op. cit.

⁹⁰ Disponível em: <http://corpus927.enfam.jus.br/>. Acesso em 5 nov. 2019.

⁹¹ Em sentido restrito e a depender do contexto, jurisprudência vem sendo utilizada para se referir à linha de entendimento mais recorrente sobre um tema. Nesta pesquisa, quando se mencionado o termo *jurisprudência*, referir-se-á a seu sentido amplo. Cf. Oliveira, *Precedentes no Direito Administrativo*, op. cit., p. 45.

⁹² Cf. *ibid.*, p. 43-45.

⁹³ Cf. Bustamante, *Teoria do Precedente Judicial*, op. cit., p. 276; Cf. Luvizotto, *Precedentes administrativos & a vinculação da atividade administrativa*, op. cit., p. 59. Na mesma linha, Cf. Oliveira, *Precedentes no Direito Administrativo*, op. cit., p. 35-36.

Em um estudo de Direito Comparado⁹⁴, é possível dizer que a tradição jurídica da *civil law*, na qual a brasileira se enquadra, tem por característica, além das influências romana e germânica na sua concepção, a supremacia de normas legislativas como fonte do direito. Em contrapartida, as famílias jurídicas de *commom law*, de origem inglesa, possuem na jurisprudência, por meio da força vinculativa dos precedentes judiciais, a condição de fonte prioritária, sendo que o Direito evolui por pequenos incrementos gerados pelas decisões judiciais e não mediante processos legislativos.⁹⁵

Destaca-se o *diferente papel ocupado pela jurisprudência nos dois sistemas jurídicos*. Nas famílias jurídicas de *civil law*, a jurisprudência institui regras sobretudo frágeis, mas que garantem graus de uniformidade e segurança na aplicação da lei.⁹⁶ Na *commom law*, por sua vez, a jurisprudência possui papel principal e a regra do precedente vinculante (*stare decisis*), aliada à técnica das distinções⁹⁷, impõe aos juízes a observância das regras criadas pelos seus predecessores.⁹⁸ Independente das vias utilizadas para se chegar às regras do Direito (lógica na *civil law* e empirismo na *commom law*), os dois sistemas possuem a mesma concepção de *Direito como razão*.⁹⁹ É nesse contexto que se materializa a aproximação entre os dois sistemas jurídicos e que foi detalhada por Thomas da Rosa de Bustamante.

Por uma perspectiva pós-positivista, Bustamante demonstra que *o processo mental realizado pelo juiz ao criar regras individuais com base em regras gerais é qualitativamente o mesmo – o racional – tanto para sistemas jurídicos de civil law quanto de commom law*. A diferença está na origem na norma geral que vincula a tomada de decisão, sendo normas produzidas pelo legislador no primeiro caso e de origem costumeira e descentralizada para o segundo sistema. De toda sorte, o autor assevera que o juiz é metodologicamente livre para decidir, desde que dentro dos limites do direito preexistente.¹⁰⁰ Quando se analisa as decisões pelo prisma da justificação, fica ainda mais claro o fato de que, do ponto de vista do raciocínio empregado, não há diferenças consideráveis entre o *commom law* e o *civil law*.¹⁰¹

⁹⁴ Segundo Dário Moura Vicente, “Chama-se Direito Comparado o ramo da Ciência Jurídica que tem por objeto o Direito na sua pluralidade e diversidade de expressões culturais e procede ao estudo comparativo destas. Melhor se diria, pois, a fim de designá-lo, *comparação de Direitos*”. VICENTE, Dário Moura. *Direito comparado: introdução, sistemas jurídicos em geral, volume 1*. 4. ed. São Paulo: Almedina, 2018, p. 18.

⁹⁵ Cf. *ibid.*, p. 69-71.

⁹⁶ Cf. DAVID, René David. *Os grandes sistemas do direito contemporâneo*. Trad. Hermínio A. Carvalho. 5. ed. São Paulo: Martins Fontes, 2014, p. 145-162.

⁹⁷ René David apresenta a técnica das distinções como sendo fundamental nos sistemas jurídicos de *commom law*. Análoga à interpretação da lei na *civil law*, a técnica das distinções permite a evolução da estrutura jurisprudencial nas famílias jurídicas anglo-saxônicas. Cf. *ibid.*, p. 443-444.

⁹⁸ Cf. *ibid.*, p. 427-436.

⁹⁹ Cf. *ibid.*, p. 445-446.

¹⁰⁰ Cf. Bustamante, *Teoria do Precedente Judicial*, op. cit., p. 102-103.

¹⁰¹ Cf. *ibid.*, p. 110.

Nesse cenário, a partir de teorias da argumentação jurídica que se traduzem em justificação racional das decisões,¹⁰² Bustamante constrói uma *teoria argumentativa dos precedentes judiciais*, de caráter normativo e universal, aplicável independentemente do sistema jurídico vigente.¹⁰³ Para a Teoria dos Precedentes de Bustamante, o fundamento de respeitar o precedente judicial advém do *princípio da universalizabilidade* e, consequentemente, de fatores essencialmente racionais.¹⁰⁴ Nesse sentido, sua teoria procura institucionalizar “parâmetros, regras e procedimentos de argumentação que se destinam a tornar o mais racional possível a prática de se seguir precedentes judiciais e utilizá-los como argumentos de justificação de decisões concretas”¹⁰⁵.

A propósito, a obrigação de motivar as decisões jurídicas é recente nas tradições romano-germânicas, tendo iniciado a partir do século XVI e solidificado apenas no século XX. Essa exigência surgiu como uma garantia de decisões maduras e conformes ao direito, ou seja, contra decisões arbitrárias.¹⁰⁶

Thomas da Rosa de Bustamante, no desenvolvimento de sua teoria, aborda diversos aspectos relacionados à técnica do precedente, mormente quanto à identificação das *rationes decidendi* a partir do *entendimento da decisão judicial encadeada por silogismos que, nada mais são que normas jurisprudenciais de diferentes graus de vinculatividade*¹⁰⁷ e que, por isso, caracterizam-se como precedentes.¹⁰⁸

Nessa esteira, o autor conceitua precedentes judiciais como se fossem “enunciados legislativos, textos dotados de autoridade que carecem de interpretação. É trabalho do aplicador

¹⁰² Cf. *ibid.*, p. 254.

¹⁰³ Cf. *ibid.*, p. 93.

¹⁰⁴ “[...] o fundamento do dever de respeitar o precedente não está apenas na autoridade, na força jurídica atribuída ao ato de criação judicial do Direito. Pelo contrário, parece claro que a exigência de universalizabilidade – núcleo da concepção kantiana de racionalidade prática – é o principal fundamento para a técnica do precedente judicial. Como a universalizabilidade é o critério mais seguro de correção de uma norma, o dever de levar em conta o precedente judicial pode ser fundamentado pelo método pragmático-universal de Habermas e Alexy: é um pressuposto para o entendimento do Direito como uma empresa racional”. Bustamante, *Teoria do Precedente Judicial*, op. cit., p. 268.

Na mesma linha, mas aplicado aos precedentes administrativos: “[...] a partir de um modelo de princípios que exige daqueles que pretendem afastar o precedente, o ônus argumentativo necessário à sua justificação, Robert Alexy (2001, p. 259) lembra que uma das ‘razões básicas para seguir os precedentes é o princípio de universalizabilidade, a exigência de que tratemos casos iguais de modo semelhante, o que está por trás da justiça como qualidade formal”. SCHIEFLER, E. A. C.; CRISTOVAM, J. S. S.; PEIXOTO, F. H. A inteligência artificial aplicada à criação de uma central de jurisprudência administrativa: o uso das novas tecnologias no âmbito da gestão de informações sobre precedentes em matéria administrativa. *Revista do Direito*. v. 3, n. 50, 2020, p. 21.

¹⁰⁵ Bustamante, *Teoria do Precedente Judicial*, op. cit., p. 21.

¹⁰⁶ Cf. David, *Os grandes sistemas do direito contemporâneo*, op. cit., p. 157.

¹⁰⁷ Cf. Bustamante, *Teoria do Precedente Judicial*, op. cit., p. 282-284.

¹⁰⁸ Cf. *ibid.*, p. 258.

do Direito extrair a *ratio decidendi* – o elemento vinculante – do caso a ser utilizado como paradigma”¹⁰⁹.

A *ratio decidendi*, para os países que adotam o sistema de *commom law*, pode ser atribuída como “aquilo que constitui o suporte necessário da decisão”¹¹⁰, diferenciando do que se chama de *obter dictum*, ou seja, “aquilo que o juiz pode ter declarado sem necessidade absoluta”¹¹¹. Logo, a *ratio decidendi* “constitui uma regra jurisprudencial que se incorpora [...] e que deve, a este título, ser seguida no futuro”¹¹².

Ademais, Bustamante assevera que, para a maioria dos casos, poder-se-á encontrar mais de uma *ratio decidendi*:

Se definirmos a *ratio decidendi* como uma norma universalizável que pode ser extraída de um precedente judicial, percebemos que na maioria dos casos se pode encontrar não apenas uma, mas diferentes *rationes decidendi*, que podem ter graus diferentes de vinculatividade em casos futuros. Ao invés de tentar encontrar uma e única *ratio decidendi* em cada caso, devemos admitir que há sempre um grau de indeterminação no Direito jurisprudencial (*case law*) e que nós podemos extrair diferentes *rationes decidendi* de um mesmo caso. Esse é o ponto de partida para uma teoria dos precedentes dotada de maior fecundidade.

Nessa nova teoria dos precedentes devemos abandonar não apenas a tese de que há apenas uma *ratio decidendi*, mas também a afirmação de que a *ratio* é sempre absolutamente vinculante.¹¹³

Nessa lógica, Bustamante estuda o conceito de *ratio decidendi* a partir de várias visões doutrinárias.¹¹⁴ Um conceito importante para o autor, é o de *normas adscritas*: à luz de Robert Alexy, “são normas criadas no processo de concretização do direito”¹¹⁵, ou seja, “uma norma concreta que decide um caso particular”¹¹⁶:

É nas normas adscritas, encontradas na fundamentação que o juiz dá à sua decisão, que se deve buscar as regras que podem servir como paradigmas para resolver casos futuros. Cada uma das normas adscritas é uma *ratio decidendi* do caso paradigma. Portanto, haverá uma *ratio decidendi* útil para a solução de casos futuros não apenas quando a corte decida determinada questão pontual acerca das consequências do caso particular – tal como “x deve fazer A” –, mas também quando essa mesma corte tenha estabelecido – com clareza e de forma justificada – uma regra geral que possa abarcar, além de x, os indivíduos y, z e outros que se achem na mesma situação. Pode-se falar, portanto, em uma pluralidade de *rationes decidendi* em um mesmo caso concreto [...].¹¹⁷

¹⁰⁹ Ibid., p. 259.

¹¹⁰ David, *Os grandes sistemas do direito contemporâneo*, op. cit., p. 430.

¹¹¹ Ibid., p. 430.

¹¹² Ibid., p. 430.

¹¹³ Bustamante, *Teoria do Precedente Judicial*, op. cit., p. 257-258.

¹¹⁴ Cf. *ibid.*, p. 259-282.

¹¹⁵ Ibid., p. 270.

¹¹⁶ Ibid., p. 237.

¹¹⁷ Ibid., p. 271.

Nessa conjuntura, Bustamante justifica a reconstrução da relação lógica entre a decisão e premissas argumentativas utilizadas pelo juiz no caso paradigmático por meio de um modelo silogístico que decompõe os passos seguidos na justificação da decisão e explicita, assim, cada uma das normas adscritas utilizadas.¹¹⁸

Em outras palavras, *para o autor, o método mais racional de se determinar as rationes decidendi é por meio da reconstrução das normas adscritas, encontradas na construção silogística dos argumentos utilizados para a fundamentação de uma decisão.*¹¹⁹

Ora, é possível estabelecer uma *relação entre os modelos jurídicos jurisdicionais* de Miguel Reale e o conceito de *rationes decidendi* aduzido na Teoria dos Precedentes de Thomas da Rosa de Bustamante. Afinal, *o conjunto das normas adscritas que compõem as decisões judiciais na visão de Bustamante, constroem os modelos¹²⁰ jurisdicionais de Reale, na medida em que servem para a adequação das normas legais aos casos concretos.*

Aprofundando um pouco mais na Teoria dos Precedentes, salienta-se que a regra do *stare decisis* não significa um Direito estático, pois há mecanismos de correção e adaptação do sistema, criando novos precedentes.¹²¹ O contrabalanceamento é dado pelo uso da *técnica de distinção (distinguishing)*, em que o operador do direito demonstra diferenças entre as situações fáticas que exigem a adoção de solução jurídica diversa do precedente. Também, a *técnica da superação (overruling)* possibilita a consideração do precedente como ultrapassado para solucionar problemas semelhantes, em razão de transformações da sociedade.¹²²

Esses conceitos de *distinguishing* e *overruling*, aliados aos anteriormente delimitados de *decisão*, de *precedente* e de *jurisprudência*, permitem a representação gráfica do que para esta pesquisa optou-se por denominar de *linha jurisprudencial de entendimento*, conforme Figura 1, abaixo:

¹¹⁸ Cf. *ibid.*, p. 277-278.

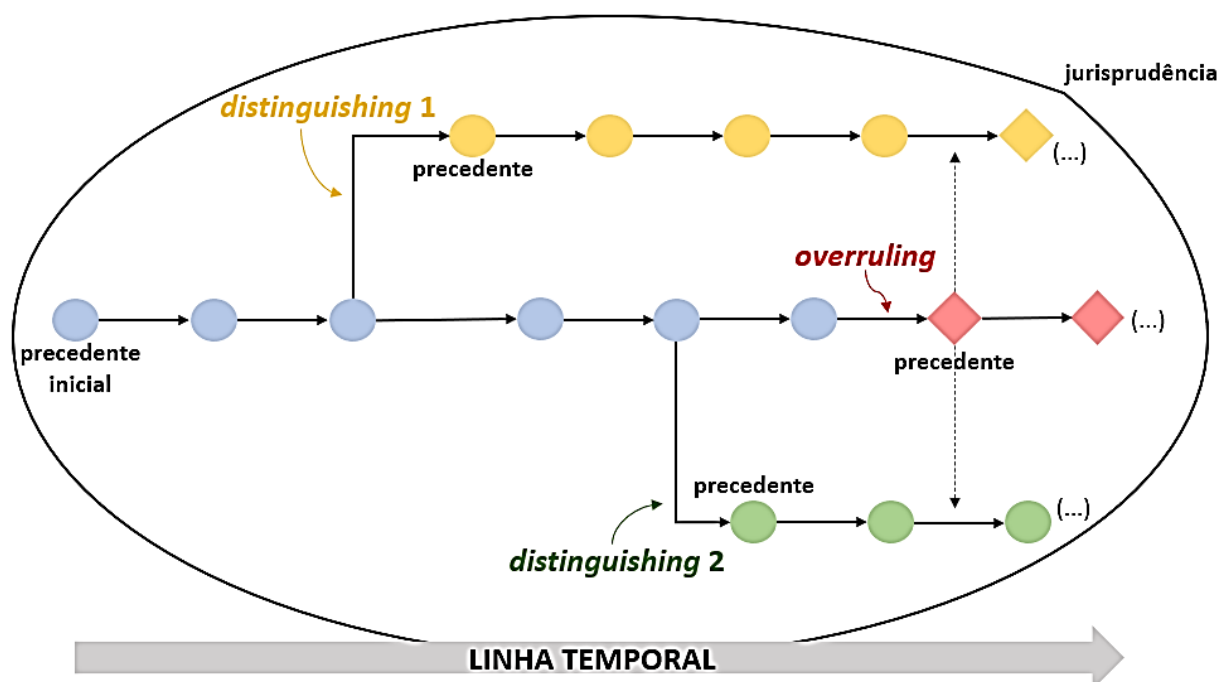
¹¹⁹ Cf. *ibid.*, p. 279.

¹²⁰ Importante lembrar que, para Miguel Reale, um “modelo jurídico resulta de uma pluralidade de normas entre si articuladas compondo um todo irreduzível às suas partes componentes”. Reale, *Fontes e modelos do Direito: para um novo paradigma hermenêutico*, op. cit., p. 30.

¹²¹ Cf. Oliveira, *Precedentes no Direito Administrativo*, op. cit., p. 62-63.

¹²² *Ibid.*, p. 51.

Figura 1 – Representação gráfica da *linha jurisprudencial de entendimento* sobre um tema hipotético ao longo do tempo. Cada bola ou losango representa uma *decisão* ocorrida em um determinado tempo.



Fonte: Elaborado pelo autor.

Quanto à origem, os precedentes podem ser classificados como *judiciais*, caso provenientes de decisões judiciais, ou *administrativos*.¹²³ Esses últimos podem ser conceituados como “a norma jurídica retirada da decisão administrativa anterior, válida e de acordo com o interesse público, que, após decidir determinado caso concreto, deve ser observada em casos futuros e semelhantes pela Administração Pública”.¹²⁴

Em países de jurisdição única (inafastabilidade de controle jurisdicional), como é o caso do Brasil, existem diferenças marcantes entre processos administrativos e judiciais. Entretanto, não é desarrozoado considerar a proximidade, em termos de fundamentação, importância de utilização, entre precedentes decorrentes de decisões judiciais e as de caráter administrativo, de modo que, com as devidas adaptações, há espaço para adoção da teoria dos precedentes administrativos.¹²⁵

¹²³ Ibid., p. 41.

¹²⁴ Ibid., p. 95.

¹²⁵ Cf. Oliveira, *Precedentes no Direito Administrativo*, op. cit., p. 97.

Ademais, ao discorrer sobre as características da tradição jurídica romano-germânica, René David, afirma: “Entre uma e outra não existe senão uma diferença sutil, quando são consideradas as decisões proferidas por certos organismos administrativos que podem não ser, tecnicamente falando, ‘jurisdições’.” Cf. David, *Os grandes sistemas do direito contemporâneo*, op. cit., p. 161-162.

Luvizotto, em seu estudo sobre os precedentes administrativos, conclui que, pelo aspecto da tomada de decisão – e, conseqüentemente, quanto ao uso dos precedentes – há muito mais semelhanças que diferenças entre as atividades administrativas e jurisdicionais:¹²⁶

Ao se reconhecer que é possível e legítimo o uso de precedentes judiciais como forma de garantir a igualdade, coerência e segurança jurídica, por meio das decisões proferidas pelos juízes, permitindo o controle de suas atividades, não se pode desconsiderar que as autoridades administrativas, ao também proferirem decisões no âmbito administrativo – especialmente no âmbito do contencioso administrativo –, estão sujeitas aos mesmos princípios informadores da atividade decisória adotada no âmbito judicial, especificamente no tocante à preocupação em conferir uniformidade às suas decisões, imparcialidade nos julgamentos, coerência e igualdade de tratamento. Por serem também intérpretes do ordenamento, predicam de instrumentos que deem concreção a tais objetivos.¹²⁷

Nessa lógica que Rafael Carvalho Rezende Oliveira apresenta os fundamentos de uma *Teoria dos Precedentes Administrativos*, à luz de princípios constitucionais¹²⁸ e pautados no conceito de autovinculação administrativa:

Os precedentes administrativos podem ser inseridos na teoria das autolimitações ou autovinculações administrativas. Isso porque em razão dos princípios da segurança jurídica, da razoabilidade e da isonomia, não há dúvida de que, no Estado Democrático de Direito, a atuação administrativa deve ser coerente e não contraditória.¹²⁹

A *autovinculação administrativa* é conceituada pelo autor como sendo a “vinculação decorrente dos próprios atos e condutas da Administração”.¹³⁰ Essa autovinculação, além de trazer coerência e previsibilidade aos atos administrativos, acarreta a eficiência às demandas repetitivas, bem como maior legitimidade e aceitação das decisões administrativas, com conseqüente redução da litigiosidade.¹³¹

¹²⁶ Cf. Luvizotto, *Precedentes administrativos & a vinculação da atividade administrativa*, op. cit., p. 68-76.

¹²⁷ Ibid., p. 71.

¹²⁸ Para o autor: (i) o princípio da *isonomia/igualdade* indica que situações fáticas e jurídicas similares devem ser tratados de forma isonômica; (ii) os princípios da *segurança jurídica*, *boa-fé* e *proteção da confiança legítima* exige a previsibilidade, lealdade e coerência da ação administrativa, garantindo as expectativas dos interessados; (iii) os princípios da *razoabilidade* e da *proporcionalidade* evitam a prática de arbitrariedades, uma vez que ordenam o respeito às decisões anteriores. Cf. Oliveira, *Precedentes no Direito Administrativo*, op. cit., p. 108.

¹²⁹ Ibid., p. 104

¹³⁰ Ibid., p. 102.

¹³¹ Cf. Ibid., p. 104-105.

Do ponto de vista normativo, antes mesmo da consolidação pelo CPC/2015,¹³² o uso de precedentes administrativos encontra amparo na Lei nº 9.784, de 29 de janeiro de 1999,¹³³ que trata do processo administrativo em âmbito federal.¹³⁴ No art. 50, inciso VII¹³⁵ da aludida lei estabelece-se que os atos administrativos devem ser motivados, com indicação dos fatos e dos fundamentos jurídicos, quando deixem de aplicar jurisprudência firmada sobre determinada questão.

Mais recentemente, as alterações sofridas pela Lei de Introdução às Normas do Direito Brasileiro – LINDB (Decreto-Lei nº 4.657, de 4 de setembro de 1942)¹³⁶ por meio da Lei nº 13.655, de 25 de abril de 2018,¹³⁷ *conferiram mais força aos precedentes administrativos*.¹³⁸ Especialmente, art. 24¹³⁹ da retromencionada lei veda que se declarem inválidas situações constituídas com base em mudança “jurisprudência judicial ou administrativa majoritária”. Por sua vez, o art. 30¹⁴⁰ da LINDB atribui aos agentes públicos um dever de atuar para fortalecer a segurança jurídica quando da aplicação das normas, por meio de regulamentos, súmulas

¹³² Lei 13.105/2015, art. 15, *caput*: “Na ausência de normas que regulem processos eleitorais, trabalhistas ou administrativos, as disposições deste Código lhes serão aplicadas supletiva e subsidiariamente.” Brasil, *Lei nº 13.105, de 16 de março de 2015*, op. cit.

¹³³ BRASIL. *Lei nº 9.784, de 29 de fevereiro de 1999 (Lei do Processo Administrativo Federal)*. Regula o Processo Administrativo no âmbito da Administração Pública Federal. Disponível em: http://www.planalto.gov.br/Ccivil_03/leis/L9784.htm. Acesso em: 23 out. 2020.

¹³⁴ Cf. Oliveira, *Precedentes no Direito Administrativo*, op. cit., p. 108-109.

¹³⁵ Lei 9784/1999: “Art. 50. Os atos administrativos deverão ser motivados, com indicação dos fatos e dos fundamentos jurídicos, quando:
[...]

VII - deixem de aplicar jurisprudência firmada sobre a questão ou discrepem de pareceres, laudos, propostas e relatórios oficiais;”. Brasil, *Lei nº 9.784, de 29 de fevereiro de 1999 (Lei do Processo Administrativo Federal)*, op. cit.

¹³⁶ BRASIL. *Decreto-Lei nº 4.657, de 4 de setembro de 1942*. Lei de Introdução às normas do Direito brasileiro. Disponível em: http://www.planalto.gov.br/ccivil_03/decreto-lei/del4657.htm. Acesso em: 23 out. 2020.

¹³⁷ BRASIL. *Lei nº 13.655, de 25 de abril de 2018*. Inclui no Decreto-Lei nº 4.657, de 4 de setembro de 1942 (Lei de Introdução às Normas do Direito Brasileiro), disposições sobre segurança jurídica e eficiência na criação e na aplicação do direito público. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13655.htm#:~:text=LEI%20N%C2%BA%2013.655%2C%20DE%2025%20DE%20ABRIL%20DE%202018.&text=Inclui%20no%20Decreto%2DLei%20n%C2%BA,na%20aplica%C3%A7%C3%A3o%20do%20direito%20p%C3%BAblico. Acesso em: 23 out. 2020.

¹³⁸ Cf. Schiefler et al., *A inteligência artificial aplicada à criação de uma central de jurisprudência administrativa: o uso das novas tecnologias no âmbito da gestão de informações sobre precedentes em matéria administrativa*, op. cit., p. 19.

¹³⁹ LINDB: “ Art. 24. A revisão, nas esferas administrativa, controladora ou judicial, quanto à validade de ato, contrato, ajuste, processo ou norma administrativa cuja produção já se houver completado levará em conta as orientações gerais da época, sendo vedado que, com base em mudança posterior de orientação geral, se declarem inválidas situações plenamente constituídas.

Parágrafo único. Consideram-se orientações gerais as interpretações e especificações contidas em atos públicos de caráter geral ou em jurisprudência judicial ou administrativa majoritária, e ainda as adotadas por prática administrativa reiterada e de amplo conhecimento público.” Brasil, *Decreto-Lei nº 4.657, de 4 de setembro de 1942*, op. cit.

¹⁴⁰ LINDB: “Art. 30. As autoridades públicas devem atuar para aumentar a segurança jurídica na aplicação das normas, inclusive por meio de regulamentos, súmulas administrativas e respostas a consultas. Parágrafo único. Os instrumentos previstos no caput deste artigo terão caráter vinculante em relação ao órgão ou entidade a que se destinam, até ulterior revisão.” Ibid.

administrativas e respostas a consultas, atribuindo a esses instrumentos força vinculativa em relação ao órgão ou entidade a que se destinam.

Em suma, *a solidificação dos precedentes em matérias administrativas contribui para a segurança jurídica, estabilidade do regime jurídico-administrativo (redução da litigiosidade) e previsibilidade das decisões da Administração Pública.*¹⁴¹ Ainda, considerando a presença cada vez mais recorrente de legislações de baixa densidade¹⁴² e utilização de conceitos jurídicos indeterminados, a cultura dos precedentes administrativos tem o condão de *contribuir para a interpretação jurídica coerente e delimitar a discricionariedade administrativa.*¹⁴³

De acordo com os fundamentos teóricos apresentados nesta seção, *a partir do significado lógico das rationes decidendi que congregam decisões correlatas, almeja-se delimitar a jurisprudência do TCU que trate sobre cláusulas restritivas ao caráter competitivo em editais de licitações.* Como a IA pode contribuir para isso é o assunto da seção 2.3.

Analogamente, *a partir das funções exercidas pelas expressões linguísticas no contexto do inteiro teor de cada decisão do TCU pertencente à jurisprudência delimitada, almeja-se caracterizar os modelos jurídicos jurisdicionais utilizados por meio da identificação os precedentes administrativos que dão concretude ao conceito aberto de cláusulas restritivas ao caráter competitivo em editais de licitação.*

2.3 Uso de Inteligência Artificial para identificação de precedentes administrativos

A *Inteligência Artificial* (IA) teve seu início na década de 1950 a partir do pressuposto de que qualquer aspecto da inteligência ou aprendizagem humana poderiam ser descritos com precisão de sorte que uma máquina conseguiria executá-los. Ao longo de sua trajetória, passou três momentos de entusiasmo (*booms*) e outros dois de ceticismo (*winters*) em relação as suas possibilidades de aplicação e de resultado.¹⁴⁴

¹⁴¹ Cf. Schiefler et al., *A inteligência artificial aplicada à criação de uma central de jurisprudência administrativa: o uso das novas tecnologias no âmbito da gestão de informações sobre precedentes em matéria administrativa*, op. cit., p. 21.

¹⁴² Como exemplo, tem-se a edição de leis marcadamente abertas (*leis-quadro*) para posterior regulação pela autoridade técnica competente. Cf. PALMA, Juliana Bonacorsi de. *Atividade normativa da Administração Pública: estudo do processo administrativo normativo*. Tese (Doutorado em Direito) – Faculdade de Direito da Universidade de São Paulo. São Paulo, 2014, p. 94.

¹⁴³ Cf. Oliveira, *Precedentes no Direito Administrativo*, op. cit., p. 99;112-116.

¹⁴⁴ Cf. GOANTA, Catalina; VAN DIJCK, Gijis; SPANAKIS, Gerasimos. Back to the Future: Waves of Legal Scholarship on Artificial Intelligence. In: RANCHORDÁS, Sofia; ROZNAI, Yaniv. *Time, Law and Change*. 1. ed. Londres: Hart Publishing, 2020, p. 2-3.

A empolgação inicial teve seu primeiro declínio devido a problemas como a falta de dados e capacidade computacional, o que resultou em desaceleração dos investimentos em pesquisas na área. Já no início da década de 1980, novas oportunidades de financiamento surgiram na medida em que apareceram abordagens voltadas para a aplicações em domínios específicos, como a criação dos sistemas especialistas. Entretanto, outra estagnação adveio quando os ganhos obtidos não refletiam mais as expectativas dos investidores.¹⁴⁵

O terceiro *boom*, que perdura até os dias atuais, iniciou-se na década de 1990 com o surgimento amplo conjunto de algoritmos¹⁴⁶ para extrair padrões significativos de dados existentes e aplicar esses padrões para tomar decisões ou previsões sobre dados futuros, chamados genericamente de aprendizado de máquina (*machine learning* – ML). Outra mudança de paradigma ocorreu a partir de 2010, com a ascensão das arquiteturas de *redes neurais artificiais*¹⁴⁷ com diversas camadas interagindo entre si, construindo a noção de aprendizado de máquina profundo (*deep learning* – DL¹⁴⁸).¹⁴⁹

Apesar de inicialmente limitado a disciplinas específicas como a ciência da computação, a IA recentemente tem ganhado popularidade considerável em outras disciplinas.¹⁵⁰

Uma das áreas que se beneficiou diretamente dos avanços da IA foi a de Processamento de Linguagem Natural (NLP, como a área é conhecida em decorrência da sigla em inglês *Natural Language Processing*), que consiste em uma disciplina desenvolvida a partir de elementos de linguística com intuito de entender como a linguagem humana, seja escrita, sonora

¹⁴⁵ Cf. *ibid.*, p. 3.

¹⁴⁶ “Um algoritmo pode ser definido, de modo simplificado, como um conjunto de regras que define precisamente uma sequência de operações, para várias finalidades, tais como modelos de previsão, classificação, especializações.” Peixoto; Silva, *Inteligência Artificial e Direito*, op. cit., p. 71.

¹⁴⁷ As redes neurais artificiais, inspiradas pela neurociência, são conjuntos de neurônios artificiais (*perceptrons*): “As redes neurais se referem a uma família de modelos que são definidos por uma camada de entrada (uma representação vetorial dos dados de entrada), uma camada oculta que consiste em neurônios e sinapses e uma camada de saída com os valores previstos. Dentro da camada oculta, as sinapses transmitem sinais entre os neurônios, que contam com uma função de ativação para proteger os sinais de entrada. As sinapses aplicam pesos aos valores de entrada e a função de ativação determina se as entradas ponderadas são suficientemente altas para ativar o neurônio e passar os valores para a próxima camada da rede.” Tradução livre de Bengfort et al., *Applied Text Analysis with Python*, op. cit., p. 298.

¹⁴⁸ *Deep learning* é uma técnica que permite que máquinas ou computadores aprendam e façam previsões a partir de dados brutos, em contraste com o método tradicional de regras ou algoritmos codificados. É um ramo do aprendizado de máquina (*machine learning*) que, por sua vez, é um ramo da Inteligência Artificial. Cf. ARUMUGAM, Rajesh; SHANMUGAMANI, Rajalingappaa. *Hands-On Natural Language Processing with Python*. 1. ed. Birmingham: Packt Publishing, 2018, p. 52-53.

¹⁴⁹ Cf. Goanta et al., *Back to the Future Waves of Legal Scholarship on Artificial Intelligence*, p. 4-5.

¹⁵⁰ Cf. ROSCA, Constanta; COVRIG, Bogdan; GOANTA, Catalina; VAN DIJCK, Gijs; SPANAKIS, Gerasimos. Return of the AI: An Analysis of Legal Research on Artificial Intelligence Using Topic Modeling. In: *Natural Legal Language Processing Workshop (NLLP)*, II., 2020, San Diego. *Proceedings*, p. 3-10. Association for Computing Machinery (ACM), 2020, p. 3.

ou por imagem, é aprendida e como pode ser modelada pelo computador.¹⁵¹ Apesar de existir desde os anos 1950, esse campo de conhecimento teve seu primeiro crescimento com o estabelecimento da NLP estatística, que compreende as abordagens quantitativas para encontrar padrões comuns que ocorrem no uso da linguagem, como modelagens probabilísticas e álgebra linear que utilizam a frequência de palavras.¹⁵² Nas últimas décadas, o *boom* de *machine learning* e *deep learning* trouxeram o NLP para um novo patamar, proporcionando aplicações para análise de sentimentos de textos em redes sociais, tradução de textos, reconhecimento de imagens etc.¹⁵³

No que concerne ao Direito, por meio de pesquisa bibliométrica e utilizando-se de técnicas de IA, Rosca et al. buscaram respostas para entender a evolução da pesquisa científica combinada entre IA e o Direito.¹⁵⁴

Em um nível conceitual, o algoritmo utilizado pelos autores buscou descobrir tópicos capazes de representar a coleção de documentos utilizada para a pesquisa, qual seja, 3931 artigos de periódicos obtidos da *HeinOnline*, uma das bases de dados internacionais líderes em materiais jurídicos. Assim, conseguiu-se estratificar o histórico de tópicos e subtópicos publicados que envolveram essas duas áreas do conhecimento simultaneamente (Figura 2).¹⁵⁵

Como resultado, 18 assuntos principais foram identificados: *governança da internet; tecnologia militar; privacidade; ambiente/espço; patentes; direito autoral; psicologia e neurociência; de humanos para máquinas; tomada de decisão algorítmica e métodos quantitativos; vigilância; contrato e resolução de disputas; sistemas baseados em conhecimento; regulamentação financeira e tecnologia; regulação da inovação; recuperação de informação; veículos autônomos; licenciamento de software; e tendências em tecnologia legal*.¹⁵⁶

¹⁵¹ Cf. MINER, G. et al. *Practical text mining and statistical analysis for non-structured text data applications*. 1. ed. Waltham: Elsevier, 2012, p. 7-8.

¹⁵² Cf. Manning; Schutze, *Foundations of statistical natural language processing*, op. cit., p. 4.

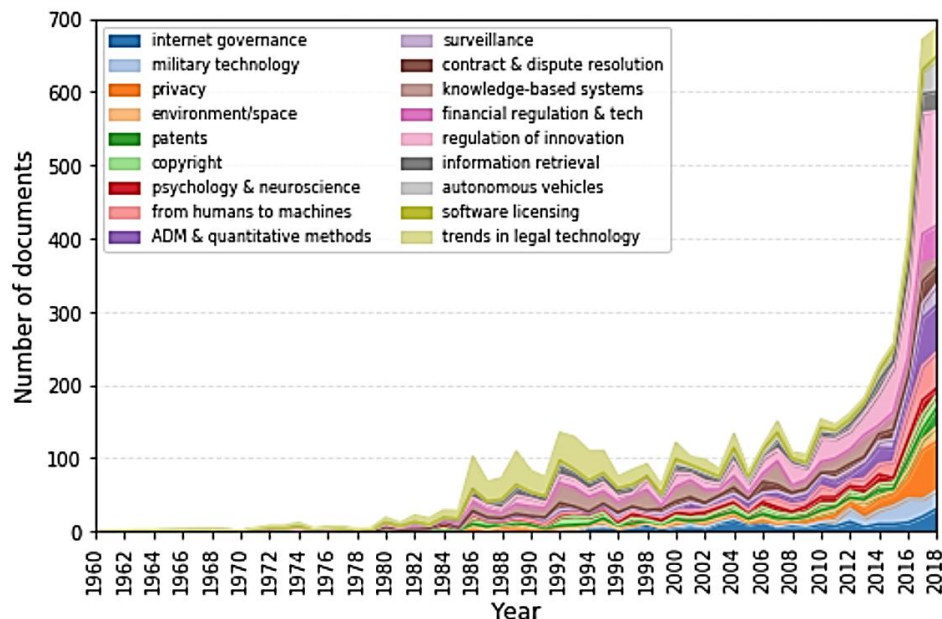
¹⁵³ Cf. Arumugam; Shanmugamani, *Hands-On Natural Language Processing with Python*, op. cit., p. 8-27.

¹⁵⁴ Cf. Rosca et al., *Return of the AI: An Analysis of Legal Research on Artificial Intelligence Using Topic Modeling*, op; cit., p. 3.

¹⁵⁵ Cf. *ibid.*, p. 4.

¹⁵⁶ Cf. *ibid.*, p. 6.

Figura 2 – Evolução anual da quantidade de artigos científicos publicados entre 1960 e 2018, por tópico encontrado na pesquisa de Rosca et al.



Fonte: Retirado de Rosca et al., *Return of the AI: An Analysis of Legal Research on Artificial Intelligence Using Topic Modeling*, op; cit., p. 7.

Dentre esses 18 assuntos, destacam-se três que guardam relação com o problema desta pesquisa: *recuperação de informação*¹⁵⁷, *sistemas baseados em conhecimento*¹⁵⁸ e *tomada de decisão algorítmica e métodos quantitativos*¹⁵⁹, e que estão entre os 10 tópicos mais encontrados pela pesquisa de Rosca et al., com grande aumento no número de documentos associados a partir no *terceiro boom* da IA, sobretudo após o advento do aprendizado profundo (2010, aproximadamente).

A respeito dessas temáticas, o objetivo de muitas pesquisas tem sido desenvolver modelos computacionais capazes de replicar o raciocínio e argumentação jurídicos humanos, com intuito de facilitar buscas jurisprudenciais, prever resultados de decisões, dentre outras atividades jurídicas. Embora os pesquisadores de AI e Direito tenham feito grandes avanços, um gargalo de representação do conhecimento permanece: na grande maioria dos casos, o conhecimento jurídico empregado pelos modelos atuais teve que ser extraído manualmente dos casos, estatutos, regulamentos, contratos e outros textos que os profissionais do Direito utilizam

¹⁵⁷ As palavras-chave que caracterizam o tópico *recuperação de informação* são: informação, pesquisa, documento, biblioteca, jurídico, pesquisa, banco de dados, caso, acesso, eletrônico. Cf. Ibid., p. 6.

¹⁵⁸ As palavras-chave que caracterizam o tópico *sistemas baseados em conhecimento* são: legal, regra, caso, sistema, razão, conhecimento, base, modelo, argumento, fato. Cf. Ibid., p. 6.

¹⁵⁹ As palavras-chave que caracterizam o tópico *tomada de decisão algorítmica e métodos quantitativos* são: algoritmo, modelo, decisão, análise, resultado, método, previsão, estudo, risco. Cf. Ibid., p. 6.

diariamente. Isto significa que especialistas tiveram que ler os textos legais e selecionar partes relevantes de seu conteúdo para servirem de informação de entrada para que os modelos computacionais pudessem as utilizar e obter os resultados desejados.¹⁶⁰

No capítulo 3 da obra “*Artificial Intelligence and Legal Analytics: new tools for law practice in the digital age*” Kevin Ashley apresenta três modelos de *raciocínio jurídico baseado em casos* que, cada um à sua maneira, são voltados para eleger decisões relevantes, compará-las em termos de semelhança e distingui-las em suas situações de fato. Uma vez que esses modelos devem ser capazes de decidir se tratam os casos de maneira análoga do ponto de vista jurídico, as semelhanças e diferenças entre os casos devem ser apresentadas de forma que o algoritmo possa processá-las e analisá-las.¹⁶¹

Os três modelos relatados, todavia, apresentam limitações em comum. Primeiro, requerem a modelagem manual de argumentos a favor e contra a aplicação da regra a uma situação de fato, o que perpassa, necessariamente, pela descrição dos próprios fatos. Ainda, as interpretações são invariavelmente retiradas dos textos por especialistas humanos e desempenham importante papel na extração de informações legalmente relevantes utilizadas pelos modelos. Assim, é possível afirmar que o desafio a ser superado é a extensão das técnicas de representação de conhecimento para conexão direta com os textos dos casos, de modo a identificar as semelhanças e as diferenças de maneira automatizada.¹⁶²

Nesse cenário, Ashley apresenta o conceito de *computação cognitiva* como esperança de solução. Para o autor, esse tipo de modelagem possibilitará um novo tipo de aplicativo legal, capaz de permitir a atividade colaborativa entre humanos e computadores, em que cada um realiza os tipos de atividades que melhor executam. *Em vez de utilizar-se da incorporação de regras construídas manualmente por especialistas no processo de aquisição de conhecimento, esse novo conceito de modelo computacional, em contraste, adquiriria o conhecimento a partir do próprio conjunto de textos.*¹⁶³

Os mais recentes avanços no caminho para a computação cognitiva apoiam-se em procedimentos de *text analytics* (ou *text mining*). As atividades de *text mining* foram iniciadas pela necessidade de catalogar documentos de texto, sobretudo para organização em bibliotecas, avançando gradativamente o foco para a extração informações em textos de diversas áreas usando técnicas de IA e NLP. Especificamente, tratam-se de um *conjunto de técnicas*

¹⁶⁰ Cf. ASHLEY, Kevin D. *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*. 1. ed. Cambridge: Cambridge University Press, 2017, p. 4-5.

¹⁶¹ Cf. *ibid.*, p. 73-77.

¹⁶² Cf. *ibid.*, p. 104-106.

¹⁶³ Cf. *ibid.*, p. 12-14.

linguísticas, estatísticas, aprendizado de máquina e aprendizado profundo que modelam e estruturam o conteúdo de informações de fontes textuais para inteligência de negócios, análise exploratória de dados, pesquisa ou investigação.¹⁶⁴ No caso de sua aplicação em textos jurídicos, deriva-se o termo *legal text analytics*.¹⁶⁵

Pensando nessas possibilidades, Schiefler et al. propõem o uso de inteligência artificial para criação de uma *central de jurisprudência administrativa*, que seria alimentada com decisões de todos os órgãos e entidades da Administração Pública brasileira.¹⁶⁶ Em síntese, a IA auxiliaria na sua construção por meio da *classificação* automática em *categorias pré-estabelecidas*¹⁶⁷ dos documentos remetidos a essa *central*, bem como na pesquisa de *jurisprudência*¹⁶⁸, pelos cidadãos e pela própria Administração Pública.¹⁶⁹

O projeto *Corpus 927*, citado anteriormente na seção 2.2, é um exemplo desse tipo de abordagem aplicada a decisões judiciais de um órgão específico, o STJ.¹⁷⁰

Outro trabalho recentemente publicado foi o exitoso Sistema *VICTOR*, que tem como uma de suas atividades classificar se um processo para lá remetido decorre de: *Acórdão*, para decisões de tribunais de primeira instância sob revisão; *Recurso Extraordinário* (RE), para pedidos de recurso; *Agravo de Recurso Extraordinário* (ARE), para impugnar o pedido de recurso; *Despacho*, para ordens judiciais; *Sentença*, para julgamentos; ou *Outros*, para documentos não incluídos nas categorias anteriores. Ademais, para os casos rotulados como RE e ARE, o Sistema promove a classificação de acordo com 28 temas de Repercussão Geral já declarados pelo STF, podendo ser classificado em uma ou mais dessas classes ou, ainda, definir que não se trata de nenhuma delas, atribuindo-lhe um rótulo genérico.¹⁷¹

¹⁶⁴ Cf. Miner et al., *Practical text mining and statistical analysis for non-structured text data applications*, op. cit., p. 3-17.

¹⁶⁵ Cf. Ashley, *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, op. cit., p. 5.

¹⁶⁶ Cf. Schiefler et al., *A inteligência artificial aplicada à criação de uma central de jurisprudência administrativa: o uso das novas tecnologias no âmbito da gestão de informações sobre precedentes em matéria administrativa*, op. cit., p. 23.

¹⁶⁷ “No caso da Central de Jurisprudência Administrativa, essas categorias predefinidas podem ser subáreas do direito público, a exemplo de: servidor público, concurso público, licitação pública, contratação pública, tributação, etc., assim como as numerosas ramificações dentro de cada uma dessas áreas.” Ibid., p. 29.

¹⁶⁸ Para os autores, o termo *jurisprudência* é empregado em “sentido amplo e para designar, como normalmente ocorre na cultura jurídica brasileira, o conjunto mais ou menos estabilizado de decisões de tribunais e órgãos julgadores em geral acerca de determinada questão”. Ibid., p. 20.

¹⁶⁹ Cf. *ibid.*, p. 31.

¹⁷⁰ Disponível em: <http://corpus927.enfam.jus.br/>. Acesso em 5 nov. 2019.

¹⁷¹ Cf. ARAUJO, P. H. L.; CAMPOS, T. E.; BRAZ, F. A.; SILVA, N. C. *VICTOR: a dataset for Brazilian legal documents classification*. In: *Conference on Language Resources and Evaluation (LREC)*, XII., 2020, Marseille. *Proceedings*. European Language Resources Association, 2020, p. 1450.

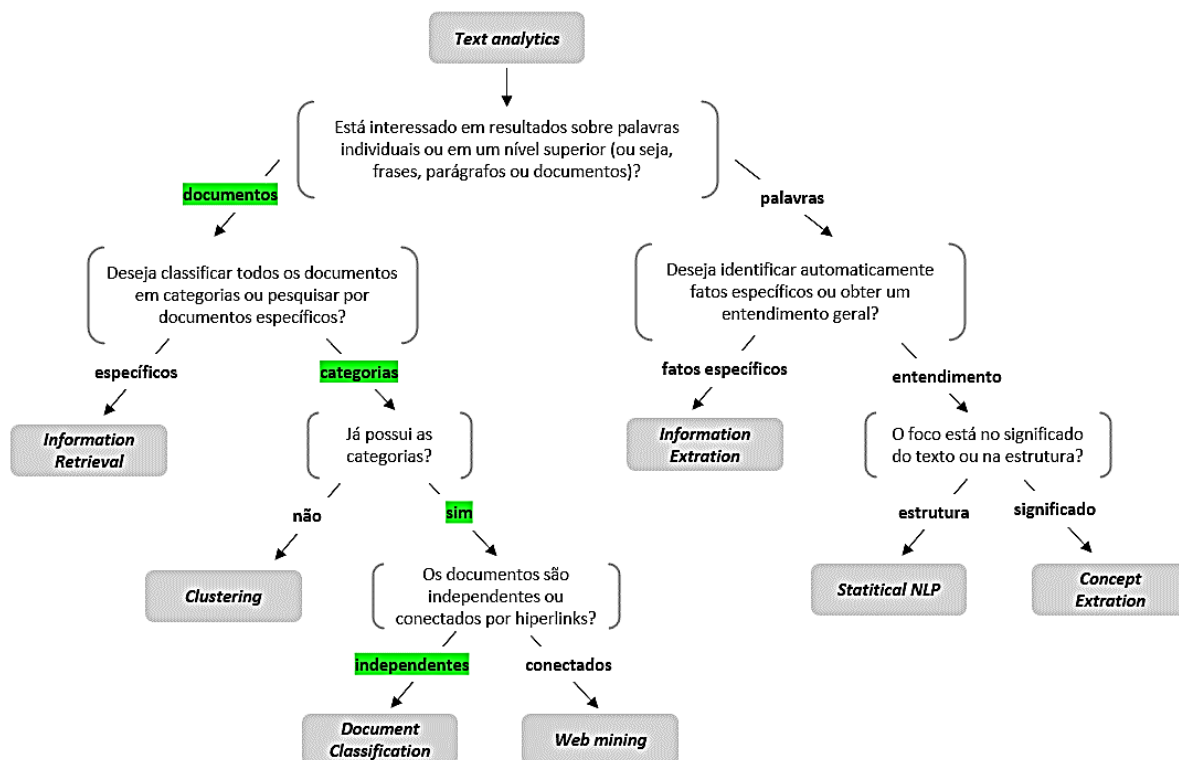
Todavia, para cada problema jurídico em que se busque uma solução de IA, faz-se necessário identificar qual conjunto de técnicas de *text mining* é mais adequado. Miner et al., dividem as aplicações de *text mining* em sete áreas práticas.¹⁷² São elas:

1. Recuperação de informações (*information retrieval*): armazenamento e recuperação de documentos de texto, incluindo mecanismos pesquisa por palavras-chave;
2. Agrupamento de documentos (*clustering*): agrupar e categorizar termos, fragmentos, parágrafos ou documentos, usando métodos de *clustering* de mineração de dados, sem a apresentação de rótulos prévios;
3. Classificação de documentos (*document classification*): agrupar e categorizar fragmentos, parágrafos ou documentos, usando métodos de classificação de mineração de dados, com base em modelos treinados em exemplos rotulados;
4. Mineração da web (*web mining / web crawling*): mineração de dados e textos na Internet, com foco específico em ganho de escala;
5. Extração de informação (*information extration*): identificação e extração de fatos e relações relevantes de textos não estruturados, transformando-os em dados estruturados ou semiestruturados;
6. Processamento de Linguagem Natural estatística (*statitital NLP*): processamento de linguagem com uso unicamente de aplicações probabilísticas para tarefas de compreensão. Por exemplo, *part of speech tagging (POS-tagging)*, que consiste na marcação gramatical de palavras.
7. Extração de conceitos (*concept extration*): agrupamento de palavras e frases em grupos semanticamente semelhantes.

Os autores também descrevem uma árvore decisória para ajudar no direcionamento para escolha da tarefa prática aplicável a cada necessidade, como mostra a Figura 3.

Figura 3 – Árvore de decisão para encontrar a área de prática de mineração de texto correta. Em destaque, as respostas escolhidas para o problema desta pesquisa.

¹⁷² Cf. Miner et al., *Practical text mining and statistical analysis for non-structured text data applications*, op. cit., p. 32.



Fonte: Traduzido e adaptado pelo autor, com base em: Miner et al., *Practical text mining and statistical analysis for non-structured text data applications*, op. cit., p. 33.

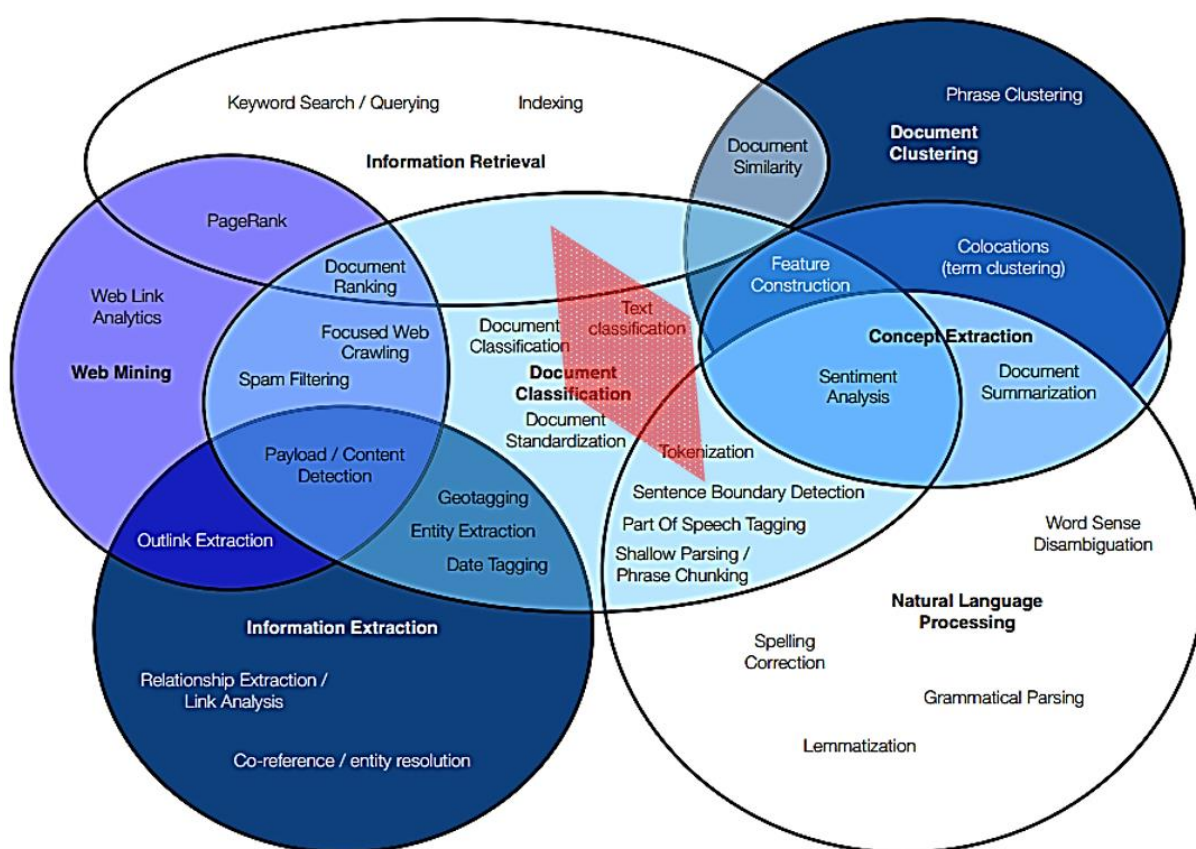
Buscou-se responder aos questionamentos direcionadores propostos por Miner et al. com base no primeiro problema prático desta pesquisa – delimitar via IA, a jurisprudência do TCU que verse sobre patologias corruptivas relacionadas a cláusulas restritivas ao caráter competitivo em editais de licitação. Uma vez que se deseja obter todas as decisões colegiadas do TCU que se relacionam com as patologias em estudo, teve-se a resposta para a primeira bifurcação. Em seguida, tendo em mente que partir-se-á de decisões do TCU pertencentes a classes de cláusulas restritivas ao caráter competitivo em editais de licitações previamente identificadas por pesquisas anteriores (como será detalhado no Capítulo 3), responde-se aos dois outros questionamentos. Para a pergunta derradeira, explica-se que cada decisão é única, apesar de poderem citar umas às outras ao longo do texto.

Por conseguinte, conclui-se que o primeiro problema prático da pesquisa, a ser solucionado por técnicas de IA e mais especificamente de *text mining*, constitui-se de um problema de *classificação de documentos* e os resultados obtidos subsidiarão a identificação dos precedentes administrativos das patologias corruptivas relacionadas à temática escolhida para estudo.

Todavia, Miner et al. asseveram que as retromencionadas sete áreas práticas se sobrepõem consideravelmente, uma vez que muitas tarefas computacionais situam-se em suas

intersecções. Assim, para se obter uma solução adequada, faz-se necessário aplicar técnicas de mais de uma área prática. Por isso, procurou-se situar a primeira problemática da pesquisa desenhando a hachura em formato de losango na cor vermelha no diagrama apresentado na Figura 4, em que Miner et al. apresentam uma visualização da sobreposição entre as sete áreas práticas.

Figura 4 – As sete áreas práticas de *text mining* (formas ovais) e como as tarefas específicas de *text mining* (rótulos dentro das formas ovais) existem em suas intersecções. Na região destacada pela hachura em formato de losango da cor vermelha, situa-se o primeiro problema prático desta pesquisa.



Fonte: Adaptado pelo autor, retirado de: Miner et al., *Practical text mining and statistical analysis for non-structured text data applications*, op. cit., p. 38.

Os procedimentos adotados para a pesquisa empírica deste trabalho serão descritos no Capítulo 4. Por ora, é possível concluir que o problema de delimitar precedentes administrativos com uso do computador perpassa por aplicações que envolvem domínio interdisciplinar de, pelo menos, Inteligência Artificial, Direito e Linguagem. Nessa acepção, à luz dos marcos teóricos linguísticos e jurídicos aliados à aplicação de técnicas de *text mining* para classificação de documentos jurídicos, almeja-se criar um método de extração jurisprudencial que se aproxime do conceito de computação cognitiva.

3 CLÁUSULA RESTRITIVA AO CARÁTER COMPETITIVO EM LICITAÇÕES: UMA PATOLOGIA CORRUPTIVA DE CONCEITO JURÍDICO INDETERMINADO

Este capítulo apresenta o recorte temático escolhido para aplicar a extração jurisprudencial com uso de Inteligência Artificial. Inicialmente, caracteriza-se a *fraude em licitações* como uma *patologia corruptiva*. Em seguida, o tipo de fraude *edital restritivo* é escolhido para ressaltar o papel dos *precedentes administrativos* do TCU para a *concreção do conceito jurídico indeterminado de cláusula restritiva ao caráter competitivo em licitações*. Por fim, é apresentado um panorama do uso de IA pelo TCU na busca por fraudes em licitações e como esta pesquisa pode contribuir para os algoritmos por ele desenvolvidos.

3.1 Fraude em licitações públicas: uma patologia corruptiva

A depender do estudioso sobre o assunto, o termo *corrupção* ganha diferentes conceituações, menos ou mais abrangentes, de acordo com o enfoque que se pretende.¹⁷³ Nesse universo, Luiz Fernando Miranda constrói um conceito de *corrupção* que buscou convergir diferentes abordagens de diversos autores, de modo a obter um razoável equacionamento entre abrangência e detalhamento.¹⁷⁴

O autor se embasa em dois pontos de partida: o primeiro, de que o ato de corromper envolve ao menos uma troca entre o corruptor e o corrompido; o segundo, de exclusão de qualquer percepção moral sobre o tema, a fim de que o fenômeno seja observado pelo prisma da legalidade/ilegalidade e consiga ter caráter prático e instrumental.¹⁷⁵

Desse modo, Miranda apresenta um esquema estrutural em que alicerça sua construção: “o sujeito A (com motivação pessoal ou política) que deseje corromper o sujeito B (com

¹⁷³ Cf. Miranda, *Unificando os conceitos de corrupção: uma abordagem através da nova metodologia dos conceitos*, op. cit., p. 238; 265.

Também sobre a pluralidade de conceitos para a corrupção, cf.:

Tribunal de Contas da União, *Referencial de combate à fraude e corrupção: aplicável a órgãos e entidades da Administração Pública*, op. cit., p. 13-14;

HEINEN, Juliano. *Comentários à Lei *Anticorrupção: Lei nº 12.846/2013*. 1. ed. Belo Horizonte: Fórum, 2015, p. 20;

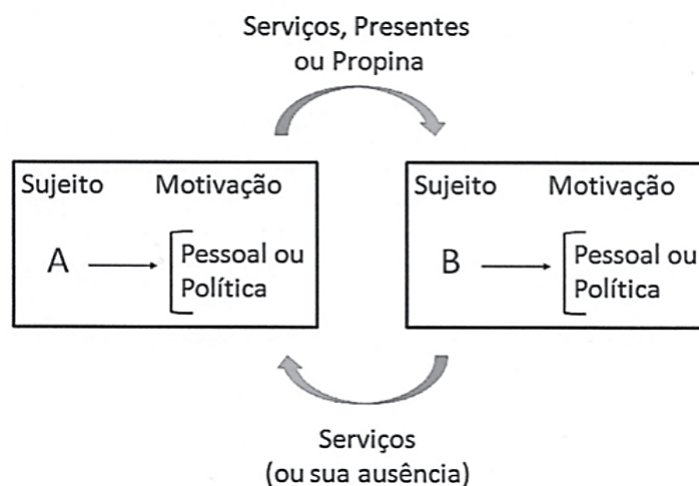
e FURTADO, Lucas Rocha. *As raízes da corrupção no Brasil: estudos de caso e lições para o Futuro*. 1. ed. Belo Horizonte: Editora Fórum, 2015, p. 26-30.

¹⁷⁴ Cf. Miranda, *Unificando os conceitos de corrupção: uma abordagem através da nova metodologia dos conceitos*, op. cit., p. 237; 266.

¹⁷⁵ Cf. *ibid.*, p. 260.

motivação pessoal ou política) oferece um serviço ou presente ou propina e obtém em contrapartida um serviço (ou sua ausência)”¹⁷⁶, conforme representado na Figura 5.

Figura 5 – Esquema estrutural da corrupção que fundamenta o conceito de *corrupção* de Luiz Fernando Miranda.



Fonte: Retirado de: Miranda, *Unificando os conceitos de corrupção: uma abordagem através da nova metodologia dos conceitos*, op. cit., p. 256.

A relação estrutural acima representada, extraída da análise de inúmeras conceituações sobre corrupção, conduziu o autor ao seguinte conceito, elaborado com base em metodologia¹⁷⁷ para esse fim:

Corrupção é o pagamento ilegal (financeiro ou não) para a obtenção, aceleração ou para que haja ausência de um serviço feito por um funcionário público ou privado. A motivação da corrupção pode ser pessoal ou política tanto para quem corrompe quanto para quem é corrompido.¹⁷⁸

O autor ainda classifica a corrupção em três tipos: (i) a *grande corrupção*, que normalmente movimenta grandes recursos financeiros e é praticada pela elite para direcionamento de políticas que beneficiam determinados grupos de interesse; (ii) a *pequena corrupção* ou *corrupção burocrática*, em que o foco é o pagamento de pequenas propinas a

¹⁷⁶ Ibid., p. 256.

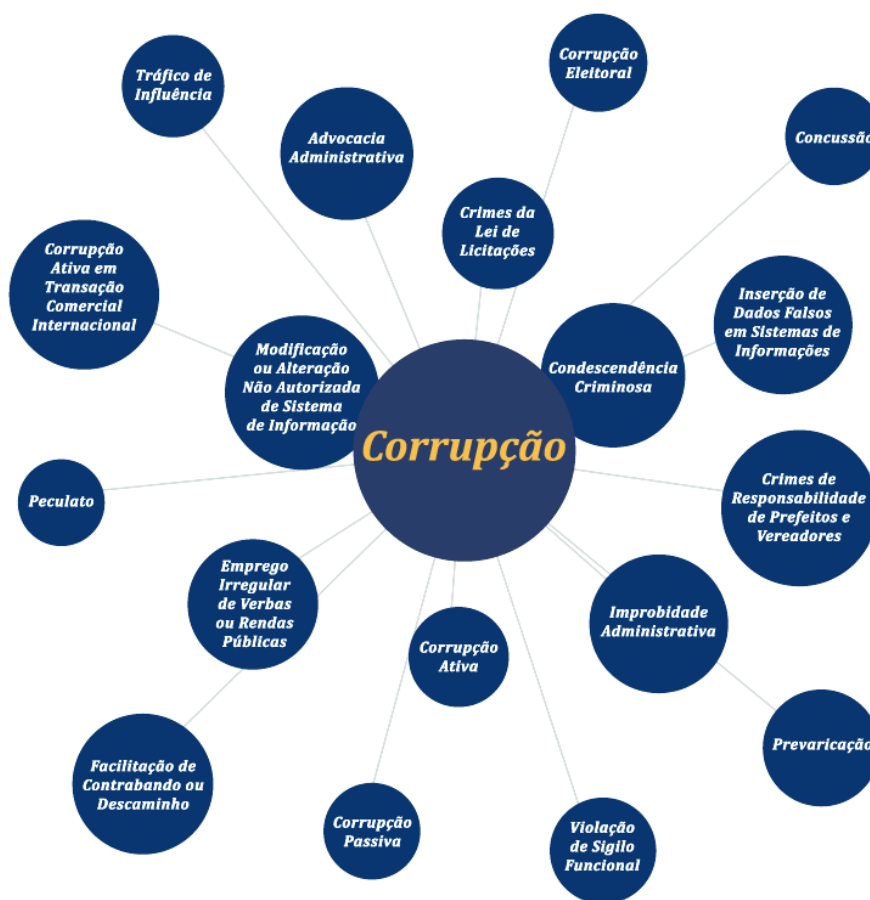
¹⁷⁷ O conceito de corrupção elaborado por Miranda leva em consideração que a qualidade dos conceitos deve ser entendida como uma tentativa de mediação entre oito critérios: *coerência, operacionalização, validade, utilidade de campo, ressonância, extensão contextual, parcimônia e utilidade analítica*. Ainda, considera a necessidade de *multidimensionalidade/multiníveis* na criação de conceitos. Cf. *ibid.*, p. 246-254.

¹⁷⁸ Ibid., p. 256.

quem detém poder decisório; e (iii) a *corrupção legislativa*, que permeia a influência dos votos dos legisladores.¹⁷⁹

O conceito criado por Luiz Fernando Miranda consegue abarcar muitas práticas corruptas, que nesta pesquisa denominam-se genericamente de *patologias corruptivas*. O Ministério Público Federal (MPF) apresenta um diagrama¹⁸⁰ (Figura 6) que busca relacionar as condutas que se caracterizam como tipos de corrupção no ordenamento jurídico nacional, incluindo tanto infrações penais, quanto civis e administrativas.

Figura 6 – Tipos de corrupção enquadrados no ordenamento jurídico brasileiro.



Fonte: Retirado de: <http://combateacorrupcao.mpf.mp.br/tipos-de-corrupcao>. Acesso em: 23 out. 2020.

Detalha-se o que o MPF classificou como “Crimes da Lei Geral de Licitações”: para a instituição, esses crimes estão especificados nos artigos 89 e 90 da Lei 8.666/1993.

¹⁷⁹ Cf. *ibid.*, p. 258-259.

¹⁸⁰ Disponível em: <http://combateacorrupcao.mpf.mp.br/tipos-de-corrupcao>. Acesso em: 23 out. 2020. Para cada tipo, o MPF possibilita a consulta dos dispositivos normativos associados e de pelo menos um exemplo de caso em que foram aplicados.

Art. 89. Dispensar ou inexigir licitação fora das hipóteses previstas em lei, ou deixar de observar as formalidades pertinentes à dispensa ou à inexigibilidade:

Pena - detenção, de 3 (três) a 5 (cinco) anos, e multa.

Parágrafo único. Na mesma pena incorre aquele que, tendo comprovadamente concorrido para a consumação da ilegalidade, beneficiou-se da dispensa ou inexigibilidade ilegal, para celebrar contrato com o Poder Público.

Art. 90. Frustrar ou **fraudar**, mediante ajuste, combinação ou qualquer outro expediente, o caráter competitivo do procedimento licitatório, com o intuito de obter, para si ou para outrem, vantagem decorrente da adjudicação do objeto da licitação:

Pena - detenção, de 2 (dois) a 4 (quatro) anos, e multa.¹⁸¹ (grifos nossos)

Considerando a delimitação de *fraude*, aduzida no supracitado artigo 90 da Lei Geral de Licitações, as *fraudes em licitações* são *espécies* do *gênero corrupção* e normalmente estão intrinsecamente relacionadas ao caráter competitivo das compras públicas.¹⁸²

Outra legislação que versa sobre *fraudes em licitações*, mas com foco em punições de pessoas jurídicas,¹⁸³ é a *Lei Anticorrupção* (Lei nº 12.846, de 1º de agosto de 2013)¹⁸⁴, que, no tocante a licitações e contratos, especifica sete atos lesivos à Administração Pública, todos contendo o termo *fraude* ou variante:

Art. 5º Constituem atos lesivos à administração pública, nacional ou estrangeira, para os fins desta Lei, todos aqueles praticados pelas pessoas jurídicas mencionadas no parágrafo único do art. 1º, que atentem contra o patrimônio público nacional ou estrangeiro, contra princípios da administração pública ou contra os compromissos internacionais assumidos pelo Brasil, assim definidos:

[...]

IV - no tocante a licitações e contratos:

- a) frustrar ou **fraudar**, mediante ajuste, combinação ou qualquer outro expediente, o caráter competitivo de procedimento licitatório público;
- b) impedir, perturbar ou **fraudar** a realização de qualquer ato de procedimento licitatório público;
- c) afastar ou procurar afastar licitante, por meio de **fraude** ou oferecimento de vantagem de qualquer tipo;
- d) **fraudar** licitação pública ou contrato dela decorrente;
- e) criar, de modo **fraudulento** ou irregular, pessoa jurídica para participar de licitação pública ou celebrar contrato administrativo;
- f) obter vantagem ou benefício indevido, de modo **fraudulento**, de modificações ou prorrogações de contratos celebrados com a administração pública, sem autorização em lei, no ato convocatório da licitação pública ou nos respectivos instrumentos contratuais; ou
- g) manipular ou **fraudar** o equilíbrio econômico-financeiro dos contratos celebrados com a administração pública;¹⁸⁵ (grifos nossos)

¹⁸¹ Brasil, *Lei nº 8.666, de 21 de junho de 1993 (Lei Geral de Licitações)*, op. cit.

¹⁸² Cf. Santos; Souza, *Como combater a corrupção em licitações: detecção e prevenção de fraudes*, op. cit., p. 29.

¹⁸³ Sobre o caráter punitivo “penalista” da Lei Anticorrupção, cf. NUCCI, Guilherme de Souza. *Corrupção e anticorrupção*. 1 ed. São Paulo: Forense, 2015, p. 85-91.

¹⁸⁴ BRASIL. *Lei nº 12.846, de 1º de agosto de 2013 (Lei Anticorrupção)*. Dispõe sobre a responsabilização administrativa e civil de pessoas jurídicas pela prática de atos contra a administração pública, nacional ou estrangeira, e dá outras providências. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2013/lei/12846.htm. Acesso em: 23 out. 2020.

¹⁸⁵ Ibid.

Ainda, a *Lei de Improbidade Administrativa* (Lei nº 8.429, de 2 de junho de 1992)¹⁸⁶ em seu art. 3º c/c art. 10, inciso VII, considera *como ato de improbidade administrativa a fraude em processos licitatórios* praticados por agentes públicos ou agentes que induzam ou se beneficiem do ato.¹⁸⁷

É fato que as *fraudes em licitações merecem atenção por parte dos gestores e controladores públicos, pois as compras públicas envolvem grande volume de recursos financeiros*. Em estudo realizado por pesquisadores do Instituto de Pesquisa Econômica Aplicada (IPEA), evidenciou-se que o volume de recursos empregados em compras governamentais, entre 2006 e 2017 foi, em média, de R\$ 5.994 bilhões, o que representa 12,5% do Produto Interno Bruto (PIB) brasileiro. Ainda, o Governo Federal representa pouco mais da metade desse mercado de compras governamentais (6,8% do PIB), enquanto a participação média das compras governamentais no PIB para o período foi de 3,2% para os municípios e 2,2% para os estados (Figura 7).¹⁸⁸

¹⁸⁶ BRASIL. *Lei nº 8.429, de 2 de junho de 1992*. Dispõe sobre as sanções aplicáveis aos agentes públicos nos casos de enriquecimento ilícito no exercício de mandato, cargo, emprego ou função na administração pública direta, indireta ou fundacional e dá outras providências. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l8429.htm. Acesso em: 23 out. 2020.

¹⁸⁷ Lei 8.429/1992: “Art. 3º As disposições desta lei são aplicáveis, no que couber, àquele que, mesmo não sendo agente público, induza ou concorra para a prática do ato de improbidade ou dele se beneficie sob qualquer forma direta ou indireta.

[...]

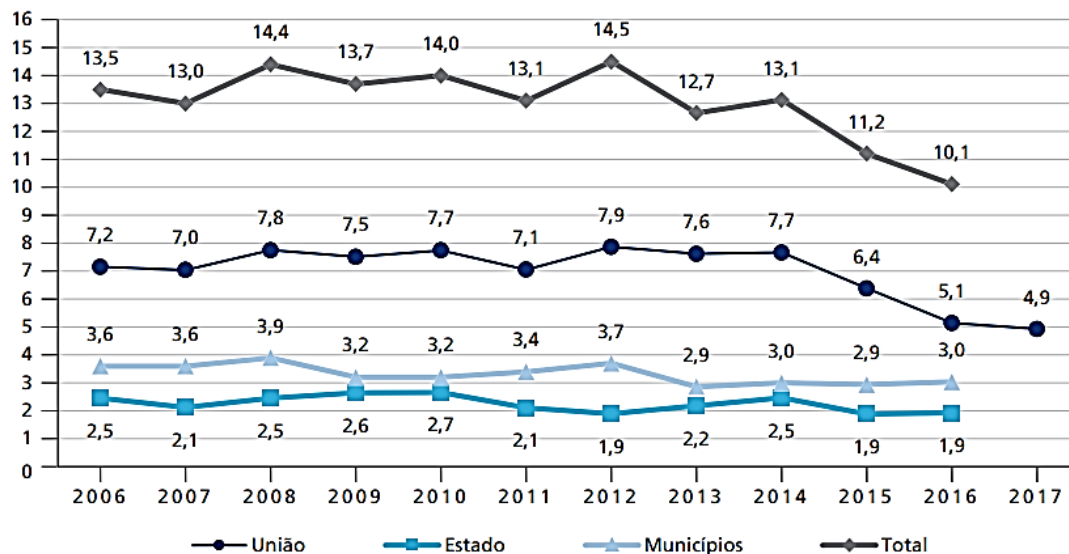
Art. 10. Constitui ato de improbidade administrativa que causa lesão ao erário qualquer ação ou omissão, dolosa ou culposa, que enseje perda patrimonial, desvio, apropriação, malbaratamento ou dilapidação dos bens ou haveres das entidades referidas no art. 1º desta lei, e notadamente:

[...]

VIII - frustrar a licitude de processo licitatório ou de processo seletivo para celebração de parcerias com entidades sem fins lucrativos, ou dispensá-los indevidamente;”. Ibid.

¹⁸⁸ Cf. Ribeiro; Inácio Júnior, *O mercado de compras governamentais brasileiro (2006-2017): mensuração e análise*, op. cit. p. 18-19.

Figura 7 – Compras governamentais do Brasil, em % do PIB, segundo entes da Federação (2006-2017).



Fonte: Retirado de: Ribeiro; Júnior, *O mercado de compras governamentais brasileiro (2006-2017): mensuração e análise*, op. cit., p. 19.

Considerando o conceito de políticas públicas de Maria Paula Dallari Bucci¹⁸⁹ e abordagem metodológica denominada de ciclo de política pública¹⁹⁰, é razoável afirmar que as compras públicas constituem parte importante da fase de implementação das políticas. Corroborar essa afirmação os achados de um levantamento realizado pelo TCU em 2013 sobre o impacto das contratações públicas para as atividades das organizações públicas federais, em que se obteve como resultado que quase metade (47%) das entidades interromperiam alguma de suas três ações orçamentárias consideradas mais relevantes caso não ocorressem aquisições, enquanto outros 24% parariam em até um mês após a interrupção das compras governamentais.¹⁹¹

¹⁸⁹ “Política pública é o programa de ação governamental que resulta de um processo ou conjunto de processos juridicamente regulados – processo eleitoral, processo de planejamento, processo de governo, processo orçamentário, processo legislativo, processo administrativo, processo judicial – visando coordenar os meios à disposição do Estado e as atividades privadas para a realização de objetivos socialmente relevantes e politicamente determinados. Como tipo ideal, a política pública deve visar a realização de objetivos definidos expressando a seleção de prioridades, a reserva de meios necessários à sua consecução e o intervalo de tempo em que se espera o atingimento dos resultados.” BUCCI, Maria Paula Dallari. *O conceito de política pública em direito*. In: BUCCI, Maria Paula Dallari (Org.) *Políticas Públicas: reflexões sobre o conceito jurídico*. São Paulo: Saraiva, 2006, p. 39.

¹⁹⁰ Para mais detalhes sobre o ciclo de políticas públicas, cf.: SCHMIDT, João Pedro. Para estudar políticas públicas: aspectos conceituais, metodológicos e abordagens teóricas. *Revista do Direito*, v. 3, n. 56, 2018, p. 130-139.

¹⁹¹ Cf. TRIBUNAL DE CONTAS DA UNIÃO. *Acórdão nº 2622/2015 – Plenário*. Disponível em: https://pesquisa.apps.tcu.gov.br/#/documento/acordao-completo/*/NUMACORDAO%253A2622%2520ANOACORDAO%253A2015/DTRELEVANCIA%2520desc%252C%2520NUMACORDAOINT%2520desc/0/%2520?uuid=76a1b960-167f-11eb-abea-2d035dec77ea. Acesso em: 23 out. 2020.

Por isso, de acordo com a retromencionada classificação de Luiz Fernando Miranda, pode-se considerar que as fraudes em licitações se enquadram no rol das grandes corrupções, dado o risco de desvios financeiros decorrentes.¹⁹²

Franklin Brasil Santos e Kleberon Roberto de Souza, a partir de suas experiências profissionais como auditores federais da Controladoria Geral da União (CGU), publicaram extensa obra¹⁹³ sobre detecção e prevenção de fraudes em licitações. Os autores classificam cinco principais tipos de fraudes em licitações: (i) projeto mágico; (ii) edital restritivo; (iii) publicidade precária; (iv) julgamento negligente, conivente ou deficiente; (v) contratação direta indevida.¹⁹⁴

O tipo projeto mágico (i) normalmente ocorre quando as especificações do objeto licitado não são suficientemente precisas, sendo genéricas, incompletas e incompreensíveis, ou ainda, quando essas especificações são exageradamente detalhadas com intuito de direcionar a licitação a um fornecedor específico. Em suma, esse tipo fraude decorre de uma apropriação inadequada da licitação para se conseguir uma solução previamente escolhida ao invés de suprir uma necessidade da Administração Pública.¹⁹⁵

O tipo (ii) acontece quando são previstas cláusulas no edital de licitação que comprometem a livre concorrência e de maneira ilegal, isto é, para além das exigências previstas no art. 27 da Lei Geral de Licitações¹⁹⁶, induzindo o resultado para licitantes favorecidos.¹⁹⁷

Por sua vez, o tipo de fraude denominada publicidade precária (iii) desenvolve-se a partir de publicações incompletas, ou forjadas, ou sem observância da antecedência mínima para abertura das propostas ou, ainda, de artimanhas que dificultem o acesso ao edital aos participantes.¹⁹⁸

¹⁹² Cf. Miranda, *Unificando os conceitos de corrupção: uma abordagem através da nova metodologia dos conceitos*, op. cit., p. 259.

¹⁹³ Santos; Souza, *Como combater a corrupção em licitações: detecção e prevenção de fraudes*, op. cit.

¹⁹⁴ Cf. *ibid.*, p. 45.

¹⁹⁵ Cf. *ibid.*, p. 46-75.

¹⁹⁶ Lei 8.666/1993: “Art. 27. Para a habilitação nas licitações exigir-se-á dos interessados, exclusivamente, documentação relativa a:

I - habilitação jurídica;

II - qualificação técnica;

III - qualificação econômico-financeira;

IV – regularidade fiscal e trabalhista;

V – cumprimento do disposto no inciso XXXIII do art. 7º da Constituição Federal.” Brasil, *Lei nº 8.666, de 21 de junho de 1993 (Lei Geral de Licitações)*, op. cit.

¹⁹⁷ Cf. Santos; Souza, *Como combater a corrupção em licitações: detecção e prevenção de fraudes*, op. cit., p. 75-110.

¹⁹⁸ Cf. *ibid.*, p. 110-128.

As fraudes do tipo (iv) ocorrem: por *negligência*, quando passam despercebidos pela comissão de licitação falhas facilmente visíveis e grosseiras, no momento de análise da documentação pelos licitantes; por *conivência*; quando atos da fase externa da licitação são deliberadamente forjados pelos responsáveis do órgão contratante; ou por *deficiência* no julgamento decorrente de adulteração sofisticada na documentação apresentada pelos participantes, dificultando ou inviabilizando a detecção pelos responsáveis pelo processamento do certame.¹⁹⁹

Por fim, as fraudes do tipo *contratação direta indevida* (v) acontecem quando são forjadas situações para fugir da regra geral de contratação via licitação pública, caracterizando o uso inadequado das hipóteses legais de dispensa ou inexigibilidade de licitação, ao contrário do que estabelece os artigos 17, 24 e 25 da Lei 8.666/1993.²⁰⁰

Para fins desta pesquisa, tendo em vista a problemática que a norteia, na próxima seção detalham-se as *fraudes do tipo edital restritivo em licitações*.

3.2 Cláusula restritiva ao caráter competitivo das licitações: um conceito jurídico indeterminado

Em um processo de compra governamental por licitação, *o edital é o instrumento que fixa as condições de participação dos licitantes*, desenvolvimento da disputa e julgamento, estando a Administração e os participantes, vinculados a este documento. É o chamado *princípio da vinculação ao instrumento convocatório*, estabelecido no art. 3º da Lei Geral de Licitações²⁰¹.

Outros princípios, também regidos pelo mesmo dispositivo legal são os da *isonomia*, da *seleção da proposta mais vantajosa* e da *imparcialidade* no julgamento que, combinados, resultam na *condição fundamental de efetiva concorrência* para que as licitações resultem em uso racional dos recursos públicos. Ao mesmo tempo, a Administração Pública se vê na

¹⁹⁹ Cf. *ibid.*, p. 129-189.

²⁰⁰ Cf. *ibid.*, p. 189-201.

²⁰¹ Lei 8.666/1993: “Art. 3º. A licitação destina-se a garantir a observância do princípio constitucional da isonomia, a seleção da proposta mais vantajosa para a administração e a promoção do desenvolvimento nacional sustentável e será processada e julgada em estrita conformidade com os princípios básicos da legalidade, da impessoalidade, da moralidade, da igualdade, da publicidade, da probidade administrativa, da vinculação ao instrumento convocatório, do julgamento objetivo e dos que lhes são correlatos.” Brasil, *Lei nº 8.666, de 21 de junho de 1993 (Lei Geral de Licitações)*, op. cit.

necessidade de mitigar os riscos de contratar uma solução incapaz de atender às necessidades iniciais.²⁰²

Essa dicotomia exige do gestor público um *equilíbrio nas especificações do objeto* para atender ao interesse público. Para Cláudio Sarian Altounian, as exigências feitas pelo gestor devem restringir-se aos limites da razoabilidade e “balancear as preocupações de afastar empresas sem condições de realizar o empreendimento com o interesse em ter o maior número de participantes, de forma a aumentar a competitividade e receber a proposta mais vantajosa”²⁰³.

Existem *maneiras legítimas de proteger a Administração e reduzir os riscos* de não cumprir com os objetivos da aquisição.²⁰⁴

Destaca-se o contorno dado pelo disposto no art. 37, inciso XXI, da Constituição Federal, que define que *as exigências para a qualificação técnica e econômica devem ser as mínimas possíveis*.²⁰⁵ Ainda, o art. 3, §1º da Lei nº 8.666/93 estabelece que é proibido a um agente público “admitir, prever, incluir ou tolerar cláusulas ou condições que comprometam, restrinjam ou frustrem o caráter competitivo da licitação”²⁰⁶. De mais a mais, a Seção II do Capítulo II da Lei Geral de Licitações rege as condições de habilitação nas licitações.

O art. 27 da Lei 8.666/1993 preceitua naturezas de documentação a serem exigidas para a habilitação dos licitantes:

Art. 27. Para a habilitação nas licitações exigir-se-á dos interessados, exclusivamente, documentação relativa a:
I- habilitação jurídica;
II- qualificação técnica;
III- qualificação econômico-financeira;
IV- regularidade fiscal e trabalhista
V- cumprimento do disposto no inciso XXXIII do art. 7º da Constituição Federal.²⁰⁷

²⁰² Cf. Santos; Souza, *Como combater a corrupção em licitações: detecção e prevenção de fraudes*, op. cit., p. 75-77.

²⁰³ ALTOUNIAN, Cláudio Sarian. *Obras públicas: licitação, contratação, fiscalização e utilização*. 5. ed. Belo Horizonte: Fórum, 2016, p. 243.

²⁰⁴ Cf. Santos; Souza, *Como combater a corrupção em licitações: detecção e prevenção de fraudes*, op. cit., p. 76.

²⁰⁵ Art. 37, inciso XXI da CF/88: “ressalvados os casos especificados na legislação, as obras, serviços, compras e alienações serão contratados mediante processo de licitação pública que assegure igualdade de condições a todos os concorrentes, com cláusulas que estabeleçam obrigações de pagamento, mantidas as condições efetivas da proposta, nos termos da lei, o qual somente permitirá as exigências de qualificação técnica e econômica indispensáveis à garantia do cumprimento das obrigações.” Brasil, *Constituição Federal de 5 de outubro de 1988*, op. cit.

²⁰⁶ Brasil, *Lei nº 8.666, de 21 de junho de 1993 (Lei Geral de Licitações)*, op. cit.

²⁰⁷ Ibid.

Nos artigos seguintes da mesma Seção II (art. 28 a art. 31), a Lei 8.666/1993 define as documentações exigíveis para cada um dos incisos do art. 27 acima transcrito.²⁰⁸ Em suma, a Administração Pública, “ao definir os requisitos de habilitação no edital, deve não só observar os limites legais, como também a razoabilidade das exigências, que, dentro da segurança de execução contratual pretendida, precisam representar o menor cerceamento à competição”²⁰⁹.

Nessa lógica, *fraudes do tipo edital restritivo em licitações constituem-se na inclusão de cláusulas que visem excluir a participação de um ou mais concorrentes*. Mas o que determina uma cláusula como restritiva?

Maria Sylvia Zanella Di Pietro define *conceito jurídico indeterminado* como aquelas expressões linguísticas que são “empregadas para designar vocábulos ou expressões que não têm um sentido preciso, objetivo, determinado, mas que são encontrados com grande frequência nas regras jurídicas dos vários ramos do direito”²¹⁰. Como *exemplos*, a autora menciona: *boa-fé, bem comum, conduta irrepreensível, pena adequada, interesse público, notório saber*, dentre outros termos.²¹¹

Uma característica de um conceito jurídico indeterminado é sua *mutabilidade*, isto é, a possibilidade de variação do seu significado no tempo e no espaço, de acordo com as mudanças sociais.²¹²

Outra característica é que todo conceito indeterminado remete a uma significação reconhecida uniformemente por um grupo social, que é o chamado de *núcleo conceitual* ou *zona de certeza*. Esta pode ser uma *certeza positiva*, ou seja, aquilo que não se duvida do que o conceito representa, ou uma *certeza negativa*, representada por aquilo que se sabe que o conceito não representa. Entre esses dois polos de certeza, existe um espaço de dúvidas quanto à aplicação ou não do conceito, denominado de *zona de incerteza* ou *halo conceitual*.²¹³

Em virtude da zona de incerteza, os conceitos legais indeterminados levam a uma discussão, no âmbito do Direito Administrativo, sobre a sua relação com a *discrecionabilidade*

²⁰⁸ Para obter um detalhamento completo, cf. Naves, *TCU e stare decisis administrativo: o caso das cláusulas restritivas em editais de licitação de obras de engenharia*, op. cit., p. 79-83.

²⁰⁹ Ibid., p. 80.

²¹⁰ Di Pietro, *Discrecionabilidade administrativa na Constituição de 1988*, op. cit., p. 90.

²¹¹ Cf. ibid., p. 90.

²¹² Cf. ibid., p. 90.

²¹³ Cf. ROZAS, Luiza Barros. Conceitos jurídicos indeterminados e discrecionabilidade administrativa. *Cadernos Jurídicos da Escola Paulista da Magistratura*, v. 20, n. 47, 2019, p. 192.

*administrativa*²¹⁴.²¹⁵ Entretanto, a depender do caso concreto, os conceitos indeterminados podem gerar discricionariedade ou vinculação da ação administrativa.²¹⁶ Veja-se.

Existem dois tipos de conceitos jurídicos indeterminados: *conceitos de valor* e *conceitos de experiência (ou empíricos)*.²¹⁷ O primeiro tipo permite que a discricionariedade seja aplicada, uma vez que, após o processo de interpretação, conclui-se por estar fora das zonas de certeza positiva ou negativa, situação em que restará certa margem de apreciação para a Administração Pública decidir.²¹⁸ Já no caso dos *conceitos de experiência*, “a discricionariedade fica afastada porque existem critérios objetivos, práticos, extraídos da experiência comum, que permitem concluir qual a única solução possível”²¹⁹.

Nessa acepção, entende-se que a expressão *cláusula restritiva é um conceito jurídico indeterminado de experiência*. Nela, há a presença de um núcleo de certeza comum a todos, qual seja, de um trecho que pertence a um determinado edital de licitação, que se constitui de uma exigência não razoável e sem justificativa para mitigar riscos relevantes, e que, por isso, frustra os princípios da competitividade e da isonomia do certame. Também existe um núcleo de incerteza que deve ser delimitado a partir da *praxe administrativa*²²⁰, que estabelece critérios práticos que permitem concluir por uma solução possível diante de um caso concreto.

Franklin Santos e Kleberson Souza definiram ao menos 11 classes de cláusulas restritivas em editais de licitação. A partir exemplos retirados de quase 300 decisões colegiadas do TCU que versaram sobre o tema, os autores detalham cada uma dessas classes de maneira a buscar concretude ao conceito de cláusula restritiva de editais de licitação.²²¹ Elas são referentes a:

1. *Garantia de proposta para controle de interessados*: a Lei nº 8.666/1993 prevê a possibilidade de a Administração exigir garantia de participação, com o objetivo de substituir a comprovação de capacidade econômico-financeira. Editais restritivos podem exigir o recolhimento dessa garantia com antecedência em relação à abertura do certame,

²¹⁴ “[...] definir a discricionariedade administrativa como a faculdade de que a lei confere à Administração para apreciar o caso concreto, segundo critérios de oportunidade e conveniência, e escolher uma dentre duas ou mais soluções, todas válidas perante o direito.” Di Pietro, *Discricionariedade administrativa na Constituição de 1988*, op. cit., p. 62.

²¹⁵ Cf. *ibid.*, p. 91.

²¹⁶ Rozas, *Conceitos jurídicos indeterminados e discricionariedade administrativa*, op. cit., p. 199-200.

²¹⁷ Cf. Di Pietro, *Discricionariedade administrativa na Constituição de 1988*, op. cit., p. 115.

²¹⁸ Cf. *ibid.*, p. 119.

²¹⁹ *Ibid.*, p. 118.

²²⁰ “A praxe administrativa é a atividade interna, reiterada e uniforme da Administração Pública na aplicação das normas e dos atos jurídicos”. Oliveira, *Precedentes no Direito Administrativo*, op. cit., p. 118.

²²¹ Cf. Santos; Souza, *Como combater a corrupção em licitações: detecção e prevenção de fraudes*, op. cit., p. 75-110.

ou ainda em percentual acima do limite legal. Essa classe foi exemplificada com 20 acórdãos do TCU.²²²

2. *Capacidade econômica exagerada*: é permitido fixar percentual mínimo para Capital Social ou Patrimônio Líquido, desde que proporcional aos riscos inerentes à contratação, como a inexecução ou falha no contrato, devendo considerar o valor estimado do contrato, a essencialidade do objeto, o tempo de duração do contrato, as características do mercado fornecedor em potencial etc. Editais restritivos podem fixar índices contábeis não usuais e injustificados, exigir, simultaneamente, garantia de proposta junto com capital social ou patrimônio líquido ou, ainda, demandar Capital Social mínimo “integralizado”. Essa classe foi exemplificada com 41 acórdãos do TCU.²²³
3. *Idoneidade financeira ou bancária*: editais restritivos podem exigir declaração ou atestado de idoneidade financeira ou idoneidade bancária em nome da empresa licitante de seus sócios e administradores, o que não está previsto em lei. Essa classe foi exemplificada com 2 acórdãos do TCU.²²⁴
4. *“Quitação” em vez de “regularidade” fiscal*: editais restritivos podem exigir que os licitantes apresentem certidões de “quitação” de obrigações fiscais emitidas pelas Fazendas Federal, Estadual ou Municipal. Entretanto, não há previsão legal para a requisição de documentação que visa atribuir situação de “quitação”, mas tão somente, para demandar a comprovação da “regularidade” junto a essas instituições. Essa classe foi exemplificada com 4 acórdãos do TCU.²²⁵
5. *Certidões ou certificações indevidas*: editais restritivos podem conter exigências para apresentação, pelos licitantes, de certidões ou certificações que não possuam previsão legal e que não foram devidamente justificadas no processo. Como exemplo, tem-se a demanda por certidão negativa de débito salarial, certidão negativa de infrações trabalhistas e certidão de regularidade sindical. Essa classe foi exemplificada com 22 acórdãos do TCU.²²⁶
6. *Alvará do município da licitação*: salvo se devidamente justificada a sua necessidade, a exigência de alvará de funcionamento no município em que será realizada a contratação como

²²² Cf. *ibid.*, p. 77-79.

²²³ Cf. *ibid.*, p. 79-85.

²²⁴ Cf. *ibid.*, p. 85.

²²⁵ Cf. *ibid.*, p. 85-86.

²²⁶ Cf. *ibid.*, p. 86-87. Esclarece-se que, originalmente, os autores apresentam essa classe em três classes distintas: “Certidão negativa de débito salarial e infrações trabalhistas”; “Certidão de regularidade sindical”; e “Outras certidões ou certificações indevidas”. Para fins desta pesquisa, como todas tratam da exigência de certidões ou certificações indevidas, resolveu-se unificar essas três classes em apenas uma.

documentação relativa à qualificação técnica, restringe o caráter competitivo do certame. Essa classe foi exemplificada com 3 acórdãos do TCU.²²⁷

7. *Capacidade técnica irregular*: a qualificação técnica divide-se em capacidade técnico-operacional, ligada à aptidão da empresa, e capacidade técnico-profissional, relacionada à aptidão dos profissionais que participam do quadro da empresa. Os autores a decompõem nas seguintes subclasses, exemplificadas por meio de 121 acórdãos distintos²²⁸ do TCU:²²⁹
 - 7.1. *Atestados em número mínimo, máximo ou fixo*: que limitam a quantidade ou somatório de atestados para comprovação da capacidade técnica; (17 acórdãos)
 - 7.2. *Quantitativo exagerado de experiência mínima*: que exigem experiência da empresa (capacidade técnico-operacional) em percentual superior a 50% dos quantitativos previstos na licitação; (12 acórdãos)
 - 7.3. *Experiência em parcela irrelevante*: que exigem experiência em parcelas de pouca relevância técnica ou valoração do objeto licitado; (6 acórdãos)
 - 7.4. *Experiência específica demais*: que requerem tipologia específica ou idêntica de experiência como, por exemplo, um tipo particular de obra; (19 acórdãos)
 - 7.5. *Experiência genérica demais*: que constituem em critérios de qualificação técnica devem não objetivos e genéricos; (10 acórdãos)
 - 7.6. *Limitação de tempo e época nos atestados*: que estabelecem limitação de época para experiência prévia (últimos cinco anos, por exemplo); (1 acórdão)
 - 7.7. *Atestado do mesmo grupo econômico*: que vedam, injustificadamente, atestados apresentados por empresas que pertençam ao mesmo grupo econômico; (2 acórdãos)
 - 7.8. *Reorganização empresarial (fusão, cisão e incorporação)*: que vedam a possibilidade de transferência de capacidade técnica nos casos de fusão, cisão e incorporação; (4 acórdãos)
 - 7.9. *Transferência de capacidade técnico-profissional para operacional*: que permitam transferir acervo técnico de pessoa física para pessoa jurídica, fazendo confusão com os termos capacidade técnico-profissional e técnico-operacional; (1 acórdão)
 - 7.10. *Matriz e filial*: que vedam a os atestados de capacidade ou de responsabilidade técnica possam ser apresentados em nome da matriz ou da filial da empresa licitante; (3 acórdãos)

²²⁷ Cf. *ibid.*, p. 87.

²²⁸ Alguns acórdãos se repetem dentre as subclasses.

²²⁹ Cf. *ibid.*, p. 88-102.

- 7.11. *Visto do CREA da Seção Local*: que exigem, como requisito de habilitação, visto do Conselho Regional de Engenharia e Agronomia (CREA) do local da obra na certidão de registro da licitante; (10 acórdãos)
- 7.12. *Certificados de Qualidade*: que demandam certificação de qualidade como requisito de habilitação em procedimentos licitatórios, a exemplo de certificados “ISO 9000”, selo da Associação Brasileira da Indústria de Café (ABIC) para o café ou nível específico de certificação PBQPH – Programa Brasileiro da Qualidade e Produtividade do Habitat e Certificado de Boas Práticas de Fabricação e Controle. Essas certificações são admitidas apenas como condição para classificação ou como requisito contratual; (20 acórdãos)
- 7.13. *Profissional do “quadro permanente”*: que restrinjam a forma de comprovação de vínculo do profissional com a licitante ou, ainda, tempo mínimo de vínculo antes da publicação do edital; (17 acórdãos)
- 7.14. *Quantitativos mínimos em experiência do profissional*: que exigem quantitativos mínimos para experiência do profissional (capacidade técnico-profissional); (11 acórdãos)
- 7.15. *CAT em nome da empresa licitante*: que exigem Certidão de Acervo Técnico (CAT) em nome do licitante, sendo que certidões desse tipo são emitidas em nome de profissionais; (2 acórdãos)
8. *Visita técnica restritiva*: em caso de licitações que possibilitem vistoria ao local do serviço (normalmente obra ou serviço de engenharia), ela deve ser entendida como um direito do licitante e não uma obrigação. Assim, é restritivo exigir visita obrigatória ou especificar o profissional para se realizar a vistoria. Essa classe foi exemplificada com 27 acórdãos do TCU.²³⁰
9. *Carta de solidariedade do fabricante*: a exigência de comprovação de que o concorrente seja representante autorizado do produto ou de declaração de solidariedade do fabricante para com o licitante é considerada como restritiva. Essa classe foi exemplificada com 3 acórdãos do TCU.²³¹
10. *Amostras subjetivas, indevidas, onerosas*: exigir amostras como condições de habilitação restringe a competitividade da licitação. Essa possibilidade existe, no entanto, em fase

²³⁰ Cf. *ibid.*, p. 102-105.

²³¹ Cf. *ibid.*, p. 105.

posterior, de classificação das propostas. Essa classe foi exemplificada com 10 acórdãos do TCU.²³²

11. *Empresas com sócios em comum*: não há vedações legais para a participação de empresas com sócios em comum, a não ser em casos de vínculo entre as licitantes e autor do projeto; ou de empresa que fiscaliza outra. Por isso, cláusulas que fogem dessas possibilidades são consideradas restritivas. Essa classe foi exemplificada com 6 acórdãos do TCU.²³³

Em suma, os autores afirmam que “quaisquer exigências especiais de habilitação devem estar previstas na lei geral de licitações ou devidamente justificadas no processo, sob pena de serem consideradas restritivas à competitividade do certame”²³⁴.

Em pesquisa similar, mas aplicada a obras e serviços de engenharia²³⁵, Fernanda de Moura Ribeiro Naves realizou “levantamento bibliográfico das principais cláusulas restritivas ao caráter competitivo, referentes à qualificação técnica e à qualificação econômico-financeira (arts. 30 e 31 da Lei Geral de Licitações)”²³⁶.

Assim, baseada na doutrina especializada, a autora identificou ao menos 14 classes específicas de cláusulas restritivas para editais de licitação de obras públicas, para as quais desenvolveu palavras-chave para busca de decisões colegiadas correlatas no portal de jurisprudência do TCU.²³⁷ Dessas quatorze, a pesquisadora *elegeu 4 classes para aprofundamento qualitativo no texto das decisões encontradas*,²³⁸ a fim de “traçar inferências descritivas, no sentido de observar se o TCU consegue reconhecer seus precedentes ao decidir casos análogos em que existam cláusulas restritivas ao caráter competitivo em editais de licitação de obras”²³⁹.

²³² Cf. *ibid.*, p. 106-107.

²³³ Cf. *ibid.*, p. 107-108.

²³⁴ *Ibid.*, p. 87.

²³⁵ Conforme a Orientação Técnica nº 002/2009 do Instituto Brasileiro de Auditoria de Obras Públicas (IBRAOP): “Obra de engenharia é a ação de construir, reformar, fabricar, recuperar ou ampliar um bem, na qual seja necessária a utilização de conhecimentos técnicos específicos envolvendo a participação de profissionais habilitados conforme o disposto na Lei Federal nº 5.194/66.

[...]

Serviço de engenharia é toda atividade que necessite da participação e acompanhamento de profissional habilitado conforme o disposto na Lei Federal nº 5.194/66, tais como: consertar, instalar, montar, operar, conservar, reparar, adaptar, manter, transportar, ou ainda, demolir. Incluem-se nesta definição as atividades profissionais referentes aos serviços técnicos profissionais especializados de projetos e planejamentos, estudos técnicos, pareceres, perícias, avaliações, assessorias, consultorias, auditorias, fiscalização, supervisão ou gerenciamento.” INSTITUTO BRASILEIRO DE AUDITORIA DE OBRAS PÚBLICAS. *Orientação Técnica OT – IBR 002/2009: obra e serviço de engenharia*, p. 2. Disponível em: <http://www.ibraop.org.br/wp-content/uploads/2013/06/OT-IBR-02-2009-Ibraop-01-07-10.pdf>. Acesso em: 27 out. 2020.

²³⁶ Naves, *TCU e stare decisis administrativo: o caso das cláusulas restritivas em editais de licitação de obras de engenharia*, op. cit., p. 83.

²³⁷ Cf. *ibid.*, p. 87.

²³⁸ Cf. *ibid.*, p. 88.

²³⁹ *Ibid.*, p. 77.

O estudo qualitativo de Naves também teve por objetivo expandir a quantidade de acórdãos encontrados, com uso exclusivo do critério de citação, isto é, a partir da leitura de cada acórdão, descobriu-se outros que foram referenciados por aqueles e que não haviam sido encontrados na busca inicial por palavras-chave no portal do TCU.²⁴⁰ Em síntese, as *quatro classes delineadas* foram:

1. “obra” e “atestado” e “tipologia específica” e “experiência” e “restrição”: análoga à classe 7, subclasse 4, de Franklin Santos e Kleberson Souza, em que se exige, como critério de habilitação, atestado de qualificação técnica comprovando experiência em tipologia específica de obra. Foram encontrados 19 acórdãos para essa classe.²⁴¹
2. “obra” e “exigência de quitação” e “crea”: em que se buscou decisões relativa à exigência de quitação de anuidade no CREA, considerada restritiva por não ser prevista em lei. Essa classe é similar à classe 5 de Franklin Santos e Kleberson Souza, pois exige-se uma certidão de regularidade indevida. A pesquisadora identificou 29 acórdãos dessa classe.²⁴²
3. “obra” e “capital social mínimo integralizado” e “restrição”: equivalente à classe 2 de Franklin Santos e Kleberson Souza, uma vez que se exige valor mínimo integralizado (quando os recursos já entraram na empresa) de Capital Social. Foram identificados 36 acórdãos.²⁴³
4. “obra” e “exigência de visto” e “crea” e “restrição”: correspondente à classe 7, subclasse 11, de Franklin Santos e Kleberson Souza, em que se refere à exigência de exigência de visto no CREA local como condição de habilitação. Foram encontrados 40 acórdãos para essa classe.²⁴⁴

Considerando as classes criadas por Franklin Santos e Kleberson Souza, seu trabalho, juntamente com de Fernanda Naves, *identificaram 330 acórdãos*²⁴⁵ do TCU, *divididos em 11 classes diferentes*, que versaram sobre a ilegalidade de cláusulas de editais de licitação, em casos concretos, por terem sido consideradas restritivas ao caráter competitivo do certame. Uma lista contendo todos os 330 acórdãos e suas respectivas classes encontra-se no Apêndice A.

²⁴⁰ Cf. *ibid.*, p. 88-89.

²⁴¹ Cf. *ibid.*, p. 89-94.

²⁴² Cf. *ibid.*, p. 94-101.

²⁴³ Cf. *ibid.*, p. 101-107.

²⁴⁴ Cf. *ibid.*, p. 107-114.

²⁴⁵ Ressalta-se que uma mesma decisão pode pertencer a mais de uma classe, uma vez que *mais de uma patologia* de restrição à competitividade *pode ser encontrada em um determinado certame* auditado pelo TCU. Houve, também, situações de acórdãos que foram identificados duplamente por Santos e Souza e por Naves para a mesma classe, e outras situações que os autores identificaram o mesmo acórdão para classes distintas. Por isso, o número de acórdãos distintos é menor que o número de classificações totais realizadas pelos autores. Ver a Tabela 6.

Dela, desprende-se que *a uma mesma decisão colegiada, pode pertencer a mais de uma classe relativa a cláusulas restritivas ao caráter competitivo em editais de licitação.*

Portanto, à luz dos modelos jurisdicionais de Miguel Reale e dos exemplos supracitados, é possível dizer que *é a partir dos precedentes administrativos que o conceito jurídico indeterminado de cláusula restritiva ganha efetiva concreção.* A jurisprudência do TCU, em especial, tem importância ainda maior, visto que suas decisões concernentes à aplicação de normas gerais de licitação servem como parâmetro para os demais entes federados, conforme Súmula nº 222 daquela Corte²⁴⁶. *Justifica-se, assim, o recorte temático escolhido para esta pesquisa, que utilizará os 330 acórdãos do Tribunal de Contas da União previamente classificados para seus testes empíricos.*

O desafio enfrentado por esta pesquisa consiste na *organização sistemática* das decisões colegiadas proferidas pelo TCU, de maneira a *possibilitar consulta inequívoca da jurisprudência e dos precedentes administrativos* da Corte de Contas da União sobre o recorte temático escolhido e, de tal modo, *proporcionar maior previsibilidade e coerência* nas ações administrativas, isto é, *maior segurança jurídica.*

Isso porque, na lógica abordada por Héctor Mairal, *a insegurança jurídica é um campo fértil para a corrupção* na medida em que coaduna com o desconhecimento de normas e de suas interpretações.²⁴⁷ Desta feita, *a sedimentação de uma cultura de precedentes administrativos no Brasil possui o condão de subsidiar futuras condutas de gestores públicos, “funcionando como instrumento de indução de boas práticas, evitando a ocorrência de irregularidades e corrigindo a assimetria de informações”*²⁴⁸.

3.3 Uso de Inteligência Artificial pelo TCU para detecção de fraudes em licitações

Dada a função constitucional dos Tribunais de Contas no Brasil de realizar o controle externo da Administração Pública ao fiscalizar a legalidade e legitimidade dos gastos, bem

²⁴⁶ “SÚMULA Nº 222. As Decisões do Tribunal de Contas da União, relativas à aplicação de normas gerais de licitação, sobre as quais cabe privativamente à União legislar, devem ser acatadas pelos administradores dos Poderes da União, dos Estados, do Distrito Federal e dos Municípios.” TRIBUNAL DE CONTAS DA UNIÃO. *Súmulas Nº 001 a 289*, op. cit.

²⁴⁷ Cf. MAIRAL, Héctor A. *As raízes legais da corrupção: ou como o direito público fomenta a corrupção em vez de combatê-la*. KRAEMER, Susan M. B. (trad.). MUKAI, Toshio (comentários ed. brasileira). São Paulo: Editora Contracorrente, 2018, p. 49-67.

²⁴⁸ Cf. Naves, *TCU e stare decisis administrativo: o caso das cláusulas restritivas em editais de licitação de obras de engenharia*, op. cit., p. 74-75.

como o bom funcionamento das políticas públicas,²⁴⁹ condutas que se encaixam como patologias corruptivas são eventualmente identificadas em suas atuações.²⁵⁰

Considerando que rotinas automatizadas que visam realizar tarefas humanas e que ensejam detectar tipologias corruptivas estão entre as múltiplas aplicações da inteligência artificial,²⁵¹ nos últimos anos, o TCU tem desenvolvido diversas ferramentas tecnológicas que, com utilização de algoritmos inteligentes, proporcionam maior eficiência, racionalidade e principalmente tempestividade para o controle externo na identificação de fraudes em licitações, a exemplo dos robôs abaixo.²⁵²

O robô *Adele* (Análise de Disputa em Licitações Eletrônicas) consiste em um painel gerencial que tem sido utilizado pela Corte de Contas da União no acompanhamento da dinâmica dos pregões eletrônicos. Seu objetivo é detectar a existência de possíveis fraudes, como o conluio entre licitantes, por meio da análise dos lances efetuados e do cruzamento de bases de dados diversas que o TCU possui acesso.²⁵³

Já o *Alice* (Análise de Licitações e Editais) foi desenvolvido inicialmente pela CGU com o objetivo de realizar uma avaliação preventiva e automatizada dos certames do poder Executivo da União. Em 2016, por intermédio de uma parceria entre os órgãos federais de controle interno e externo, o código do *Alice* foi transferido para o TCU e as instituições passaram a utilizar e desenvolver melhorias na ferramenta conjuntamente.²⁵⁴

Basicamente, o *Alice* funciona por meio de expressões regulares²⁵⁵ que identificam padrões associados a indícios de irregularidade nos editais de licitação eletrônico publicados diariamente no Portal de Compras do Governo Federal (*ComprasNet*). Dentre as tipologias buscadas, estão possíveis restrições à competitividade “previstas na jurisprudência do TCU”²⁵⁶

²⁴⁹ Ver o art. 71 da Constituição Federal de 1988. Brasil, *Constituição Federal de 5 de outubro de 1988*, op. cit.

²⁵⁰ Cf. NEVES, F. R., SILVA, P. B., CARVALHO, H. L. M. Artificial ladies against corruption: searching for legitimacy at the Brazilian Supreme Audit Institution. *Revista de Contabilidade e Organizações*, v. 13, 2019, p. 34.

²⁵¹ Cf. FREITAS, Juarez. Direito Administrativo e inteligência artificial. *Interesse Público – IP*, n. 114, 2019, p. 19.

²⁵² Cf. COSTA, Marcos Bemquerer. BASTOS, Patrícia Reis Leitão. Alice, Monica, Adele, Sofia, Carina e Ágata: o uso da inteligência artificial pelo Tribunal de Contas da União. *Controle Externo: Revista do Tribunal de Contas do Estado de Goiás*, v. 2, n. 3, 2020, p. 32-33.

²⁵³ Cf. *ibid.*, p. 25.

²⁵⁴ Cf. Naves, *TCU e stare decisis administrativo: o caso das cláusulas restritivas em editais de licitação de obras de engenharia*, op. cit., p. 56-63;165-168. Cf. Costa e Bastos, *Alice, Monica, Adele, Sofia, Carina e Ágata: o uso da inteligência artificial pelo Tribunal de Contas da União*, op. cit., p. 14-21.

²⁵⁵ Expressão regular, também chamada de *Regex*, é uma linguagem computacional utilizada para identificação e transformação de cadeias de caracteres de interesse.

²⁵⁶ Costa e Bastos, *Alice, Monica, Adele, Sofia, Carina e Ágata: o uso da inteligência artificial pelo Tribunal de Contas da União*, op. cit., p. 13.

e valores da licitação, de modo que essas inconsistências são enviadas automaticamente por e-mail para as unidades técnicas do TCU, possibilitando ações de controle tempestivas.²⁵⁷

A partir de 2019, por iniciativa da Associação dos Membros dos Tribunais de Contas do Brasil (Atricon), iniciou-se o projeto *Alice Nacional* para ampliação do uso deste robô também por tribunais de contas estaduais e municipais.²⁵⁸ Esses tribunais encaminham os editais de licitação sob sua jurisdição ao TCU para consumo do *Alice* e recebem os indícios de irregularidades por mensagens eletrônicas.²⁵⁹ À luz do *conceito de políticas públicas* de Maria Paula Dallari Bucci²⁶⁰, a iniciativa do *Alice Nacional* tem o condão de esboçar uma política pública nacional de combate e prevenção a fraudes em licitações, visto que se constitui um conjunto de ações de controle externo coordenadas para a ação tempestiva a tipologias corruptivas comuns a aquisições públicas.

Com vistas a aprimorar os algoritmos do *Alice*, o TCU desenvolveu recentemente o *Ágata* (Aplicação para Geração de Análise Textual Acelerada) baseado em algoritmos de aprendizado de máquina com intuito de construir pesquisas textuais de forma simples e ágil, com uso de inteligência artificial.²⁶¹

O *Ágata* funciona em duas etapas: primeiramente, exige-se do usuário que haja a definição de um termo de busca que é a fonte inicial de informações; em seguida, o robô apresenta excertos de editais e solicita ao usuário que confirme se representa a situação procurada. Após esse processo, o usuário pode optar por receber e-mails com informações sobre novos editais de licitação publicados que atendam aos requisitos estabelecidos.²⁶²

Em suma, *essas inovações tecnológicas utilizam a expertise humana sobre fraudes em licitações para treinar um sistema baseado em Inteligência Artificial para automatizar a busca*

²⁵⁷ Cf. Naves, *TCU e stare decisis administrativo: o caso das cláusulas restritivas em editais de licitação de obras de engenharia*, op. cit., p. 56-63. Cf. Costa e Bastos, *Alice, Monica, Adele, Sofia, Carina e Ágata: o uso da inteligência artificial pelo Tribunal de Contas da União*, op. cit., p. 14-21.

²⁵⁸ Cf. Costa e Bastos, *Alice, Monica, Adele, Sofia, Carina e Ágata: o uso da inteligência artificial pelo Tribunal de Contas da União*, op. cit., p. 19-21.

²⁵⁹ O TCE-GO participa do projeto *Alice Nacional*. Disponível em: <https://portal.tce.go.gov.br/-/robo-alice-automatiza-analise-de-licitacoes>. Acesso em: 23 jan. 2021.

²⁶⁰ “Política pública é o programa de ação governamental que resulta de um processo ou conjunto de processos juridicamente regulados – processo eleitoral, processo de planejamento, processo de governo, processo orçamentário, processo legislativo, processo administrativo, processo judicial – visando coordenar os meios à disposição do Estado e as atividades privadas para a realização de objetivos socialmente relevantes e politicamente determinados. Como tipo ideal, a política pública deve visar a realização de objetivos definidos expressando a seleção de prioridades, a reserva de meios necessários à sua consecução e o intervalo de tempo em que se espera o atingimento dos resultados.” BUCCI, Maria Paula Dallari. *O conceito de política pública em direito*. In: BUCCI, Maria Paula Dallari (Org.) *Políticas Públicas: reflexões sobre o conceito jurídico*. São Paulo: Saraiva, 2006, p. 39.

²⁶¹ Cf. Costa e Bastos, *Alice, Monica, Adele, Sofia, Carina e Ágata: o uso da inteligência artificial pelo Tribunal de Contas da União*, op. cit., p. 14.

²⁶² Cf. *ibid.*, p. 26-27.

de possíveis irregularidades em aquisições públicas, como a existência de cláusulas que restrinjam o caráter competitivo do certame.

Todavia, não existe no TCU um trabalho de sistematização de sua jurisprudência de modo a dar suporte inequívoco ao conhecimento consumido pelos robôs que buscam irregularidades. O sítio eletrônico do TCU, apesar de ser referência para os demais tribunais de contas quanto a pesquisa nas bases de dados jurisprudenciais, limita-se às buscas por palavras-chave em qualquer parte do texto das decisões colegiadas²⁶³ ou à visualização de acórdãos previamente selecionados pela própria Corte²⁶⁴.

Assim, o um método de extração jurisprudencial que se aplique universalmente às Cortes de Contas, pode contribuir para que exista maior segurança jurídica não apenas aos gestores públicos em suas ações administrativas, mas também aos auditores de controle externo no aprimoramento dos algoritmos utilizados pelos robôs do TCU. A sistemática aqui delineada lança um novo olhar sobre como a IA pode inovar a *organização jurisprudencial e consequente segurança jurídica dos algoritmos de IA* utilizados na busca por fraudes em licitações, em especial, aquelas relacionadas a cláusulas restritivas à competitividade.

²⁶³ O Portal de Jurisprudência de acórdãos do TCU está disponível em: <https://pesquisa.apps.tcu.gov.br/#/pesquisa/acordao-completo>. Acesso em 18 out. 2019.

²⁶⁴ O Portal de Jurisprudência Selecionada do TCU está disponível em: <https://pesquisa.apps.tcu.gov.br/#/pesquisa/jurisprudencia-selecionada>. Acesso em 18 out. 2019.

4 ESTRATÉGIA EMPÍRICA: TENTATIVA DE CONCEPÇÃO DE UM MÉTODO DE EXTRAÇÃO SISTEMÁTICA DE JURISPRUDÊNCIA

Este capítulo destina-se à *apresentação dos resultados empíricos* deste estudo. Primeiramente, é trazida uma *revisão sistemática da literatura, com enfoque meta-analítico*, associada ao estado da arte computacional aplicado ao problema da pesquisa (seção 4.1). Em seguida, a seção 4.2 contém os *detalhes operacionais adotados, os resultados obtidos e suas análises críticas* pela ótica da discussão teórica detalhada nos capítulos anteriores. Por fim, na seção 4.3, apresenta-se um ensaio para *identificação de precedentes administrativos* e suas *rationes decidendi*, bem como uma *provável linha jurisprudencial de entendimento* acerca de cláusulas consideradas restritivas em editais de licitação por exigirem capacidade técnica dos licitantes de maneira irregular.

4.1 Revisão sistemática da literatura com enfoque meta-analítico do estado da arte aplicado ao problema de pesquisa

Como demonstrado pelas pesquisas de Franklin Santos e Kleberon Souza e de Fernanda Naves no capítulo anterior, a um mesmo acórdão pode ser atribuída mais de uma das 11 classes relativas a cláusulas restritivas ao caráter competitivo em editais de licitação (ver Apêndice A). Essa característica impõe ao problema de pesquisa que ele seja tratado como uma classificação *multi-label* de documentos, isto é, quando cada instância analisada pode receber mais de um rótulo ao mesmo tempo.²⁶⁵

Assim, primeiramente foi realizada uma revisão sistemática da literatura²⁶⁶, com enfoque meta-analítico²⁶⁷, a fim de auxiliar na descoberta do estado da arte em Inteligência

²⁶⁵ Um problema de classificação do tipo *multi-class* ocorre quando a cada instância d pertencente ao conjunto D é atribuída a uma única classe y pertencente ao conjunto Y . Por outro lado, um problema de classificação do tipo *multi-label* ocorre quando a cada instância d pertencente ao conjunto D pode ser atribuída mais de uma classe y pertencente ao conjunto Y . Cf. THABTAH, F. A.; COWLING, P.; PENG, Y. MMAC: a new multi-class, multi-label associative classification approach. In: *IEEE International Conference on Data Mining (ICDM)*, IV., 2004, Brighton. *Proceedings*, p. 217-224. Institute of Electrical and Electronics Engineers (IEEE), 2004, p. 218.

²⁶⁶ “Uma revisão sistemática da literatura é um meio de avaliar e interpretar todas as pesquisas disponíveis relevantes para uma determinada questão de pesquisa, área de tópico ou fenômeno de interesse. As revisões sistemáticas visam apresentar uma avaliação justa de um tópico de pesquisa usando uma metodologia confiável, rigorosa e auditável.” (tradução livre). KITCHENHAM, Barbara; CHARTERS, Stuart. *Guidelines for performing Systematic Literature Reviews in Software Engineering*. *EBSE Technical Report EBSE-2007-01*. Keele University and University of Durham, version 2.3, 2007, p. 6.

²⁶⁷ Trata-se de um método desenhado em etapas, que utiliza o critério de impacto de revistas, citações de autores e artigos e frequência de palavras-chaves, aliado a um longo trabalho manual para escolha do material a ser utilizado. Como resultado, combinam bases de dados conceituadas, apresentando um conjunto de material

Artificial e Processamento de Linguagem Natural (técnicas de *text analytics*), para problemas de classificação *multi-label*²⁶⁸ em textos jurídicos.

Para tanto, contou-se com o apoio da ferramenta on-line projetada para auxiliar pesquisadores na condução de revisões sistemáticas da literatura, o *Parsifal*. Essa plataforma serve como um guia, contendo os passos para realização de estudos sistemáticos conforme diretrizes previamente descritas por autores de referência.²⁶⁹

O primeiro passo foi definir a questão de pesquisa que será respondida ao final dessa revisão sistemática: *qual é o estado da arte para as abordagens em classificação multi-label em textos jurídicos?*

Em seguida, a partir de uma análise exploratória prévia de estudos primários de interesse, definiu-se as palavras-chave mais usuais, bem como seus sinônimos, para construir o argumento de busca (*string*) padrão a ser utilizado nas diferentes bases de dados de literatura acadêmica: “((“*natural language processing*” AND “*multi-label*”) AND *classification*) AND (((*legal* OR *laws* OR *legislation*) AND *domain*) OR (*supreme* AND *court*) OR (*jurisprudence*))”. Ainda, como o foco consiste em descobrir o estado da arte do problema, definiu-se como escopo temporal trabalhos realizados a partir de 2015.

Os bancos de dados bibliográficos escolhidos para a pesquisa foram: *Google Scholars*, pela abrangência e representatividade linguística²⁷⁰; *Scopus*, *Science Direct* e *Web of Science*, por suas importâncias científicas²⁷¹; e *IEEE Xplore*, pela sua cobertura em tecnologia e ciência da computação²⁷².

Como cada mecanismo de busca nas retromencionadas bases de dados funciona de maneira distinta, precisou-se construir uma *string* a ser utilizada para cada base. Os argumentos de busca utilizados e a quantidade de resultados decorrentes encontrados em cada base de dados

confiável. Portanto, o enfoque meta-analítico visa oferecer uma técnica objetiva de escolha da literatura para respaldar o trabalho. Cf. MARIANO, A. M; ROCHA, M. S. Revisão da Literatura: Apresentação de uma Abordagem Integradora. In: *AEDM International Conference – Economy, Business and Uncertainty: Ideas for a European and Mediterranean Industrial Policy*, XXVI., 2017, Reggio Calabria. *Proceedings*, p. 427-443. Asociación Española de Dirección y Economía de la Empresa (AEDM), 2017, p. 433.

²⁶⁸ A atividade de classificação de textos pode envolver problemas restritos a apenas um rótulo por padrão (*multi-class*) ou vários rótulos associados simultaneamente a um mesmo documento (*multi-label*). Tendo em vista, como demonstrado pelas pesquisas de Franklin e

²⁶⁹ Disponível em: <https://parsif.al/>. Acesso em: 26 ago. 2020.

²⁷⁰ Cf. Mariano, A. M; Rocha, M. S. *Revisão da Literatura: Apresentação de uma Abordagem Integradora*, op. cit. p. 429.

²⁷¹ Cf. *ibid.*, p. 429.

²⁷² Cf. KLOCK, A. C. T. Mapeamentos e Revisões Sistemáticas da Literatura: um Guia Teórico e Prático. *Revista Cadernos de Informática*, Porto Alegre, v. 10, n. 1, 2018, p. 7. Disponível em: <https://seer.ufrgs.br/cadernosdeinformatica/article/view/v10n1201801-09/49901>. Acesso em: 02 set. 2018,

pesquisada encontra-se na Tabela 1. Ao todo, extraiu-se 430 trabalhos científicos das cinco bases de dados bibliográficas utilizadas.

Tabela 1 – Resumo dos argumentos de busca e resultados dos trabalhos científicos encontrados para as bases de busca bibliográficas escolhidas.

Banco de dados	Argumento de busca (<i>string</i>) utilizado	Resultados encontrados
<i>Google Scholars</i>	((("natural language processing" AND "multi-label") AND classification) AND (((legal OR laws OR legislation) AND domain) OR (supreme AND court) OR (jurisprudence))) Filtros aplicados: 2015 a 2021	161
<i>Scopus</i>	((("natural language processing" AND "multi-label") AND classification) AND (((legal OR laws OR legislation) AND domain) OR (supreme AND court) OR (jurisprudence)) AND (LIMIT-TO (SUBJAREA , "COMP") OR LIMIT-TO (SUBJAREA , "SOCI") OR LIMIT-TO (SUBJAREA , "MULT")) AND (LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015)))	119
<i>Science Direct</i>	<i>String</i> 01: (("natural language processing" AND "multi-label") AND classification) AND (((legal OR laws OR legislation) AND domain) OR (supreme AND court)) Filtros aplicados: 2015 a 2021 <i>String</i> 02: (("natural language processing" AND "multi-label") AND classification) AND (((legal OR laws OR legislation) AND domain) OR (jurisprudence)) Filtros aplicados: 2015 a 2021	115
<i>Web of Science</i>	ALL = (("natural language processing" AND "multi-label") AND classification) AND ALL = (((legal OR laws OR legislation) AND domain) OR (supreme AND court) OR (jurisprudence))	0
<i>IEEE Xplore</i>	((("Full Text & Metadata": "natural language processing" AND "Full Text & Metadata": "multi-label") AND "Full Text & Metadata": classification) AND ((("Full Text & Metadata": legal OR laws OR legislation) AND "Full Text & Metadata": domain) OR ("Full Text & Metadata": supreme AND "Full Text & Metadata": court) OR ("Full Text & Metadata": jurisprudence)))) Filtros aplicados: 2015 a 2021	35

Fonte: Elaborado pelo autor; consultas realizadas nos mecanismos de busca na data de 29 de agosto de 2020.

Os metadados – como, por exemplo, nome dos autores, nome do periódico publicado, palavras-chave, resumo, referências utilizadas etc – dos 430 trabalhos científicos, extraídos das cinco bases de dados bibliográficas utilizadas, foram importados para a plataforma *Parsifal*. Dentre eles, 116 eram duplicados, restando 314. Com o intuito de certificar a aderência dos

trabalhos extraídos à questão de pesquisa, obteve-se a nuvem de palavras dos seus resumos (*abstracts*), conforme ilustra a Figura 8.

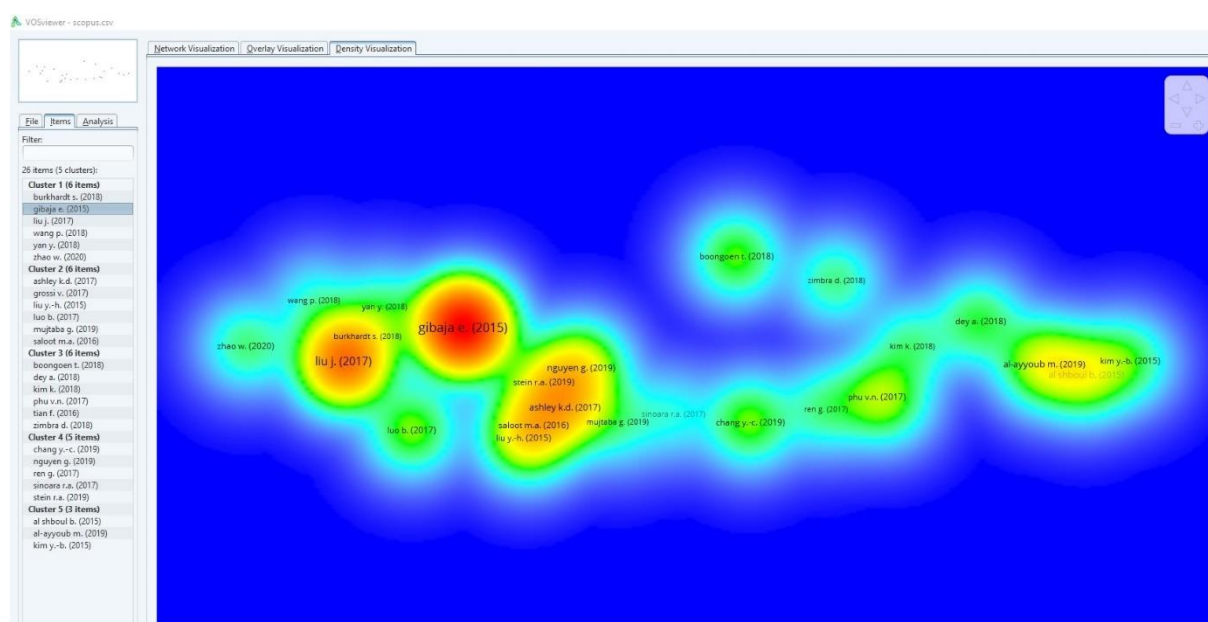
Figura 8 – Nuvem das 30 mais frequentes palavras dos resumos dos 314 trabalhos científicos únicos extraídos das cinco bases de dados bibliográficos utilizadas.



Fonte: Elaborado pelo autor com uso do website <https://tagcrowd.com/>

Ainda, os metadados obtidos da base de dados *Scopus* foram submetidos à análise de *coupling*²⁷³ (Figura 9), isto é, artigos que compartilham referências entre si, de modo a identificar as principais frentes de pesquisas dentre os trabalhos científicos extraídos dessa base.

Figura 9 – Análise de *coupling* dos 119 trabalhos científicos extraídos do *Scopus*.



Fonte: Elaborado pelo autor com uso do *software* VOSviewer.

²⁷³ Cf. Mariano, A. M.; Rocha, M. S. *Revisão da Literatura: Apresentação de uma Abordagem Integradora*, op. cit., p. 438. Por limitação dos metadados extraídos das diferentes bases acadêmicas, essa análise não pôde ser replicada para todos os bancos de dados utilizados.

Em seguida, a partir da leitura do título e resumo de cada um dos 314 trabalhos, foram aplicados a eles critérios de inclusão e exclusão.²⁷⁴ Os critérios de inclusão foram: ser um trabalho sobre classificação *multi-label*; ser um trabalho sobre classificação de textos jurídicos; ser um trabalho sobre classificação *multi-label* de textos jurídicos; conter aplicação das técnicas em textos grandes; ou pertencer a uma frente de pesquisa do *Scopus* (análise de *coupling*). Já os critérios de exclusão: não ser um documento válido ou um trabalho sobre classificação; ser um trabalho classificação *multi-label* em elementos não textuais; ser um trabalho de classificação não *multi-label* em textos não jurídicos; ou conter aplicação das técnicas em textos curtos.

Como resultado, obteve-se 103 trabalhos científicos incluídos e aptos, portanto, para a próxima etapa de avaliação da qualidade. Para essa próxima fase, foi necessário estabelecer os critérios de qualidades usados para considerar ou não o estudo, conforme a Tabela 2:

Tabela 2 – Critérios de qualidade aplicados aos 103 trabalhos científicos incluídos.

Critério de qualidade	Pontuação das possíveis repostas
1) O trabalho envolve classificação de textos jurídicos?	Sim = 1 ponto Parcialmente = 0,5 ponto Não = 0 ponto
2) O trabalho envolve classificação <i>multi-label</i> ?	Sim = 1 ponto Parcialmente = 0,5 ponto Não = 0 ponto
3) Trabalho utiliza <i>dataset</i> em português?	Sim = 1 ponto Parcialmente = 0,5 ponto Não = 0 ponto
4) Trabalho utiliza <i>dataset</i> com poucos exemplos ou possui código disponível?	Sim = 1 ponto Parcialmente = 0,5 ponto Não = 0 ponto
5) O trabalho é um <i>survey</i> ?	Sim = 1 ponto Parcialmente = 0,5 ponto Não = 0 ponto
6) Trabalho pertence aos mais citados (<i>índice do Google Metrics</i>)?	(num. citações >=100) = 1 ponto (40<= num. citações <99) = 0,6 ponto (39<= num. citações <0) = 0,3 ponto (num. citações = 0) = 0 ponto
7) Publicação em periódico entre os melhores (<i>índice h5 mediano</i>) no <i>Google Metrics</i> ?	(h5 mediano >=100) = 1 ponto (40<= h5 mediano <99) = 0,6 ponto (39<= h5 mediano <0) = 0,3 ponto (h5 mediano = 0) = 0 ponto

Fonte: Elaborado pelo autor.

²⁷⁴ Os critérios de inclusão definem uma característica que o estudo retornado pela busca deve possuir para continuar para as próximas da etapa do estudo sistemático. Já os critérios de exclusão definem uma característica que o estudo não deve possuir para que continue sendo considerado. Cf. Klock, *Mapeamentos e Revisões Sistemáticos da Literatura: um Guia Teórico e Prático*, op. cit., p. 5.

Os critérios de 1 a 4 procuram medir a aderência do trabalho à questão foco dessa revisão sistemática. A pontuação para trabalhos do tipo *survey*, destina-se a proporcionar a qualificação de pesquisas que apresentem uma visão geral sobre o tema pesquisado. Já as métricas de *citação* e *índice h5 mediano* (perguntas 6 e 7) visam indicar indiretamente a qualidade do trabalho científico em análise, com base em sua repercussão no meio científico e nível de rigor exigido para publicação, respectivamente.

Como a pontuação máxima de aderência aos critérios de qualidade foi de 7,0 pontos, considerou-se como ponto de corte trabalhos com pontuação maiores ou iguais a 3,0, resultando em 33 pesquisas qualificadas, sendo 31 artigos científicos, 01 dissertação de mestrado e 01 tese de doutorado. A Tabela 3 apresenta os trabalhos qualificados, em ordem alfabética de seus títulos.

Tabela 3 – Detalhe dos 33 trabalhos científicos qualificados após revisão sistemática.

ID	Autores	Título	Ano	Tipo do trabalho
1	Mironczuk M. M., Protasiewicz J.	A recent overview of the state-of-the-art elements of text classification	2018	Artigo Científico; teórico
2	Li Q., Peng H., Li J., Xia C., Yang, R., Sun, L., Yu, P. S., He, L.	A Survey on Text Classification: From Shallow to Deep Learning	2020	Artigo Científico; teórico
3	Gibaja E., Ventura S.	A tutorial on multilabel learning	2015	Artigo Científico; teórico
4	Salaün O., Langlais P., Lou A., Westermann H., Benyekhlef K.	Analysis and multilabel classification of quebec court decisions in the domain of housing law	2020	Artigo Científico; aplicado
5	Bao Q., Zan H., Gong P., Chen J., Xiao Y.	Charge Prediction with Legal Attention	2019	Artigo Científico; aplicado
6	Liu J., Chang W.-C., Wu Y., Yang Y.	Deep learning for extreme multi-label text classification	2017	Artigo Científico; aplicado
7	Bezerra D. A., Inazawa P. H. G., Zumblik R., Campos T. E., Silva N. C., Braz F. A., Peixoto F. H.	Discovery of terms that characterize legal pieces	2019	Artigo Científico; aplicado
8	Zhou P., El-Gohary N.	Domain-specific hierarchical text classification for supporting automated environmental compliance checking	2016	Artigo Científico; aplicado
9	Chalkidis I., Fergadiotis M., Malakasiotis P.,	Extreme Multi-Label Legal Text Classification: A case study in EU Legislation	2019	Artigo Científico; aplicado

	Aletras N., Androutsopoulos I.			
10	Wang Y., Yao Q., Kwok J.T., Ni L.M.	Generalizing from a Few Examples: A Survey on Few-shot Learning	2020	Artigo Científico; teórico
11	Zhao W., Gao H., Chen S., Wang N.	Generative Multi-Task Learning for Text Classification	2020	Artigo Científico; aplicado
12	Wang P., Fan Y., Niu S., Yang Z., Zhang Y., Guo J.	Hierarchical Matching Network for Crime Classification	2019	Artigo Científico; aplicado
13	Martija M. A., Domoguen J., Naval P.	How Deep is Your Law? Predicting Associations between Cases in Philippine Jurisprudence	2019	Artigo Científico; aplicado
14	Landthaler J.	Improving Semantic Search in the German Legal Domain	2020	Tese de Doutorado; aplicado
15	Sousa A. W., Fabro M. D. D.	Iudicium Textum Dataset A Base of Legal Texts for NLP	2019	Artigo Científico; aplicado
16	Xiao L., Huang X., Chen B., Jing L.	Label-specific document representation for multi-label text classification	2020	Artigo Científico; aplicado
17	Chalkidis I., Fergadiotis M., Malakasiotis P., Androutsopoulos I.	Large-scale multi-label text classification on EU legislation	2020	Artigo Científico; aplicado
18	Luo B., Feng Y., Xu J., Zhang X., Zhao D.	Learning to predict charges for criminal cases with legal basis	2017	Artigo Científico; aplicado
19	Howe J. S. T., Khang L. H., Chai I. E.	Legal Area Classification: A Comparative Study of Text Classifiers on Singapore Supreme Court Judgments	2019	Artigo Científico; aplicado
20	Shaffer R., Mayhew S.	Legal Linking: Citation Resolution and Suggestion in Constitutional Law	2019	Artigo Científico; aplicado
21	Peng D., Wu Q.	LegalCap: a model for complex case discrimination based on capsule neural network	2020	Artigo Científico; aplicado
22	Carvalho D. S., Nguyen M.-T., Tran C.-X., Nguyen, M.-L.	Lexical-Morphological Modeling for Legal Text Analysis	2017	Artigo Científico; aplicado
23	Nguyen G., Dlugolinsky S., Bobák M., Tran V., García Á. L., Heredia I., Malík P., Hluch L.	Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey	2019	Artigo Científico; teórico
24	Rocha A. C. P.	Mining of texts for classification of labor lawsuits	2019	Dissertação de Mestrado; aplicado
25	Wang P., Yang Z., Niu S., Zhang Y., Zhang L., Niu S.	Modeling dynamic pairwise attention for crime classification over legal articles	2018	Artigo Científico; aplicado

26	Dong H., Yang F., Wang X.	Multi-label charge predictions leveraging label co-occurrence in imbalanced data scenario	2020	Artigo Científico; aplicado
27	Alkhatib W., Rensing C., Silberbauer J.	Multi-label text classification using semantic features and dimensionality reduction with autoencoders	2017	Artigo Científico; aplicado
28	Chalkidis, I., Androutsopoulos, I., Aletras, N.	Neural legal judgment prediction in English	2020	Artigo Científico; aplicado
29	Liu Y.-H., Chen Y.-L., Ho W.-L.	Predicting associated statutes for legal problems	2015	Artigo Científico; aplicado
30	Polo F. M., Ciochetti I., Bertolo E.	Predicting Legal Proceedings Status: an Approach Based on Sequential Text Data	2020	Artigo Científico; aplicado
31	Xu D., Shi Y., Tsang I. W., Ong Y.-S., Gong C., Shen X.	Survey on Multi-Output Learning	2020	Artigo Científico; teórico
32	Kowsari K., Meimandi K. J., Heidarysafa M., Mendu S., Barnes L. E., Brown D. E.	Text Classification Algorithms: A Survey	2019	Artigo Científico; teórico
33	Araujo P. H. L., Campos T. E., Braz F. A., Silva N. C.	VICTOR: a Dataset for Brazilian Legal Documents Classification	2020	Artigo Científico; aplicado

Fonte: Elaborado pelo autor. Cada trabalho científico encontra-se devidamente referenciado em seção específica desta pesquisa.

Conforme indicado pela última coluna da Tabela 3, foi possível dividir os trabalhos qualificados em 2 categorias: aqueles que abordam a *teoria* sobre classificação de textos, sendo os ID's 1, 2, 3, 10, 23, 31 e 32; e os demais, que são estudos de caso onde modelos de classificação são *aplicados* em um conjunto de textos, para solução de problemas jurídicos.

Primeiramente, os sete trabalhos teóricos foram utilizados para trazer uma contextualização recente sobre a tarefa de classificação de textos. Em seguida, os resultados das pesquisas aplicadas foram integrados para responder ao questionamento inicial dessa revisão sistemática.

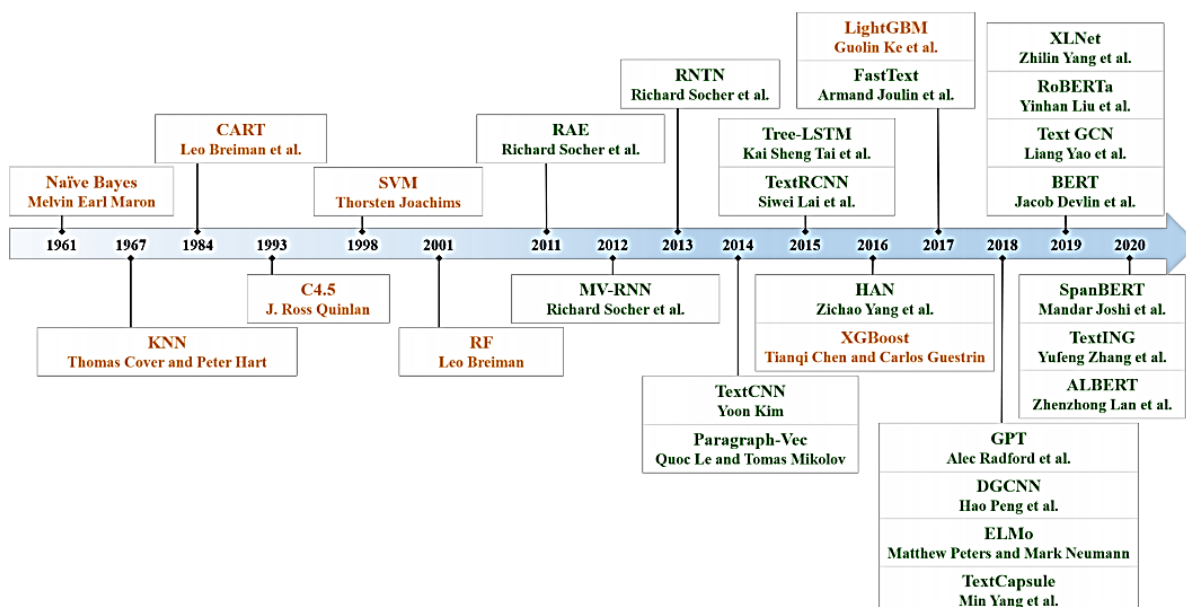
4.1.1 Contextualização sobre a atividade de classificação de textos

Classificação de textos é a atividade de rotular documentos com base na extração de informações de textos brutos, podendo ser separado em diferentes atividades como a análise de

sentimento (*sentiment analysis*), a rotulagem de tópicos (*topic labeling*), resposta a perguntas (*question answering*), recuperação de informação (*information retrieval*), dentre outros.²⁷⁵

Já os modelos, também chamados de métodos ou de algoritmos construídos para se atingir um resultado, podem ser classificados em modelos de *machine learning* e de *deep learning* (modelos de linguagem), tendo o primeiro predominado até a década de 2010, conforme ilustração esquemática apresentada na Figura 10.

Figura 10 – Ilustração esquemática dos métodos de classificação de texto de 1961 a 2020. Antes de 2010, quase todos os métodos existentes eram baseados em *machine learning* (cor laranja); desde 2010, a maioria dos trabalhos nesta área tem se concentrado em esquemas de *deep learning* (cor verde).



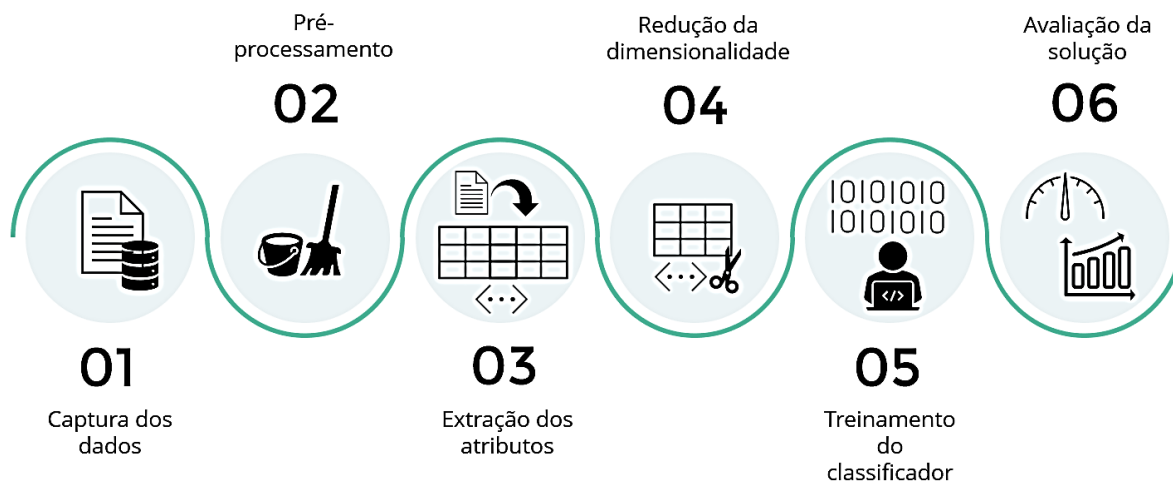
Fonte: Li, Q et. al. *A Survey on Text Classification: From Shallow to Deep Learning*, op. cit., p. 3.

Com algumas adequações, o processo de classificação de texto consiste nas seguintes fases: (1) captura dos dados, (2) pré-processamento, (3) extração dos atributos, (4) redução da dimensionalidade (5) treinamento do classificador e (6) avaliação da solução.²⁷⁶

Figura 11 – Fases do processo de classificação de texto.

²⁷⁵ Cf. LI, Q.; PENG, H.; LI, J.; XIA, C.; YANG, R.; SUN, L.; YU, P. S.; HE, L. A Survey on Text Classification: From Shallow to Deep Learning. *ACM Computing Surveys*, v. 37, n. 4, art.35, jul. 2020, p. 1.

²⁷⁶ Cf. Li, Q et. al. *A Survey on Text Classification: From Shallow to Deep Learning*, op. cit., p. 2;
Cf. MIRONCZUK, M. M.; PROTASIEWICZ, J. A recent overview of the state-of-the-art elements of text classification. *Expert Systems with Applications*, v. 106, n. 15, p. 36-54, set. 2018, p. 1-3.;
Cf. KOWSARI, K.; MEIMANDI, K. J.; HEIDARYSAFA, M.; MENDU, S.; BARNES, L. E.; BROWN, D. E. Text Classification Algorithms: A Survey. *Information*, v. 10, art. 150, 2019, p. 2.



Fonte: Elaborado pelo autor com base em: Li, Q et. al. *A Survey on Text Classification: From Shallow to Deep Learning*, op. cit., p. 2; Mironczuk, M. M.; Protasiewicz, J. *A recent overview of the state-of-the-art elements of text classification*, op. cit., p. 38; KOWSARI, K. et. al. *Text Classification Algorithms: A Survey*, op. cit., p. 2.

Mais detalhadamente, o processo de classificação se inicia com a captura dos dados, que deve conter o conjunto de textos (*corpus*²⁷⁷) que possuem a classificação desejada. Também deve ser coletado o *corpus* referente aos textos que se pretende classificar. Em seguida, esses *corpora* são pré-processados com intuito de promover uma limpeza no texto, removendo palavras indesejadas, por exemplo.

A extração de atributos é a fase em que o texto, que é um conjunto de dados não estruturados, é transformado em uma representação estruturada de *tokens*, isto é, significantes únicos que podem ser palavras ou conjunto de palavras. Ao final dessa fase, cada documento que compõe os *corpora* de pesquisa passa a ser representado por um vetor de *tokens*, ponderados por medidas relacionadas à frequência de ocorrência deles em cada documento.

Como cada documento pode ter diversas palavras únicas, pode ser necessário a redução de dimensionalidade dessas representações matriciais, de modo proporcionar melhor performance computacional (menor uso de memória de processamento).

Em seguida, após escolha do método de classificação a ser utilizado, seu algoritmo é utilizado para treinar a máquina a ser capaz de reconhecer as diferentes classes (também chamadas de rótulos, etiquetas ou *labels*). Uma vez treinado o modelo de classificação, é o momento de avaliar a solução por meio de indicadores de medição de desempenho. Quando um modelo de classificação é desenvolvido de forma adequada, ele consegue classificar dados de entrada por ele desconhecidos, que devem ser representados de maneira semelhante aos dados

²⁷⁷ “Um conjunto de textos é chamado de *corpus* - *corpus* significa simplesmente "corpo" em latim, e quando você possui várias dessas coleções de textos, é chamado de *corpora*”. Cf. Manning; Schutze, *Foundations of statistical natural language processing*, op. cit., p. 6.

utilizados na fase de treinamento. Consequentemente, o classificador produz uma decisão (tecnicamente, uma probabilidade) que define a classe de cada texto de entrada.

A atividade de classificação pode envolver problemas restritos a apenas um rótulo por padrão ou vários rótulos associados simultaneamente a um mesmo documento, sendo esse último chamado de *multi-label* (múltiplos rótulos). Resolver um problema de classificação *multi-label* envolve desafios como o crescimento exponencial de combinações de rótulos a serem levados em conta, bem como o custo computacional de construção e consulta dos modelos. Além disso, situações *multi-label* geralmente apresentam características como: *rótulos não balanceados*, ou seja, diferentes quantidades de documentos por classe, o que pode gerar viés no treinamento do modelo; *etiquetas correlacionadas entre si*, que podem diminuir a eficácia do modelo treinado; e *rótulos de saída não aplicáveis*, resultantes de classes novas que apareceram no conjunto de treinamento.²⁷⁸

Para solucionar a problemática de classes desbalanceadas, abordagens como a *Few-Shot Learning* (FSL) e a de *Zero-Shot Learning* (ZSL) foram desenvolvidas. A FSL incorpora conhecimentos prévios para promover o aprendizado pela máquina, mesmo contendo alguns exemplos rotulados de algumas classes. Já a ZSL proporciona solução adequada para a saída do modelo frente a rótulos para os quais ele não viu exemplares durante sua etapa de treinamento.²⁷⁹

Outra problemática recorrente da atualidade, presente nas tarefas de classificação de texto, é o amplo volume, variedade, velocidade e variedade de dados (*Big Data*), o que exige arquiteturas computacionais (*softwares* e *hardwares*) que permitam o tratamento eficiente para conjuntos de dados em grande escala. Como exemplo, cita-se os conjuntos de bibliotecas de algoritmos (*frameworks*) – a grande maioria desenvolvida em linguagem Python, a mais popular nesta área – associados ao uso de placas de vídeo dedicadas (Unidades de Processamento Gráfico – GPU): *Keras*, *PyTorch* e *TensorFlow*.²⁸⁰

4.1.2 Descobertas sobre o estado da arte em classificação *multi-label* de textos jurídicos

²⁷⁸ Cf. GIBAJA, E.; VENTURA, S. A tutorial on multilabel learning. *ACM Computing Surveys*, v. 47, n. 3, artigo 52, abr. 2015, p. 1-3; XU, D.; SHI, Y.; TSANG, I. W.; ONG, Y.-S.; GONG, C.; SHEN, X. A Survey on Multi-output Learning. *IEEE Transactions on Neural Networks and Learning Systems*, v. 31, n. 7, p. 2409-2429, jul. 2020, p. 5.

²⁷⁹ Cf. WANG, Y.; YAO, Q.; KWOK, J.T.; NI, L.M. Generalizing from a Few Examples: A Survey on Few-shot Learning. *ACM Computing Surveys*, v. 53, n. 3, artigo 63, jun. 2020, p. 1; 27-28.

²⁸⁰ Cf. NGUYEN, G.; DLUGOLINSKY, S.; BOBÁK, M.; TRAN, V.; GARCÍA, Á. L.; HEREDIA, I.; MALÍK, P.; HLUCH, L. Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey. *Artificial Intelligence Review*, v. 52, n. 1, jun. 2019, p. 78-80; 85; 97-99.

Os 26 trabalhos científicos considerados como estudos práticos (Tabela 3) tiveram seus resultados integrados a fim de trazer luz ao questionamento inicial dessa revisão sistemática, qual seja: *qual é o estado da arte para as abordagens em classificação multi-label em textos jurídicos?*

A **Tabela 4** apresenta o objetivo resumido, a base de dados utilizada (*dataset*), as técnicas de pré-processamento, os modelos e algoritmos utilizados, bem como os resultados de mensuração encontrados. Foi possível agrupar os trabalhos em cinco macro objetivos:

1. Classificar processos jurídicos;
2. Criar um corpus jurídico em Língua Portuguesa para pesquisas de IA em Direito;
3. Melhorar pesquisas jurisprudenciais;
4. Predição de decisões;
5. Testar *deep learning* para classificação *multi-label*.

Dentre os trabalhos detalhados, destacam-se aqueles que possuem como macro objetivo “melhorar pesquisas jurisprudenciais”, pois se relacionam intimamente com o objetivo da presente pesquisa. Discorrer-se-á brevemente sobre eles (ID’s da Tabela 3: 13, 14, 19, 20, 22 e 29).

O artigo científico ID 13 da Tabela 3, intitulado “*How Deep is Your Law? Predicting Associations Between Cases in Philippine Jurisprudence*” buscou resolver o problema de prever decisões relevantes da Suprema Corte filipina. Para tanto, dividiu-se o problema em duas abordagens: (i) identificar as associações de leis relevantes para um determinado caso (*case-to-law*); e (ii) associar casos a outras jurisprudências, isto é, dado um caso e sua exposição de fatos, encontrar decisões de casos que estão intimamente relacionados a ele (*case-to-case*). Os estudos concentraram-se em casos criminais de 1987 a 2013 (coleção *Philippine Jurisprudence* do *LawPhil Project* da *Arrellano Law Foundation*). Em ambos os enfoques do problema, partiu-se do critério de citação, seja a uma determinada legislação ou a uma outra decisão para distribuir os rótulos e, a partir disso, identificar casos análogos por meio de classificação de textos. Os autores afirmam que resultados encontrados possuem grande espaço para melhorias, uma vez que a precisão e o *recall* não foram satisfatórios, sobretudo para o enfoque *case-to-case* (ver **Tabela 4**). Entretanto, concluem pela importância da pesquisa para o incentivo a pesquisas semelhantes, que visam solucionar o problema de encontrar precedentes úteis para a solução de um caso.²⁸¹

²⁸¹ Cf. MARTIJA, M. A.; DOMOGUEN, J.; NAVAL, P. How Deep is Your Law? Predicting Associations between Cases in Philippine Jurisprudence. In: *IEEE Region 10 Conference (TENCON)*, 2019, Singapore. *Proceedings*, p. 886-891. Institute of Electrical and Electronics Engineers (IEEE), 2019.

O trabalho de ID 14 da Tabela 3, “*Improving Semantic Search in the German Legal Domain with Word Embeddings*”, corresponde à tese de doutorado de Jörg Landthaler. Dentre as diversas abordagens empíricas realizadas pelo autor, a de correspondência semântica de textos (*semantic text matching*) consiste em identificar segmentos de texto de vários tipos de documentos que estão semanticamente vinculados. Utilizou-se de um *corpus* contendo informações condensadas de leis e decisões judiciais alemãs sobre locação para treinar um sistema que, a partir da inserção segmentos de texto arbitrários pelos usuários, aplica os algoritmos testados e retorna trechos de normativos e decisões que guardam relação com aquele trecho inserido pelo usuário. Os melhores resultados foram com algoritmos *FastTest* com a abordagem de quebra dos textos por sentença e indicam que os termos do domínio jurídico são importantes para o treinamento do algoritmo. Por fim, o estudo sugere que a busca por seleção por trechos pode ser útil como um método de busca complementar à busca por palavras-chave e auxiliar juristas na pesquisa de legislação correlata e jurisprudencial.²⁸²

O artigo científico de ID 19 da Tabela 3, cujo título é “*Legal Area Classification: A Comparative Study of Text Classifiers on Singapore Supreme Court Judgments*”, usa um conjunto de dados de julgamentos da Suprema Corte de Cingapura para estudar comparativamente o desempenho de várias abordagens de classificação de documentos aplicados a *corpora* jurídicos de textos longos e com poucos documentos rotulados (apenas 588). O resultado obtido surpreendeu os autores, uma vez que, os modelos estatísticos tradicionais (*machine learning*) superaram modelos de linguagem (*deep learning*) pré-treinados em textos não jurídicos. Concluem pela indicação de duas frentes de pesquisas futuras promissoras: (i) criação de conjuntos de dados específicos de leis e linhas de base para treinamento e *benchmarking* de classificadores de textos jurídicos e (ii) exploração de técnicas de aprendizagem de representação que alavancam métodos de aprendizagem por transferência, mas escalam bem em textos longos.²⁸³

O artigo de ID 20 da Tabela 3, nomeado “*Legal Linking: Citation Resolution and Suggestion in Constitutional Law*”, apresenta um método para vincular parágrafos de casos da Suprema Corte com seções da Constituição dos Estados Unidos da América. Os resultados obtidos com o modelo de rede neural superaram os dos modelos clássicos testados. O modelo opera no nível de trechos curtos de texto, como parágrafos, e dá uma decisão binária entre pares

²⁸² Cf. LANDTHALER, J. *Improving Semantic Search in the German Legal Domain*. Tese de Doutorado (doutorado em Ciências Naturais) – Technische Universität München, Munique, Alemanha, 2020, p. 97-134.

²⁸³ Cf. HOWE, J. S. T.; KHANG, L. H.; CHAI, I. E. *Legal Area Classification: A Comparative Study of Text Classifiers on Singapore Supreme Court Judgments*. In: *Natural Legal Language Processing Workshop (NLLP)*, I., 2019, Minneapolis. *Proceedings*, p. 67-77. Association for Computational Linguistics (ACL), 2019.

de textos, marcando a presença ou ausência de uma relação, mesmo quando não há referência direta à emenda constitucional no parágrafo pesquisado.²⁸⁴

O artigo científico de ID 29 da Tabela 3, intitulado “*Predicting associated statutes for legal problems*”, apresenta o método denominado TPP (*Three Phase Prediction*) focado na identificação de leis e regulamentos relacionados a um problema especificado. O desempenho do método TPP foi comparado a outros três algoritmos de recuperação de informação e apresentou performance superior. Os autores concluem que a abordagem TPP apresentou-se como inovadora para problemas de previsão de estatutos legais.²⁸⁵ Esse modelo, inclusive, foi considerado como paradigma para o trabalho de ID 22 da Tabela 3, relatado abaixo.

Por fim, o trabalho “*Lexical-Morphological Modeling for Legal Text Analysis*” (ID 22 da Tabela 3), no contexto de uma competição sobre extração e registro de informações jurídicas (*Competition on Legal Information Extraction/Entailment – COLIEE*) e a partir do Código Civil japonês traduzido para o inglês, propôs um modelo que compreende as etapas necessárias para encontrar os documentos relevantes a uma questão jurídica e decidir sobre as evidências de vinculação textual para fornecer uma resposta correta. A principal contribuição desse artigo é a adaptação aprimorada de um método de classificação textual, baseado em conjuntos de aprendizado de máquina (*Adaboost*), incluindo um recurso de similaridade semântica (*Word2Vec*), que resultou em um método mais simples quando comparado ao TPP²⁸⁶ e que opera com consideravelmente menos dados de treinamento. Os autores consideram o modelo criado como competitivo com abordagens de última geração em recuperação de informações legais (*information retrieval*) e resposta a perguntas (*question answering*).²⁸⁷

Ademais, as técnicas de pré-processamento dos textos, os algoritmos e as medidas de mensuração utilizados por todos esses 26 trabalhos (ver Tabela 4) serão tomados como referência para direcionamento das abordagens empíricas adotadas por esta pesquisa.

²⁸⁴ Cf. SHAFFER, R.; MAYHEW, S. Legal Linking: Citation Resolution and Suggestion in Constitutional Law. In: *Natural Legal Language Processing Workshop (NLLP)*, I., 2019, Minneapolis. *Proceedings*, p. 39-44. Association for Computational Linguistics (ACL), 2019.

²⁸⁵ Cf. LIU, Y.-H.; CHEN, Y.-L.; HO, W.-L. Predicting associated statutes for legal problems. *Information Processing & Management*, v. 51, n. 1, p. 194-211, 2015.

²⁸⁶ Cf. LIU, Y.-H.; CHEN, Y.-L.; HO, W.-L., *Predicting associated statutes for legal problems*, op. cit., p. 194-211.

²⁸⁷ Cf. CARVALHO, D. S.; NGUYEN, M.-T.; TRAN, C.-X.; NGUYEN, M.-L. Lexical-Morphological Modeling for Legal Text Analysis. In: *JSAI International Symposium on Artificial Intelligence (JSAI-isAI)*, 2015, Tsukuba. *Lecture Notes in Computer Science*, v. 10091, p. 295-311. Springer, 2017.

Tabela 4 – Integração dos resultados dos 26 trabalhos científicos de estudos práticos em classificação *multi-label* de textos jurídicos.

ID da Tab. 3	Objetivo	Base de dados utilizada (dataset)	Pré-processamento	Modelo	Algoritmo	Medidas
4	Predição de decisões com uso de classificação <i>multi-label</i>	Decisões na língua francesa da <i>Régie du Logement du Québec</i> , relativas à legislação habitacional: total de 667.305 textos com um comprimento médio de 363 tokens	Não informado.	<i>Deep learning</i>	<i>FlauBERT</i> , uma variante de <i>BERT</i> pré-treinada em francês	1) Cenário "senhorio versus inquilino": F1 score (micro average) = 93.7%; Acurácia = 85.2%; 2) Cenário "inquilino versus senhorio": F1 score (micro average) = 84.9%; Acurácia = 74.6%;
5	Predição de decisões com uso de classificação <i>multi-label</i>	Conjunto de dados jurídicos chineses <i>CAIL2018</i> , que contém mais de 2,6 milhões de processos criminais, com 202 acusações criminais e 183 artigos legais relevantes relacionados. Foram escolhidos aleatoriamente 203.823 casos para treinamento, 20.000 para validação e 40.000 para teste	Não informado	<i>Deep learning</i>	<i>LegalAtt</i> (attention-based neural model). Código: https://github.com/nlp208/legal_attention	1) Extração de artigo legal: F1 score (micro) = 80,3%; F1 score (macro) = 78,7%; 2) Predição de acusação penal: F1 score (micro) = 81,0%; F1 score (macro) = 77,4%;
6	Testar <i>deep learning</i> para classificação <i>multi-label</i> extrema	<i>EUR-Lex</i> , que inclui tratados, legislação e jurisprudência União Europeia	Não informado	<i>Deep learning</i>	<i>SLEEC</i>	P@1 = 78,21%; P@2 = 64,33%; P@3 = 52,47%; G@1 = 78,21%; G@2 = 67,86%; G@3 = 61,57%;
7	Criar um corpus jurídico em Língua Portuguesa para	Conjunto de dados do <i>Projeto Victor</i> , sendo 70% dos registros	1) Remoção de caracteres especiais, tais como #, @, \$, e afins;	<i>Machine learning</i>	<i>Naive Bayes</i> (com χ^2)	F1 score = 96%;

ID da Tab. 3	Objetivo	Base de dados utilizada (dataset)	Pré-processamento	Modelo	Algoritmo	Medidas
	pesquisas de IA em Direito – encontrar as palavras mais relevantes em um <i>corpus</i> jurídico em português	para treino, 10% para teste e 20% para validação	2) Remoção de termos alfanuméricos, isto é, com números e letras mesclados; 3) Transformação dos termos que sejam e-mails ou links nos símbolos léxicos “EMAIL” e “LINK”; 4) Transformação dos termos referentes a números de leis e artigos nos símbolos léxicos “LEI_X” e “ARTIGO_X”, onde o X representa a respectiva lei ou artigo citado;			
8	Classificar processos jurídicos – verificação automática de conformidade ambiental de projetos, de acordo com cláusulas de documentos regulatórios	1.200 cláusulas foram coletadas de 10 documentos regulatórios ambientais, como o Código Internacional de Conservação de Energia de 2012, e foram classificadas em 10 subtópicos do tópico eficiência energética (uma subclasse do tópico ambiental na hierarquia de tópicos semânticos)	1) Remoção de <i>stopwords</i> específicas para o domínio;	<i>Machine learning</i>	<i>Support Vector Machine (SVM)</i>	Precisão = 84,3%; Recall = 97,3%;
9	Testar <i>deep learning</i> para classificação <i>multi-label</i> extrema	<i>EURLEX57K</i> , um subconjunto de 57.000 documentos selecionados da <i>EUR-Lex</i>	Não informado	<i>Deep learning</i>	<i>BIGRU-LWAN</i>	RP@5 = 76,6%; nDCG@5 = 79,6%; F1 score (micro) = 69,8%;

ID da Tab. 3	Objetivo	Base de dados utilizada (<i>dataset</i>)	Pré-processamento	Modelo	Algoritmo	Medidas
11	Testar <i>deep learning</i> para classificação <i>multi-label</i>	Conjunto de textos de casos policiais. Foram utilizados 127.142 textos de casos policiais de janeiro a agosto de 2018	Não informado	<i>Deep learning</i>	<i>Generative Multitask Learning based Classification (GMLC)</i>	Acurácia = 95,9%; Precisão = 94,0%; F1 score (micro) = 92,3%;
12	Predição de decisões com uso de classificação <i>multi-label</i>	1) <i>Fraud and Civil Action</i> : 40.256 processos criminais relacionados com fraude e ações civis diversas. Dados coletados do <i>China Judgment Online</i> (jan. a jun. 2016); 2) <i>CAIL</i> : contém processos criminais publicados pelo Supremo Tribunal Popular Chinês. Possui descrição do fato e o resultado do julgamento correspondente (incluindo leis, artigos e penalidades)	Não informado	<i>Deep learning</i>	<i>Hierarchical Matching Network (HMN)</i>	1) <i>Fraud and Civil Action</i> : Precisão (macro) = 73,1%; Recall (macro) = 37,1%; F1 score (macro) = 45,2%; Jaccard = 68,9%; 2) <i>CAIL</i> : Precisão (macro) = 80,9%; Recall (macro) = 61,9%; F1 score (macro) = 66,5%; Jaccard = 77,4%;
13	Melhorar pesquisas jurisprudenciais – encontrar decisões semelhantes	Casos criminais de 1987 a 2013, extraídos do repositório da Suprema Corte filipina (<i>LawPhil Project da Arellano Law Foundation</i>)	Ocultação de seções no texto que mencionam qualquer um dos rótulos	<i>Machine learning e Deep learning</i>	1) <i>case-to-law</i> : <i>Convolutional Neural Network (CNN)</i> 2) <i>case-to-case</i> : <i>Doc2Vec</i> ; <i>Convolutional Neural Network (CNN)</i>	1) <i>case-to-law (top 25 labels)</i> : Precisão = 71.43%; Recall = 62.81%; F1 score = 66.84%; 1) <i>case-to-case (top 100 labels)</i> : <i>Doc2Vec</i> : Recall = 50%; <i>CNN</i> : Precisão = 41,91%; Recall = 32,59%;

ID da Tab. 3	Objetivo	Base de dados utilizada (dataset)	Pré-processamento	Modelo	Algoritmo	Medidas
						F1 score = 36,67%;
14	Melhorar pesquisas jurisprudenciais – instituir uma "busca por seleção", em que um trecho selecionado é utilizado para encontrar situações semelhantes na legislação (<i>semantic text matching</i>)	Seis modelos alemães de contratos de locação e três conjuntos de comentários sobre a lei de locação alemã	1) Remoção de <i>tags</i> XML do <i>dataset</i> de comentários de texto legal; 2) Substituição de nova linha; 3) <i>Tokenização</i> de espaço em branco; 4) Substituição do símbolo "&"; 5) Remoção de caracteres especiais (exceto hífen); 6) Remoção de token de um e dois caracteres; 7) Transformação em letras minúsculas; 8) Remoção de <i>stopwords</i> ; 9) <i>Stemming</i> ;	<i>Deep learning</i>	<i>FastText</i> com segmentação em sentenças	Precisão = 63%; <i>Recall</i> = 0,17%;
15	Criar um corpus jurídico em Língua Portuguesa para pesquisas de IA em Direito	<i>Iudicium Textum Dataset</i> (IDT), uma base de textos jurídicos em Língua Portuguesa composta por documentos dos acórdãos do Supremo Tribunal Federal (http://dadosabertos.c3sl.ufpr.br/acordaos/tools/extract_txt_html.sh)	1) Limpeza de dados de assinatura digital, bem como dos elementos do cabeçalho e rodapé dos documentos;	Não aplicável	Não aplicável	Não aplicável
16	Testar <i>deep learning</i> para classificação <i>multi-label</i>	<i>EUR-Lex</i> , que inclui tratados, legislação e jurisprudência União Europeia	Não informado	<i>Deep learning</i>	<i>Label Specific Attention Network</i> (LSAN)	P@1 = 79,17%; P@3 = 64,99%; P@5 = 53,67%;

ID da Tab. 3	Objetivo	Base de dados utilizada (dataset)	Pré-processamento	Modelo	Algoritmo	Medidas
					https://github.com/EMNLP2019LSAN/LSAN/	nDCG@3 = 68,32%; nDCG@5 = 62,47%;
17	Testar <i>deep learning</i> para classificação <i>multi-label</i> extrema	EURLEX57K, um subconjunto de 57.000 documentos selecionados da <i>EUR-Lex</i>	Não informado	Deep learning	BIGRU-LWAN	RP@5 = 76,6%; nDCG@5 = 79,6%; F1 score (micro) = 69,8%;
18	Predição de decisões com uso de classificação <i>multi-label</i>	Documentos de julgamento desde 2013, coletados do <i>China Judgments Online</i> (http://wenshu.court.gov.cn). Foram escolhidos aleatoriamente 50.000 documentos para treinamento, 5.000 para validação e 5.000 para teste	Não informado	Deep learning	Neural Network Fact Gold Art	Precisão (micro) = 98,78%; Precisão (macro) = 95,26%; Recall (Micro) = 98,24%; Recall (Macro) = 95,57%; F1 score (micro) = 98,51%; F1 score (macro) = 95,42%;
19	Melhorar pesquisas jurisprudenciais – classificação de decisões jurídicas em áreas, com ênfase em classes desbalanceadas	6.227 sentenças da Suprema Corte de Singapura, escritas em inglês e emitidas entre 3 de janeiro de 2000 e 18 de fevereiro de 2019. Foram baixadas de http://www.singaporelawwatch.sg	Não informado	Machine learning	Lsac, que usa a abordagem <i>one-vs-rest</i> para <i>Linear Support Vector Machines</i> (linSVMs) e <i>Latent Semantic Analysis</i> (LSA)	Precisão (micro) = 81,7%; Precisão (macro) = 83,4%; F1 score (micro) = 73,3%; F1 score (macro) = 63,2%; Recall (micro) = 66,5%; Recall (macro) = 57,8%;
20	Melhorar pesquisas jurisprudenciais – inferir ligações entre parágrafos de casos da Suprema Corte e seções da Constituição	Repositório do <i>Cornell Legal Information Institute</i> (Cornell LII)	Não informado	Deep learning	Neural Network (stripped) https://github.com/mayhewsw/legal-linking	Precisão = 76,5%; Recall = 56,2%; F1 score = 64,8%;

ID da Tab. 3	Objetivo	Base de dados utilizada (<i>dataset</i>)	Pré-processamento	Modelo	Algoritmo	Medidas
21	Predição de decisões com uso de classificação <i>multi-label</i>	<i>CAIL</i> : contém processos criminais publicados pelo Supremo Tribunal Popular Chinês. Possui descrição do fato e o resultado do julgamento correspondente (incluindo leis, artigos e penalidades)	Não informado	<i>Deep learning</i>	<i>LegalCap</i> (https://github.com/Ezra521/LegalCap.git)	F1 score (micro) = 87,8%; F1 score (macro) = 73,3%;
22	Melhorar pesquisas jurisprudenciais – encontrar textos legais associados a questões jurídicas	Código Civil japonês	1) <i>Tokenização</i> e <i>POS-tagging</i> ; 2) Remoção das palavras irrelevantes: determinantes, conjunções, preposições e pontuação; 3) <i>Lematização</i> ;	<i>Machine learning</i> e <i>deep learning</i>	<i>AdaBoost-DecSt e R2NC</i> (<i>Word2Vec</i>)	Acurácia = 58,2%;
24	Classificar processos jurídicos – Classificação de processos trabalhistas por assuntos com uso de classificação <i>multi-label</i>	Processos trabalhistas dos 24 Tribunais Regionais do Trabalho (TRTs)	1) Redução da dimensionalidade com limpeza e normalização do conteúdo; 2) Remoção das marcações HTML; 3) Remoção de acentos e caracteres especiais; 4) Remoção de números e <i>stopwords</i> ; 5) Transformação em minúsculas;	<i>Machine learning</i>	<i>Support Vector Machine</i> (SVM)	Acurácia (micro) = 60,42%; Precisão (micro) = 73,63%; F1 score (micro) = 55,72%;
25	Predição de decisões com uso de classificação <i>multi-label</i>	40.256 documentos de julgamento relacionados com o Crime de Fraude e Ação Civil durante o	Não informado	<i>Deep learning</i>	<i>Dynamic Pairwise Attention Model</i> (DPAM)	Precisão (macro) = 60,4%; <i>Recall</i> (macro) = 37,7%; F1 score (macro) = 43,3%; Jaccard = 77,2%;

ID da Tab. 3	Objetivo	Base de dados utilizada (dataset)	Pré-processamento	Modelo	Algoritmo	Medidas
		período de janeiro a junho de 2016			https://github.com/yangze01/DPAM	
26	Predição de decisões com uso de classificação <i>multi-label</i> (ênfase em classes desbalanceadas)	<i>CAIL</i> : contém processos criminais publicados pelo Supremo Tribunal Popular Chinês. Possui descrição do fato e o resultado do julgamento correspondente (incluindo leis, artigos e penalidades)	1) <i>Uppersampling</i> and <i>Undersampling</i> 2) Esquema de peso de instância é introduzido para aliviar ainda mais o problema de dados desequilibrados	<i>Deep learning</i>	CBLLC (CNN + BiLSTM)	1) Predição de penalidade: F1 score (macro) = 92,9%; 2) Predição de artigo relacionado: F1 score (macro) = 86,6%;
27	Classificar processos jurídicos – incorporação de análise semântica na extração e seleção de atributos para classificação <i>multi-label</i>	<i>EUR-Lex</i> , que inclui tratados, legislação e jurisprudência União Europeia	1) <i>Tokenização</i> e <i>POS-tagging</i> ; 2) Extração de frase nominal usando filtro linguístico; 3) Remoção de <i>stopwords</i> ; 4) Diminuição de atributos com <i>auto-encoders</i> ;	<i>Machine learning</i>	<i>Multi-label k Nearest Neighbors</i> (ML-KNN)	Acurácia = 36%; F1 score (micro) = 67%; F1 score (macro) = 31%; Precisão = 73%;
28	Predição de decisões com uso de classificação <i>multi-label</i>	Subconjunto extraído da <i>European Court of Human Rights</i> (ECHR), contendo aproximadamente 11,5 mil casos do banco de dados público. https://archive.org/details/ECHR-ACL2019	Não informado	<i>Deep learning</i>	HIER-BERT	Precisão = 65,9% <i>Recall</i> = 55,1% F1 score = 60,0%
29	Melhorar pesquisas jurisprudenciais – encontrar textos legais associados a questões jurídicas	Julgamentos criminais de Taiwan armazenados no Sistema de recuperação de leis e regulamentos. No total, o	1) Remoção de <i>stopwords</i> (chinês); 2) <i>POS-tagging</i> ;	<i>Machine learning</i>	<i>libSVM</i>	Coverage = 91,7%

ID da Tab. 3	Objetivo	Base de dados utilizada (<i>dataset</i>)	Pré-processamento	Modelo	Algoritmo	Medidas
		conjunto de dados contém 1518 julgamentos criminais diferentes				
30	Classificar processos jurídicos – classificação de processos judiciais por situação de andamento com uso de classificação <i>multi-label</i>	Amostras aleatórias de processos dos Tribunais de Justiça do Estado de São Paulo e do Estado Rio de Janeiro	1) Conversão de maiúsculas para minúsculas; 2) Remoção de <i>stopwords</i> ; 3) Remoção de ruído e padronização de expressões;	<i>Deep learning</i>	<i>Recurrent Neural Network</i> (RNN) com <i>Long Short-Term Memory units</i> (LSTM) https://bit.ly/36yJZY3	F1 score (micro) = 93%; Precisão (micro) = 93%; Recall (micro) = 93%; F1 score (macro) = 89%; Precisão (macro) = 92%; Recall (macro) = 87%;
33	Criar um corpus jurídico em Língua Portuguesa para pesquisas de IA em Direito; Classificar processos jurídicos – classificar documentos e assuntos processuais	Conjunto de dados <i>VICTOR</i> , composto por 45.532 Recursos Extraordinários do Supremo Tribunal Federal (STF). Cada ação, por sua vez, contém diversos documentos distintos, que vão desde o próprio recurso até certidões e decisões, totalizando 692.966 documentos com 4.603.784 páginas	1) <i>Stemming</i> ; 2) Remoção de <i>stopwords</i> ; 3) Transformação em minúsculas; 4) <i>Tokenização</i> de e-mails e <i>URLs</i> e <i>tokenização</i> específica de artigos de lei (Ex.: Lei 11.419 para LEI_11419);	<i>Deep learning</i>	<i>XGBoost</i>	F1 score (average) = 88,82%

Fonte: Elaborado pelo autor.

4.2 Procedimentos empíricos: aplicação das fases do processo de classificação de texto ao problema da pesquisa

Uma pesquisa envolvendo IA e Direito, como a que se apresenta nesta dissertação, impõe a realização de um trabalho integrado por equipes de tecnologia e de Direito.²⁸⁸ Inicialmente, contou-se com a colaboração informal do mestrando em Computação Aplicada pela Universidade de Brasília, Maurício Barros de Jesus²⁸⁹, que trabalhou em conjunto com este pesquisador com vistas a esboçar os primeiros resultados preliminares.

Posteriormente, formalizou-se a realização de Estágio em Pesquisa em Nível de Mestrado com o mestrando Wesley Modanez Freitas²⁹⁰, do Programa de Pós-graduação em Ciência da Computação (PPGCC) vinculado ao Instituto de Informática da Universidade Federal de Goiás, e cuja pesquisa envolve classificação de textos jurídicos. Assim, em uma colaboração sinérgica, o mestrando do PPGCC atuou especialista em tecnologia, enquanto o pesquisador do Programa de Pós-graduação em Direito e Políticas Públicas (PPGDP) – no âmbito do qual esta pesquisa foi desenvolvida – atuou como analista de negócio.

A seguir, serão descritos os procedimentos realizados no âmbito de cada fase do processo de classificação de texto, conforme anteriormente representado na Figura 11²⁹¹. Os *scripts* (conjunto de algoritmos) foram construídos em linguagem *Python*, utilizando os *frameworks*²⁹² *Scikit-Learn*²⁹³ para modelos de *machine learning* e *TensorFlow*²⁹⁴ para os de *deep learning*. Esses *scripts* podem ser acessados em <https://github.com/wmodanez/acordaos>, e estão disponíveis nos Apêndices B, C, D, E, F e G.

Para o processamento dos dados, foi utilizado um computador tipo notebook com as seguintes configurações: processador i7 de sétima geração, 16 GB de memória RAM, 480 GB SSD, placa de vídeo GTX 1070 com 8GB de memória e sistema operacional Linux.

4.2.1 Captura dos dados

²⁸⁸ Cf. Peixoto; Silva, *Inteligência Artificial e Direito*, op. cit., p. 136.

²⁸⁹ Currículo Lattes disponível em: <http://lattes.cnpq.br/0131709712990645>. Acesso em 01 nov. 2020.

²⁹⁰ Currículo Lattes disponível em: <http://lattes.cnpq.br/5149778767750720>. Acesso em 01 nov. 2020.

²⁹¹ Relembrando, as fases do processo de classificação de texto consiste nas seguintes fases: (1) captura dos dados, (2) pré-processamento, (3) extração dos atributos, (4) redução da dimensionalidade (5) treinamento do classificador e (6) avaliação da solução.

²⁹² Frameworks são bibliotecas de algoritmos de código aberto para desenvolvimento de aplicações de IA.

²⁹³ Disponível em: <https://scikit-learn.org/stable/index.html>. Acesso em 11 nov. 2020.

²⁹⁴ Disponível em: <https://www.tensorflow.org/?hl=pt>. Acesso em 11 nov. 2020.

Quanto à natureza (tipologia) dos dados, utilizou-se de dados primários,²⁹⁵ quais sejam, todas as decisões daquele Tribunal. Para obtenção dessas decisões, rastreou-se duas bases de dados disponíveis:

- i. Uma base de dados obtida em mídia digital pela ex-pesquisadora do PPGDP Fernanda de Moura Ribeiro Naves, por meio de consulta formal à Ouvidoria do TCU, conforme descrito no Anexo B de sua dissertação de mestrado.²⁹⁶ Segundo a Corte de Contas, o arquivo coletado possui todas as decisões colegiadas do TCU entre 1992 e 17/01/2018, bem como todos os atributos presentes na base de Acórdãos da instituição, explicitados no Manual Completo de Pesquisa de Jurisprudência do TCU²⁹⁷; ou
- ii. Uma base de dados pública disponível para download na plataforma *Kaggle* do perfil do usuário José Ferraz Neto,²⁹⁸ que realizou a coleta por intermédio de *web crawling*²⁹⁹, isto é, um processo algorítmico iterativo que automatiza a busca e salvamento de informações em sítios eletrônicos. Em teoria, esta base diz respeito aos acórdãos proferidos pelo TCU entre os anos de 1992 até 30/08/2019, sendo que o conteúdo está armazenado em um arquivo *sqlite3* (.db) e possui aproximadamente 3.8GB.³⁰⁰

Considerando o período compreendido pelas duas bases de dados, era esperado que (ii) contivesse mais decisões colegiadas que (i). Entretanto, em comparação, aferiu-se que a base (i) possui 309.806 acórdãos, enquanto a base (ii) contém 298.942 acórdãos. Ademais, constatou-se que a base (ii) armazenou as informações dos atributos de texto sem a separação entre parágrafos, o que pode comprometer os treinamentos dos modelos de IA, a depender do algoritmo escolhido. Por essas razões, optou-se pela utilização da base de dados (i) que, mesmo abarcando um período menor, aparentou mais confiabilidade e qualidade em seus registros.

²⁹⁵ Gustin e Dias, *(Re)Pensando a pesquisa jurídica*, op. cit., p. 29-30 e 74-76.

²⁹⁶ Naves, *TCU e stare decisis administrativo: o caso das cláusulas restritivas em editais de licitação de obras de engenharia*, op. cit., p. 169-170.

²⁹⁷ TRIBUNAL DE CONTAS DA UNIÃO. *Manual Completo de Pesquisa de Jurisprudência*, p. 27. Disponível em: https://portal.tcu.gov.br/data/files/10/94/CC/39/F8610710F5680BF6F18818A8/Manual_Completo_Pesquisa_Jurisprudencia_TCU.PDF. Acesso em: 31 out. 2020.

²⁹⁸ Disponível em: <https://www.kaggle.com/ferraz/acordaos-tcu>. Acesso em 05 nov. 2019.

²⁹⁹ “O processo foi iterativo e envolveu mais de uma abordagem. A primeira se deu pelo uso da biblioteca *Selenium* implementado pela classe *AcordaosTCU* localizada em *scripts.crawler*. O segundo *approach* foi feito pelo uso do *scrapy* por meio da classe *ApiSpider* localizado em *crawlers.projects.api_acordaos.apiacordao.spiders.api*”. Cf.: <https://github.com/netoferraz/acordaos-tcu>. Acesso em 05 nov. 2019.

³⁰⁰ Disponível em: <https://github.com/netoferraz/acordaos-tcu>. Acesso em 05 nov. 2019.

Os *atributos*³⁰¹ da base de dados escolhida e que foram utilizados por esta pesquisa encontram-se explicitados na Tabela 5.³⁰²

Tabela 5 – Atributos utilizados dos acórdãos contidos na base de dados obtida por Fernanda Naves após requerimento formal ao TCU.

Nome do atributo	Descrição
ACORDAO	Parte dispositiva da deliberação
ANOACORDAO	Ano em que foi proferido o julgamento
COLEGIADO	Colegiado responsável pelo julgamento
DECISAO	Parte dispositiva da deliberação (forma antiga de deliberação)
NUMACORDAO	Número do acórdão
RELATORIO	Inteiro teor do relatório
VOTO	Inteiro teor do voto
VOTOCOMPLEMENTAR	Inteiro teor dos votos complementares
VOTOMINISTROREVISOR	Inteiro teor do voto do ministro revisor

Fonte: Dados da pesquisa.

Após a seleção dos atributos, verificou-se a validade dos registros. Foram encontrados 2.871 acórdãos que possuíam registros vazios, simultaneamente, para os atributos *ACORDAO*, *DECISAO*, *RELATORIO* e *VOTO*, ou que continham registros marcados como conteúdo sigiloso para as colunas selecionadas para o trabalho. Assim, *restaram 306.935 registros únicos*.

Em sequência, foi preciso standardizar a forma de identificação cada registro, de modo a ter um código único que fosse capaz de remeter ao tipo, número, ano e colegiado correspondente a cada acórdão. Considerando todas as possibilidades observadas, padronizou-se um identificador (*ID*) para cada acórdão:

TT-NNNNNN-AAAA-SS, sendo:

- *TT* → a depender do tipo de acórdão: AC para “acórdão”; DE para “decisão”; e AR para “acórdão de relação”;
- *NNNNNN* → seis dígitos para representar o número do acórdão;
- *AAAA* → quatro dígitos para representar o ano do acórdão;
- *SS* → a depender do colegiado: PL para “Plenário”; 1C para “Primeira Câmara”; e 2C para “Segunda Câmara”.

³⁰¹ Atributos são as informações de cada registro de dado. Em uma tabela, por exemplo, as linhas são os registros, enquanto as colunas, os atributos.

³⁰² Para informações sobre os outros atributos da base de dados utilizada, cf. Tribunal de Contas da União, *Manual Completo de Pesquisa de Jurisprudência*, op. cit, p. 27.

Ademais, considerando que as *rationes decidendi* são o conjunto das normas adscritas identificáveis a partir do modelo silogístico utilizado na justificação das decisões, buscou-se consolidar, em um único atributo, todo o inteiro teor do relatório, voto e parte dispositiva de uma decisão. Para tanto, concatenou-se na coluna *DECISAO*, os conteúdos dos atributos *ACORDAO* (ou *DECISAO*)³⁰³, *RELATORIO*, *VOTO*, *VOTOCOMPLEMENTAR* e *VOTOMINISTROREVISOR*.

Como resultado, armazenou-se um *dataset* (conjunto de dados) denominado *df_acordaos*, contendo duas colunas de atributos, *ID* e *DECISAO*, conforme representado na Figura 12.

Figura 12 – Detalhes do *dataset* *df_acordaos*, contendo 306.935 acórdãos e dois atributos para cada registro: um identificador único (ID) que remete ao tipo, ano, número e colegiado do acórdão; e uma coluna contendo o inteiro teor do relatório, voto e parte dispositiva da decisão colegiada (*DECISAO*).

df_acordaos.info()		
<pre><class 'pandas.core.frame.DataFrame'> Int64Index: 306935 entries, 0 to 309804 Data columns (total 2 columns): # Column Non-Null Count Dtype --- - 0 ID 306935 non-null object 1 DECISAO 306935 non-null object dtypes: object(2) memory usage: 7.0+ MB</pre>		
df_acordaos.sample(2)		
	ID	DECISAO
26207	DE-000356-2000-PL	O Tribunal Pleno, diante das razões expostas ...
95172	AC-000353-2008-1C	VISTOS, relatados e discutidos estes autos de ...
df_acordaos.tail(5)		
	ID	DECISAO
309800	AC-000058-2018-PL	<p class="tCU_- _Ac_- _item_9_- _\$\$_1">Vistos, re...
309801	AC-000031-2018-PL	<p class="normal_1">VISTOS, relatados e discut...
309802	AC-000033-2018-PL	<p class="normal_1">VISTOS, relatados e discut...
309803	AC-000030-2018-PL	<p class="normal_1">VISTOS, relatados e discut...
309804	AR-000061-2018-PL	Documento classificado como sigiloso com funda...

Fonte: Dados da pesquisa.

³⁰³ A depender da decisão colegiada, o atributo preenchido pode ser o *decisao* ou o *acordao* para a parte dispositiva da decisão colegiada, uma vez que o TCU se utilizava do atributo *decisao* para as mais antigas.

Para treinar a máquina a identificar o padrão de decisão referente ao conceito aberto de cláusula restritiva ao caráter competitivo em editais de licitação, recorreu-se aos resultados dos estudos realizados por Franklin Santos e Kleberon Souza e por Fernanda Naves que, juntos, identificaram 330 decisões do TCU, divididas em 11 classes diferentes, conforme descrito na seção 3.2.

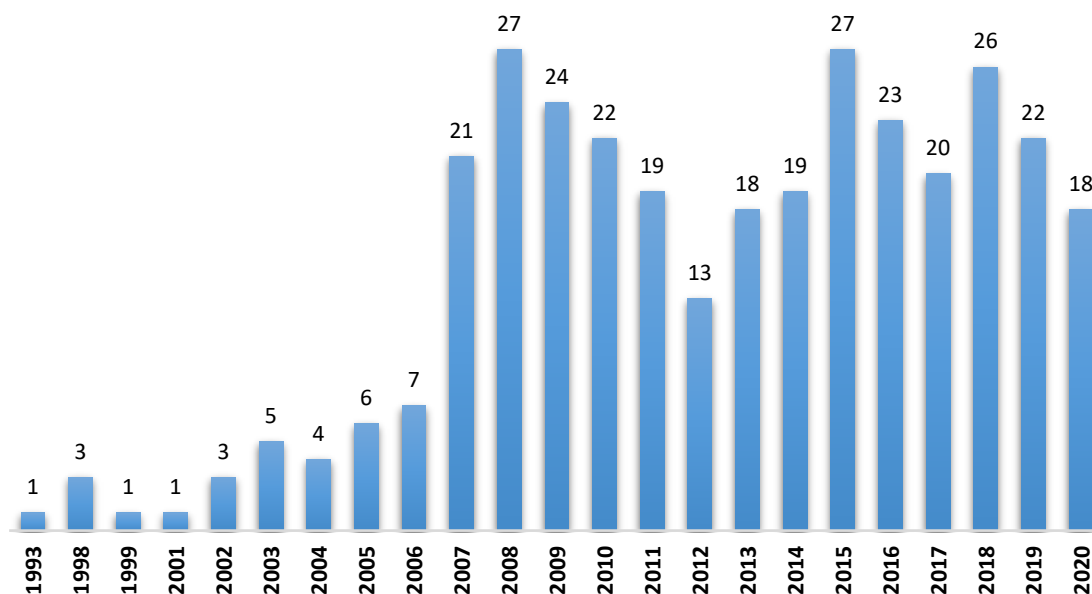
Esses acórdãos resultaram em 383 classificações, distribuídas conforme descrito na Tabela 6), uma vez que foram encontrados 25 acórdãos pertencentes a duas classes distintas e 1 acórdão classificado em 3 classes ao mesmo tempo (ver Apêndice A).

Tabela 6 – Quantidade de acórdãos, identificados nos trabalhos de Franklin Santos e Kleberon Souza e de Fernanda Naves. Os acórdãos foram classificados conforme as 11 classes estabelecidas por Santos e Souza e correspondem a diferentes tipos de cláusulas restritivas ao caráter competitivo em editais de licitação extraídas a partir de decisões do TCU que as consideraram restritivas à competitividade do certame.

Classe	Quantidade de acórdãos
1) Garantia de proposta para controle de interessados	20
2) Capacidade econômica exagerada	71
3) Idoneidade financeira ou bancária	2
4) “Quitação” em vez de “regularidade” fiscal	4
5) Certidões ou certificações indevidas	51
6) Alvará do município da licitação	3
7) Capacidade técnica irregular	168
8) Visita técnica restritiva	27
9) Carta de solidariedade do fabricante	3
10) Amostras subjetivas, indevidas, onerosas	10
11) Empresas com sócios em comum	6
Total de acórdãos classificados (rotulados)	365
Total de acórdãos distintos	330

Fonte: Dados da pesquisa.

Figura 13 – Quantidade de acórdãos rotulados por ano de julgamento.



Fonte: Dados da pesquisa.

Utilizando o identificador único (*ID*), almejou-se localizar os 330 acórdãos rotulados no *dataset df_acordaos*, tendo sido encontrados 275 deles. Isso porque, conforme apresentado na Figura 13, dentre os 330 acórdãos classificados, 30 foram emanados em 2019 e 2020, e outros 26 em 2018; desses 26, apenas um estava contido no *df_acordaos*, por ter sido julgado antes de 17/01/2018. Por isso, recorreu-se a consulta ao *sítio eletrônico de jurisprudência do TCU*³⁰⁴ para obtenção dos textos correspondentes ao inteiro teor dos 55 acórdãos não presentes inicialmente no *df_acordaos*, de modo a incrementá-lo. Feito isso, promoveu-se a identificação de todos os 330 acórdãos rotulados e textos do atributo *DECISAO* correspondentes no *dataset df_acordaos*.

Então, derivou-se um *dataset*, denominado de *df_rotulados*, que contém os 330 acórdãos previamente classificados e seus atributos referentes ao identificador único, inteiro teor e respectivos rótulos, conforme representado na Figura 14.

³⁰⁴ Disponível em: <https://pesquisa.apps.tcu.gov.br/#/pesquisa/acordao-completo>. Acesso em: 01 nov. 2020.

Figura 14 – Detalhes do *dataset df_rotulados*, contendo 330 acórdãos previamente classificados em 11 classes possíveis. Às últimas colunas, foi atribuído 1 se o acórdão pertence aquela classe ou 0, se não pertence.

	ID	DECISAO	LABELS	1	2	3	4	5	6	7	8	9	10	11
0	AC-000035-2018-PL	<p>VISTOS, relatados e discutidos estes autos ...	1,2,0	1	1	0	0	0	0	0	0	0	0	0
1	AC-000043-2008-PL	Vistos, relatados e discutidos estes autos Rep...	7,0,0	0	0	0	0	0	0	1	0	0	0	0
2	AC-000103-2009-PL	<p>VISTOS, relatados e discutidos estes autos ...	7,0,0	0	0	0	0	0	0	1	0	0	0	0
3	AC-000107-2009-PL	<p class="body_Text_Indent_3">VISTOS, relatado...	2,0,0	0	1	0	0	0	0	0	0	0	0	0
4	AC-000111-2016-PL	<p>VISTOS, relatados e discutidos estes autos ...	7,0,0	0	0	0	0	0	0	1	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
325	DE-000512-2002-2C	A Segunda Câmara, diante das razões expostas ...	7,0,0	0	0	0	0	0	0	1	0	0	0	0
326	DE-000681-1998-PL	O Tribunal Pleno, diante das razões expostas ...	2,0,0	0	1	0	0	0	0	0	0	0	0	0
327	DE-000767-1998-PL	O Tribunal Pleno, diante da razões exposta pe...	2,0,0	0	1	0	0	0	0	0	0	0	0	0
328	DE-001025-2001-PL	O Tribunal Pleno, diante das razões expostas ...	5,0,0	0	0	0	0	1	0	0	0	0	0	0
329	DE-001224-2002-PL	O Tribunal Pleno, diante das razões expostas ...	7,0,0	0	0	0	0	0	0	1	0	0	0	0

Fonte: Dados da pesquisa.

4.2.2 Pré-processamento

Algumas transformações foram realizadas com uso de expressões regulares³⁰⁵ no texto da coluna *DECISAO*. A exemplo do pré-processamento realizado pelo trabalho de ID 3 da Tabela 3, realizou-se a conversão dos termos referentes a citações de artigos, incisos e parágrafos de legislações para significantes únicos, conforme Figura 15.

Analogamente, realizou-se a transformação dos termos referentes a citações acórdãos anteriores do próprio TCU para significantes únicos (ver Figura 16).

Figura 15 – Representação da transformação de leis, artigos e incisos em significantes únicos.

VISTOS, relatados e discutidos estes autos de auditoria de conformidade realizada em obras de penitenciárias sob a responsabilidade do Departamento Penitenciário Nacional (Depen/MJ), por determinação do DE-450-2006-PL, objetivando aferir se os custos das obras estão de acordo com os preços de mercado e se as determinações propostas por este Tribunal por meio do DE-496-2004-2C estão sendo cumpridas, ACORDAM os Ministros do Tribunal de Contas da União, reunidos em sessão do Plenário, ante as razões expostas pelo Relator, em: 9.1. determinar ao Ministério da Justiça, com fulcro no ARTIGO 43, INCISO I, da LEI 8443/92 c/c o ARTIGO 250, INCISO II, do Regimento Interno do TCU, que: 9.1.1. promova o ajuste dos Contratos 01/2003, de 22/08/2003, e 03/2004, de 27/12/2004, firmados com a CEF, caso ainda vigentes, com a finalidade de que sejam elaboradas planilhas orçamentárias detalhadas de serviços, contemplando os respectivos quantitativos e composições de todos os seus custos unitários, possibilitando, assim, discriminar o que deve ser pago à contratada, em atendimento ao ARTIGO 7, PARAGRAFO 2, INCISO II, e ARTIGO 40, PARAGRAFO 2, INCISO II, da LEI 8666/93; 9.1.2. certifique-se que a CEF efetue a análise da documentação jurídica referente às propostas dos beneficiários selecionadas pelo Departamento Penitenciário Nacional (Depen/MJ), incluída a regularidade do processo licitatório do objeto financiado, conforme previsto na Cláusula Segunda, item II, alínea j, do Contrato 03/2004, de 27/12/2004; 9.1.3. ao firmar, nos termos do ARTIGO 2, parágrafo único, do Decreto 1819/96, instrumento com fulcro no ARTIGO 43, INCISO I, da LEI 8443/92 c/c ARTIGO 250, INCISO II, do Regimento Interno do TCU, do Acórdão de relação 496/2004-2ª Câmara, prevista nos PARAGRAFO 1, 7º e 9º do ARTIGO 2 da IN/STN 01/97 c/c o INCISO IX do ARTIGO 6 da LEI 8666/93, sob pena, nos termos do ARTIGO 58, INCISO VII, da LEI 8443/92; 9.2.2. AC-496-2004-2C, no sentido de que somente celebre convênios e seus aditivos ou libere os respectivos recursos, conforme o caso, após se assegurar da compatibilidade dos preços do objeto com os de mercado, em observância ao princípio da eficiência, insculpido no ARTIGO 37 da CONSTITUICAO FEDERAL, e ao ARTIGO 101 da Lei de Diretrizes Orçamentárias (LEI 10707/2003), assim como ao contido nas LDO, nos termos do ARTIGO 58, INCISO VII, da LEI 8443/92

³⁰⁵ Expressão regular, também chamada de *RegEx*, é uma linguagem computacional utilizada para identificação e transformação de cadeias de caracteres de interesse.

Fonte: Dados da pesquisa.

Figura 16 – Representação da transformação de citações de acórdãos anteriores do próprio TCU para significantes únicos.

VISTOS, relatados e discutidos estes autos de auditoria de conformidade realizada em obras de penitenciárias sob a responsabilidade do Departamento Penitenciário Nacional (Depen/MJ), por determinação do Decisão 450/2006-TCU-Plenário, objetivando aferir se os custos das obras estão de acordo com os preços de mercado e se as determinações proferidas por este Tribunal por meio do Decisão 496/2004-TCU-2ª Câmara estão sendo cumpridas, ACORDAM os Ministros do Tribunal de Contas da União, reunidos em sessão do Plenário, ante as razões expostas pelo Relator, em: 9.1.1. determinar ao Ministério da Justiça, com fulcro no art. 43, inciso I, da Lei 8.443/92 c/c o art. 250, inciso II, do Regimento Interno do TCU, que: 9.1.1.1. promova o ajuste dos Contratos 01/2003, de 22/08/2003, e 03/2004, de 27/12/2004, firmados com a Caixa Econômica Federal, caso ainda vigentes, com a finalidade de que sejam elaboradas planilhas orçamentárias de talhadas de serviços, contemplando os respectivos quantitativos e composições de todos os seus custos unitários, possibilitando, assim, discriminar o que deve ser pago à contratada, em atendimento ao artigo 7º, § 2º, inciso II, e artigo 40, § 2º, inciso II, da Lei 8.666/93; 9.1.2. certifique-se que a Caixa Econômica Federal efetue a análise da documentação jurídica referente às propostas dos beneficiários selecionados pelo Departamento Penitenciário Nacional (Depen/MJ), incluída a regularidade do processo licitatório do objeto financiado, conforme previsto na Cláusula Segunda, item II, alínea *z*, do Contrato 03/2004, de 27/12/2004; 9.1.3. ao firmar, nos termos do art. 2º, parágrafo único, do Decreto 1.819/96, instrumento com fulcro no art. 43, inciso I, da Lei 8.443/92 c/c o art. 250, inciso II, do Regimento Interno do TCU, do Acórdão de relação 496/2004-2ª Câmara, prevista nos § 1º, 7º e 9º do art. 2º da IN/STN 01/97 c/c o inciso IX do art. 6º da Lei 8.666/93, sob pena, nos termos do art. 58, inciso VII, da Lei 8.443/92; 9.2.2. **AC-496-2004-2C**, no sentido de que somente celebre convênios e seus aditivos ou libere os respectivos recursos, conforme o caso, após se assegurar da compatibilidade dos preços do objeto com os de mercado, em observância ao princípio da eficiência, insculpido no art. 37 da Constituição Federal, e ao art. 101 da Lei de Diretrizes Orçamentárias (Lei 10.707/2003), assim como ao contido nas LDO, nos termos do art. 58, inciso VII, da Lei 8.443/92

VISTOS, relatados e discutidos estes autos de auditoria de conformidade realizada em obras de penitenciárias sob a responsabilidade do Departamento Penitenciário Nacional (Depen/MJ), por determinação do **DE-450-2006-PL**, objetivando aferir se os custos das obras estão de acordo com os preços de mercado e se as determinações proferidas por este Tribunal por meio do Decisão 496/2004-TCU-2ª Câmara estão sendo cumpridas, ACORDAM os Ministros do Tribunal de Contas da União, reunidos em sessão do Plenário, ante as razões expostas pelo Relator, em: 9.1.1. determinar ao Ministério da Justiça, com fulcro no art. 43, inciso I, da Lei 8.443/92 c/c o art. 250, inciso I, do Regimento Interno do TCU, que: 9.1.1.1. promova o ajuste dos Contratos 01/2003, de 22/08/2003, e 03/2004, de 27/12/2004, firmados com a Caixa Econômica Federal, caso ainda vigentes, com a finalidade de que sejam elaboradas planilhas orçamentárias detalhadas de serviços, contemplando os respectivos quantitativos e composições de todos os seus custos unitários, possibilitando, assim, discriminar o que deve ser pago à contratada, em atendimento ao artigo 7º, § 2º, inciso II, e artigo 40, § 2º, inciso II, da Lei 8.666/93; 9.1.2. certifique-se que a Caixa Econômica Federal efetue a análise da documentação jurídica referente às propostas dos beneficiários selecionados pelo Departamento Penitenciário Nacional (Depen/MJ), incluída a regularidade do processo licitatório do objeto financiado, conforme previsto na Cláusula Segunda, item II, alínea *z*, do Contrato 03/2004, de 27/12/2004; 9.1.3. ao firmar, nos termos do art. 2º, parágrafo único, do Decreto 1.819/96, instrumento com fulcro no art. 43, inciso I, da Lei 8.443/92 c/c o art. 250, inciso II, do Regimento Interno do TCU, do Acórdão de relação 496/2004-2ª Câmara, prevista nos § 1º, 7º e 9º do art. 2º da IN/STN 01/97 c/c o inciso IX do art. 6º da Lei 8.666/93, sob pena, nos termos do art. 58, inciso VII, da Lei 8.443/92; 9.2.2. Acórdão 496/2004-2ª Câmara, no sentido de que somente celebre convênios e seus aditivos ou libere os respectivos recursos, conforme o caso, após se assegurar da compatibilidade dos preços do objeto com os de mercado, em observância ao princípio da eficiência, insculpido no art. 37 da Constituição Federal, e ao art. 101 da Lei de Diretrizes Orçamentárias (Lei 10.707/2003), assim como ao contido nas LDO, nos termos do art. 58, inciso VII, da Lei 8.443/92

VISTOS, relatados e discutidos estes autos de auditoria de conformidade realizada em obras de penitenciárias sob a responsabilidade do Departamento Penitenciário Nacional (Depen/MJ), por determinação do Decisão 450/2006-TCU-Plenário, objetivando aferir se os custos das obras estão de acordo com os preços de mercado e se as determinações proferidas por este Tribunal por meio do Decisão 496/2004-TCU-2ª Câmara estão sendo cumpridas, ACORDAM os Ministros do Tribunal de Contas da União, reunidos em sessão do Plenário, ante as razões expostas pelo Relator, em: 9.1.1. determinar ao Ministério da Justiça, com fulcro no art. 43, inciso I, da Lei 8.443/92 c/c o art. 250, inciso II, do Regimento Interno do TCU, que: 9.1.1.1. promova o ajuste dos Contratos 01/2003, de 22/08/2003, e 03/2004, de 27/12/2004, firmados com a Caixa Econômica Federal, caso ainda vigentes, com a finalidade de que sejam elaboradas planilhas orçamentárias de talhadas de serviços, contemplando os respectivos quantitativos e composições de todos os seus custos unitários, possibilitando, assim, discriminar o que deve ser pago à contratada, em atendimento ao artigo 7º, § 2º, inciso II, e artigo 40, § 2º, inciso II, da Lei 8.666/93; 9.1.2. certifique-se que a Caixa Econômica Federal efetue a análise da documentação jurídica referente às propostas dos beneficiários selecionados pelo Departamento Penitenciário Nacional (Depen/MJ), incluída a regularidade do processo licitatório do objeto financiado, conforme previsto na Cláusula Segunda, item II, alínea *z*, do Contrato 03/2004, de 27/12/2004; 9.1.3. ao firmar, nos termos do art. 2º, parágrafo único, do Decreto 1.819/96, instrumento com fulcro no art. 43, inciso I, da Lei 8.443/92 c/c o art. 250, inciso II, do Regimento Interno do TCU, do **AR-496-2004-2C**, prevista nos § 1º, 7º e 9º do art. 2º da IN/STN 01/97 c/c o inciso IX do art. 6º da Lei 8.666/93, sob pena, nos termos do art. 58, inciso VII, da Lei 8.443/92; 9.2.2. Acórdão 496/2004-2ª Câmara, no sentido de que somente celebre convênios e seus aditivos ou libere os respectivos recursos, conforme o caso, após se assegurar da compatibilidade dos preços do objeto com os de mercado, em observância ao princípio da eficiência, insculpido no art. 37 da Constituição Federal, e ao art. 101 da Lei de Diretrizes Orçamentárias (Lei 10.707/2003), assim como ao contido nas LDO, nos termos do art. 58, inciso VII, da Lei 8.443/92

Fonte: Dados da pesquisa.

Similar ao que foi feito por Martija et al.³⁰⁶, a padronização dessas expressões visou a criação de *tokens* únicos, possibilitando encontrar associações entre os acórdãos da base pela abordagem *case-to-law*, quando citarem dispositivos legais em comum, e pela abordagem *case-to-case*, quando se referenciam às mesmas decisões colegiadas do TCU anteriores.

Ademais, promoveu-se a remoção ou substituição de palavras que não agregam valor ao texto e possuem elevada frequência (*stopwords*), como:

³⁰⁶ Cf. Martija et al., *How Deep is Your Law? Predicting Associations between Cases in Philippine Jurisprudence*, op. cit.

Para essa etapa, inicialmente realizou-se a conversão de todas as palavras do atributo *DECISAO* para caracteres minúsculos, bem como a *tokenização* por palavra, a fim de obter maior congregação de *tokens* únicos.

Optou-se pela criação da representação vetorial de cada documento no tipo *n-gram*, com $n=1$, em que cada *token* (palavra) de um registro (acórdão) transforma-se em uma coluna na matriz que representa todos os textos.

4.2.4 Redução da dimensionalidade

Boa parte do pré-processamento realizado, descrito na seção 4.2.2, contribuiu para a redução da dimensionalidade da matriz de atributos. Além disso, para modelos de *machine learning* (ML), utilizou-se a técnica de *tf-idf* (*term frequency-inverse document frequency*), que mede a relevância de cada palavra com base na sua frequência em cada documento e em comparação com o restante do *corpus*.

Já para testes em modelos de *deep learning* (DL), agregou-se à transformação vetorial a técnica de *word embedding*³⁰⁸, que é capaz de extrair mínimas associações entre as palavras que possuem conexões semânticas.

4.2.5 Treinamento dos classificadores

Inicialmente, dividiu-se o *dataset df_rotulados* em duas partes, sendo 70% para treinamento dos modelos e 30% para testes deles. Com isso, exigiu-se a utilização, dentre as 11 classes existentes, das que tinham dez ou mais registros rotulados (classes 7, 2, 5, 8, 1 e 10, ver Tabela 6), visto que, não se tinham registros suficientes para a tal separação para as demais classes.

Em se tratando de uma classificação *multi-label*, empregou-se a abordagem comparativa *one-vs-rest*³⁰⁹, em que se divide as checagens em vários problemas de classificação binários. Ainda, em decorrência do desbalanceamento de dados rotulados entre as classes, aplicou-se pesos maiores às classes com menos registros, de modo a mitigar o enviesamento do aprendizado pelos algoritmos.

³⁰⁸ Disponível em: <http://www.nilc.icmc.usp.br/nilc/index.php/repositorio-de-word-embeddings-do-nilc>. Acesso em 11 nov. 2020.

³⁰⁹ Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.multiclass.OneVsRestClassifier.html>. Acesso em 16 no. 2020. Essa também foi a abordagem utilizada no artigo Araújo et al., *VICTOR: a dataset for Brazilian legal documents classification*, op. cit., p. 1455.

Assim sendo, foram testados modelos de ML e DL, utilizando como métrica de avaliação a *área sob a curva de desempenho ROC* (*area under the ROC curve – AUC*), que consiste no cálculo da área embaixo de uma curva traçada com base na quantidade de falsos positivos (aqueles que foram rotulados pelo modelo como pertencentes a determinada classe, mas que não deveriam ter sido) e de falsos negativos (aqueles que deveriam ter sido rotulados para uma determinada classe, mas não foram). A escolha dessa métrica deu-se pelo fato de conseguir equilibrar, em um valor único, o bom e mau desempenho de cada algoritmo. Além disso, uma vez que se pretendeu delinear toda a jurisprudência correlata, prezou-se pela menor quantidade de falsos negativos possível.

Ressalta-se que foi priorizado o emprego de algoritmos de código aberto e, com base naqueles utilizados nos trabalhos qualificados na revisão da literatura (ver Tabela 4), optou-se pela treinamento, para fins comparativos, dos seguintes: *Logistic Regression*, *Multinomial Naive Bayes (Multinomial NB)*, *Linear Support Vector Classification (Linear SVC)* e *Random Forest Classifier*, para modelos de ML; e *Deep Neural Network (DNN)*, *Convolutional Neural Network (CNN)*, *Long Short-Term Memory (LSTM)* e *Bidirectional Long Short-Term Memory (Bi-LSTM)*, para modelos de DL.

No geral, os resultados dos testes dos algoritmos foram positivos, mormente para os modelos de DL, conforme demonstrado pelas medidas de AUC constantes na Tabela 7. O classificador de melhor desempenho foi o treinado com *Convolutional Neural Network (CNN)*.

Tabela 7 – Resultados (AUC) do treinamento para modelos de ML e DL aplicados à classificação *multi-label* das classes 7, 2, 5, 8 e 10.

Modelos de machine learning			
<i>Logistic Regression</i>	<i>Multinomial NB</i>	<i>Linear SVC</i>	<i>Random Forest</i>
72,63%	70,23%	48,34%	45,98%
Modelos de deep learning			
<i>DNN</i>	<i>CNN</i>	<i>LSTM</i>	<i>Bi-LSTM</i>
79,82%	82,69%	78,15%	82,24%

Fonte: Dados da pesquisa.

4.2.6 Avaliação da solução

O classificador com melhores resultados da etapa anterior, o *Convolutional Neural Network (CNN)*, foi utilizado para prever classes para cada acórdão contido no *dataset df_acordaos*. Como resultado, obteve-se a probabilidade de cada uma dessas decisões colegiadas corresponder às classes 7, 2, 5, 8, 1 e 10.

Entretanto, *o modelo não correspondeu em bons resultados para a fase de validação com supervisão humana*, uma vez que, mesmo para os acórdãos que o modelo já conhecia, ele não foi capaz de predizer a classificação provável correta. Basicamente, mesmo tendo sido aplicado pesos no treinamento, as probabilidades obtidas espelharam o desbalanceamento das classes, indicando para todos os acórdãos da base que a classificação mais provável seria a classe 7, a segunda mais provável, a classe 2, e assim por diante.

Por isso, *conclui-se que a solução desenvolvida de classificação multi-label de documentos foi inconclusiva para a solução do problema de pesquisa*.

Um dos motivos deve-se à pequena quantidade utilizada de acórdãos rotulados para cada classe, bem como aos diminutos exemplos de registros múltiplos rótulos, uma vez que, dos 330 acórdãos rotulados extraídos das pesquisas de Franklin Santos e Kleberson Souza e de Fernanda Naves, apenas 32 pertenciam a mais de uma classe.

O artigo científico de ID 19 da Tabela 3, encontrado na revisão da literatura realizada e intitulado “*Legal Area Classification: A Comparative Study of Text Classifiers on Singapore Supreme Court Judgments*”, acusou resultados razoáveis mesmo com poucos documentos rotulados (588), mas que representavam aproximadamente 10% do universo pesquisável.³¹⁰ Em comparação, a quantidade de acórdãos rotulados utilizados para a abordagem testada nesta pesquisa correspondia a apenas 0,1% da quantidade total de acórdãos pesquisáveis (306.935 decisões colegiadas do *df_acordaos*).

Apesar dos resultados inconclusivos, é possível traçar possibilidades de prosseguimentos dos testes empíricos com intuito de contribuir com pesquisas futuras.

Por exemplo, o artigo científico de ID 26 da Tabela 3, encontrado na revisão da literatura e intitulado “*Multi-label charge predictions leveraging label co-occurrence in imbalanced data scenario*”, fez uso de técnicas de *Few-Shot Learning* (FSL) para superar problemas de classes pouco rotuladas (com menos de 50 rótulos).

Uma outra saída é incrementar a quantidade de acórdãos rotulados por intermédio da utilização de associações *case-to-law* e *case-to-case*. Explica-se.

Uma vez que foram criados *tokens* únicos com expressões regulares, para padronização de dispositivos normativos e decisões anteriores do próprio TCU, é possível buscar no atributo *DECISAO* criado – que contém o relatório, o voto e a parte dispositiva de cada acórdão – referências a esses significantes. Exemplificando, para incremento da classe 7, que trata de

³¹⁰ Cf. Howe et al., *Legal Area Classification: A Comparative Study of Text Classifiers on Singapore Supreme Court Judgments*, op. cit.

capacidade técnica irregular, pode-se buscar pelos *tokens* “((LEI_8666_1993 OU LEI_8666_93) E (ARTIGO_30))”³¹¹ com intuito de retornar um conjunto (*conjunto A*) de acórdãos que fazem menção à mesma especificação legal relativa à qualificação técnica em licitações, ou seja, que apresentam a mesma associação *case-to-law*. Analogamente, a busca por significantes, no atributo *DECISAO*, contidos na lista das 168 decisões conhecidas já rotuladas na classe 7 (ex.: AC-002192-2007-PL), possibilita o retorno de um conjunto (*conjunto B*) de acórdãos que as referenciaram e que, por isso, provavelmente possuem uma associação *case-to-case* com essas decisões reconhecidamente classificadas como pertencentes à classe 7. Com isso, da união dos conjuntos A e B, após supervisão humana, é factível a extração de novas decisões da Classe 7, de modo a ampliar a lista existente.

Um outro motivo que pode ter influenciado nos resultados inconclusivos é o fato de uma mesma decisão colegiada poder pertencer a mais de uma classe relativa a cláusulas restritivas ao caráter competitivo em editais de licitação. Essa foi a razão de terem sido utilizadas técnicas de classificação *multi-label*, mas é possível que as menções às patologias associadas a cada classe, no decorrer do texto, ocorram de forma breve ou sucinta, não sendo suficiente para que a máquina, em meio a um texto longo, identifique o padrão dos rótulos.

Por isso, outra possibilidade é migrar da área prática de *classificação de documentos* para problemas de *text mining* (ver Figura 4), para a área prática de *recuperação de informações* (*information retrieval*).³¹² Como mostrado na seção 2.3, a partir da árvore de decisão proposta por Miner et al.³¹³ (Figura 3), concluiu-se que o primeiro problema prático da pesquisa se constituía de um problema de *classificação de documentos*. Um fato que corrobora o caminho adotado é que esta foi também a abordagem prática escolhida pelo trabalho científico de ID 13 da Tabela 3, encontrado na revisão da literatura realizada e intitulado “*How Deep is Your Law? Predicting Associations Between Cases in Philippine Jurisprudence*”³¹⁴. Esse artigo teve objetivo muito semelhante ao desta pesquisa, qual seja, o de encontrar precedentes úteis para um determinado caso, e utilizou-se de técnicas de classificação de documentos com resultados

³¹¹ O art. 30 da Lei 8.666/1993 versa, especificamente, sobre a documentação relativa à qualificação técnica.

³¹² Relembra-se que, conforme classificado por Miner et al, são sete as áreas práticas de *text analysis*: recuperação de informações (*information retrieval*); agrupamento de documentos (*clustering*); classificação de documentos (*document classification*); mineração da web (*web mining / web crawling*); extração de informação (*information extration*); Processamento de Linguagem Natural estatística (*statistical NLP*); extração de conceitos (*concept extration*). Ver Figura 4. Cf. Miner et al., *Practical text mining and statistical analysis for non-structured text data applications*, op. cit., p. 32.

³¹³ Cf. *ibid.*, p. 33.

³¹⁴ Cf. Martija et al., *How Deep is Your Law? Predicting Associations between Cases in Philippine Jurisprudence*, op. cit.

também nada animadores (menos de 50% de precisão) quando comparados com demais trabalhos de classificação *multi-label* de documentos (ver Tabela 4).

Nessa lógica, na revisão sistemática da literatura realizada, encontrou-se evidências de que os avanços da área prática de *recuperação de informações* vão para além da indexação por palavras-chave e podem ser úteis ao problema desta pesquisa. Na tese de doutorado “*Improving Semantic Search in the German Legal Domain with Word Embeddings*” (ID 14 da Tabela 3), Jörg Landthaler utilizou o procedimento por correspondência semântica de textos (*semantic text matching*) para identificar segmentos de texto semanticamente vinculados, concluindo ser essa uma alternativa viável para pesquisas de legislação correlatas e de jurisprudência.³¹⁵

Ora, considerando que as *rationes decidendi*, conforme aduzido por Thomas da Rosa de Bustamante, constituem-se do conjunto das normas adscritas utilizadas para a justificação da decisão, uma sistemática de busca de decisões similares a partir da identificação, a priori, de possíveis *rationes decidendi* relacionadas à temática buscada, parece ser uma alternativa razoável a ser testada para o problema desta pesquisa futuramente.

Esses parecem ser os caminhos mais prováveis para a criação de algoritmos que se aproximam do conceito de *computação cognitiva*³¹⁶ de Kevin Ashley e que podem ser a solução para a criação de uma *central de jurisprudência administrativa*³¹⁷, idealizada por Schiefler et al.

4.3 Ensaio para identificação de precedentes administrativos que dão concretude ao conceito aberto de cláusulas restritivas em editais de licitação

Dado o resultado insatisfatório para o primeiro problema prático da pesquisa, em que não foi possível, com uso de *text mining* para a classificação de documentos jurídicos, delinear a jurisprudência acerca do recorte temático escolhido, não se pôde responder, inequivocamente, ao segundo problema prático, que visava *identificar os precedentes administrativos* que dão concretude ao conceito jurídico indeterminado de cláusulas restritivas ao caráter competitivo em editais de licitação.

³¹⁵ Cf. Landthaler, *Improving Semantic Search in the German Legal Domain*, op. cit., p. 97-134.

³¹⁶ Cf. Ashley, *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, op. cit., p. 12-14.

³¹⁷ Cf. Schiefler et al., *A inteligência artificial aplicada à criação de uma central de jurisprudência administrativa: o uso das novas tecnologias no âmbito da gestão de informações sobre precedentes em matéria administrativa*, op. cit., p. 23.

Todavia, tendo em vista o objetivo geral desta pesquisa, a fim de contribuir para a concreção prática de cláusulas, em editais de licitação, consideras restritivas por exigirem capacidade técnica dos licitantes de maneira irregular (classe 7, ver Tabela 6), apresenta-se um ensaio da construção da *linha jurisprudencial de entendimento* com indicação dos prováveis *precedentes administrativos* para a classe 7.

Importante rememorar que, para a essa classe, foram identificadas 15 subclasses que, em correspondência ao conceito de *linha jurisprudencial de entendimento* (ver Figura 1), podem ser consideradas ramos resultantes da técnica de distinção (*distinguishing*). A Tabela 8 apresenta a quantidade de acórdãos para cada uma dessas 15 subclasses (ver Apêndice A para listagem do identificador único dessas decisões):

Tabela 8 – Quantidade de acórdãos para as subclasses da classe 7, identificados nos trabalhos de Franklin Santos e Kleberson Souza e de Fernanda Naves. Os acórdãos foram classificados conforme as subclasses estabelecidas por Santos e Souza e correspondem a decisões que versaram sobre cláusulas, em editais de licitação, que exigiram capacidade técnica dos licitantes de maneira irregular e que, por isso, foram consideradas restritivas ao caráter competitivo.

Subclasses da classe 7 (<i>capacidade técnica irregular</i>)	Quantidade de acórdãos
7.1) Atestados em número mínimo, máximo ou fixo	18
7.2) Quantitativo exagerado de experiência mínima	12
7.3) Experiência em parcela irrelevante	6
7.4) Experiência específica demais	35
7.5) Experiência genérica demais	10
7.6) Limitação de tempo e época nos atestados	1
7.7) Atestado do mesmo grupo econômico	2
7.8) Reorganização empresarial: fusão, cisão e incorporação	4
7.9) Transferência de capacidade técnico-profissional para operacional	1
7.10) Matriz e filial	3
7.11) Visto do CREA da Seção Local	42
7.12) Certificados de Qualidade	20
7.13) Profissional do “quadro permanente”	17
7.14) Quantitativos mínimos em experiência do profissional	11
7.15) CAT em nome da empresa licitante	2
Total de acórdãos classificados (rotulados)	184
Total de acórdãos distintos	168

Fonte: Dados da pesquisa.

A partir dos 168 acórdãos encontrados pelas pesquisas de Franklin Santos e Kleberson Souza e de Fernanda Naves e classificados como pertencentes à classe 7 e suas 15 subclasses, buscou-se identificar os possíveis *precedentes administrativos* mediante a leitura dos acórdãos

mais antigos de cada subclasse. Caso, no decorrer dessa leitura, fossem identificadas citações a decisões colegiadas anteriores do TCU, elas foram verificadas para constatar sua pertinência, ou não, com a temática da subclasse em exame. Esse procedimento foi repedido iterativamente para cada nova decisão encontrada, até que não houvessem menções no texto a prováveis precedentes anteriores.

Como resultado, identificou-se quatro novas decisões, elevando o número de acórdãos da classe 7 de 168 para 172. Dessas quatro, três correspondiam a novos prováveis precedentes para as subclasses 7.3, 7.10 e 7.12.

Em seguida, de posse dos possíveis precedentes administrativos de cada subclasse, *buscou-se operacionalizar o conceito de ratio decidendi de Thomas da Rosa de Bustamante, extraindo as normas adscritas*³¹⁸ *dos modelos silogísticos utilizados na justificação de cada precedente.*

Desta feita, os precedentes administrativos identificados e as *rationes decidendi* que os caracterizam encontram-se especificados na tabela abaixo (apresentados por ordem cronológica do precedente mais antigo para o mais novo).³¹⁹

Tabela 9 – Precedentes administrativos (com suas respectivas *rationes decidendi*) do TCU identificados que versaram sobre cláusulas, em editais de licitação, que exigiram capacidade técnica dos licitantes de maneira irregular e que, por isso, foram consideradas restritivas ao caráter competitivo.

Descrição do precedente	<i>Rationes decidendi</i> extraídas dos relatórios, votos e partes dispositivas dos acórdãos
Precedente inicial: DE-000434-1993-PL, que dá concreção à classe 7.11 - Visto do CREA da Seção Local	RELATÓRIO: “Ante o exposto: entendemos, data máxima vênua, que: a. o artigo 69 da Lei 5.194, de 24.12.66, encontra-se revogado pelo Decreto-lei 2.300/86, e suas alterações posteriores produzidas pelos Decretos-leis 2.348/87 2.360/87, por estar em conflito com: a.1. - o caráter simplificador e agilizador que norteou a edição do citado Estatuto das Licitações e Contratos Administrativos; a.2 - o artigo 25, II c/c o seu parágrafo 2º, 1, bem como o parágrafo 1º, I, do artigo 3º, todos do Decreto-lei nº 2.300/86 e suas alterações posteriores; b. a empresa MARIALVA CONSTRUTORA LTDA tinha Capacidade genérica, na medida em que apresentou as certidões às fls. 89/92, fornecidas pelo CREA do Estado de Minas Gerais, Unidade da Federação onde se encontra sediada a aludida empresa. c. a obrigatoriedade de visar o registro na entidade profissional competente, na jurisdição onde a obra deva ser executada, é da empresa ganhadora do certame licitatório, e não das

³¹⁸ Rememora-se que as normas adscritas são normas são normas criadas no processo de concretização do direito e utilizadas na justificação das decisões. Para Thomas da Rosa de Bustamante, as *rationes decidendi* constituem-se no conjunto das normas adscritas utilizadas no encadeamento silogístico realizado pelo operador do Direito. Ver sessão 2.2.

³¹⁹ Ressalta-se que, como não foram lidas todas as 171 decisões, não foi possível identificar os precedentes decorrentes de possíveis técnicas de superação (*overruling*) ocorridas.

	empresas participantes, conforme se pode inferir do artigo 58, da Lei 5.194/66, transcrito no item 9, desta instrução. E a empresa ganhadora realmente o fez, ainda que intempestivamente (fls. 15)." VOTO: “O ponto fulcral é a exigência de comprovação de capacidade técnica para participar de processos licitatórios por empresas ou profissionais do ramo da engenharia, arquitetura e agronomia. Como destacado no Relatório acima, o art. 69 da Lei nº 5.194/66 exige das empresas e profissionais a apresentação de "prova de quitação de débito ou visto do Conselho Regional da jurisdição onde a obra, o serviço técnico ou projeto deva ser executado". É, portanto, um dispositivo restritivo e protecionista. Ocorre que o Decreto-lei nº 2.300/86, sob cujo império se efetivou a licitação, dispôs, em seu art. 25, II, que "para a habilitação nas licitações exigir-se-á dos interessados, exclusivamente, documentação relativa a... capacidade técnica...". Tal Documentação consiste, simplesmente, no "registro ou inscrição na entidade profissional competente", conforme disposição contida no citado dispositivo legal. Frente, pois, este Plenário a uma questão em que a uma Lei Especial se sucede uma Lei Geral regulando a mesma matéria. É sabido que se a uma Lei Geral se sucede uma Especial, normalmente, aquela continua a vigir, visto que pode coexistir com a outra. Já o contrário é muito duvidoso. Neste passo, como concluiu o analista informante, o art. 69 da Lei nº 5.194/69 parece ter sido revogado pelos dispositivos citados do Decreto-lei nº 2.300/86. É de notar, por outro lado, que abrogação tácita não resulta, apenas, de incompatibilidade entre dois dispositivos legais: opera-se, também, quanto uma Lei nova regula toda a matéria disciplinada pela Lei anterior. Deduz-se, portanto, no caso, a vontade do legislador de liquidar o passado, estabelecendo um novo ordenamento completo e autônomo, ou seja, um reordenamento jurídico que não tolera desvios de leis precedentes. O Decreto-lei nº 2.300/86 criou, à época, um ordenamento jurídico completo em matéria de licitação e contrato, como agora, novamente, o faz a Lei nº 8.666/93.” DECISÃO: Não há trecho específico correspondente.
Precedente decorrente da técnica de distinção: DE-000679-1997-PL, que dá concreção à classe 7.10 – Matriz e filial	RELATÓRIO: “Entre as questões analisadas na oportunidade, destacava-se o questionamento suscitado pela equipe de auditoria da SECEX/SE, referente à "inabilitação indevida da empresa XEROX do Brasil Ltda., na Tomada de Preços nº 16/95, em virtude da apresentação de registro de CGC divergentes entre o CND, o FGTS e a Relação de Empregados, sem levar em consideração o recurso interposto pela empresa no qual comprova a centralização das arrecadações das contribuições do INSS e do FGTS na filial do Rio de Janeiro (infringência aos princípios da isonomia e da igualdade preconizados no art. 3º, "caput", da Lei nº 8.666/93)". A SECEX/SE refutou a justificativa apresentada pela DAMF/SE em resposta à diligência então efetuada, assinalando que, ao contrário do que afirmava o órgão, havia correspondência entre os números do CGC da empresa nos documentos apresentados pela XEROX, cujos recolhimentos de FGTS são centralizados em sua sede no Rio de Janeiro. A Unidade Técnica considerou, em consequência, que a desclassificação da empresa na Tomada de Preços por esse motivo constituía infração ao princípio da igualdade.” VOTO: “Com efeito, é procedente o argumento da recorrente quando requer que, ante o exposto reconhecimento da ilegalidade do procedimento seguido pela DAMF/SE, do qual resultou sua inabilitação na tomada de preços, o TCU formule determinação ao órgão para que evite desqualificar empresas inscritas em licitações em razão de apresentarem comprovantes de regularidade de sua situação perante o INSS e o FGTS com base em documentação referente à arrecadação centralizada das contribuições, pois a lei admite tal forma de pagamento.” DECISÃO: “m) evite inabilitar participantes de processos licitatórios em razão somente de diferenças entre números de registro de CGC das respectivas matriz e filiais, nos comprovantes pertinentes ao CND, ao FGTS, INSS e Relação de Empregados,

	quando a empresa interessada comprovar a centralização do recolhimento de contribuições, tendo em vista a legalidade desse procedimento;"
Precedente decorrente da técnica de distinção: DE-000020-1998-PL, que dá concreção à classe 7.12 – Certificados de Qualidade	RELATÓRIO: “5. No tocante ao Certificado ISO 9001, a Unidade Técnica, nas instruções de fls. 21/4 e 35/8, entendeu irregular a exigência, visto que, de um total de quatro milhões de empresas operantes no país, aproximadamente 1.200 possuem o certificado, de forma que esse critério de qualificação prévia constitui uma afronta ao art. 3º, § 1º, inciso I, da Lei n. 8.666/93.” [...] “8. A douta Procuradoria, representada nos autos pelo ilustre Subprocurador-Geral Lucas Rocha Furtado, teceu diversas considerações sobre a questão em análise (fls. 40/1), que me permito transcrever, em parte: "Com relação ao Certificado ISO 9001, entendemos que a desclassificação da proposta técnica ante a ausência de sua apresentação não se conforma ao Direito. Não apenas pelo aspecto fático apontado pela SECEX/SP à fl. 22, relativo à pequena quantidade de empresas brasileiras certificadas, o que implicaria restringir ou frustrar o caráter competitivo da licitação, nos termos do art. 3º, § 1º, inciso I da Lei n. 8.666/93, mas, antes, por ser dada importância exagerada àquele certificado.” VOTO: “Relativamente à exigência da apresentação do Certificado ISO 9001, sob pena de desclassificação da proposta técnica, realmente justifica-se considerá-la como condição que restringe o caráter competitivo da licitação, tendo em vista o reduzido número de empresas que já obtiveram tal Certificado.” DECISÃO: “1 - determinar ao Conselho Regional de Medicina do Estado de São Paulo que, nas futuras licitações para aquisição de bens ou serviços de informática, não exija que, sob pena de desclassificação da proposta, seja apresentado Certificado da série ISO 9000;”
Precedente decorrente da técnica de distinção: AC-000167-2001-PL, que dá concreção à classe 7.3 – Experiência em parcela irrelevante	RELATÓRIO: “Restrição ao caráter competitivo das Tomadas de Preços n. 086/99, 087/99, 088/99 e 089/99, uma vez que houve a exigência de prova de qualificação técnica para a execução de redes elétricas de 69 KV, cujos valores orçados correspondiam a apenas 3,7 % do valor total dos objetos licitados, não sendo, portanto, essa execução considerada parcela de maior relevância e valor significativo dos objetos das licitações, em inobservância ao disposto nos arts. 3º, § 1º, inciso I, e 30, §§ 1º e 2º, inciso I, da Lei n. 8.666/93. Além da pequena relevância do item mencionado, não se coadunou com o princípio da razoabilidade a inclusão, no objeto de certames que visavam à construção de redes de distribuição de energia em 13,8 KV e 34,5 KV, de execução de pequeno trecho (cerca de 1 Km) de rede de 69 KV, uma vez que essa última espécie de rede caracteriza-se como linha de transmissão, sendo, portanto, destinada a ligar duas subestações, ao passo que não há subestações distantes entre de si por apenas 1 Km.” [...] “10. A situação é análoga à verificada, por exemplo, na construção de um edifício de 30 andares. De fato, para a adequada utilização desse prédio, a instalação de elevadores é fundamental. No entanto, haveria clara restrição à competitividade de uma licitação que tivesse por objeto essa construção, se fosse vedado à Contratada a faculdade de subcontratar os serviços atinentes à instalação dos elevadores, que têm natureza específica, em relação aos demais serviços atinentes à construção do edifício, bem assim não se caracterizam como "parcelas de maior relevância e valor significativo da licitação", a ensejar a obrigatoriedade de a própria Contratada comprovar a capacidade técnica pertinente, conforme o disposto no art. 30, § 1º, inciso I, da Lei n. 8.666/93. 11. Esta Unidade Técnica destacou a pequena relevância (3,7 % dos totais), nos orçamentos da CEPISA, do valor dos materiais e serviços relativos à rede de 69 KV, em relação ao valor total dos objetos licitados. Esse percentual é ainda menor nos Contratos, incluindo-se os valores dos termos aditivos contratuais (os três contratos tiveram os seus valores aumentados praticamente no percentual máximo prescrito pela lei, de 25 %, sendo que, no entanto, não houve acréscimo de nenhum dos itens referentes à rede de 69 KV);”

	[...] “d.1) vedação à inclusão, nos atos convocatórios, de cláusulas e condições que comprometam, restrinjam ou frustrem o caráter competitivo da licitação (arts. 3º, § 1º, inciso I), tal como a exigência de capacidade técnica do licitante para a execução de parcelas de serviços de natureza especializada que não tem maior relevância e valor significativo, nos termos do 30, §§ 1º e 2º, inciso I;” VOTO: “2. A limitação constante do referido inciso I do § 1º do art. 30, que grifamos, tem o nítido propósito de não possibilitar a inclusão, em editais, de exigências desarrazoadas que venham a frustrar ou restringir o caráter competitivo do certame. Nesse sentido, trazemos à colação os ensinamentos de Marçal Justen Filho: ‘Também não se admitem requisitos que, restritivos à participação no certame, sejam irrelevantes para a execução do objeto licitado. Deve-se considerar a atividade principal e essencial a ser executada, sem maiores referências a especificações ou detalhamentos. Isso não significa afirmar que tais peculiaridades sejam irrelevantes. São significativas para a execução do objeto, mas não para a habilitação.’ (in Comentários à Lei de Licitações e Contratos Administrativos, 5ª ed., p. 312). 3. Veja-se que a exigência de que fosse apresentada comprovação de habilitação técnica para a execução de rede de 69 KV se deu com inobservância à limitação constante do referido inciso I do § 1º do art. 30 e revelou-se restritiva e inoportuna: restritiva, porque resultou na inabilitação de licitantes; inoportuna, porque os serviços relativos à rede de 69KV não representavam, nos contratos originais, sequer 3,8% de seu valor total. Computando-se os aditamentos contratuais realizados em momentos posteriores, a representatividade de tais serviços tornou-se ainda menor, situando-se, nos três contratos, abaixo dos 3,4%. 4. A desnecessidade daquela exigência pode ser comprovada mediante a análise da execução dos contratos: em 11.01.2001 já haviam sido executados 99,08%, 94,75% e 96,82% dos contratos decorrentes das Tomadas de Preços nºs 086/99, 087/99 e 089/99, respectivamente, sem que tivesse sido cumprido nenhum dos itens de materiais e serviços relativos às redes de 69KV, inicialmente previstos. 5. A verificação do conjunto dos fatos trazidos aos autos fornece-me a convicção de que a inclusão de tal exigência se prestava exclusivamente a restringir o certame a um grupo restrito de empresas.” ACÓRDÃO: “b) restrição ao caráter competitivo dos certames, caracterizada pela exigência de prova de qualificação técnica para a execução de parcelas de serviços de menor relevância e valor, contrariando o comando contido no inciso I do § 1º do art. 30 da Lei nº 8.666/93;”
Precedente decorrente da técnica de distinção: AC-000244-2003-PL, que dá concreção à classe 7.1 – Atestados em número mínimo, máximo ou fixo	RELATÓRIO: “III.3 Exigência de que os licitantes comprovassem sua qualificação técnica com apenas três atestados, limitação essa não prevista em lei (infração ao art. 30, § 5º, da Lei nº 8.666/93)” [...] “O art. 30, § 5º, da Lei nº 8.666/93 veda que sejam feitas quaisquer exigências relativas à comprovação de atividade ou de aptidão que não estejam previstas em lei:” [...] “O item 5.5.1.3 do edital estabeleceu que as empresas proponentes deveriam comprovar com, no máximo, três atestados a execução dos seguintes serviços (fls. 28/29):” [...] “Observa-se que, além do estabelecimento de quantidades mínimas, houve a limitação do número de atestados com os quais se deveria comprovar a execução dessas quantidades. Não se vai entrar aqui na discussão a respeito da exigência de quantidades mínimas, pois aceita-se que, em certos casos, na análise da capacidade técnico-operacional de uma empresa, deva-se considerar não somente a qualidade, mas também o volume de serviços já executados em obras similares. No caso em análise, os interessados deveriam satisfazer quantidades mínimas de sete itens com apenas três atestados. Faria sentido inabilitar uma empresa que pudesse

	comprovar todos os requisitos, não com três, mas com quatro ou cinco atestados? Supondo que uma empresa tivesse realizado extração de pedras (item b) num contrato, transporte terrestre (item c) noutra e colocação de pedras (item d) num terceiro. Haveria alguma dificuldade de realizar os três itens numa única obra? O bom senso diz que não. Supondo que uma empresa tivesse executado os três itens num só contrato, ainda teria de comprovar a execução das quantidades mínimas dos outros quatro itens em, no máximo, mais dois atestados. Isso sem contar que foi exigida a execução prévia de transporte marítimo de pedras, quando o edital não colocava esse serviço como obrigatório, mas facultativo. Entendo que os dispositivos legais citados pelo responsável não suportam essa exigência. O art. 40, XVII, diz que o edital conterá XVII - outras indicações específicas ou peculiares da licitação. Já o caput do art. 41 estabelece que a Administração não pode descumprir as normas e condições do edital, ao qual se acha estritamente vinculada. É evidente que a Administração vincula-se às normas estabelecidas no instrumento convocatório; isso não significa que ela possa criar quaisquer normas, ainda mais se inibidoras da participação. Também aqui não se pode avaliar o efeito dessa limitação no resultado da concorrência. Duas empresas (Estacon Engenharia S/A e Bandeirantes Dragagens e Construção Ltda.) chegaram a apresentar quatro atestados, mas esse fato não foi a causa determinante de suas inabilitações, pois, mesmo se considerados, não conseguiriam comprovar todas as exigências da habilitação técnica. A primeira chegou a recorrer da decisão de inabilitação e teve seu recurso julgado procedente na parte que atacava o número máximo de atestados, mas improcedente noutros itens. Essa empresa conseguiu, por meio de medida judicial, que sua proposta comercial fosse aberta, mas, mesmo assim, não logrou ter o melhor preço. Mesmo que não se possa aferir o efeito real da exigência no resultado da licitação - pois jamais saberemos se alguma empresa simplesmente deixou de participar da licitação por esse motivo - não há como negar que um edital que limita o número máximo de atestados é mais restritivo que aquele que não faz limitação alguma, ou aceita um número maior que o primeiro. E essa restrição, no entendimento deste analista, é ilegal. Como já disse Marçal Justen Filho, comentando o art. 30 da Lei nº 8.666, ao analisar o problema dos limites a exigências quanto à capacitação técnico-operacional: ‘(...) o intérprete/aplicador tem de considerar que o processo de produção normativa, disciplinado pela Constituição, orienta-se pelo princípio de restrição mínima possível. A Constituição não defere ao administrador a faculdade de, ao discriminar as condições de habilitação, optar pela maior segurança possível.’ (obra citada, pág. 309)” VOTO: “42.Haja vista a possibilidade de restrição à competitividade, em razão da exigência no Edital de Concorrência nº 039/00 de que os licitantes possuísem em seu quadro funcional permanente responsável técnico, na data da publicação do edital, item 5.5.1.2. (fl. 27, V.p.), contrariando o art. 30, II c/c seu §1º, I da Lei 8.666/93, e comprovassem sua qualificação técnica com no máximo três atestados, item 5.5.1.3., “OBS” (fls. 28/29, V.p.), em detrimento do art. 30, §5º c/c o art. 3º, §1º, I da referida Lei, determinei a oitiva da licitante vencedora para que se pronunciasse, tendo em vista a possibilidade de anulação do contrato, caso configurada a prefalada restrição.” ACÓRDÃO: “9.2.4.quando realizar licitações envolvendo a utilização de recursos federais, observe rigorosamente a Lei nº 8.666/93, evitando que se repitam as seguintes práticas observadas na Concorrência 039/00:” [...] 9.2.4.3.instituir limitações não previstas em lei que inibam a participação na licitação, a exemplo da exigência de que os licitantes comprovem sua qualificação técnica com apenas três atestados (art. 30, § 5º da Lei nº 8.666/93);
Precedente decorrente da técnica de distinção: AC-001284-2003-PL, que dá concreção à	RELATÓRIO: “1.a.2) quando da fixação dos quantitativos mínimos já executados, não estabeleça percentuais mínimos acima de 50% (cinquenta por cento) dos quantitativos dos itens de maior relevância da obra/serviço, salvo em casos excepcionais, cujas justificativas para tal extrapolação deverão estar tecnicamente explicitadas, ou no processo

classe 7.2 – Quantitativo exagerado de experiência mínima	licitatório, previamente ao lançamento do respectivo Edital, ou no próprio Edital e seus anexos, em respeito ao inciso XXI do art. 37 da Constituição Federal; inciso I do § 1º do art. 3º e inciso II do art. 30 da Lei nº 8.666/93;” VOTO: “Também estou de acordo com a determinação àquela secretaria para que, em futuras licitações envolvendo recursos federais, observe procedimentos estabelecidos na Lei 8.666/93 e art. 37 da Constituição Federal por ocasião da avaliação da qualificação técnico-operacional e da qualificação econômico-financeira das empresas licitantes, objetivando permitir a participação de maior número de empresas potencialmente interessadas, e assim, garantir a observância do princípio constitucional da isonomia e selecionar a proposta mais vantajosa para a administração.” ACÓRDÃO: “9.1. determinar à Secretaria de Estado da Justiça e Cidadania de Santa Catarina - SSPDC/SC que:” [...] “9.1.2.1.2. em relação à fixação dos quantitativos mínimos já executados, não estabeleça percentuais mínimos acima de 50% dos quantitativos dos itens de maior relevância da obra ou serviço, salvo em casos excepcionais, cujas justificativas para tal extrapolação deverão estar tecnicamente explicitadas, ou no processo licitatório, previamente ao lançamento do respectivo edital, ou no próprio edital e seus anexos, em observância ao inciso XXI do art. 37 da Constituição Federal; inciso I do § 1º do art. 3º e inciso II do art. 30 da Lei 8.666/93;”
Precedente decorrente da técnica de distinção: AC-001140-2005-PL, que dá concreção à classe 7.4 – Experiência específica demais	RELATÓRIO: “d) exigência de que os atestados de execução de serviços anteriores demonstrem que a empresa atuou como contratada principal, prevista no item 9.2 do edital;” [...] “4.28 O item 9.2 do edital impugnado exige que os atestados que comprovam a capacitação técnico-profissional devem estar em nome da licitante e como contratada principal. A Lei 8.666/93 não permite que se faça esta distinção.” [...] “4.29 Destarte, os requisitos que o legislador reputou importantes para a comprovação da qualificação técnica são a pertinência e a compatibilidade entre os atestados e o objeto da licitação. Ou seja, os atestados devem mostrar que o licitante executou obras parecidas, e não iguais, em quantidade e prazos compatíveis com aquela que está sendo licitada. Quaisquer outras exigências que limitem a competitividade são vedadas pela Lei 8.666/93. Então, a exigência de que os atestados demonstrem que as licitantes executaram obras como contratadas principais é vedada pela lei. O importante é que a empresa tenha executado obras semelhantes, não sendo relevante se como contratada principal ou como subcontratada. 4.30 Assim, com relação a este item do edital, devem ser rejeitadas as razões de justificativa do responsável ouvido em audiência e determinado ao DNIT, em consonância com o art. 45 da Lei 8.443/92 e art. 251 do Regimento Interno do TCU, que adote providências com vistas à adequação do edital da Concorrência n.º 042/02-00 de forma a suprimir a exigência de que os atestados mostrem que a empresa atuou como contratada principal, prevista no item 9.2 da referida concorrência, pois isto fere o art. 37, inciso XXI da Constituição Federal e os arts. 3º, § 1º, inciso I e 30, § 5º, da Lei 8.666/93.” VOTO: “8. Quanto às demais questões apresentadas nesta representação, compartilho e incorporo a meu Voto o entendimento manifestado pela 1ª Secex, no sentido de serem ilegais as seguintes exigências previstas no Edital do DNIT:” [...] “b) necessidade de que os atestados de execução de serviços anteriores demonstrem que a licitante atuou anteriormente como contratada principal, por contrariar além dos dispositivos citados no item anterior o disposto no art. 30 da Lei nº 8.666/93;” ACÓRDÃO: “9.2. determinar ao Departamento Nacional de Infra-Estrutura de Transportes - DNIT que abstenha-se de estabelecer em suas licitações as seguintes exigências ou condições:”

	[...] “9.2.2. necessidade de que os atestados de execução de serviços anteriores demonstrem que a licitante atuou anteriormente como contratada principal, por contrariar além dos dispositivos citados no item anterior o disposto no art. 30 da Lei nº 8.666/93;”
Precedente decorrente da técnica de distinção: AC-001517-2005-PL, que dá concreção à classe 7.8 – Reorganização empresarial: fusão, cisão e incorporação	RELATÓRIO: “A justificativa, entretanto, não deve ser acatada. Em primeiro lugar, não consta do texto do edital da concorrência (fls. 334/351) a previsão contida na referida cláusula décima-nona, item IV, do contrato (fl. 249), que admite a transferência do contrato no todo ou em parte, desde que com a anuência da contratante. Logo, tal previsão contratual não pode ser considerada válida por caracterizar inovação em relação às normas editalícias.” [...] “inconstitucionalidade e ilegalidade da divisão das responsabilidades da contratada, mesmo que de forma solidária, tornando inócua qualquer previsão divergente porventura contida no edital ou no contrato.” VOTO: Não há trecho específico correspondente. ACÓRDÃO: Não há trecho específico correspondente.
Precedente decorrente da técnica de distinção: AC-002297-2005-PL, que dá concreção à classe 7.13 – Profissional do “quadro permanente”	RELATÓRIO: “17. As justificativas apresentadas e os documentos juntados nos permite acreditar que a cláusula restritiva, objeto do subitem 4.4.1, letra 'b' do edital da Tomada de Preços 2005/1909/1149, não é incompatível com o objeto da licitação. A exigência de que a licitante deva possuir, em seu quadro permanente de pessoal, na data da licitação, ENGENHEIRO, com curso de formação elétrica e/ou eletrônica e experiência comprovada através de cópia autenticada de certidão (ões) de Acervo Técnico - C.A.T é aceitável, e não restringiu à competitividade do certame, haja vista o número de empresas que atenderam a essa exigência. 18. As informações contidas nos autos nos leva a pensar que a empresa Sentran pretende subcontratar os serviços objetos da licitação. A possibilidade de subcontratação é conferida pela art. 72 da Lei nº 8.666/93. Em seus comentários a este artigo, Marçal Justen Filho assim se posiciona: 'A execução da prestação pelo próprio contratado não se impõe como exigência meramente subjetiva da Administração. (...) O que interessa à Administração é o recebimento da prestação exatamente idêntica àquela ofertada na proposta vencedora (...) Os riscos de receber uma prestação mal executada (...) conduzem a Administração a exigir que o próprio licitante desempenhe as tarefas(...)A lei autoriza, porém, que a Administração, em cada caso, avalie a conveniência de permitir a subcontratação, respeitados limites predeterminados. (...) A escolha da Administração deve ser orientada pelos princípios que regem a atividade privada. Se, na iniciativa privada, prevalece a subcontratação na execução de certas prestações, o ato convocatório deverá albergar permissão para que idênticos procedimentos sejam adotados(...) Assim se impõe porque, estabelecendo regras diversas das praticadas entre os particulares, a Administração reduziria a competitividade do certame.' 19. O edital em questão não prevê a subcontratação. Ora se o Banco do Brasil pretende que a empresa a ser contratada desempenhe as tarefas objeto do certame licitatório em análise, por razões de segurança e em obediência ao disposto no art. 2º da Lei n. 7.012/83, a representante não poderia executar os serviços a serem contratados, haja vista não dispor, em seu quadro, de profissional qualificado na área de circuito fechado de TV. Posto isso, considerando que a cláusula restritiva, objeto do subitem 4.4.1, letra 'b' do edital da Tomada de Preços n. 2005/1909/1149, não é exigência supérflua ou excessiva que tenha reduzido o universo de licitantes, por conseguinte, não teria o condão de invalidar o certame licitatório ora em análise;” VOTO: “6. Contudo, a questão que ora se põe não se encontra adstrita a esse ponto. O que se discute no presente feito, e o que originou a representação sob exame, é a exigência feita pelo Banco do Brasil de que as licitantes possuam engenheiro em seu quadro permanente mediante registro na Carteira de Trabalho, no Livro de Registro de Funcionários ou, ainda, no Contrato Social, no caso de sócio da Empresa.

	7. Conforme ressaltei na oportunidade em que apreciei a medida cautelar, a Lei nº 8.666/93 disciplinou de modo minucioso a matéria referente à qualificação técnica e uma das características mais marcantes dessa norma foi a redução da margem de liberdade da Administração Pública nesse campo e a limitação do âmbito das exigências. Com isso, buscou-se evitar que exigências formais e desnecessárias acerca da qualificação técnica se tornassem instrumento de restrição indevida à liberdade de participação em procedimentos licitatórios. 8. O artigo 30, § 1º, inciso I, da Lei nº 8.666/93, utiliza a expressão "qualificação técnico-profissional" para indicar a existência, nos quadros permanentes de uma empresa, de profissionais em cujo acervo técnico conste a responsabilidade pela execução de obras ou serviços similares àqueles aspirados pelo órgão ou entidade da Administração. 9. Todavia, há que se atentar para o fato de que a Lei nº 8.666/93 não define o que seja "quadro permanente". Assim, essa expressão poderia ser compreendida como o conjunto de pessoas ligadas à empresa de modo permanente, sem natureza eventual, por meio de vínculos de natureza trabalhista e/ou societária. Esse conceito, entretanto, reclama certa ampliação nas hipóteses em que a autonomia no exercício da profissão descaracteriza o vínculo empregatício sem afastar a qualificação do sujeito como integrante do quadro permanente, como é o caso dos profissionais da área de engenharia. 10. A exigência de que as empresas concorrentes possuam vínculo empregatício, por meio de carteira de trabalho assinada, com o profissional técnico qualificado mostra-se, ao meu ver, excessiva e limitadora à participação de eventuais interessados no certame, uma vez que o essencial, para a Administração, é que o profissional esteja em condições de efetivamente desempenhar seus serviços no momento da execução de um possível contrato. Em outros termos, o sujeito não integrará o quadro permanente quando não estiver disponível para prestar seus serviços de modo permanente durante a execução do objeto do licitado.” ACÓRDÃO: “9.3.1. observe, no momento da abertura de novo procedimento licitatório, os dispositivos da Lei nº 8.666/1993 relativos aos princípios norteadores e ao caráter competitivo dos procedimentos licitatórios, de modo a se evitar que exigências formais e desnecessárias acerca da qualificação técnica se tornem instrumento de restrição indevida à liberdade de participação de possíveis interessados;”
Precedente decorrente da técnica de distinção: AC-002081-2007-PL, que dá concreção à classe 7.14 – Quantitativos mínimos em experiência do profissional	RELATÓRIO: “7.7.2. Analisando o disposto nesse artigo da Lei 8.666/93, verifica-se, no que se refere-se à capacitação técnico-profissional, uma certa inadequação do comando para outros serviços que não sejam os vinculados à área de engenharia, a exemplo de serviços de advocacia, que nem sempre podem ser comprovados mediante atestados de responsabilidade técnica. Porém, podem ser comprovados mediante listagens obtidas junto a escritanias judiciais de varas e tribunais, em que fiquem registradas os nomes dos advogados que comporão a equipe técnica que executará os trabalhos, a exemplo do que foi previsto no subitem 9.1.2 do Edital do presente Pregão. 7.7.3. Nas contratações de serviços advocatícios, também se aplicam as vedações quanto ao estabelecimento de quantidades mínimas ou prazos máximos, podendo, contudo, a Administração exigir documentação que comprove possuir a licitante aptidão para desempenho da atividade, pessoal técnico adequado e disponível para a realização do objeto da licitação, bem como a qualificação de cada um dos membros da equipe técnica que se responsabilizará por trabalhos considerados de alta complexidade e relevância. 7.7.4 Ou seja, a Administração ao optar pela licitação do tipo menor preço, pode fazer constar do edital a exigência de apresentação de documentação, inclusive indicando tipo e natureza, por parte da licitante, a exemplo do contido no item 9.1.2 do Anexo 1 do Edital, que comprovem: a) aptidão para desempenho de atividade pertinente e compatível em características, quantidades e prazos com o objeto da licitação, desde que não sejam exigidos quantitativos mínimos e prazos máximos; b) indicação das instalações e do aparelhamento e do pessoal técnico adequados e disponíveis para a realização do objeto da licitação; e c) qualificação de cada um dos membros da equipe técnica que se responsabilizará pelos trabalhos.

	8. Analisando os termos dos subitens 9.1.2.1 e 9.1.2.2 do Edital, já com as alterações propostas pelos responsáveis, verificamos que permanecem as estipulações de quantidades mínimas ou de prazos máximos vedadas no Art. 30, § 1º e inciso I, da Lei 8.666/93, vez que foi mantida a restrição à comprovação de aptidão que se refira à atividades anteriores aos últimos 3 anos, bem como a exigência de que a sociedade comprove ter patrocinado 400 (quatrocentas) ações, sendo 50 na área trabalhista e 30 na área cível.” VOTO: “4. Verifico que as exigências de quantidade mínima de ações patrocinadas pela sociedade e por advogado (400 ações), prazo (nos últimos 3 anos) e titulação de pós-graduação ou cursos assemelhados (no mínimo 4 advogados) são todas relativas à capacidade técnico-profissional, incidindo, portanto, a vedação disposta no art. 30, § 1º, inciso I, da Lei nº 8.666/1993. 5. Por serem restritivas, atentando contra o princípio da isonomia, da legalidade, da competitividade e da razoabilidade, insculpidos no art. 37, inciso XXI, da Constituição Federal de 1988 e no art. 3º, caput e § 1º, inciso I, da Lei de Licitações e Contratos, podem conduzir à anulação do processo licitatório.” ACÓRDÃO: “9.3.1. suprimir, nos itens 9.1.2.1 e 9.1.2.2 do Anexo I - Termo de Referência, as exigências de quantidades mínimas, referentes à capacidade técnico-profissional, dada sua vedação disposta no art. 30, § 1º, inciso I, da Lei nº 8.666/1993;”
Precedente decorrente da técnica de distinção: AR-000451-2010-PL, que dá concreção à classe 7.7 – Atestado do mesmo grupo econômico	RELATÓRIO: Não há trecho específico correspondente. VOTO: Não há trecho específico correspondente. ACÓRDÃO: “Considerando que a representante alegou que: i) os documentos de habilitação de outra licitante teriam sido apresentados fora do prazo definido no edital; ii) o tipo de licitação deveria ter sido o de melhor preço e técnica em razão da complexidade da atividade; iii) o atestado de capacidade técnica da vencedora seria imprestável ao fim a que se destinava, pois não teria atendido ao edital, pois não constava a época em que as atividades haviam sido realizadas; iv) a apresentação de atestado emitido por empresa pertencente ao mesmo grupo econômico seria ilegítima; v) os critérios utilizados pela Caixa Econômica Federal teriam sido discrepantes daqueles do Pregão Eletrônico nº 116/7855-2009, quando a representante teria sido inabilitada por formalismo exagerado.” [...] “Considerando que, em relação à alegação de que o atestado de capacidade técnica não poderia ter sido emitido por empresa do mesmo grupo econômico, tendo sido observado que não havia vedação na Lei de Licitações nem no edital do pregão e que controlada e controladora conservam personalidade e patrimônio distintos.”
Precedente decorrente da técnica de distinção: AC-000584-2013-PL, que dá concreção à classe 7.5 – Experiência genérica demais	RELATÓRIO: “Item 12.5.4. A empresa deverá apresentar pelo menos um Atestado de Capacidade Técnica, expedido por entidades de direito público ou privado. Deverá ser comprovada a capacidade técnica da licitante para a execução dos serviços objeto deste edital, através de um atestado que contenha serviço de porte e complexidade SEMELHANTE, limitadas estas às parcelas de maior relevância e valor significativo do objeto da licitação.” [...] “Quanto ao mérito, entendeu a unidade técnica que: (...) assiste razão à representante no que diz respeito à descrição genérica atribuída ao item 12.5.4 do edital, o qual estabelece as exigências para comprovação da capacidade técnica dos licitantes. A redação atribuída ao item guarda conformidade com a Súmula 263 deste Tribunal, que pela sua aplicação geral, contempla de forma genérica as exigências admitidas para comprovação da capacidade técnica. Em cada caso específico, no entanto, tais exigências devem ser estabelecidas de forma objetiva e devem ser proporcionais à dimensão e à complexidade do objeto a ser executado.” VOTO: “Como se vê, realizando-se o cotejo entre os motivos do ato administrativo que inabilitou a empresa representante e o teor do item 12.5.4 do edital do Pregão

	Eletrônico nº 481/2012, verifica-se que as exigências impostas pela administração no que diz respeito ao porte da empresa e ao quantitativo de funcionários não encontram amparo no próprio edital, que, como se sabe, é a lei do certame, vinculando tanto a administração quanto as licitantes.” ACÓRDÃO: “[...] que a análise da qualificação técnica se dê por meio de atestados que comprovem a capacidade da empresa em executar os serviços objeto do contrato em quantitativo e prazo semelhantes [...]”
Precedente decorrente da técnica de distinção: AC-002205-2014-2C, que dá concreção à classe 7.6 – Limitação de tempo e época nos atestados	RELATÓRIO: “41. De acordo com a publicação do TCU intitulada “Licitações e Contratos”, na esteira do art. 30, § 5º, da Lei 8.666/1993 é vedada a exigência de atestados com limitação de época a não ser “quando a tecnologia a ser adotada só se tornou disponível a partir do período indicado” (BRASIL. Tribunal de Contas da União, ed. 4, Brasília/Secretaria Especial de Editoração e Publicações do Senado Federal, 210, p. 407). 42. Portanto, a limitação de época deve ser admitida excepcionalmente e quando devidamente justificada, a qual, no caso, foi exigida em atividades de “reestruturação empresarial ou de reengenharia de processos, ou, ainda, de transformação/reconversão do negócio em portos localizados no Brasil ou exterior”. 43. Não se vislumbra, a princípio, grandes inovações surgidas nessas áreas nos últimos cinco anos, razão pela qual, em juízo de cognição sumária, a limitação de época em tela soa desarrazoada.” VOTO: 14. Sobre a limitação de época para os atestados de habilitação técnica, o edital estabeleceu que a experiência comprovada deveria se referir a trabalhos realizados nos últimos cinco anos, o que foi inicialmente considerado como exigência indevida, incompatível com o art. 30, § 5º, da Lei 8.666/1993. 15. Na linha da análise realizada pela SefidTransporte, considero que os esclarecimentos prestados pela SEP podem ser acolhidos. A vedação à exigência de atestados com limitação de época pode ser temporizada nas situações em que a tecnologia envolvida só se tornou disponível a partir do período indicado. Conforme apontado pela unidade técnica, mudanças tecnológicas nos processos desenvolvidos nas áreas portuárias, a exemplo da containerização de produtos agrícolas, “ganham força em um passado recente, de modo que a restrição dos atestados de capacidade técnica a atividades prestadas pelos interessados nos últimos cinco anos pode ser considerada razoável.” 16. Demonstrada a adequação e pertinência da exigência em relação ao objeto licitado, não se caracterizou a restrição ao caráter competitivo da licitação (p.ex., 1.417/2008 – Plenário). No caso concreto, as alegações trazidas aos autos são plausíveis e não interferiram diretamente no certame, já que a empresa que ofertou o menor lance foi tecnicamente habilitada. 17. No entanto, é essencial que justificativas dessa natureza, por seu caráter excepcional, sejam especificadas e fundamentadas em estudos técnicos que constem do processo de licitação. Por essa razão, apresento proposta de ciência à unidade sobre esse assunto, para aprimoramento de futuros certames.” ACÓRDÃO: “9.2.2. ausência de justificativas específicas e fundamentadas em estudos técnicos que constem do processo de licitação para exigência de comprovação de atividades com limitações de tempo ou de época, o que caracteriza violação do §5º do art. 30 da Lei 8.666/1993;”
Precedente decorrente da técnica de distinção: AC-002208-2016-PL, que dá concreção à classe 7.9 – Transferência de capacidade técnico-profissional para operacional	RELATÓRIO: “31. Ademais, a transferência de acervo técnico de pessoa física à pessoa jurídica pode ensejar o possível ‘comércio’ de acervo, permitindo assim que empresas aventureiras participem de licitação sem que possuam a real capacidade de executar o objeto, apenas pela simples formalização de contrato com responsável técnico detentor da qualificação requerida.” VOTO: “10. No que se refere ao segundo ponto, também foi confirmado o posicionamento da Secex/BA quanto à improcedência das alegações da representante. Segundo registrado na instrução, não há fundamento para se aceitar a transferência do acervo

	técnico da pessoa física para a pessoa jurídica, como permitido pelo CFA. Foi defendido que a capacitação técnico-operacional da empresa não se confunde com a capacitação do profissional, uma vez que a primeira abrange também as instalações, o aparelhamento, as metodologias de trabalho e os processos internos de controle de qualidade, entre outros aspectos. A unidade instrutiva considerou que, nesse contexto, não há garantia de que o simples fato de a empresa contar com o profissional irá resultar na execução satisfatória do serviço, já que outros fatores são necessários para a adequada prestação. 11. A unidade técnica ponderou que, no tocante às licitações públicas, a aplicação da norma pode resultar em afronta ao interesse público ante a possibilidade de que empresas tecnicamente despreparadas apresentem atestados de qualificação de titularidade de outra empresa em razão da transferência do profissional.” ACÓRDÃO: Não há trecho específico correspondente.
Precedente decorrente da técnica de distinção: AC-002208-2016-PL, que dá concreção à classe 7.15 – CAT em nome da empresa licitante	RELATÓRIO: “21. Nos termos da Lei 8.666/1993, a comprovação de capacidade técnica se dá a partir do fornecimento, por pessoa jurídica de direito público ou privado, de documento que ateste que a licitante desempenhou satisfatoriamente serviços similares aos que estão sendo licitados. 22. Conforme consignado em instrução da preliminar, a capacitação técnico-profissional e a capacitação técnico-operacional não se confundem e a certificação emitida pelos CRA’ s conjugando acervos técnicos de diferentes naturezas poderá levar a Administração Pública a contratar empresas que não tenham a qualificação necessária para executar o contrato satisfatoriamente. 23. Enquanto a capacitação técnico-profissional está relacionada à qualificação do corpo técnico, a capacitação técnico-operacional, por sua vez, é bem mais ampla e alcança requisitos empresariais, tais como estrutura administrativa, métodos organizacionais, processos internos de controle de qualidade, etc. Na prática, a qualificação comprovada de um profissional não é suficiente para garantir a experiência operacional da empresa à qual esse profissional esteja vinculado, seja na condição de prestador de serviço ou na condição de sócio, e, conseqüentemente, a qualidade da execução contratual poderá ser comprometida. 24. Em síntese, a previsão de que o acervo profissional possa ser acrescido ao acervo técnico de pessoa jurídica, como qualificação técnico-operacional, para fins de atendimento da exigência contida no art. 30, II, da Lei 8.666/1993, não tem amparo legal, uma vez que tornaria inócuos os dispositivos legais que estabelecem clara distinção entre os dois institutos.” VOTO: “15. Observa-se que o cerne da justificativa tomada pelo CFA reside na confusão entre os conceitos de capacidade técnico-operacional (art. 30, inciso II, da Lei 8.666/1993) e de capacidade técnico-profissional (art. 30, §1º, inciso I). Ressalte-se, contudo, que a distinção entre esses dois conceitos apresenta-se estabelecida na Lei de Licitações. 16. A qualificação técnico-operacional corresponde à capacidade da empresa, visto que o dispositivo que trata do assunto, o art. 30, inciso II, da lei, refere-se a aspectos típicos desse ente, como instalações, equipamentos e equipe” ACÓRDÃO: Não há trecho específico correspondente.

Fonte: Dados da pesquisa.

Por fim, com a jurisprudência (171 acórdãos) acerca de cláusulas restritivas em editais de licitação que exigiram capacidade técnica dos licitantes de maneira irregular, construiu-se a provável *linha jurisprudencial de entendimento* para essa patologia corruptiva, apresentado no Apêndice H, com os precedentes destacados na cor verde.

5 CONCLUSÃO

Esta pesquisa empírica teve como objetivos principais: (a) *criar um método replicável de delineamento de jurisprudência e identificação de precedentes com uso de IA* aplicável, sobretudo, a decisões do TCU que versem sobre cláusulas restritivas ao caráter competitivo em editais de licitação; e (b) *contribuir para a sedimentação da cultura de precedentes administrativos no país e para a delimitação jurisprudencial de conceitos jurídicos indeterminados*, mormente aqueles ligados a práticas corruptas.

Para tanto, testou-se a aplicação de técnicas de IA nos acórdãos do TCU (entre 1992 e 2018), com intuito de *sistematizar a delimitação jurisprudencial e consequente identificação inequívoca dos precedentes administrativos que dão concreção ao conceito aberto de cláusulas restritivas ao caráter competitivo em editais de licitação*.

Visto tratar-se de um estudo interdisciplinar, que congrega os campos de conhecimento da *Linguística*, do *Direito* e da *Ciência da Computação*, inicialmente foi preciso aprofundar nos marcos teóricos que nortearam o olhar para a problemática da pesquisa.

Este trabalho ancorou-se nos *dois momentos de Ludwig Wittgenstein*: o primeiro, em que o filósofo austríaco descreve a *função representativa da linguagem*, para a qual as expressões linguísticas possuem uma forma lógica da qual se extrai seus significados; o segundo, em que o autor estabelece a “*the use theory of meaning*”, na qual o significado de uma expressão linguística é definido pelas circunstâncias de seu uso e não apenas de sua forma lógica.

Do ponto de vista jurídico, apoiou-se na *Teoria dos Modelos do Direito* de Miguel Reale, uma vez que os *modelos jurídicos jurisdicionais* promovem soluções para eficácia quando da aplicação das diferentes regras jurídicas que compõem o ordenamento e que se relacionam com o caso concreto. Ainda, transportou-se da *Teoria dos Precedentes* de Thomas da Rosa de Bustamante, o *conceito de ratio decidendi* como o *conjunto de normas adscritas identificáveis a partir do modelo silogístico utilizado na justificação das decisões*.

Buscou-se, também, a conformação teórica da problemática da pesquisa no âmbito nas áreas práticas de *text mining* para Inteligência Artificial, concluindo tratar-se de um problema de *classificação de documentos* para delimitação jurisprudencial, cujos resultados subsidiariam a identificação dos precedentes administrativos almejados.

Dessa forma, *com o uso de IA e a partir do significado lógico das rationes decidendi que congregam decisões correlatas*, buscou-se delimitar a jurisprudência que versa sobre

cláusulas restritivas ao caráter competitivo em editais de licitações para, em seguida, considerando a função exercida por essas expressões linguísticas dentro do contexto do inteiro teor de cada acórdão, caracterizar os modelos jurídicos por meio da identificação dos precedentes administrativos que dão concretude ao conceito jurídico indeterminado em estudo.

Assentados os pilares teóricos da pesquisa, partiu-se, então, para o estudo do recorte temático escolhido para aplicação das técnicas de IA na extração jurisprudencial. Constatou-se que *a fraude em licitações é uma espécie do gênero corrupção e que uma de suas patologias é a existência de cláusulas nos editais que restringem a competitividade do certame.* Ademais, verificou-se que a expressão *cláusula restritiva é um conceito jurídico indeterminado de experiência*, que possui um núcleo de incerteza que deve ser delimitado a partir da praxe administrativa.

Nessa lógica, as *decisões colegiadas do TCU ganham destaque pois estabelecem critérios práticos* que permitem concluir por uma solução possível diante de um caso concreto. Sabendo disso, identificou-se em trabalhos empíricos anteriores 330 *acórdãos previamente rotulados em 11 classes* relacionadas à cláusulas de editais de licitação consideradas restritivas ao caráter competitivo de certames.

Após revisão do estado da arte computacional aplicado a problemas semelhantes, essas 330 decisões foram utilizadas para *treinar modelos de machine learning e deep learning* para classificação *multi-label* dos acórdãos do TCU, com vistas a constatar se a máquina seria capaz de apontar quais outras decisões do TCU (entre 1993 e 2018) se relacionavam a alguma dessas 11 classes.

Os resultados obtidos para as fases de treinamento e teste do modelo de DL utilizando redes neurais convolucionais (CNN) mostraram-se razoáveis, pois apresentaram métrica de avaliação de 82,69%. Entretanto, na etapa de avaliação supervisionada, o algoritmo de aprendizado profundo treinado resultou inconclusivo para a delimitação jurisprudencial almejada.

Considera-se que esta pesquisa logrou êxito, ao menos parcialmente, no atingimento de seus objetivos gerais. Explica-se.

Com relação ao primeiro objetivo geral, não obstante esses resultados insatisfatórios para a utilização de IA no delineamento da jurisprudência acerca do recorte temático escolhido, *foi possível traçar possíveis caminhos para estudos futuros.* Para tanto, analisou-se os resultados obtidos da validação das técnicas de *text mining* para a *classificação de documentos jurídicos* pelo prisma dos marcos teóricos balizadores da pesquisa, de modo a detalhar

procedimentos que podem ser aplicados para robustecer o *corpus* de acórdãos rotulados, tendo em vista que o baixo número de decisões classificadas pode ter sido um dos motivos do insucesso.

Apontou-se, também, para a possibilidade de migrar da área prática de *classificação de documentos* para problemas de *text mining*, para a área prática de *recuperação de informações* (*information retrieval*), com o propósito de estabelecer uma sistemática de busca de acórdãos similares a partir da identificação, a priori, de possíveis *rationes decidendi* relacionadas à temática buscada.

Portanto, constata-se que o *mapeamento do estado da arte computacional* aplicado ao problema da pesquisa (seção 4.1) e os *detalhes operacionais descritos* (seção 4.2 e apêndices) proporcionam *transferência de know-how para pesquisas que vislumbram a utilização de IA para sistematização jurisprudencial*.

Quanto ao segundo objetivo geral, a pesquisa *contribuiu para a sedimentação de uma cultura de precedentes administrativos no Brasil*, visto que demonstrou a *importância de rastrear sistematicamente os precedentes administrativos para que haja maior segurança jurídica* não apenas aos *gestores públicos* em suas ações administrativas, mas também aos *auditores de controle externo no aprimoramento dos algoritmos utilizados por aplicações de IA que vasculham possíveis irregularidades*, como o robô Alice do TCU.

Ademais, o trabalho realizado *contribuiu para a concreção de um conceito jurídico indeterminado associado fraudes em licitações*, uma vez que *foi apresentado um ensaio da construção da linha jurisprudencial de entendimento* (ver Apêndice H) com *indicação dos prováveis precedentes administrativos* e suas *rationes decidendi* para os casos de cláusulas, em editais de licitação, consideradas restritivas por exigirem capacidade técnica dos licitantes de maneira irregular (ver Tabela 9).

Em suma, esta pesquisa empírica deixa sua contribuição para aplicações tecnológicas que se aproximem do conceito de *computação cognitiva*³²⁰ e que podem ser a solução para a criação de uma futura *central de jurisprudência administrativa*³²¹.

³²⁰ Cf. Ashley, *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, op. cit., p. 12-14.

³²¹ Cf. Schiefler et al., *A inteligência artificial aplicada à criação de uma central de jurisprudência administrativa: o uso das novas tecnologias no âmbito da gestão de informações sobre precedentes em matéria administrativa*, op. cit., p. 23.

REFERÊNCIAS

- ADEODATO, João Maurício. *A retórica constitucional (sobre tolerância, direitos humanos e outros fundamentos éticos do direito positivo)*. 2. ed. São Paulo: Saraiva, 2010.
- _____. *Ética & Retórica: Para uma teoria da dogmática jurídica*. 4. ed. São Paulo: Saraiva, 2012.
- _____. *Uma teoria retórica da norma jurídica e do direito subjetivo*. 2. ed. São Paulo: Saraiva, 2014.
- ALKHATIB, W.; RENSING, C.; SILBERBAUER, J. Multi-label Text Classification using Semantic Features and Dimensionality Reduction with Autoencoders. In: *International Conference on Language, Data and Knowledge (LDK)*, I., 2017, Galway. *Proceedings*, p. 380-394. Springer, 2017.
- ALTOUNIAN, Cláudio Sarian. *Obras públicas: licitação, contratação, fiscalização e utilização*. 5. ed. Belo Horizonte: Fórum, 2016.
- ARAUJO, P. H. L.; CAMPOS, T. E.; BRAZ, F. A.; SILVA, N. C. VICTOR: a dataset for Brazilian legal documents classification. In: *Conference on Language Resources and Evaluation (LREC)*, XII., 2020, Marseille. *Proceedings*, p. 1449-1458. European Language Resources Association, 2020.
- ARUMUGAM, Rajesh; SHANMUGAMANI, Rajalingappaa. *Hands-On Natural Language Processing with Python*. 1. ed. Birmingham, UK: Packt Publishing, 2018.
- ASHLEY, Kevin D. *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*. 1. ed. Cambridge: Cambridge University Press, 2017.
- BANCO INTERAMERICANO DE DESENVOLVIMENTO. De estruturas a serviços: o caminho para uma melhor infraestrutura na América Latina e no Caribe. 2020. Disponível em: <https://publications.iadb.org/publications/portuguese/document/De-estruturas-a-servicos-O-caminho-para-uma-melhor-infraestrutura-na-America-Latina-e-no-Caribe.pdf>. Acesso em: 24 out. 2020.
- BAO Q., ZAN H., GONG P., CHEN J., XIAO Y. Charge Prediction with Legal Attention. In: *International Conference on Natural Language Processing and Chinese Computing (NLPCC)*, VIII., 2019, Dunhuang. *Lecture Notes in Computer Science*, v. 11838, p. 447-458. Springer, 2019.
- BENGFORT Benjamin; BILBRO, Rebecca; OJEDA, Tony. *Applied Text Analysis with Python*. 1. ed. Sebastopol: O'Reilly Media, 2018.
- BEZERRA, D. A.; INAZAWA, P. H. G.; ZUMBLIK, R.; CAMPOS, T. E.; SILVA, N. C.; BRAZ, F. A.; PEIXOTO, F. H. Descoberta de termos que caracterizam peças jurídicas. In: *The International Conference on Forensic Computer Science and Cyber Law (ICFCS)*, XI., 2019, São Paulo. *Proceedings*, p. 7-14. High Technology Crime Investigation Association, 2019.

BIRD, S.; KLEIN, E.; LOPER, E. *Natural Language Processing with Python*. 1. ed. Sebastopol: O'Reilly Media, 2009.

BOBBIO, Norberto. *Teoria do ordenamento jurídico*. Tradução Maria Celeste Cordeiro Leite Santos. 10. ed. Brasília: Editora UnB, 1999.

BRASIL. *Constituição Federal de 5 de outubro de 1988*. Disponível em: http://www.planalto.gov.br/ccivil_03/Constituicao/Constituicao.htm. Acesso em: 23 out. 2020.

_____. *Decreto-Lei nº 4.657, de 4 de setembro de 1942*. Lei de Introdução às normas do Direito brasileiro. Disponível em: http://www.planalto.gov.br/ccivil_03/decreto-lei/del4657.htm. Acesso em: 23 out. 2020.

_____. *Lei nº 8.666, de 21 de junho de 1993 (Lei Geral de Licitações)*. Regulamenta o art. 37, inciso XXI, da Constituição Federal, institui normas para licitações e contratos da Administração Pública e dá outras providências. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/L8666cons.htm. Acesso em: 23 out. 2020.

_____. *Lei nº 9.784, de 29 de fevereiro de 1999 (Lei do Processo Administrativo Federal)*. Regula o Processo Administrativo no âmbito da Administração Pública Federal. Disponível em: http://www.planalto.gov.br/Ccivil_03/leis/L9784.htm. Acesso em: 23 out. 2020.

_____. *Lei nº 13.655, de 25 de abril de 2018*. Inclui no Decreto-Lei nº 4.657, de 4 de setembro de 1942 (Lei de Introdução às Normas do Direito Brasileiro), disposições sobre segurança jurídica e eficiência na criação e na aplicação do direito público. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13655.htm#:~:text=LEI%20N%C2%BA%2013.655%2C%20DE%2025%20DE%20ABRIL%20DE%202018.&text=Inclui%20no%20Decreto%2DLei%20n%C2%BA,na%20aplica%C3%A7%C3%A3o%20do%20direito%20p%C3%BAblico. Acesso em: 23 out. 2020.

_____. *Lei nº 13.105, de 16 de março de 2015*. Código de Processo Civil. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2015/Lei/L13105.htm. Acesso em: 23 out. 2020.

_____. *Lei nº 12.846, de 1º de agosto de 2013 (Lei Anticorrupção)*. Dispõe sobre a responsabilização administrativa e civil de pessoas jurídicas pela prática de atos contra a administração pública, nacional ou estrangeira, e dá outras providências. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2013/lei/112846.htm. Acesso em: 23 out. 2020.

_____. *Lei nº 8.429, de 2 de junho de 1992*. Dispõe sobre as sanções aplicáveis aos agentes públicos nos casos de enriquecimento ilícito no exercício de mandato, cargo, emprego ou função na administração pública direta, indireta ou fundacional e dá outras providências. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l8429.htm. Acesso em: 23 out. 2020.

BUCCI, Maria Paula Dallari. O conceito de política pública em direito. In: BUCCI, Maria Paula Dallari (Org.) *Políticas Públicas: reflexões sobre o conceito jurídico*. São Paulo: Saraiva, 2006.

BUSTAMANTE, Thomas da Rosa de. *Teoria do Precedente Judicial*. 1. ed. São Paulo: Noeses, 2012.

CARVALHO, D. S.; NGUYEN, M.-T.; TRAN, C.-X.; NGUYEN, M.-L. Lexical-Morphological Modeling for Legal Text Analysis. In: *JSAI International Symposium on Artificial Intelligence (JSAI-isAI)*, 2015, Tsukuba. *Lecture Notes in Computer Science*, v. 10091, p. 295-311. Springer, 2017.

CHALKIDIS, I.; ANDROUTSOPOULOS, I.; ALETRAS, N. Neural Legal Judgment Prediction in English. In: *Annual Meeting of the Association for Computational Linguistics*, LVII., 2019, Florence. *Proceedings*, p. 4317-4323. Association for Computational Linguistics (ACL), 2019.

CHALKIDIS, I.; FERGADIOTIS, M.; MALAKASIOTIS, P.; ALETRAS, N.; ANDROUTSOPOULOS, I. Extreme Multi-Label Legal Text Classification: A case study in EU Legislation. *Computing Research Repository (CoRR)*, v. abs/1905.10892, 2019. Disponível em: <http://arxiv.org/abs/1905.10892>.

CHALKIDIS, I.; FERGADIOTIS, E.; MALAKASIOTIS, P.; ANDROUTSOPOULOS, I. Large-Scale Multi-Label Text Classification on EU Legislation. In: *Annual Meeting of the Association for Computational Linguistics*, LVII., 2019, Florence. *Proceedings*, p. 6314-6322. Association for Computational Linguistics (ACL), 2019.

COSTA, Marcos Bemquerer. BASTOS, Patrícia Reis Leitão. Alice, Monica, Adele, Sofia, Carina e Ágata: o uso da inteligência artificial pelo Tribunal de Contas da União. *Controle Externo: Revista do Tribunal de Contas do Estado de Goiás*, v. 2, n. 3, p. 11-34, 2020.

DENG, L.; YANG, L. (org.) *Deep Learning in Natural Language Processing*. 1. ed. Singapura: Springer, 2017.

DEVLIN, J.; CHANG, M.-W.; LEE, K.; TOUTANOVA, K. Bert: Pretraining of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, v. 2, 2018.

DI PIETRO, Maria Sylvia Zanella. *Discrecionalidade administrativa na Constituição de 1988*. 3. ed. São Paulo: Atlas, 2012.

DONG, H.; YANG, F.; WANG, X. Multi-label charge predictions leveraging label co-occurrence in imbalanced data scenario. *Soft Comput*, v. 24, p. 1-26, 2020.

EPSTEIN, Lee; KING, Gary. Pesquisa empírica em Direito. São Paulo: Direito GV, 2013. Disponível em: <http://bibliotecadigital.fgv.br/dspace/handle/10438/11444>. Acesso em: 8 fev. 2019.

FERRAZ JR., Tercio Sampaio. *Introdução ao Estudo do Direito: técnica, decisão, dominação*. 10. ed. São Paulo: Atlas, 2018.

FREITAS, Juarez. Direito Administrativo e inteligência artificial. *Interesse Público – IP*, n. 114, p. 15-29, 2019.

FURTADO, Lucas Rocha. *As raízes da corrupção no Brasil: estudos de caso e lições para o Futuro*. 1. ed. Belo Horizonte: Editora Fórum, 2015.

GIBAJA, E.; VENTURA, S. A tutorial on multilabel learning. *ACM Computing Surveys*, v. 47, n. 3, art. 52, abr. 2015.

GOANTA, Catalina; VAN DIJCK, Gijs; SPANAKIS, Gerasimos. Back to the Future: Waves of Legal Scholarship on Artificial Intelligence. In: RANCHORDÁS, Sofia; ROZNAI, Yaniv. *Time, Law and Change*. 1. ed. Londres: Hart Publishing, 2020.

GUSTIN, Miracy Barbosa de Sousa; DIAS, Maria Teresa Fonseca. *(Re)Pensando a pesquisa jurídica*. 4. ed. rev. e atual. Belo Horizonte: Del Rey, 2013.

HEINEN, Juliano. *Comentários à Lei Anticorrupção: Lei nº 12.846/2013*. 1. ed. Belo Horizonte: Editora Fórum, 2015.

HOWE, J. S. T.; KHANG, L. H.; CHAI, I. E. Legal Area Classification: A Comparative Study of Text Classifiers on Singapore Supreme Court Judgments. In: *Natural Legal Language Processing Workshop (NLLP)*, I., 2019, Minneapolis. *Proceedings*, p. 67-77. Association for Computational Linguistics (ACL), 2019.

INSTITUTO BRASILEIRO DE AUDITORIA DE OBRAS PÚBLICAS. *Orientação Técnica OT – IBR 002/2009: obra e serviço de engenharia*, p. 1-9. Disponível em: <http://www.ibraop.org.br/wp-content/uploads/2013/06/OT-IBR-02-2009-Ibraop-01-07-10.pdf>. Acesso em: 27 out. 2020.

KITCHENHAM, Barbara; CHARTERS, Stuart. Guidelines for performing Systematic Literature Reviews in Software Engineering. *EBSE Technical Report EBSE-2007-01*. Keele University and University of Durham, version 2.3, 2007.

KLOCK, A. C. T. Mapeamentos e Revisões Sistemáticos da Literatura: um Guia Teórico e Prático. *Revista Cadernos de Informática*, v. 10, n. 1, p. 1-9, 2018. Disponível em: <https://seer.ufrgs.br/cadernosdeinformatica/article/view/v10n1201801-09/49901>. Acesso em: 02 set. 2018.

KOWSARI, K.; MEIMANDI, K. J.; HEIDARYSAFA, M.; MENDU, S.; BARNES, L. E.; BROWN, D. E. Text Classification Algorithms: A Survey. *Information*, v. 10, art. 150, 2019.

LANDTHALER, J. *Improving Semantic Search in the German Legal Domain*. Tese de Doutorado (doutorado em Ciências Naturais) – Technische Universität München, Munique, Alemanha, 2020.

LEAL, Rogério Gesta; SCHNEIDER, Yuri. Os efeitos deletérios da corrupção em face dos Direitos Humanos e Fundamentais. *Revista da AJURIS*, v. 41, n. 136, p. 415-435, dez. 2014.

LI, Q.; PENG, H.; LI, J.; XIA, C.; YANG, R.; SUN, L.; YU, P. S.; HE, L. A Survey on Text Classification: From Shallow to Deep Learning. *ACM Computing Surveys*, v. 37, n. 4, artigo 35, jul. 2020.

LIU, J.; CHANG, W-C.; WU, Y.; YANG, Y. Deep Learning for Extreme Multi-label Text Classification. In: *International Conference on Research and Development in Information Retrieval (SIGIR)*, XL., 2017, Tokyo. *Proceedings*, p. 115–124. Association for Computing Machinery (ACM), 2017.

LIU, Y.-H.; CHEN, Y.-L.; HO, W.-L. Predicting associated statutes for legal problems. *Information Processing & Management*, v. 51, n. 1, p. 194-211, 2015.

LUO, B.; FENG, Y.; XU, J.; ZHANG, X.; ZHAO, D. Learning to Predict Charges for Criminal Cases with Legal Basis. In: *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2017, Copenhagen. *Proceedings*, p. 2727-2736. Association for Computational Linguistics (ACL), 2017.

LUVIZOTTO, Juliana Cristina. *Precedentes administrativos & a vinculação da atividade administrativa*. 2. ed. Curitiba: Juruá, 2017.

MAIRAL, Héctor A. *As raízes legais da corrupção: ou como o direito público fomenta a corrupção em vez de combatê-la*. KRAEMER, Susan M. B. (trad.). MUKAI, Toshio (comentários ed. brasileira). São Paulo: Editora Contracorrente, 2018.

MANNING, Christopher D.; SCHUTZE, Hinrich. *Foundations of statistical natural language processing*. 6. ed. Cambridge: MIT Press, 2003.

MARCONDES, Danilo. *Textos básicos de linguagem: de Platão a Foucault*. 1 ed. Rio de Janeiro: Zahar, 2010.

MARIANO, A. M; ROCHA, M. S. Revisão da Literatura: Apresentação de uma Abordagem Integradora. In: *AEDM International Conference – Economy, Business and Uncertainty: Ideas for a European and Mediterranean Industrial Policy*, XXVI., 2017, Reggio Calabria. *Proceedings*, p. 427-443. Asociación Española de Dirección y Economía de la Empresa (AEDM), 2017.

MARTIJA, M. A.; DOMOGUEN, J.; NAVAL, P. How Deep is Your Law? Predicting Associations between Cases in Philippine Jurisprudence. In: *IEEE Region 10 Conference (TENCON)*, 2019, Singapore. *Proceedings*, p. 886-891. Institute of Electrical and Electronics Engineers (IEEE), 2019.

MINER, G.; DELEN, D.; ELDER, J.; FAST, A.; HILL, T.; NISBET, R. *Practical text mining and statistical analysis for non-structured text data applications*. 1. ed. Waltham: Elsevier, 2012.

MIRANDA, Luiz Fernando. Unificando os conceitos de corrupção: uma abordagem através da nova metodologia dos conceitos. *Revista Brasileira de Ciência Política*, n. 25, p. 237-272, 2018.

MIRONCZUK, M. M.; PROTASIEWICZ, J. A recent overview of the state-of-the-art elements of text classification. *Expert Systems with Applications*, v. 106, n. 15, p. 36-54, set. 2018.

NAVES, Fernanda de Moura Ribeiro. *TCU e stare decisis administrativo: o caso das cláusulas restritivas em editais de licitação de obras de engenharia*. Dissertação (Mestrado em Direito e Políticas Públicas) – Universidade Federal de Goiás, Goiânia, 2019.

NEVES, F. R., SILVA, P. B., CARVALHO, H. L. M. Artificial ladies against corruption: searching for legitimacy at the Brazilian Supreme Audit Institution. *Revista de Contabilidade e Organizações*, v. 13, p. 31-50, 2019.

NGUYEN, G.; DLUGOLINSKY, S.; BOBÁK, M.; TRAN, V.; GARCÍA, Á. L.; HEREDIA, I.; MALÍK, P.; HLUCH, L. Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey. *Artificial Intelligence Review*, v. 52, n. 1, p. 77-124, jun. 2019.

NUCCI, Guilherme de Souza. *Corrupção e anticorrupção*. 1 ed. São Paulo: Forense, 2015.

OLIVEIRA, Rafael Carvalho Rezende. *Precedentes no Direito Administrativo*. 1. ed. Rio de Janeiro: Forense, 2018.

PALMA, Juliana Bonacorsi de. *Atividade normativa da Administração Pública: estudo do processo administrativo normativo*. Tese (Doutorado em Direito) – Faculdade de Direito da Universidade de São Paulo. São Paulo, 2014.

PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. *Inteligência Artificial e Direito*. v. 1, 1. ed. Curitiba: Alteridade Editora, 2019.

PENG, D.; WU, Q. LegalCap: a model for complex case discrimination based on capsule neural network. *Soft Comput*, v. 24, n. 21, p. 16043-16055, 2020.

POLO, F. M.; CIOCHETTI, I.; BERTOLO, E. Predicting Legal Proceedings Status: an Approach Based on Sequential Texts. In: *International Conference on Machine Learning (ICLR)*, XXXVII., 2019, Vienna. *Proceedings*, PMLR 108, 2020.

POPPER, Karl Raimund. *A lógica da pesquisa científica*. Tradução Leônidas Hegenberg; Octanny Silveira da Mota. 2. ed. São Paulo: Cultrix, 2013 [1972].

RAO, D.; MCMAHAN, B.. *Natural Language Processing with PyTorch*. 1. ed. Sebastopol: O'Reilly Media, 2019.

REALE, Miguel. *Fontes e modelos do Direito: para um novo paradigma hermenêutico*. 1 ed. São Paulo: Saraiva, 1994.

_____. *Teoria Tridimensional do Direito*. 5. ed. São Paulo: Saraiva, 1994.

RIBEIRO, Cássio Garcia; INÁCIO JÚNIOR, Edmundo. O mercado de compras governamentais brasileiro (2006-2017): mensuração e análise. *Texto para Discussão*, n. 2476,

p. 7-31, 2019. Disponível em:
http://repositorio.ipea.gov.br/bitstream/11058/9315/1/td_2476.pdf. Acesso em 23 out. 2020.

ROCHA, Ana Carolina Pereira. *Mineração de Textos para Classificação de Processos Judiciais Trabalhistas*. Dissertação (Mestrado Profissional em Computação Aplicada) – Universidade de Brasília, Brasília, 2019.

ROSCA, Constanta; COVRIG, Bogdan; GOANTA, Catalina; VAN DIJCK, Gijs; SPANAKIS, Gerasimos. Return of the AI: An Analysis of Legal Research on Artificial Intelligence Using Topic Modeling. In: *Natural Legal Language Processing Workshop (NLLP)*, II., 2020, San Diego. *Proceedings*, p. 3-10. Association for Computing Machinery (ACM), 2020.

ROZAS, Luiza Barros. Conceitos jurídicos indeterminados e discricionariedade administrativa. *Cadernos Jurídicos da Escola Paulista da Magistratura*, v. 20, n. 47, p. 191-201, 2019.

SALAÜN O., LANGLAIS P., LOU A., WESTERMANN H., BENYEKHLEF K. Analysis and Multilabel Classification of Quebec Court Decisions in the Domain of Housing Law. In: *International Conference on Applications of Natural Language to Information Systems (NLDB)*, XXV., 2020, Saarbrücken. *Lecture Notes in Computer Science*, v. 12089, p. 135-143. Springer, 2020.

SANTOS, Franklin Brasil; SOUZA, Kleberon Roberto de. *Como combater a corrupção em licitações: detecção e prevenção de fraudes*. 3. ed. Belo Horizonte: Fórum, 2020.

SCHIEFLER, E. A. C.; CRISTÓVAM, J. S. S.; PEIXOTO, F. H. A inteligência artificial aplicada à criação de uma central de jurisprudência administrativa: o uso das novas tecnologias no âmbito da gestão de informações sobre precedentes em matéria administrativa. *Revista do Direito*. v. 3, n. 50, p. 18-34, 2020.

SCHMIDT, João Pedro. Para estudar políticas públicas: aspectos conceituais, metodológicos e abordagens teóricas. *Revista do Direito*, v. 3, n. 56, p. 119-149, 2018.

SHAFFER, R.; MAYHEW, S. Legal Linking: Citation Resolution and Suggestion in Constitutional Law. In: *Natural Legal Language Processing Workshop (NLLP)*, I., 2019, Minneapolis. *Proceedings*, p. 39-44. Association for Computational Linguistics (ACL), 2019.

SOUSA Willian; DEL FABRO, Marcos D. Iudicium Textum Dataset: Uma Base de Textos Jurídicos para NLP. In: *Dataset Showcase Workshop at SBBB (Brazilian Symposium on Databases)*, II., 2019, Fortaleza. *Anais*, p. 1-11. Sociedade Brasileira de Computação (SBC), 2019.

STRECK, Lênio. Novo CPC terá mecanismos para combater decisionismos e arbitrariedades? Disponível em: <<https://www.conjur.com.br/2014-dez-18/senso-incomum-cpc-mecanismoscombater-decisionismos-arbitrariedades>>. Acesso em 10 ago 2019.

THABTAH, F. A.; COWLING, P.; PENG, Y. MMAC: a new multi-class, multi-label associative classification approach. In: *IEEE International Conference on Data Mining*

(ICDM), IV., 2004, Brighton. *Proceedings*, p. 217-224. Institute of Electrical and Electronics Engineers (IEEE), 2004.

TRIBUNAL DE CONTAS DA UNIÃO. *Referencial de combate à fraude e corrupção: aplicável a órgãos e entidades da Administração Pública*. 2. ed. 2018. Disponível em: <https://portal.tcu.gov.br/referencial-de-combate-a-fraude-e-corrupcao.htm>. Acesso em: 17 jul. 2019.

_____. *Acórdão nº 2622/2015 – Plenário*. Disponível em: https://pesquisa.apps.tcu.gov.br/#/documento/acordao-completo/*/NUMACORDAO%253A2622%2520ANOACORDAO%253A2015/DTRELEVANCIA%2520desc%252C%2520NUMACORDAOINT%2520desc/0/%2520?uuid=76a1b960-167f-11eb-abea-2d035dec77ea. Acesso em: 23 out. 2020.

_____. *Súmulas Nº 001 a 289*. Disponível em: <https://portal.tcu.gov.br/lumis/portal/file/fileDownload.jsp?fileId=8A8182A25753C20F0157679AA5617071&inline=1>. Acesso em: 28 out. 2020.

_____. *Manual Completo de Pesquisa de Jurisprudência*. Disponível em: https://portal.tcu.gov.br/data/files/10/94/CC/39/F8610710F5680BF6F18818A8/Manual_Completo_Pesquisa_Jurisprudencia_TCU.PDF. Acesso em: 31 out. 2020.

WANG, P.; FAN, Y.; NIU, S.; YANG, Z.; ZHANG, Y.; GUO, J. Hierarchical Matching Network for Crime Classification. In: *International Conference on Research and Development in Information Retrieval (SIGIR)*, XLII., 2019, Paris. *Proceedings*, p. 325-334. Association for Computing Machinery (ACM), 2019.

WANG, P.; YANG, Z.; NIU, S.; ZHANG, Y.; ZHANG, L.; NIU, S. Modeling Dynamic Pairwise Attention for Crime Classification over Legal Articles. In: *International Conference on Research and Development in Information Retrieval (SIGIR)*, XLI., 2018, Ann Arbor. *Proceedings*, p. 485-494. Association for Computing Machinery (ACM), 2018.

WANG, Y.; YAO, Q.; KWOK, J.T.; NI, L.M. Generalizing from a Few Examples: A Survey on Few-shot Learning. *ACM Computing Surveys*, v. 53, n. 3, art. 63, jun. 2020.

WITTGENSTEIN, Ludwig. *Tractatus Logico-Philosophicus*. Trad. Luiz Henrique Borges dos Santos (edição bilíngue). 3. ed. São Paulo: Edusp, 2017.

_____. *Investigações filosóficas*. Trad. Emmanuel Carneiro Leão. 1. ed. São Paulo: Universitária São Francisco, 2014.

XIAO, L.; HUANG, X.; CHEN, B.; JING, L. Label-Specific Document Representation for Multi-Label Text Classification. In: *International Joint Conference on Natural Language Processing*, IX., 2019, Hong Kong. *Proceedings*, p. 466-475. Association for Computational Linguistics (ACL), 2019.

XU, D.; SHI, Y.; TSANG, I. W.; ONG, Y.-S.; GONG, C.; SHEN, X. A Survey on Multi-output Learning. *IEEE Transactions on Neural Networks and Learning Systems*, v. 31, n. 7, p. 2409-2429, jul. 2020.

ZHAO, W.; GAO, H.; CHEN, S.; WANG, N. Generative Multi-Task Learning for Text Classification. *IEEE Access*, vol. 8, p. 86380-86387, 2020.

ZHOU, Peng; EL-GOHARY Nora. Domain-Specific Hierarchical Text Classification for Supporting Automated Environmental Compliance Checking. *Journal of Computing in Civil Engineering*, v. 30, n. 4, art. 04015057, 2016.

APÊNDICE A – LISTA DAS 330 DECISÕES COLEGIADAS DO TCU
CLASSIFICADAS EM 11 CLASSES REFERENTES A PATOLOGIAS
CORRUPTIVAS DE CLÁUSULAS RESTRITIVAS AO CARÁTER COMPETITIVO
EM EDITAIS DE LICITAÇÃO

Ano	Nº da decisão	Tipo da decisão	Colegiado	Chave da decisão	Classes pertencentes	Subclasses pertencentes
1993	434	Decisão	Plenário	DE-000434-1993-PL	5 e 7	7.11
1998	681	Acórdão	Plenário	AC-000681-1998-PL	2	
1998	279	Decisão	Plenário	DE-000279-1998-PL	5 e 7	7.11
1998	767	Decisão	Plenário	DE-000767-1998-PL	2	
1999	348	Decisão	Plenário	DE-000348-1999-PL	5 e 7	7.11
2001	1025	Decisão	Plenário	DE-001025-2001-PL	5	
2002	512	Acórdão	2ª Câmara	AC-000512-2002-2C	7	7.11
2002	417	Decisão	Plenário	DE-000417-2002-PL	2	
2002	1224	Decisão	Plenário	DE-001224-2002-PL	7	7.11
2003	244	Acórdão	Plenário	AC-000244-2003-PL	7	7.1
2003	808	Acórdão	Plenário	AC-000808-2003-PL	2	
2003	1284	Acórdão	Plenário	AC-001284-2003-PL	7	7.2
2003	1708	Acórdão	Plenário	AC-001708-2003-PL	5	
2003	1914	Acórdão	Plenário	AC-001914-2003-PL	7	7.11
2004	584	Acórdão	Plenário	AC-000584-2004-PL	7	7.1
2004	657	Acórdão	Plenário	AC-000657-2004-PL	7	7.11
2004	1730	Acórdão	Plenário	AC-001730-2004-PL	7	7.11
2004	2088	Acórdão	Plenário	AC-002088-2004-PL	7	7.2
2005	979	Acórdão	Plenário	AC-000979-2005-PL	5 e 7	7.11
2005	1140	Acórdão	Plenário	AC-001140-2005-PL	2, 5 e 7	7.4
2005	1314	Acórdão	Plenário	AC-001314-2005-PL	5	
2005	1517	Acórdão	Plenário	AC-001517-2005-PL	7	7.8
2005	1871	Acórdão	Plenário	AC-001871-2005-PL	2	
2005	2297	Acórdão	Plenário	AC-002297-2005-PL	7	7.13
2006	361	Acórdão	Plenário	AC-000361-2006-PL	7	7.13
2006	697	Acórdão	Plenário	AC-000697-2006-PL	5	
2006	1107	Acórdão	Plenário	AC-001107-2006-PL	7	7.12
2006	1230	Acórdão	Plenário	AC-001230-2006-PL	2	
2006	1529	Acórdão	Plenário	AC-001529-2006-PL	5	
2006	1898	Acórdão	Plenário	AC-001898-2006-PL	2 e 7	7.1
2006	3335	Acórdão	2ª Câmara	AC-003335-2006-2C	2	
2007	170	Acórdão	Plenário	AC-000170-2007-PL	2 e 7	7.1 e 7.13
2007	291	Acórdão	Plenário	AC-000291-2007-PL	2	
2007	597	Acórdão	Plenário	AC-000597-2007-PL	7	7.13
2007	634	Acórdão	Plenário	AC-000634-2007-PL	7	7.8
2007	701	Acórdão	Plenário	AC-000701-2007-PL	2	

2007	890	Acórdão	Plenário	AC-000890-2007-PL	5	
2007	951	Acórdão	Plenário	AC-000951-2007-PL	5	
2007	992	Acórdão	1ª Câmara	AC-000992-2007-1C	5 e 7	7.11
2007	1028	Acórdão	Plenário	AC-001028-2007-PL	2	
2007	1097	Acórdão	Plenário	AC-001097-2007-PL	7	7.13
2007	1291	Acórdão	Plenário	AC-001291-2007-PL	7	7.12
2007	1332	Acórdão	Plenário	AC-001332-2007-PL	10	
2007	1636	Acórdão	Plenário	AC-001636-2007-PL	7	7.1
2007	1699	Acórdão	Plenário	AC-001699-2007-PL	4	
2007	1839	Acórdão	Plenário	AC-001839-2007-PL	7	7.4
2007	2081	Acórdão	Plenário	AC-002081-2007-PL	4 e 7	7.14
2007	2192	Acórdão	Plenário	AC-002192-2007-PL	7	7.13
2007	2194	Acórdão	Plenário	AC-002194-2007-PL	6	
2007	2462	Acórdão	Plenário	AC-002462-2007-PL	7	7.1
2007	2656	Acórdão	Plenário	AC-002656-2007-PL	7	7.12 e 7.2
2007	3130	Acórdão	1ª Câmara	AC-003130-2007-1C	10	
2008	43	Acórdão	Plenário	AC-000043-2008-PL	7	7.1
2008	313	Acórdão	2ª Câmara	AC-000313-2008-2C	2	
2008	402	Acórdão	Plenário	AC-000402-2008-PL	2	
2008	597	Acórdão	Plenário	AC-000597-2008-PL	7	7.1
2008	608	Acórdão	Plenário	AC-000608-2008-PL	7	7.12 e 7.14
2008	800	Acórdão	Plenário	AC-000800-2008-PL	7 e 8	7.13
2008	890	Acórdão	Plenário	AC-000890-2008-PL	8	
2008	1174	Acórdão	Plenário	AC-001174-2008-PL	8	
2008	1229	Acórdão	Plenário	AC-001229-2008-PL	2	
2008	1312	Acórdão	Plenário	AC-001312-2008-PL	7	7.14
2008	1547	Acórdão	Plenário	AC-001547-2008-PL	7	7.13
2008	1612	Acórdão	Plenário	AC-001612-2008-PL	7	7.12
2008	1728	Acórdão	Plenário	AC-001728-2008-PL	7	7.11
2008	1768	Acórdão	Plenário	AC-001768-2008-PL	7	7.11
2008	1908	Acórdão	Plenário	AC-001908-2008-PL	7	7.11
2008	1949	Acórdão	Plenário	AC-001949-2008-PL	7	7.13 e 7.1
2008	2056	Acórdão	Plenário	AC-002056-2008-PL	3	
2008	2122	Acórdão	1ª Câmara	AC-002122-2008-1C	5	
2008	2150	Acórdão	Plenário	AC-002150-2008-PL	8	
2008	2215	Acórdão	Plenário	AC-002215-2008-PL	7	7.2 e 7.12
2008	2255	Acórdão	Plenário	AC-002255-2008-PL	2	
2008	2264	Acórdão	Plenário	AC-002264-2008-PL	2	
2008	2882	Acórdão	Plenário	AC-002882-2008-PL	2	
2008	3056	Acórdão	Plenário	AC-003056-2008-PL	7	7.1
2008	88	Acórdão de Relação	2ª Câmara	AR-000088-2008-2C	1	
2008	366	Acórdão de Relação	Plenário	AR-000366-2008-PL	7	7.1
2008	471	Acórdão de Relação	Plenário	AR-000471-2008-PL	4	

2009	103	Acórdão	Plenário	AC-000103-2009-PL	7	7.13
2009	107	Acórdão	Plenário	AC-000107-2009-PL	2	
2009	113	Acórdão	Plenário	AC-000113-2009-PL	2	
2009	297	Acórdão	Plenário	AC-000297-2009-PL	11	
2009	311	Acórdão	Plenário	AC-000311-2009-PL	7	7.4
2009	374	Acórdão	2ª Câmara	AC-000374-2009-2C	7	7.3
2009	381	Acórdão	Plenário	AC-000381-2009-PL	7	7.12
2009	512	Acórdão	Plenário	AC-000512-2009-PL	7	7.12
2009	727	Acórdão	Plenário	AC-000727-2009-PL	7 e 8	7.14
2009	772	Acórdão	Plenário	AC-000772-2009-PL	7	7.11
2009	855	Acórdão	Plenário	AC-000855-2009-PL	6	
2009	1265	Acórdão	Plenário	AC-001265-2009-PL	2	
2009	1502	Acórdão	Plenário	AC-001502-2009-PL	7	7.4
2009	1780	Acórdão	Plenário	AC-001780-2009-PL	7	7.1
2009	1942	Acórdão	Plenário	AC-001942-2009-PL	7	7.3
2009	2013	Acórdão	Plenário	AC-002013-2009-PL	5	
2009	2073	Acórdão	Plenário	AC-002073-2009-PL	2	
2009	2099	Acórdão	Plenário	AC-002099-2009-PL	7	7.11
2009	3037	Acórdão	Plenário	AC-003037-2009-PL	5	
2009	3043	Acórdão	Plenário	AC-003043-2009-PL	7	7.11
2009	5375	Acórdão	1ª Câmara	AC-005375-2009-1C	2	
2009	6198	Acórdão	1ª Câmara	AC-006198-2009-1C	5	
2009	6613	Acórdão	1ª Câmara	AC-006613-2009-1C	2	
2009	4278	Acórdão de Relação	1ª Câmara	AR-004278-2009-1C	10	
2010	556	Acórdão	Plenário	AC-000556-2010-PL	2	
2010	1258	Acórdão	2ª Câmara	AC-001258-2010-2C	5	
2010	1265	Acórdão	Plenário	AC-001265-2010-PL	8	
2010	1328	Acórdão	Plenário	AC-001328-2010-PL	7	7.11 e 7.3
2010	1336	Acórdão	Plenário	AC-001336-2010-PL	5	
2010	1337	Acórdão	Plenário	AC-001337-2010-PL	2	
2010	1339	Acórdão	Plenário	AC-001339-2010-PL	5	
2010	1354	Acórdão	1ª Câmara	AC-001354-2010-1C	7	7.12
2010	1593	Acórdão	2ª Câmara	AC-001593-2010-2C	2	
2010	1733	Acórdão	Plenário	AC-001733-2010-PL	7	7.4 e 7.11
2010	1762	Acórdão	Plenário	AC-001762-2010-PL	7	7.11
2010	2074	Acórdão	Plenário	AC-002074-2010-PL	2	
2010	2581	Acórdão	Plenário	AC-002581-2010-PL	5	
2010	2585	Acórdão	Plenário	AC-002585-2010-PL	7	7.14
2010	3088	Acórdão	Plenário	AC-003088-2010-PL	5	
2010	3105	Acórdão	Plenário	AC-003105-2010-PL	7	7.14
2010	3156	Acórdão	Plenário	AC-003156-2010-PL	5	
2010	3314	Acórdão	Plenário	AC-003314-2010-PL	2	
2010	4606	Acórdão	2ª Câmara	AC-004606-2010-2C	7	7.11
2010	5848	Acórdão	1ª Câmara	AC-005848-2010-1C	5	

2010	451	Acórdão de Relação	Plenário	AR-000451-2010-PL	7	7.7
2010	2328	Acórdão de Relação	Plenário	AR-002328-2010-PL	7	7.11
2011	125	Acórdão	Plenário	AC-000125-2011-PL	5	
2011	213	Acórdão	Plenário	AC-000213-2011-PL	2	
2011	276	Acórdão	Plenário	AC-000276-2011-PL	7	7.14 e 7.11
2011	492	Acórdão	Plenário	AC-000492-2011-PL	7	7.12
2011	534	Acórdão	Plenário	AC-000534-2011-PL	2	
2011	549	Acórdão	Plenário	AC-000549-2011-PL	4	
2011	600	Acórdão	Plenário	AC-000600-2011-PL	7	7.13
2011	1085	Acórdão	Plenário	AC-001085-2011-PL	7	7.12
2011	1291	Acórdão	Plenário	AC-001291-2011-PL	10	
2011	1533	Acórdão	Plenário	AC-001533-2011-PL	2	
2011	1898	Acórdão	Plenário	AC-001898-2011-PL	7	7.11
2011	2179	Acórdão	Plenário	AC-002179-2011-PL	3	
2011	2272	Acórdão	Plenário	AC-002272-2011-PL	5	
2011	2780	Acórdão	2ª Câmara	AC-002780-2011-2C	10	
2011	2992	Acórdão	Plenário	AC-002992-2011-PL	7	7.4
2011	3081	Acórdão	Plenário	AC-003081-2011-PL	7	7.3
2011	8117	Acórdão	1ª Câmara	AC-008117-2011-1C	7	7.4
2011	11196	Acórdão	2ª Câmara	AC-011196-2011-2C	7	7.11
2011	8271	Acórdão de Relação	2ª Câmara	AR-008271-2011-2C	2	
2012	165	Acórdão	Plenário	AC-000165-2012-PL	7	7.14
2012	727	Acórdão	Plenário	AC-000727-2012-PL	7	7.4
2012	1226	Acórdão	Plenário	AC-001226-2012-PL	7	7.4
2012	1321	Acórdão	Plenário	AC-001321-2012-PL	2	
2012	1528	Acórdão	Plenário	AC-001528-2012-PL	7	7.8
2012	1865	Acórdão	Plenário	AC-001865-2012-PL	7	7.1
2012	2074	Acórdão	Plenário	AC-002074-2012-PL	1	
2012	2239	Acórdão	Plenário	AC-002239-2012-PL	7	7.11
2012	2241	Acórdão	Plenário	AC-002241-2012-PL	7	7.7
2012	2760	Acórdão	Plenário	AC-002760-2012-PL	7	7.4
2012	2898	Acórdão	Plenário	AC-002898-2012-PL	7	7.13
2012	5372	Acórdão	2ª Câmara	AC-005372-2012-2C	2 e 7	7.11
2012	5964	Acórdão	1ª Câmara	AC-005964-2012-1C	5	
2013	222	Acórdão	Plenário	AC-000222-2013-PL	7	7.4
2013	397	Acórdão	Plenário	AC-000397-2013-PL	7	7.2
2013	526	Acórdão	Plenário	AC-000526-2013-PL	11	
2013	529	Acórdão	Plenário	AC-000529-2013-PL	8	
2013	584	Acórdão	Plenário	AC-000584-2013-PL	7	7.5
2013	765	Acórdão	Plenário	AC-000765-2013-PL	7	7.11
2013	845	Acórdão	Plenário	AC-000845-2013-PL	7	7.11
2013	935	Acórdão	Plenário	AC-000935-2013-PL	2	
2013	1023	Acórdão	Plenário	AC-001023-2013-PL	7	7.4

2013	1214	Acórdão	Plenário	AC-001214-2013-PL	7	7.4, 7.5 e 7.14
2013	1223	Acórdão	Plenário	AC-001223-2013-PL	7	7.4
2013	1542	Acórdão	Plenário	AC-001542-2013-PL	7	7.12
2013	1842	Acórdão	Plenário	AC-001842-2013-PL	2, 7 e 8	7.13
2013	1998	Acórdão	Plenário	AC-001998-2013-PL	7	7.4
2013	2914	Acórdão	Plenário	AC-002914-2013-PL	7	7.4
2013	2991	Acórdão	Plenário	AC-002991-2013-PL	7	7.11
2013	3070	Acórdão	Plenário	AC-003070-2013-PL	7	7.2 e 7.14
2013	4050	Acórdão	1ª Câmara	AC-004050-2013-1C	5	
2014	628	Acórdão	Plenário	AC-000628-2014-PL	2	
2014	641	Acórdão	1ª Câmara	AC-000641-2014-1C	7	7.11
2014	827	Acórdão	Plenário	AC-000827-2014-PL	7	7.2
2014	1049	Acórdão	Plenário	AC-001049-2014-PL	2	
2014	1212	Acórdão	Plenário	AC-001212-2014-PL	7	7.3
2014	1215	Acórdão	1ª Câmara	AC-001215-2014-1C	8	
2014	1225	Acórdão	Plenário	AC-001225-2014-PL	7	7.12
2014	1443	Acórdão	Plenário	AC-001443-2014-PL	7	7.4
2014	1738	Acórdão	Plenário	AC-001738-2014-PL	2	
2014	1955	Acórdão	Plenário	AC-001955-2014-PL	8	
2014	2205	Acórdão	2ª Câmara	AC-002205-2014-2C	7	7.6
2014	2329	Acórdão	2ª Câmara	AC-002329-2014-2C	2	
2014	2826	Acórdão	Plenário	AC-002826-2014-PL	8	
2014	2872	Acórdão	Plenário	AC-002872-2014-PL	5	
2014	2913	Acórdão	Plenário	AC-002913-2014-PL	1 e 2	
2014	3148	Acórdão	Plenário	AC-003148-2014-PL	7	7.5
2014	3291	Acórdão	Plenário	AC-003291-2014-PL	7	7.12 e 7.13
2014	3486	Acórdão	Plenário	AC-003486-2014-PL	8	
2014	7105	Acórdão	2ª Câmara	AC-007105-2014-2C	7	7.1
2015	165	Acórdão	Plenário	AC-000165-2015-PL	7	7.12
2015	234	Acórdão	Plenário	AC-000234-2015-PL	8	
2015	244	Acórdão	Plenário	AC-000244-2015-PL	7	7.5
2015	335	Acórdão	Plenário	AC-000335-2015-PL	2	
2015	372	Acórdão	Plenário	AC-000372-2015-PL	8	
2015	667	Acórdão	Plenário	AC-000667-2015-PL	7	7.11
2015	679	Acórdão	Plenário	AC-000679-2015-PL	7	7.4
2015	966	Acórdão	2ª Câmara	AC-000966-2015-2C	7	7.11
2015	1084	Acórdão	Plenário	AC-001084-2015-PL	2	
2015	1265	Acórdão	2ª Câmara	AC-001265-2015-2C	2	
2015	1277	Acórdão	Plenário	AC-001277-2015-PL	7	7.1
2015	1350	Acórdão	Plenário	AC-001350-2015-PL	9	
2015	1447	Acórdão	Plenário	AC-001447-2015-PL	5 e 8	
2015	1452	Acórdão	Plenário	AC-001452-2015-PL	7	7.11
2015	1585	Acórdão	Plenário	AC-001585-2015-PL	7	7.4
2015	1805	Acórdão	Plenário	AC-001805-2015-PL	9	

2015	1851	Acórdão	Plenário	AC-001851-2015-PL	7	7.2
2015	1944	Acórdão	Plenário	AC-001944-2015-PL	2 e 5	
2015	2731	Acórdão	Plenário	AC-002731-2015-PL	7	7.4
2015	3014	Acórdão	Plenário	AC-003014-2015-PL	7	7.13
2015	3028	Acórdão	Plenário	AC-003028-2015-PL	5	
2015	3301	Acórdão	Plenário	AC-003301-2015-PL	8	
2015	3356	Acórdão	Plenário	AC-003356-2015-PL	7	7.4
2015	3358	Acórdão	Plenário	AC-003358-2015-PL	1	
2015	463	Acórdão de Relação	Plenário	AR-000463-2015-PL	7	7.1
2015	744	Acórdão de Relação	2ª Câmara	AR-000744-2015-2C	7	7.4
2015	2492	Acórdão de Relação	Plenário	AR-002492-2015-PL	5	
2016	111	Acórdão	Plenário	AC-000111-2016-PL	7	7.4
2016	434	Acórdão	Plenário	AC-000434-2016-PL	5 e 7	7.11
2016	445	Acórdão	Plenário	AC-000445-2016-PL	7	7.12
2016	534	Acórdão	Plenário	AC-000534-2016-PL	7	7.14
2016	553	Acórdão	Plenário	AC-000553-2016-PL	7	7.4
2016	711	Acórdão	Plenário	AC-000711-2016-PL	5	
2016	804	Acórdão	Plenário	AC-000804-2016-PL	1	
2016	806	Acórdão	Plenário	AC-000806-2016-PL	5	
2016	1176	Acórdão	Plenário	AC-001176-2016-PL	7	7.11
2016	1219	Acórdão	Plenário	AC-001219-2016-PL	11	
2016	1246	Acórdão	Plenário	AC-001246-2016-PL	7	7.12
2016	2126	Acórdão	Plenário	AC-002126-2016-PL	5 e 8	
2016	2208	Acórdão	Plenário	AC-002208-2016-PL	7	7.9 e 7.15
2016	2379	Acórdão	Plenário	AC-002379-2016-PL	1	
2016	2672	Acórdão	Plenário	AC-002672-2016-PL	8	
2016	2743	Acórdão	Plenário	AC-002743-2016-PL	1 e 2	
2016	2803	Acórdão	Plenário	AC-002803-2016-PL	11	
2016	2835	Acórdão	Plenário	AC-002835-2016-PL	7	7.11
2016	2942	Acórdão	Plenário	AC-002942-2016-PL	5	
2016	3108	Acórdão	1ª Câmara	AC-003108-2016-1C	11	
2016	3663	Acórdão	1ª Câmara	AC-003663-2016-1C	7	7.2
2016	4788	Acórdão	1ª Câmara	AC-004788-2016-1C	7	7.12
2016	5620	Acórdão	1ª Câmara	AC-005620-2016-1C	5	
2017	134	Acórdão	Plenário	AC-000134-2017-PL	7	7.4
2017	212	Acórdão	Plenário	AC-000212-2017-PL	8	
2017	217	Acórdão	Plenário	AC-000217-2017-PL	8	
2017	361	Acórdão	Plenário	AC-000361-2017-PL	7	7.3
2017	449	Acórdão	Plenário	AC-000449-2017-PL	7	7.4
2017	658	Acórdão	Plenário	AC-000658-2017-PL	2	
2017	739	Acórdão	Plenário	AC-000739-2017-PL	5 e 7	7.11
2017	866	Acórdão	Plenário	AC-000866-2017-PL	8	
2017	1474	Acórdão	Plenário	AC-001474-2017-PL	7	7.12

2017	1823	Acórdão	Plenário	AC-001823-2017-PL	8	
2017	2365	Acórdão	Plenário	AC-002365-2017-PL	1, 2 e 7	7.11
2017	2397	Acórdão	Plenário	AC-002397-2017-PL	1 e 2	
2017	2416	Acórdão	1ª Câmara	AC-002416-2017-1C	8	
2017	2470	Acórdão	Plenário	AC-002470-2017-PL	1	
2017	2552	Acórdão	Plenário	AC-002552-2017-PL	1 e 2	
2017	2883	Acórdão	2ª Câmara	AC-002883-2017-2C	2	
2017	7982	Acórdão	2ª Câmara	AC-007982-2017-2C	5 e 6	
2017	8330	Acórdão	2ª Câmara	AC-008330-2017-2C	2	
2017	10362	Acórdão	2ª Câmara	AC-010362-2017-2C	7	7.11
2017	343	Acórdão de Relação	1ª Câmara	AR-000343-2017-1C	5	
2018	35	Acórdão	Plenário	AC-000035-2018-PL	1 e 2	
2018	170	Acórdão	Plenário	AC-000170-2018-PL	8	
2018	418	Acórdão	Plenário	AC-000418-2018-PL	2	
2018	433	Acórdão	Plenário	AC-000433-2018-PL	7	7.4
2018	514	Acórdão	Plenário	AC-000514-2018-PL	1 e 2	
2018	710	Acórdão	Plenário	AC-000710-2018-PL	1 e 2	
2018	891	Acórdão	Plenário	AC-000891-2018-PL	7	7.5
2018	1095	Acórdão	Plenário	AC-001095-2018-PL	7	7.1
2018	1567	Acórdão	Plenário	AC-001567-2018-PL	7	7.4
2018	1744	Acórdão	Plenário	AC-001744-2018-PL	2	
2018	1767	Acórdão	Plenário	AC-001767-2018-PL	7	7.4
2018	1910	Acórdão	Plenário	AC-001910-2018-PL	7	7.4
2018	2017	Acórdão	2ª Câmara	AC-002017-2018-2C	2	
2018	2104	Acórdão	Plenário	AC-002104-2018-PL	7	7.4
2018	2361	Acórdão	Plenário	AC-002361-2018-PL	7	7.13
2018	2575	Acórdão	Plenário	AC-002575-2018-PL	7	7.4
2018	2740	Acórdão	Plenário	AC-002740-2018-PL	2 e 7	7.11
2018	2743	Acórdão	Plenário	AC-002743-2018-PL	2	
2018	2903	Acórdão	Plenário	AC-002903-2018-PL	1 e 2	
2018	2939	Acórdão	Plenário	AC-002939-2018-PL	8	
2018	4345	Acórdão	2ª Câmara	AC-004345-2018-2C	5	
2018	6750	Acórdão	1ª Câmara	AC-006750-2018-1C	7	7.5
2018	12879	Acórdão	1ª Câmara	AC-012879-2018-1C	5	
2018	12893	Acórdão	1ª Câmara	AC-012893-2018-1C	7	7.4
2018	282	Acórdão de Relação	2ª Câmara	AR-000282-2018-2C	1	
2018	2346	Acórdão de Relação	Plenário	AR-002346-2018-PL	2	
2019	339	Acórdão	Plenário	AC-000339-2019-PL	10	
2019	628	Acórdão	Plenário	AC-000628-2019-PL	5	
2019	697	Acórdão	Plenário	AC-000697-2019-PL	1	
2019	739	Acórdão	Plenário	AC-000739-2019-PL	2	
2019	806	Acórdão	Plenário	AC-000806-2019-PL	1 e 2	
2019	914	Acórdão	Plenário	AC-000914-2019-PL	7	7.5

2019	1042	Acórdão	Plenário	AC-001042-2019-PL	7	7.11
2019	1044	Acórdão	Plenário	AC-001044-2019-PL	1	
2019	2162	Acórdão	Plenário	AC-002162-2019-PL	5	
2019	2304	Acórdão	Plenário	AC-002304-2019-PL	2	
2019	2640	Acórdão	Plenário	AC-002640-2019-PL	10	
2019	2660	Acórdão	Plenário	AC-002660-2019-PL	8	
2019	2696	Acórdão	1ª Câmara	AC-002696-2019-1C	7	7.2
2019	8677	Acórdão	2ª Câmara	AC-008677-2019-2C	7	7.1
2019	846	Acórdão de Relação	Plenário	AR-000846-2019-PL	5	
2019	859	Acórdão de Relação	Plenário	AR-000859-2019-PL	7	7.5
2019	1396	Acórdão de Relação	1ª Câmara	AR-001396-2019-1C	7	7.4
2019	2002	Acórdão de Relação	Plenário	AR-002002-2019-PL	7	7.2
2019	2736	Acórdão de Relação	Plenário	AR-002736-2019-PL	8	
2019	11833	Acórdão de Relação	1ª Câmara	AR-011833-2019-1C	5	
2019	12754	Acórdão de Relação	1ª Câmara	AR-012754-2019-1C	7	7.1
2019	13322	Acórdão de Relação	1ª Câmara	AR-013322-2019-1C	1 e 5	
2020	753	Acórdão	Plenário	AC-000753-2020-PL	5	
2020	913	Acórdão	Plenário	AC-000913-2020-PL	11	
2020	1073	Acórdão	Plenário	AC-001073-2020-PL	7	7.2
2020	1201	Acórdão	Plenário	AC-001201-2020-PL	2	
2020	1246	Acórdão	Plenário	AC-001246-2020-PL	7	7.8
2020	1321	Acórdão	Plenário	AC-001321-2020-PL	2	
2020	1353	Acórdão	Plenário	AC-001353-2020-PL	7	7.5 e 7.12
2020	6121	Acórdão	1ª Câmara	AC-006121-2020-1C	10	
2020	238	Acórdão de Relação	Plenário	AR-000238-2020-PL	7	7.11
2020	260	Acórdão de Relação	Plenário	AR-000260-2020-PL	10	
2020	409	Acórdão de Relação	Plenário	AR-000409-2020-PL	7	7.1, 7.13 e 7.15
2020	505	Acórdão de Relação	Plenário	AR-000505-2020-PL	9	
2020	576	Acórdão de Relação	Plenário	AR-000576-2020-PL	2	
2020	654	Acórdão de Relação	Plenário	AR-000654-2020-PL	2	
2020	961	Acórdão de Relação	Plenário	AR-000961-2020-PL	5 e 7	7.5
2020	1019	Acórdão de Relação	Plenário	AR-001019-2020-PL	7	7.1
2020	1275	Acórdão de Relação	2ª Câmara	AR-001275-2020-2C	7	7.4
2020	4776	Acórdão de Relação	2ª Câmara	AR-004776-2020-2C	10	

APÊNDICE B – SCRIPT PARA GERAR O DF_ACORDAOS

Esse script está disponível no link <https://github.com/wmodanez/acordaos>.

01 - Create df_acordaos

November 16, 2020

```
[1]: import re, os, signal

import numpy as np
import pandas as pd
import modanez_tools as mt

from pathlib import Path
from datetime import datetime

[2]: # Carregar o dataframe df_acordaos
start_time = datetime.now()

df_acordaos = pd.read_csv('dados/acordaos.csv', sep=',', error_bad_lines=False,
    ↪warn_bad_lines=False, low_memory=False)
df_acordaos.rename(str.upper, axis='columns', inplace=True)

df_acordaos.insert(0, 'ID', range(1, len(df_acordaos)+1))
df_acordaos.ID = df_acordaos.ID.map(str)
df_acordaos.ANOACORDAO = df_acordaos.ANOACORDAO.map(str)
df_acordaos.insert(13, 'TIPO_DECISAO', np.NaN)

print('Duration: {}'.format(datetime.now() - start_time))
df_acordaos.info()
```

Duration: 0:00:44.030440

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 309806 entries, 0 to 309805

Data columns (total 30 columns):

#	Column	Non-Null Count	Dtype
0	ID	309806 non-null	object
1	ACORDAO	288120 non-null	object
2	ADVOGADO	223477 non-null	object
3	ANOACORDAO	309806 non-null	object
4	ASSUNTO	77524 non-null	object
5	COLEGIADO	309806 non-null	object
6	DATASESSAO	308360 non-null	object
7	DECISAO	19451 non-null	object

8	DECLARACAOVOTO	3328 non-null	object
9	ENTIDADE	253844 non-null	object
10	INTERESSADOS	255674 non-null	object
11	MINISTROALEGOUIMPEDIMENTOSSESAO	5185 non-null	object
12	MINISTROAPRESENTOUDECLARACAOVOTO	744 non-null	object
13	TIPO_DECISAO	0 non-null	float64
14	MINISTROAUTORVOTOVENCEDOR	198161 non-null	object
15	MINISTROREVISOR	4722 non-null	object
16	NUMACORDAO	309806 non-null	object
17	NUMATA	307239 non-null	object
18	PROC	307121 non-null	object
19	RECURSOS	6029 non-null	object
20	RELATOR	307262 non-null	object
21	RELATORDELIBERACAORECORRIDA	0 non-null	float64
22	RELATORIO	110737 non-null	object
23	REPRESENTANTEMP	156269 non-null	object
24	SUMARIO	115466 non-null	object
25	TIPOPROCESSO	298928 non-null	object
26	UNIDADETECNICA	245667 non-null	object
27	VOTO	110475 non-null	object
28	VOTOCOMPLEMENTAR	2517 non-null	object
29	VOTOMINISTROREVISOR	2925 non-null	object

dtypes: float64(2), object(28)
memory usage: 70.9+ MB

```
[3]: df_acordaos.drop(['REPRESENTANTEMP', 'UNIDADETECNICA', 'UNIDADETECNICA',
↳ 'MINISTROREVISOR',
↳ 'MINISTROAUTORVOTOVENCEDOR',
↳ 'MINISTROAPRESENTOUDECLARACAOVOTO',
↳ 'MINISTROALEGOUIMPEDIMENTOSSESAO', 'INTERESSADOS', 'RECURSOS',
↳ 'ENTIDADE',
↳ 'DECLARACAOVOTO', 'ADVOGADO', 'RELATORDELIBERACAORECORRIDA'],
↳ axis=1, inplace=True)

df_acordaos.rename(columns={'ANOACORDAO': 'ANO_ACORDAO', 'NUMACORDAO':
↳ 'NUMERO_ACORDAO',
↳ 'TIPOPROCESSO': 'TIPO_PROCESSO', 'DATA_SESSAO':
↳ 'DATA_SESSAO',
↳ 'NUMATA': 'NUMERO_ATA', 'PROC': 'PROCESSO'},
↳ inplace=True)

df_acordaos.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 309806 entries, 0 to 309805
Data columns (total 18 columns):
#   Column                Non-Null Count  Dtype
#   :                      :              #   :
```

```

---
0 ID 309806 non-null object
1 ACORDAO 288120 non-null object
2 ANO_ACORDAO 309806 non-null object
3 ASSUNTO 77524 non-null object
4 COLEGIADO 309806 non-null object
5 DATA_SESSAO 308360 non-null object
6 DECISAO 19451 non-null object
7 TIPO_DECISAO 0 non-null float64
8 NUMERO_ACORDAO 309806 non-null object
9 NUMERO_ATA 307239 non-null object
10 PROCESSO 307121 non-null object
11 RELATOR 307262 non-null object
12 RELATORIO 110737 non-null object
13 SUMARIO 115466 non-null object
14 TIPO_PROCESSO 298928 non-null object
15 VOTO 110475 non-null object
16 VOTOCOMPLEMENTAR 2517 non-null object
17 VOTOMINISTROREVISOR 2925 non-null object
dtypes: float64(1), object(17)
memory usage: 42.5+ MB

```

- ```

[4]: # DE -> DECISAO = 19451 : 647 marcados como SIGILOSO por esse motivo não possui
 ↳ ACORDAO, VOTO, RELATORIO, SUMARIO
 df_acordaos.loc[(df_acordaos.DECISAO.notnull() == True) & ((df_acordaos.ACORDAO.
 ↳ isnull()==True) | (df_acordaos.ACORDAO.str.contains('SIGILOSO', na=False))),
 ↳ 'TIPO_DECISAO'] = 'DE'

[5]: # AR -> ACORDAO DE RELACAO : 193963
 df_acordaos.loc[(df_acordaos.ACORDAO.notnull()) & (df_acordaos.DECISAO.
 ↳ isnull()) & (df_acordaos.RELATORIO.isnull()) & (df_acordaos.VOTO.isnull()),
 ↳ 'TIPO_DECISAO'] = 'AR'

[6]: df_acordaos.loc[df_acordaos.TIPO_DECISAO.isna(), 'TIPO_DECISAO'] = 'AC'

[7]: df_acordaos.info()

```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 309806 entries, 0 to 309805
Data columns (total 18 columns):
Column Non-Null Count Dtype
--- ---
0 ID 309806 non-null object
1 ACORDAO 288120 non-null object
2 ANO_ACORDAO 309806 non-null object
3 ASSUNTO 77524 non-null object
4 COLEGIADO 309806 non-null object
5 DATA_SESSAO 308360 non-null object

```

```

6 DECISAO 19451 non-null object
7 TIPO_DECISAO 309806 non-null object
8 NUMERO_ACORDAO 309806 non-null object
9 NUMERO_ATA 307239 non-null object
10 PROCESSO 307121 non-null object
11 RELATOR 307262 non-null object
12 RELATORIO 110737 non-null object
13 SUMARIO 115466 non-null object
14 TIPO_PROCESSO 298928 non-null object
15 VOTO 110475 non-null object
16 VOTOCOMPLEMENTAR 2517 non-null object
17 VOTOMINISTROREVISOR 2925 non-null object
dtypes: object(18)
memory usage: 42.5+ MB

```

```

[8]: start_time = datetime.now()

df_acordaos.ID = df_acordaos.apply(mt.criaChaveAcordao, axis=1)

print('Duration: {}'.format(datetime.now() - start_time))

```

Duration: 0:00:08.525057

```

[9]: # 2871 registros com os campos, ACORDAO, DECISAO, VOTO e RELATORIO nulos
↳ simultaneamente e por isso excluídos
df_acordaos.drop(df_acordaos.loc[(df_acordaos.ACORDAO.isnull()) & df_acordaos.
↳ DECISAO.isnull() & df_acordaos.RELATORIO.isnull() & df_acordaos.VOTO.
↳ isnull()].index, inplace=True)

[10]: # Registros marcados como SIGILOSO e por esse motivo convertidos para NaN a fim
↳ de padronização
df_acordaos.loc[(df_acordaos.ACORDAO.str.lower() == 'sigiloso'), 'ACORDAO'] =
↳ np.NaN
df_acordaos.loc[(df_acordaos.DECISAO.str.lower() == 'sigiloso'), 'DECISAO'] =
↳ np.NaN
df_acordaos.loc[(df_acordaos.RELATORIO.str.lower() == 'sigiloso'), 'RELATORIO']
↳ = np.NaN
df_acordaos.loc[(df_acordaos.VOTO.str.lower() == 'sigiloso'), 'VOTO'] = np.NaN

[11]: df_acordaos.fillna('', inplace=True)

[12]: df_acordaos['DECISAO'] = df_acordaos.ACORDAO + ' ' + df_acordaos.DECISAO + ' '
↳ + df_acordaos.RELATORIO + ' ' + df_acordaos.VOTO + ' ' + df_acordaos.
↳ VOTOCOMPLEMENTAR + ' ' + df_acordaos.VOTOMINISTROREVISOR

[13]: # Retira colunas que não interessam para o trabalho

```

```
df_acordaos = df_acordaos.drop(['ANO_ACORDAO', 'ASSUNTO', 'COLEGIADO',
↳ 'DATA_SESSAO', 'NUMERO_ACORDAO',
↳ 'NUMERO_ATA', 'PROCESSO', 'RELATOR', 'SUMARIO',
↳ 'TIPO_PROCESSO', 'ACORDAO',
↳ 'VOTO', 'RELATORIO', 'VOTOCOMPLEMENTAR',
↳ 'VOTOMINISTROREVISOR',
↳ 'TIPO_DECISAO'], axis=1)

df_acordaos.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 306935 entries, 0 to 309804
Data columns (total 2 columns):
Column Non-Null Count Dtype
--- ---
0 ID 306935 non-null object
1 DECISAO 306935 non-null object
dtypes: object(2)
memory usage: 7.0+ MB
```

```
[14]: df_acordaos.sample(2)
```

```
[14]:
```

|       | ID                | DECISAO                                           |
|-------|-------------------|---------------------------------------------------|
| 26207 | DE-000356-2000-PL | 0 Tribunal Pleno, diante das razões expostas ...  |
| 95172 | AC-000353-2008-1C | VISTOS, relatados e discutidos estes autos de ... |

```
[15]: df_acordaos.tail(5)
```

```
[15]:
```

|        | ID                | DECISAO                                              |
|--------|-------------------|------------------------------------------------------|
| 309800 | AC-000058-2018-PL | <p class="tCU_- _Ac_- _item_9_- _§§_1">Vistos, re... |
| 309801 | AC-000031-2018-PL | <p class="normal_1">VISTOS, relatados e discut...    |
| 309802 | AC-000033-2018-PL | <p class="normal_1">VISTOS, relatados e discut...    |
| 309803 | AC-000030-2018-PL | <p class="normal_1">VISTOS, relatados e discut...    |
| 309804 | AR-000061-2018-PL | Documento classificado como sigiloso com funda...    |

```
[16]: for path in Path('dados/').rglob('**/*.txt'):
 texto = ''
 with path.open() as f:
 texto += f.read()
 if df_acordaos.ID.str.contains(path.stem).any() == False :
 df_acordaos = df_acordaos.append({'ID': path.stem, 'DECISAO': texto},
↳ ignore_index=True)
```

```
[17]: # df_acordaos = df_acordaos.append({'ID': 'AC-000213-2011-PL', 'DECISAO':
↳ texto}, ignore_index=True)
```

```
[18]: df_acordaos.loc[df_acordaos.ID == 'AC-000213-2011-PL']
```

```
[18]:
```

|        | ID                | DECISAO                                           |
|--------|-------------------|---------------------------------------------------|
| 306935 | AC-000213-2011-PL | Relatório\nAdoto como relatório a instrução da... |

```
[19]: df_acordaos.tail(5)
```

```
[19]:
```

|        | ID                | DECISAO                                           |
|--------|-------------------|---------------------------------------------------|
| 306981 | AC-002740-2018-PL | Relatório\nAdoto, como Relatório, a instrução ... |
| 306982 | AC-002939-2018-PL | Relatório\nTrata-se de representação apresenta... |
| 306983 | AC-002743-2018-PL | Relatório\nTrata-se de apartado da auditoria d... |
| 306984 | AR-002346-2018-PL | Acórdão\nACÓRDÃO Nº 2346/2018 - TCU - Plenário... |
| 306985 | AC-012879-2018-1C | Relatório\nAdoto, como relatório, a instrução ... |

```
[20]: mt.save_file(df_acordaos, 'dados/df_acordaos.csv')
```

Duration: 0:01:16.390667

```
[]: os.kill(os.getpid(), signal.SIGKILL)
```

## APÊNDICE C – SCRIPT PARA GERAR O *DF\_ROTULADOS*

Esse script está disponível no link <https://github.com/wmodanez/acordaos>.

### 01a - Create *df\_rotulados*

November 16, 2020

```
[1]: import os, signal

import numpy as np
import pandas as pd
import modanez_tools as mt

from datetime import datetime
```

Trabalhando com o dataset de acórdãos rotulados

```
[2]: df_rotulados = mt.read_file('dados/AC_CLASS_IMPORT.csv', separator=';', encoding='iso-8859-1')
df_rotulados.info()
```

```
Duration: 0:00:00.003817
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 330 entries, 0 to 329
Data columns (total 4 columns):
Column Non-Null Count Dtype
--- ---
0 NUMACORDAO 330 non-null object
1 LABEL_A 330 non-null int64
2 LABEL_B 32 non-null float64
3 LABEL_C 3 non-null float64
dtypes: float64(2), int64(1), object(1)
memory usage: 10.4+ KB
```

```
[3]: df_rotulados.LABEL_A.value_counts()
```

```
[3]: 7 152
 2 61
 5 48
 8 22
 1 20
 10 10
 11 6
 4 4
 9 3
```

```
6 2
3 2
Name: LABEL_A, dtype: int64
```

```
[4]: df_rotulados.LABEL_B.value_counts()
```

```
[4]: 7.0 14
 2.0 10
 8.0 4
 5.0 3
 6.0 1
 Name: LABEL_B, dtype: int64
```

```
[5]: df_rotulados.LABEL_C.value_counts()
```

```
[5]: 7.0 2
 8.0 1
 Name: LABEL_C, dtype: int64
```

```
[6]: df_rotulados.rename(columns={'NUMACORDAO': 'ID'}, inplace=True)

df_rotulados.fillna(0, inplace=True)

df_rotulados.LABEL_B = df_rotulados.LABEL_B.astype(int)

df_rotulados.LABEL_C = df_rotulados.LABEL_C.astype(int)

df_rotulados.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 330 entries, 0 to 329
Data columns (total 4 columns):
 # Column Non-Null Count Dtype
--- ---
 0 ID 330 non-null object
 1 LABEL_A 330 non-null int64
 2 LABEL_B 330 non-null int64
 3 LABEL_C 330 non-null int64
dtypes: int64(3), object(1)
memory usage: 10.4+ KB
```

```
[7]: df_acordaos = mt.read_file('dados/df_acordaos.csv')
```

```
Duration: 0:00:37.803639
```

```
[8]: df_acordaos.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```

RangeIndex: 306986 entries, 0 to 306985
Data columns (total 2 columns):
Column Non-Null Count Dtype
--- ---
0 ID 306986 non-null object
1 DECISAO 306986 non-null object
dtypes: object(2)
memory usage: 4.7+ MB

```

```
[9]: # df_acordaos.drop_duplicates(subset=['ID', 'DECISAO'], keep='first',
 ↪ inplace=True)
```

```
[10]: df_acordaos = pd.merge(df_acordaos, df_rotulados, on='ID', how='left')
df_acordaos.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 306986 entries, 0 to 306985
Data columns (total 5 columns):
Column Non-Null Count Dtype
--- ---
0 ID 306986 non-null object
1 DECISAO 306986 non-null object
2 LABEL_A 330 non-null float64
3 LABEL_B 330 non-null float64
4 LABEL_C 330 non-null float64
dtypes: float64(3), object(2)
memory usage: 14.1+ MB

```

```
[11]: df_rotulados = df_acordaos.loc[df_acordaos.LABEL_A.notnull()]
df_rotulados.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 330 entries, 2830 to 306985
Data columns (total 5 columns):
Column Non-Null Count Dtype
--- ---
0 ID 330 non-null object
1 DECISAO 330 non-null object
2 LABEL_A 330 non-null float64
3 LABEL_B 330 non-null float64
4 LABEL_C 330 non-null float64
dtypes: float64(3), object(2)
memory usage: 15.5+ KB

```

```
[12]: df_rotulados.LABEL_A.value_counts()
```

```
[12]: 7.0 152
 2.0 61
 5.0 48
 8.0 22
 1.0 20
 10.0 10
 11.0 6
 4.0 4
 9.0 3
 3.0 2
 6.0 2
 Name: LABEL_A, dtype: int64
```

```
[13]: df_rotulados.LABEL_B.value_counts()
```

```
[13]: 0.0 298
 7.0 14
 2.0 10
 8.0 4
 5.0 3
 6.0 1
 Name: LABEL_B, dtype: int64
```

```
[14]: df_rotulados.LABEL_C.value_counts()
```

```
[14]: 0.0 327
 7.0 2
 8.0 1
 Name: LABEL_C, dtype: int64
```

```
[15]: df_acordaos.LABEL_A.fillna(0, inplace=True)
 df_acordaos.LABEL_A = df_acordaos.LABEL_A.astype(int)

 df_acordaos.LABEL_B.fillna(0, inplace=True)
 df_acordaos.LABEL_B = df_acordaos.LABEL_B.astype(int)

 df_acordaos.LABEL_C.fillna(0, inplace=True)
 df_acordaos.LABEL_C = df_acordaos.LABEL_C.astype(int)
 df_acordaos.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 306986 entries, 0 to 306985
Data columns (total 5 columns):
Column Non-Null Count Dtype
--- ---
0 ID 306986 non-null object
1 DECISAO 306986 non-null object
```

```
2 LABEL_A 306986 non-null int64
3 LABEL_B 306986 non-null int64
4 LABEL_C 306986 non-null int64
dtypes: int64(3), object(2)
memory usage: 14.1+ MB
```

```
[16]: # Salva o arquivo em um csv
mt.save_file(df_acordaos.loc[df_acordaos.LABEL_A != 0], 'dados/df_rotulados.
 ↳ csv')
```

Duration: 0:00:00.457998

```
[17]: mt.save_file(df_acordaos.loc[df_acordaos.LABEL_A == 0], 'dados/df_acordaos_rot.
 ↳ csv')
```

Duration: 0:01:00.498898

```
[]: os.kill(os.getpid(), signal.SIGKILL)
```

## APÊNDICE D – SCRIPT DE PRÉ-PROCESSAMENTO DO DF\_ACORDAOS

Esse script está disponível no link <https://github.com/wmodanez/acordaos>.

### 02 - Pre processing df\_acordaos

November 16, 2020

#### 0.0.1 Tratamento do texto do campo DECISAO do dataset de Acórdãos

```
[1]: import os, signal

import modanez_tools as mt

from datetime import datetime
```

```
[2]: df_acordaos = mt.read_file('dados/df_acordaos.csv')
df_acordaos.info()
```

```
Duration: 0:00:44.497757
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 306986 entries, 0 to 306985
Data columns (total 5 columns):
 # Column Non-Null Count Dtype
--- ---
 0 ID 306986 non-null object
 1 DECISAO 306986 non-null object
 2 LABEL_A 306986 non-null int64
 3 LABEL_B 306986 non-null int64
 4 LABEL_C 306986 non-null int64
dtypes: int64(3), object(2)
memory usage: 11.7+ MB
```

```
[3]: # Aplicação do regex no texto do dataset de acordaos
start_time = datetime.now()
df_acordaos.DECISAO = df_acordaos.apply(lambda x: mt.padronizacao(x), axis = 1)
print('Duration: {}'.format(datetime.now() - start_time))
```

```
Duration: 9:32:14.500669
```

```
[]: mt.save_file(df_acordaos, 'dados/df_acordaos_proc.csv')
```

```
Duration: 0:00:40.050639
```

```
[]: import os, signal
os.kill(os.getpid(), signal.SIGKILL)
```

## APÊNDICE E – SCRIPT DE PRÉ-PROCESSAMENTO DO *DF\_ROTULADOS*

Esse script está disponível no link <https://github.com/wmodanez/acordaos>.

### 02a - Pre processing *df\_rotulados*

November 16, 2020

#### 0.0.1 Tratamento do texto do campo DECISAO do dataset de Acórdãos rotulados

```
[1]: import os, signal

import pandas as pd
import pyperclip as pc
import texthero as hero
import modanez_tools as mt
import matplotlib.pyplot as plt

from datetime import datetime
from utils import highlight_regex_matches as hl

[2]: df_rotulados = mt.read_file('dados/df_rotulados.csv')
df_rotulados.info()

Duration: 0:00:00.229214
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 330 entries, 0 to 329
Data columns (total 5 columns):
Column Non-Null Count Dtype
--- -
0 ID 330 non-null object
1 DECISAO 330 non-null object
2 LABEL_A 330 non-null int64
3 LABEL_B 330 non-null int64
4 LABEL_C 330 non-null int64
dtypes: int64(3), object(2)
memory usage: 13.0+ KB

[3]: # Aplicação do regex no texto do dataset de acordaos rotulados
start_time = datetime.now()
df_rotulados.DECISAO = df_rotulados.apply(lambda x: mt.padronizacao(x), axis = 1)
print('Duration: {}'.format(datetime.now() - start_time))
df_rotulados
```

Duration: 0:02:24.939525

```
[3]:
```

|     | ID                | DECISAO \                                         |
|-----|-------------------|---------------------------------------------------|
| 0   | DE-000434-1993-PL | tribunal pleno razões postas relator decide un... |
| 1   | DE-000767-1998-PL | tribunal pleno razões aposta relator decide co... |
| 2   | DE-000681-1998-PL | tribunal pleno razões postas relator decide co... |
| 3   | DE-000279-1998-PL | plenário tribunal razões postas relator fulcro... |
| 4   | DE-000348-1999-PL | tribunal pleno razões postas relator decide co... |
| ... | ...               | ...                                               |
| 325 | AC-002740-2018-PL | relatório adoto relatório instrução auditor co... |
| 326 | AC-002939-2018-PL | relatório trata-se representação apresentada c... |
| 327 | AC-002743-2018-PL | relatório trata-se apartado auditoria conformi... |
| 328 | AR-002346-2018-PL | acórdão acórdão plenário ministros reunidos se... |
| 329 | AC-012879-2018-1C | relatório adoto relatório instrução elaborada ... |

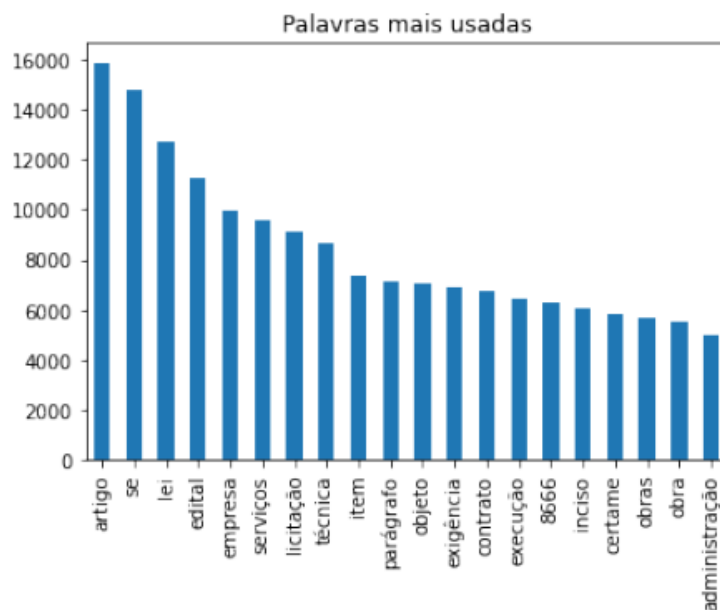
|     | LABEL_A | LABEL_B | LABEL_C |
|-----|---------|---------|---------|
| 0   | 5       | 7       | 0       |
| 1   | 2       | 0       | 0       |
| 2   | 2       | 0       | 0       |
| 3   | 5       | 7       | 0       |
| 4   | 5       | 7       | 0       |
| ... | ...     | ...     | ...     |
| 325 | 2       | 7       | 0       |
| 326 | 8       | 0       | 0       |
| 327 | 2       | 0       | 0       |
| 328 | 2       | 0       | 0       |
| 329 | 5       | 0       | 0       |

[330 rows x 5 columns]

```
[4]: top_20 = hero.visualization.top_words(hero.remove_punctuation(df_rotulados.
 ↳ DECISAO)).head(20)

top_20.plot.bar(rot=90, title='Palavras mais usadas')

plt.show(block=True)
```



```
[6]: # pc.copy(df_rotulados.iloc[177].DECISAO)

[7]: df_rotulados = pd.DataFrame(df_rotulados.groupby(['ID',
↳ 'DECISAO'])['LABEL_A', 'LABEL_B', 'LABEL_C'].apply(lambda x: tuple(x.values)))
df_rotulados.reset_index(inplace=True)
df_rotulados = df_rotulados.rename(columns={0: 'LABELS'})

for index, row in df_rotulados.iterrows():
 df_rotulados.iloc[index].LABELS = ','.join([str(elem) for elem in
↳ df_rotulados.iloc[index].LABELS[0]])

new_cols = ['1', '2', '3', '4', '5', '6', '7', '8', '9', '10', '11']
new_vals = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
df_rotulados = df_rotulados.reindex(columns=df_rotulados.columns.
↳ tolist()+new_cols)
df_rotulados[new_cols] = new_vals

for index, row in df_rotulados.iterrows():
 labels = df_rotulados.iloc[index].LABELS.split(',')
 for i in range(0, 3):
 if int(labels[i])>0:
 df_rotulados.iloc[index, int(labels[i])+2] = 1

df_rotulados

<ipython-input-7-ff75a172ede4>:1: FutureWarning: Indexing with multiple keys
(implicitly converted to a tuple of keys) will be deprecated, use a list
instead.
 df_rotulados = pd.DataFrame(df_rotulados.groupby(['ID',
'DECISAO'])['LABEL_A', 'LABEL_B', 'LABEL_C'].apply(lambda x: tuple(x.values)))

[8]: df_rotulados.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 330 entries, 0 to 329
Data columns (total 14 columns):
Column Non-Null Count Dtype
--- -
0 ID 330 non-null object
1 DECISAO 330 non-null object
2 LABELS 330 non-null object
3 1 330 non-null int64
4 2 330 non-null int64
5 3 330 non-null int64
6 4 330 non-null int64
7 5 330 non-null int64
8 6 330 non-null int64
```

```
9 7 330 non-null int64
10 8 330 non-null int64
11 9 330 non-null int64
12 10 330 non-null int64
13 11 330 non-null int64
```

```
dtypes: int64(11), object(3)
```

```
memory usage: 36.2+ KB
```

```
[9]: df_rotulados.DECISAO = hero.remove_diacritics(df_rotulados.DECISAO)
```

```
[10]: mt.save_file(df_rotulados, 'dados/df_rotulados_spel_diac.csv')
```

```
[]: os.kill(os.getpid(), signal.SIGKILL)
```

## APÊNDICE F – *SCRIPT* DE TREINAMENTOS E TESTES DE MODELOS DE *MACHINE LEARNING*

Esse script está disponível no link <https://github.com/wmodanez/acordaos>.

### 03 - Trainning - ML

November 16, 2020

#### 0.1 Fonte

<https://medium.com/swlh/multi-label-text-classification-with-scikit-learn-and-tensorflow-257f9ee30536>

```
[1]: import numpy as np
import pandas as pd
import seaborn as sns
import tensorflow as tf
import matplotlib.pyplot as plt

from sklearn import metrics
from sklearn.pipeline import Pipeline
from sklearn.base import TransformerMixin
from sklearn.metrics import accuracy_score, f1_score, precision_score, \
 roc_auc_score
from sklearn.multiclass import OneVsRestClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer

[2]: df_rotulados = pd.read_csv('dados/df_rotulados_spel_diac.csv', sep='|')
df_rotulados.drop(['LABELS'], axis=1, inplace=True)
print(df_rotulados.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 330 entries, 0 to 329
Data columns (total 13 columns):
Column Non-Null Count Dtype
--- ---
0 ID 330 non-null object
1 DECISAO 330 non-null object
2 1 330 non-null int64
3 2 330 non-null int64
4 3 330 non-null int64
5 4 330 non-null int64
6 5 330 non-null int64
7 6 330 non-null int64
8 7 330 non-null int64
```

```

9 8 330 non-null int64
10 9 330 non-null int64
11 10 330 non-null int64
12 11 330 non-null int64
dtypes: int64(11), object(2)
memory usage: 33.6+ KB
None

```

```

[3]: def acordoas_categoria(title=''):

 categories = list(df_rotulados.columns.values)
 sns.set(font_scale = 2)
 plt.figure(figsize=(15,8))

 ax= sns.barplot(categories[2:], df_rotulados.iloc[:,2:].sum().values)
 plt.title(title, fontsize=24)
 plt.ylabel('Numero de acórdãos', fontsize=18)
 plt.xlabel('Tipo de acórdão ', fontsize=18)

 #adding the text labels
 rects = ax.patches
 labels = df_rotulados.iloc[:,2:].sum().values
 for rect, label in zip(rects, labels):
 height = rect.get_height()
 ax.text(rect.get_x() + rect.get_width()/2, height + 1, label,
 ↪ha='center', va='bottom', fontsize=14)
 plt.show()

```

```

[4]: acordoas_categoria('Todos os tipos de acórdãos')

''' Remoção das classes com menos de 10 exemplos uma vez que isso interfere na
↪hora de calcular a curva ROC,
pois no momento do split dos dados para criação dos dataset de treino e teste
↪não há registros suficientes para todas as classes
'''
df_rotulados.drop(['3','4','6','9', '11'],axis=1, inplace=True)

Remoção das linhas que pelo motivo acima não possuem registros
df_rotulados.drop(df_rotulados.loc[(df_rotulados.iloc[:,2:].sum(axis=1))==0].
↪index, axis=0, inplace=True)

acordoas_categoria('Rótulos com mais de dez acórdãos exemplo')

```

/home/modanez/anaconda3/envs/servio/lib/python3.8/site-packages/seaborn/\_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will

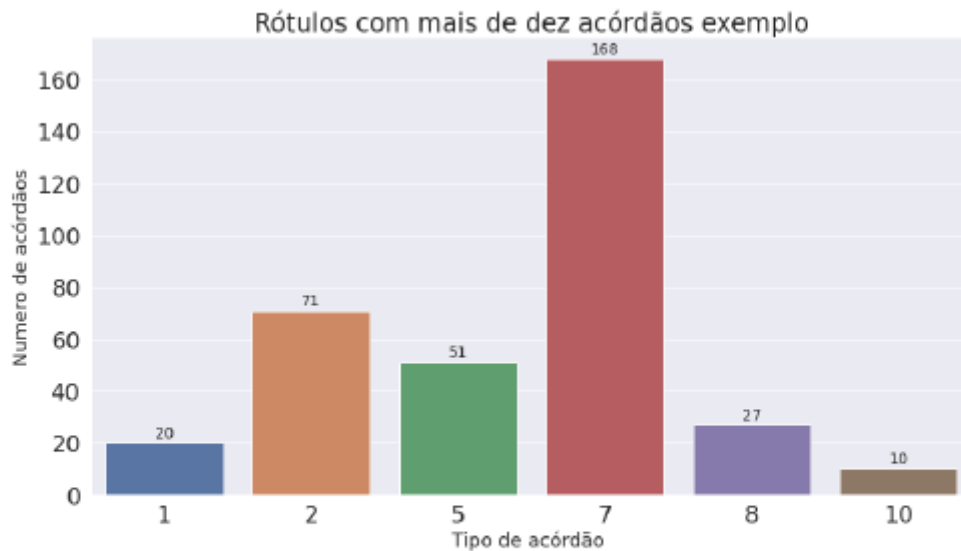
```
result in an error or misinterpretation.
warnings.warn(

```



```
/home/modanez/anaconda3/envs/servio/lib/python3.8/site-
packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variables
as keyword args: x, y. From version 0.12, the only valid positional argument
will be `data`, and passing other arguments without an explicit keyword will
result in an error or misinterpretation.
warnings.warn(

```



```
[5]: bar_plot = pd.DataFrame()
bar_plot['ACORDAOS'] = df_rotulados.columns[2:]
bar_plot['SOMA'] = df_rotulados.iloc[:,2:].sum().values
bar_plot.sort_values(['SOMA'], inplace=True, ascending=False)
bar_plot.reset_index(inplace=True, drop=True)
print(bar_plot)

threshold = 25

plt.figure(figsize=(15,8))
sns.set(font_scale = 1.5)
sns.set_style('whitegrid')

pal = sns.color_palette("Blues_r", len(bar_plot))
rank = bar_plot['SOMA'].argsort().argsort()

ax = sns.barplot(bar_plot['ACORDAOS'], bar_plot['SOMA'], palette=np.array(pal[:,
↪-1])[rank])
plt.axhline(threshold, ls='--', c='red')
plt.title("Quantidade de exemplos para treinamento", fontsize=24)
plt.ylabel('Quantidade de exemplos', fontsize=18)
plt.xlabel('Tipo de Acórdão', fontsize=18)
plt.xticks(rotation='vertical')

adding the text labels
```

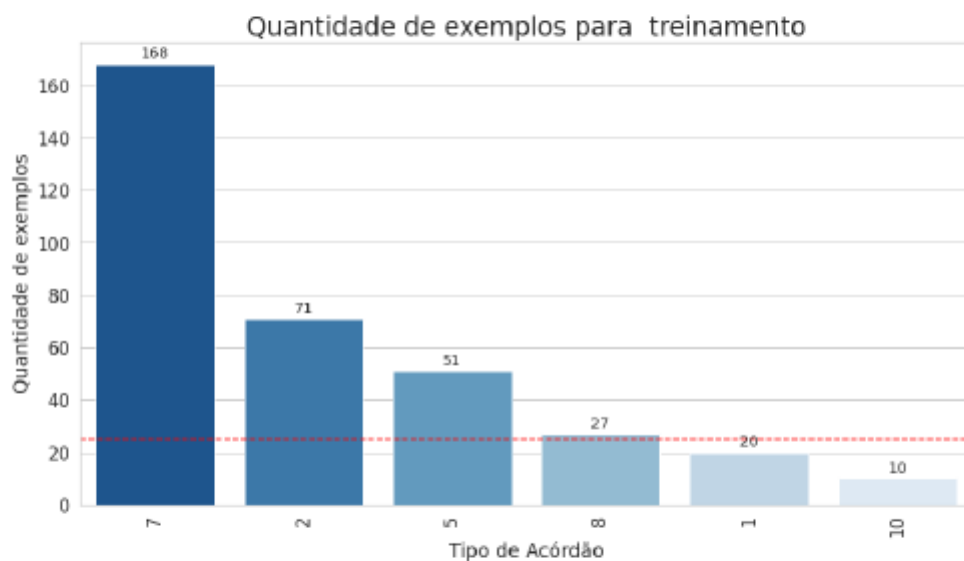
```

rects = ax.patches
labels = bar_plot.SOMA
for rect, label in zip(rects, labels):
 height = rect.get_height()
 ax.text(rect.get_x() + rect.get_width()/2, height + 1, label, ha='center',
 va='bottom', fontsize=14)
plt.show()

```

/home/modanez/anaconda3/envs/servio/lib/python3.8/site-packages/seaborn/\_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

```
warnings.warn(
```



```

[6]: rowSums = df_rotulados.iloc[:,2:].sum(axis=1)
multiLabel_counts = rowSums.value_counts()
multiLabel_counts = multiLabel_counts.iloc[:]
sns.set(font_scale = 2)
plt.figure(figsize=(15,8))
ax = sns.barplot(multiLabel_counts.index, multiLabel_counts.values)
plt.title("Acórdãos com múltiplos rótulos")
plt.ylabel('Numero de acórdãos', fontsize=18)
plt.xlabel('Numero de rótulos', fontsize=18)

#adding the text labels

```

```

rects = ax.patches
labels = multiLabel_counts.values
for rect, label in zip(rects, labels):
 height = rect.get_height()
 ax.text(rect.get_x() + rect.get_width()/2, height + 1, label, ha='center', va='bottom', fontsize=14)
plt.show()

```

/home/modanez/anaconda3/envs/servio/lib/python3.8/site-packages/seaborn/\_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

```
warnings.warn(
```



```

[7]: boxplot = df_rotulados.copy()
boxplot['len'] = df_rotulados.DECISAO.apply(lambda x: len(x))

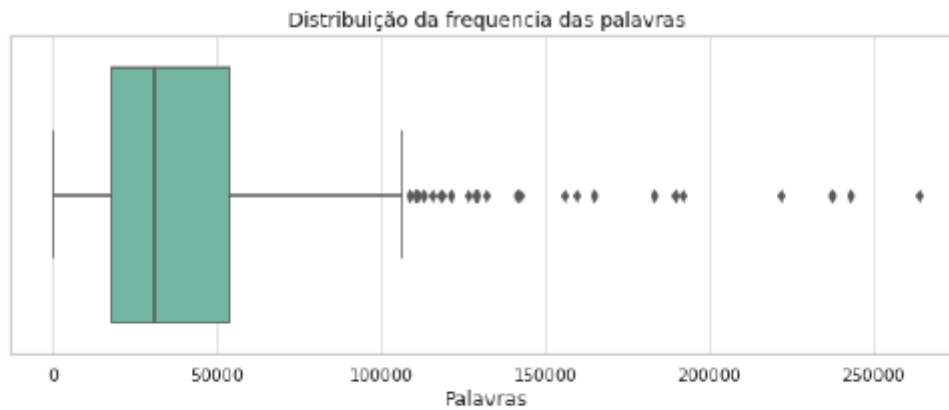
sns.set(style="whitegrid", rc={"font.size":13,"axes.labelsize":13})

plt.figure(figsize=(9, 4))

ax = sns.boxplot(x='len', data=boxplot, orient="h", palette="Set2")
plt.ylabel('')
plt.xlabel('Palavras')
plt.title("Distribuição da frequência das palavras", fontsize=13)

```

```
plt.tight_layout(h_pad=3)
```



```
[8]: columns=['OneVsAll', 'BinaryRelevance', 'ClassifierChain', 'MultipleOutput']
results = pd.DataFrame(columns = columns)
results
```

```
[8]: Empty DataFrame
Columns: [OneVsAll, BinaryRelevance, ClassifierChain, MultipleOutput]
Index: []
```

## 0.2 Formas de dividir o dataset em três partes

```
train, validate, test = np.split(df_rotulados.sample(frac=1), [int(.6len(df_rotulados)),
int(.8len(df_rotulados))])
```

```
x_train, x_test, y_train, y_test = train_test_split(df_rotulados.DECISAO,
df_rotulados[df_rotulados.columns[2:]], test_size=0.2, random_state=1977, shuffle=True)
```

```
x_train, x_valid, y_train, y_valid = train_test_split(x_train, y_train, test_size=0.25, random_state=1977, shuffle=True)
```

```
[9]: x_train, x_test, y_train, y_test = train_test_split(df_rotulados.DECISAO,
↳ df_rotulados[df_rotulados.columns[2:]],
test_size=0.3,
random_state=1977,
shuffle=True)

vectorizer = TfidfVectorizer(strip_accents='unicode', analyzer='word',
↳ ngram_range=(1,1), norm='l2')
vectorizer.fit(x_train)
```

```
x_train = vectorizer.transform(x_train)
x_test = vectorizer.transform(x_test)
```

```
[10]: LR_pipeline = Pipeline([('clf',
 ↳ OneVsRestClassifier(LogisticRegression(solver='sag'), n_jobs=-1))])

accuracy = 0
for category in df_rotulados.columns[2:]:
 print('**Processando a classe {} dos acórdãos...**'.format(category))

 # Training logistic regression model on train data
 LR_pipeline.fit(x_train, y_train[category])

 # calculating test accuracy
 prediction = LR_pipeline.predict(x_test)
 accuracy = accuracy + accuracy_score(y_test[category], prediction)
 print('AUC ROC score is {}'.format(roc_auc_score(y_test[category], prediction)))
 print("\n")

print('Test averaged AUC ROC is {}'.format(accuracy/len(df_rotulados.columns[1:
 ↳])))
```

```
Processando a classe 1 dos acórdãos...
AUC ROC score is 0.5
```

```
Processando a classe 2 dos acórdãos...
AUC ROC score is 0.5227272727272727
```

```
Processando a classe 5 dos acórdãos...
AUC ROC score is 0.5
```

```
Processando a classe 7 dos acórdãos...
AUC ROC score is 0.671985815602837
```

```
Processando a classe 8 dos acórdãos...
AUC ROC score is 0.5
```

```
Processando a classe 10 dos acórdãos...
AUC ROC score is 0.5
```

Test averaged AUC ROC is 0.7263157894736842

```
[11]: from sklearn.naive_bayes import MultinomialNB

NB_pipeline = Pipeline([('clf',
 ↳OneVsRestClassifier(MultinomialNB(fit_prior=True, class_prior=None)))]

accuracy = 0
for category in df_rotulados.columns[2:]:
 print('**Processing {} titles...**'.format(category))

 # Training logistic regression model on train data
 NB_pipeline.fit(x_train, y_train[category])

 # calculating test accuracy
 prediction = NB_pipeline.predict(x_test)
 accuracy = accuracy + accuracy_score(y_test[category], prediction)
 print('AUC ROC is {}'.format(roc_auc_score(y_test[category], prediction)))
 print("\n")

print('Test averaged AUC ROC is {}'.format(accuracy/len(df_rotulados.columns[1:
 ↳])))
```

\*\*Processing 1 titles...\*\*

AUC ROC is 0.5

\*\*Processing 2 titles...\*\*

AUC ROC is 0.5

\*\*Processing 5 titles...\*\*

AUC ROC is 0.5

\*\*Processing 7 titles...\*\*

AUC ROC is 0.5108599290780141

\*\*Processing 8 titles...\*\*

AUC ROC is 0.5

\*\*Processing 10 titles...\*\*

AUC ROC is 0.5

Test averaged AUC ROC is 0.7022556390977444

```
[12]: from sklearn.svm import LinearSVC

SVC_pipeline = Pipeline([('clf', OneVsRestClassifier(LinearSVC(), n_jobs=1))])

accuracy = 0
for category in df_rotulados.columns[2:]:
 print('**Processing {} titles...**'.format(category))

 # Training logistic regression model on train data
 SVC_pipeline.fit(x_train, y_train[category])

 # calculating test accuracy
 prediction = SVC_pipeline.predict(x_test)
 accuracy = accuracy + roc_auc_score(y_test[category], prediction)
 print('F1-score is {}'.format(roc_auc_score(y_test[category], prediction)))
 print("\n")

print('Test averaged f1-score is {}'.format(accuracy/len(df_rotulados.columns[1:
↵])))

Processing 1 titles...
F1-score is 0.5

Processing 2 titles...
F1-score is 0.6681195516811956

Processing 5 titles...
F1-score is 0.5

Processing 7 titles...
F1-score is 0.7158687943262412

Processing 8 titles...
F1-score is 0.5

Processing 10 titles...
F1-score is 0.5

Test averaged f1-score is 0.483426906572491
```

```
[13]: from sklearn.multiclass import OneVsRestClassifier
 from sklearn.ensemble import RandomForestClassifier

 RF_pipeline = Pipeline([('clf', OneVsRestClassifier(RandomForestClassifier(),
 ↪n_jobs=1))])

 accuracy = 0
 for category in df_rotulados.columns[2:]:
 print('**Processing {} titles...**'.format(category))

 # Training logistic regression model on train data
 RF_pipeline.fit(x_train, y_train[category])

 # calculating test accuracy
 prediction = RF_pipeline.predict(x_test)
 accuracy = accuracy + roc_auc_score(y_test[category], prediction)
 print('AUC ROC is {}'.format(roc_auc_score(y_test[category], prediction)))
 print("\n")

 print('Test averaged AUC ROC is {}'.format(accuracy/len(df_rotulados.columns[1:
 ↪])))
```

```
Processing 1 titles...
AUC ROC is 0.5
```

```
Processing 2 titles...
AUC ROC is 0.5454545454545454
```

```
Processing 5 titles...
AUC ROC is 0.5
```

```
Processing 7 titles...
AUC ROC is 0.6728723404255318
```

```
Processing 8 titles...
AUC ROC is 0.5
```

```
Processing 10 titles...
AUC ROC is 0.5
```

```
Test averaged AUC ROC is 0.45976098369715385
```

```
[14]: results.loc[4, 'OneVsAll'] = accuracy/len(df_rotulados.columns[1:])
 results
```

```
[14]: OneVsAll BinaryRelevance ClassifierChain MultipleOutput
 4 0.459761 NaN NaN NaN
```

```
[15]: from skmultilearn.problem_transform import BinaryRelevance
 from sklearn.naive_bayes import GaussianNB

 classifier = BinaryRelevance(GaussianNB())
 classifier.fit(x_train, y_train)
 predictions = classifier.predict(x_test)
 accuracy_score(y_test, predictions)
 print('AUC ROC is {}'.format(roc_auc_score(y_test, predictions.toarray())))
```

AUC ROC is 0.547005345718581

```
[16]: results.loc[4, 'BinaryRelevance'] = roc_auc_score(y_test, predictions.toarray())
 results
```

```
[16]: OneVsAll BinaryRelevance ClassifierChain MultipleOutput
 4 0.459761 0.547005 NaN NaN
```

```
[17]: from skmultilearn.problem_transform import ClassifierChain
 from sklearn.linear_model import LogisticRegression

 classifier = ClassifierChain(LogisticRegression())
 classifier.fit(x_train, y_train)
 predictions = classifier.predict(x_test)

 print('AUC ROC is {}'.format(roc_auc_score(y_test, predictions.toarray())))
```

AUC ROC is 0.5055609284332688

```
[18]: results.loc[4, 'ClassifierChain'] = roc_auc_score(y_test, predictions.toarray())
 results
```

```
[18]: OneVsAll BinaryRelevance ClassifierChain MultipleOutput
 4 0.459761 0.547005 0.505561 NaN
```

```
[19]: from sklearn.multioutput import MultiOutputClassifier
 from sklearn.neighbors import KNeighborsClassifier

 clf = MultiOutputClassifier(KNeighborsClassifier()).fit(x_train, y_train)
 predictions = clf.predict(x_test)

 print('AUC ROC is {}'.format(roc_auc_score(y_test, predictions)))
```

AUC ROC is 0.5797167499196225

```
[20]: results.loc[4, 'MultipleOutput'] = roc_auc_score(y_test, predictions)
 results
```

```
[20]: OneVsAll BinaryRelevance ClassifierChain MultipleOutput
 4 0.459761 0.547005 0.505561 0.579717
```

## APÊNDICE G – SCRIPT DE TREINAMENTOS E TESTES DE MODELOS DE *DEEP LEARNING*

Esse script está disponível no link <https://github.com/wmodanez/acordaos>.

### 03 - Trainning-DL

November 16, 2020

```
[1]: import numpy as np
import pandas as pd
import seaborn as sns
import tensorflow as tf
import texthero as hero
import matplotlib.pyplot as plt

from datetime import datetime
from sklearn.model_selection import train_test_split

from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing.sequence import pad_sequences

[2]: df_rotulados = pd.read_csv('dados/df_rotulados_spel_diac.csv', sep='|')
df_rotulados.drop(['LABELS'], axis=1, inplace=True)
print(df_rotulados.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 330 entries, 0 to 329
Data columns (total 13 columns):
Column Non-Null Count Dtype
--- ---
0 ID 330 non-null object
1 DECISAO 330 non-null object
2 1 330 non-null int64
3 2 330 non-null int64
4 3 330 non-null int64
5 4 330 non-null int64
6 5 330 non-null int64
7 6 330 non-null int64
8 7 330 non-null int64
9 8 330 non-null int64
10 9 330 non-null int64
11 10 330 non-null int64
12 11 330 non-null int64
dtypes: int64(11), object(2)
memory usage: 33.6+ KB
None
```

```
[3]: def acordoas_categoria(title=''):

 categories = list(df_rotulados.columns.values)
 sns.set(font_scale = 2)
 plt.figure(figsize=(15,8))

 ax= sns.barplot(categories[2:], df_rotulados.iloc[:,2:].sum().values)
 plt.title(title, fontsize=24)
 plt.ylabel('Numero de acórdãos', fontsize=18)
 plt.xlabel('Tipo de acórdão ', fontsize=18)

 #adding the text labels
 rects = ax.patches
 labels = df_rotulados.iloc[:,2:].sum().values
 for rect, label in zip(rects, labels):
 height = rect.get_height()
 ax.text(rect.get_x() + rect.get_width()/2, height + 1, label,
 ha='center', va='bottom', fontsize=14)
 plt.show()

def max_len(x):
 return len(x.split())
```

```
[4]: acordoas_categoria('Todos os tipos de acórdãos')

''' Remoção das classes com menos de 10 exemplos uma vez que isso interfere na
↳ hora de calcular a curva ROC,
pois no momento do split dos dados para criação dos dataset de treino e teste
↳ não há registros suficientes para todas as classes
'''
df_rotulados.drop(['3','4','6','9', '11'],axis=1, inplace=True)

Remoção das linhas que pelo motivo acima não possuem registros
df_rotulados.drop(df_rotulados.loc[(df_rotulados.iloc[:,2:].sum(axis=1))==0].
↳ index, axis=0, inplace=True)

acordoas_categoria('Rótulos com mais de dez acórdãos exemplo')
```

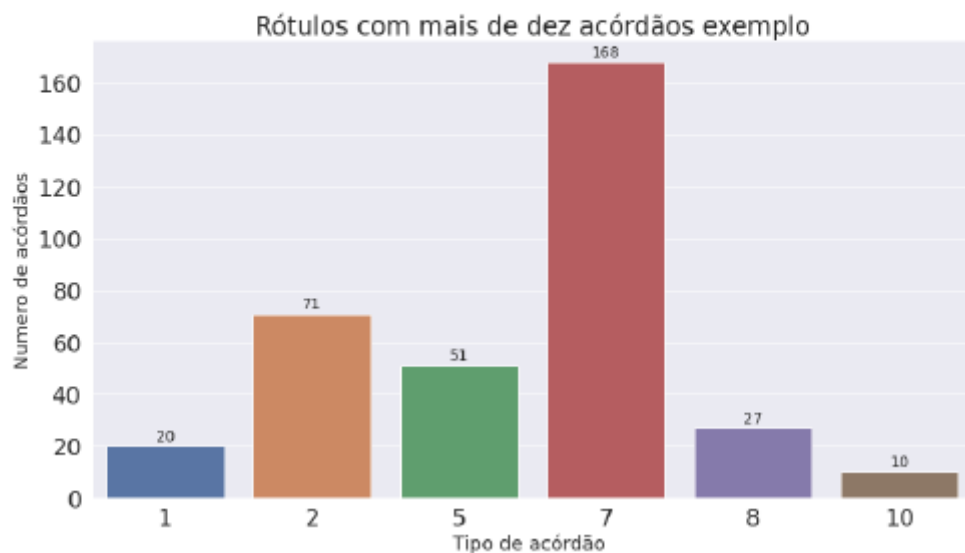
```
/home/modanez/anaconda3/envs/servio/lib/python3.8/site-
packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variables
as keyword args: x, y. From version 0.12, the only valid positional argument
will be `data`, and passing other arguments without an explicit keyword will
result in an error or misinterpretation.
 warnings.warn(
```



```

/home/modanez/anaconda3/envs/servio/lib/python3.8/site-
packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variables
as keyword args: x, y. From version 0.12, the only valid positional argument
will be `data`, and passing other arguments without an explicit keyword will
result in an error or misinterpretation.
warnings.warn(

```



```
[5]: bar_plot = pd.DataFrame()
bar_plot['ACORDAOS'] = df_rotulados.columns[2:]
bar_plot['SOMA'] = df_rotulados.iloc[:,2:].sum().values
bar_plot.sort_values(['SOMA'], inplace=True, ascending=False)
bar_plot.reset_index(inplace=True, drop=True)
print(bar_plot)

threshold = 25

plt.figure(figsize=(15,8))
sns.set(font_scale = 1.5)
sns.set_style('whitegrid')

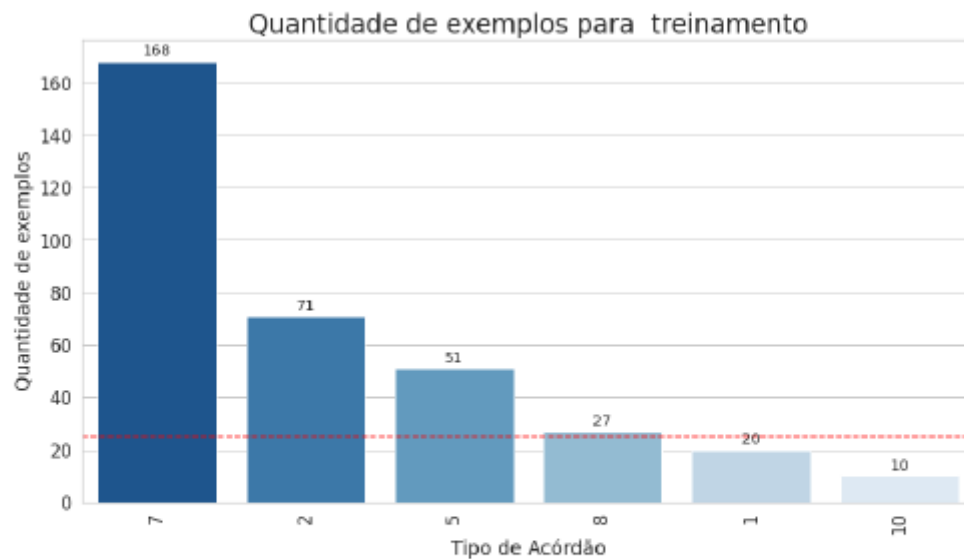
pal = sns.color_palette("Blues_r", len(bar_plot))
rank = bar_plot['SOMA'].argsort().argsort()

ax = sns.barplot(bar_plot['ACORDAOS'], bar_plot['SOMA'], palette=np.array(pal[::
 ↪-1])[rank])
plt.axhline(threshold, ls='--', c='red')
plt.title("Quantidade de exemplos para treinamento", fontsize=24)
plt.ylabel('Quantidade de exemplos', fontsize=18)
plt.xlabel('Tipo de Acórdão', fontsize=18)
plt.xticks(rotation='vertical')

adding the text labels
rects = ax.patches
labels = bar_plot.SOMA
for rect, label in zip(rects, labels):
 height = rect.get_height()
 ax.text(rect.get_x() + rect.get_width()/2, height + 1, label, ha='center', ↪
 ↪va='bottom', fontsize=14)
plt.show()
```

/home/modanez/anaconda3/envs/servio/lib/python3.8/site-packages/seaborn/\_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

```
warnings.warn(
```



```
[6]: bar_plot = pd.DataFrame()
bar_plot['ACORDAOS'] = df_rotulados.columns[2:]
bar_plot['SOMA'] = df_rotulados.iloc[:,2:].sum().values
bar_plot.sort_values(['SOMA'], inplace=True, ascending=False)
bar_plot.reset_index(inplace=True, drop=True)
print(bar_plot)
```

```
[7]: columns=['DNN', 'CNN', 'LSTM', 'Bi-LSTM']
results = pd.DataFrame(columns = columns)
results
```

```
[7]: Empty DataFrame
Columns: [DNN, CNN, LSTM, Bi-LSTM]
Index: []
```

## 1 Deep Learning

### 1.1 DNN

```
[8]: tokenizer = Tokenizer(num_words=10000, lower=True)
tokenizer.fit_on_texts(df_rotulados.DECISAO)
sequences = tokenizer.texts_to_sequences(df_rotulados.DECISAO)
x = pad_sequences(sequences, maxlen=512)

seeds = [1, 43, 678, 90, 135, 77]
```

```
[9]: x_train, x_test, y_train, y_test = train_test_split(x,
 df_rotulados[df_rotulados.
<-columns[2:]],
 test_size=0.3,
 random_state=seeds[4])

[10]: bar_plot['PESO'] = len(bar_plot) / bar_plot['SOMA']
class_weight = {}
categories = df_rotulados.columns[2:]
for index, label in enumerate(categories):
 class_weight[index] = bar_plot[bar_plot['ACORDAOS'] == categories['PESO']].
<-values[0]

bar_plot

[11]: num_classes = y_train.shape[1]
max_words = len(tokenizer.word_index) + 1
maxlen = 256
maxlen = max(df_rotulados.DECISAO.apply(max_len))

[12]: from tensorflow.keras import backend as K

def recall_m(y_true, y_pred):
 true_positives = K.sum(K.round(K.clip(y_true * y_pred, 0, 1)))
 possible_positives = K.sum(K.round(K.clip(y_true, 0, 1)))
 recall = true_positives / (possible_positives + K.epsilon())
 return recall

def precision_m(y_true, y_pred):
 true_positives = K.sum(K.round(K.clip(y_true * y_pred, 0, 1)))
 predicted_positives = K.sum(K.round(K.clip(y_pred, 0, 1)))
 precision = true_positives / (predicted_positives + K.epsilon())
 return precision

def f1_m(y_true, y_pred):
 precision = precision_m(y_true, y_pred)
 recall = recall_m(y_true, y_pred)
 return 2*((precision*recall)/(precision+recall+K.epsilon()))

[13]: from tensorflow.keras.models import load_model
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Embedding, GlobalMaxPool1D, Dropout
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.callbacks import ReduceLROnPlateau, EarlyStopping,
<-ModelCheckpoint, TensorBoard
import tensorflow as tf
```

```

model = Sequential()
model.add(Embedding(max_words, 20, input_length=maxlen))
model.add(Dropout(0.2))
model.add(GlobalMaxPool1D())
model.add(Dense(num_classes, activation='sigmoid'))

model.compile(optimizer=Adam(0.015), loss='binary_crossentropy', metrics=[tf.
↳keras.metrics.AUC()])
callbacks = [
 ReduceLROnPlateau(),
 EarlyStopping(patience=10),
 TensorBoard('./logs/dnn'),
 ModelCheckpoint(filepath='./weights/dnn/model_dnn_reg_diac.h5',
↳save_best_only=True)
]

history = model.fit(x_train, y_train,
 class_weight=class_weight,
 epochs=100,
 batch_size=32,
 validation_split=0.4,
 callbacks=callbacks,
 verbose=0)

```

WARNING:tensorflow:Model was constructed with shape (None, 256) for input Tensor("embedding\_input:0", shape=(None, 256), dtype=float32), but it was called on an input with incompatible shape (None, 512).

WARNING:tensorflow:Model was constructed with shape (None, 256) for input Tensor("embedding\_input:0", shape=(None, 256), dtype=float32), but it was called on an input with incompatible shape (None, 512).

WARNING:tensorflow:From /home/modanez/anaconda3/envs/servio/lib/python3.8/site-packages/tensorflow/python/ops/summary\_ops\_v2.py:1277: stop (from tensorflow.python.eager.profiler) is deprecated and will be removed after 2020-07-01.

Instructions for updating:  
use `tf.profiler.experimental.stop` instead.

WARNING:tensorflow:Model was constructed with shape (None, 256) for input Tensor("embedding\_input:0", shape=(None, 256), dtype=float32), but it was called on an input with incompatible shape (None, 512).

```

[14]: dnn_model = model
 metrics = dnn_model.evaluate(x_test, y_test)
 print("{}: {}".format(dnn_model.metrics_names[1], metrics[1]))

```

```

3/3 [=====] - 0s 2ms/step - loss: 0.3804 - auc: 0.7982
auc: 0.7981802225112915

```

```
[15]: results.loc[0, 'DNN'] = metrics[1]
 results
```

```
[15]: DNN CNN LSTM Bi-LSTM
 0 0.79818 NaN NaN NaN
```

## 1.2 CNN

```
[16]: from tensorflow.keras.layers import Dense, Activation, Embedding, Flatten,
 ↪ GlobalMaxPool1D, Dropout, Conv1D

 filter_length = 300

 model = Sequential()
 model.add(Embedding(max_words, 20, input_length=maxlen))
 model.add(Dropout(0.2))
 model.add(Conv1D(filter_length, 3, padding='valid', activation='relu',
 ↪ strides=1))
 model.add(GlobalMaxPool1D())
 model.add(Dense(num_classes))
 model.add(Activation('sigmoid'))

 model.compile(optimizer='adam', loss='binary_crossentropy', metrics=[tf.keras.
 ↪ metrics.AUC()])

 callbacks = [
 ReduceLROnPlateau(),
 EarlyStopping(patience=10),
 TensorBoard('./logs/cnn'),
 ModelCheckpoint(filepath='./weights/cnn/model_cnn_reg_diact.h5',
 ↪ save_best_only=True)
]

 history = model.fit(x_train, y_train,
 class_weight=class_weight,
 epochs=100,
 batch_size=32,
 validation_split=0.4,
 callbacks=callbacks,
 verbose = 0)
```

WARNING:tensorflow:Model was constructed with shape (None, 256) for input Tensor("embedding\_1\_input:0", shape=(None, 256), dtype=float32), but it was called on an input with incompatible shape (None, 512).

WARNING:tensorflow:Model was constructed with shape (None, 256) for input Tensor("embedding\_1\_input:0", shape=(None, 256), dtype=float32), but it was called on an input with incompatible shape (None, 512).

WARNING:tensorflow:Callbacks method `on\_train\_batch\_end` is slow compared to the batch time (batch time: 0.0125s vs `on\_train\_batch\_end` time: 0.0643s). Check your callbacks.

WARNING:tensorflow:Model was constructed with shape (None, 256) for input Tensor("embedding\_1\_input:0", shape=(None, 256), dtype=float32), but it was called on an input with incompatible shape (None, 512).

```
[17]: cnn_model = model
 metrics = cnn_model.evaluate(x_test, y_test)
 print("{}: {}".format(cnn_model.metrics_names[1], metrics[1]))
```

```
3/3 [=====] - 0s 4ms/step - loss: 0.3610 - auc_1:
0.8269
auc_1: 0.8268698453903198
```

```
[18]: results.loc[0, 'CNN'] = metrics[1]
 results
```

```
[18]: DNN CNN LSTM Bi-LSTM
0 0.79818 0.82687 NaN NaN
```

### 1.3 GLOVE - LSTM

```
[19]: from gensim.models import KeyedVectors
```

```
[20]: embedding = KeyedVectors.load_word2vec_format('embedding/n1lc/glove_s50.txt',
 ↪ encoding='utf8')
```

```
NUM_WORDS = 32148
EMBEDDING_DIM=embedding.vector_size
word_index = tokenizer.word_index
vocabulary_size=min(len(word_index)+1, NUM_WORDS)
embedding_matrix = np.zeros((vocabulary_size, EMBEDDING_DIM))
for word, i in word_index.items():
 if i>=NUM_WORDS:
 continue
 try:
 embedding_vector = embedding[word]
 embedding_matrix[i] = embedding_vector
 except KeyError:
 embedding_matrix[i]=np.random.normal(0,np.sqrt(0.25),EMBEDDING_DIM)
```

```
[21]: from tensorflow.keras.layers import Dense, Activation, Embedding, Flatten,
 ↪ GlobalMaxPool1D, Dropout, Conv1D, LSTM
```

```
model = Sequential()
```

```

model.add(Embedding(vocabulary_size, EMBEDDING_DIM, weights=[embedding_matrix],
↳ trainable=False))
model.add(LSTM(128))
model.add(Dropout(0.3))
model.add(Dense(num_classes))
model.add(Activation('sigmoid'))

model.compile(optimizer='adam', loss='binary_crossentropy', metrics=[tf.keras.
↳ metrics.AUC()])

callbacks = [
 ReduceLROnPlateau(),
 EarlyStopping(patience=10),
 TensorBoard('./logs/lstm'),
 ModelCheckpoint(filepath='./weights/cnn/model_lstm_reg_diact.h5',
↳ save_best_only=True)
]

model.summary()

history = model.fit(x_train, y_train,
 class_weight=class_weight,
 epochs=30,
 batch_size=32,
 validation_split=0.4,
 callbacks=callbacks,
 verbose = 0)

```

WARNING:tensorflow:Callbacks method `on\_train\_batch\_end` is slow compared to the batch time (batch time: 0.0161s vs `on\_train\_batch\_end` time: 0.0816s). Check your callbacks.

```

[22]: lstm_model = model
metrics = lstm_model.evaluate(x_test, y_test)
print("{}: {}".format(lstm_model.metrics_names[1], metrics[1]))

```

```

3/3 [=====] - 0s 7ms/step - loss: 0.3915 - auc_2:
0.7815
auc_2: 0.7815449237823486

```

```

[23]: results.loc[0, 'LSTM'] = metrics[1]
results

```

```

[23]: DNN CNN LSTM B1-LSTM
0 0.79818 0.82687 0.781545 NaN

```

## 1.4 GLOVE - Bi-LSTM

```
[24]: from tensorflow.keras.layers import Dense, Activation, Embedding, Flatten,
 ↳ GlobalMaxPool1D, Dropout, Conv1D, LSTM, Bidirectional

model = Sequential()
model.add(Embedding(vocabulary_size, EMBEDDING_DIM, weights=[embedding_matrix],
 ↳ trainable=False))
model.add(Bidirectional(LSTM(128, return_sequences=True, dropout=0.1,
 ↳ recurrent_dropout=0.3)))
model.add(Dropout(0.3))
model.add(GlobalMaxPool1D())
model.add(Dense(128, activation='relu'))
model.add(Dropout(0.3))
model.add(Dense(6, activation='sigmoid'))

model.compile(optimizer='adam', loss='binary_crossentropy', metrics=[tf.keras.
 ↳ metrics.AUC()])

callbacks = [
 ReduceLROnPlateau(),
 EarlyStopping(patience=10),
 TensorBoard('./logs/lstm'),
 ModelCheckpoint(filepath='./weights/cnn/model_lstm_reg_diact.h5',
 ↳ save_best_only=True)
]

model.summary()

history = model.fit(x_train, y_train,
 class_weight=class_weight,
 epochs=30,
 batch_size=32,
 validation_split=0.4,
 callbacks=callbacks,
 verbose = 0)
```

WARNING:tensorflow:Layer lstm\_1 will not use cuDNN kernel since it doesn't meet the cuDNN kernel criteria. It will use generic GPU kernel as fallback when running on GPU

WARNING:tensorflow:Layer lstm\_1 will not use cuDNN kernel since it doesn't meet the cuDNN kernel criteria. It will use generic GPU kernel as fallback when running on GPU

WARNING:tensorflow:Layer lstm\_1 will not use cuDNN kernel since it doesn't meet the cuDNN kernel criteria. It will use generic GPU kernel as fallback when running on GPU

```
[25]: bilstm_model = model
metrics = bilstm_model.evaluate(x_test, y_test)
print("{}: {}".format(bilstm_model.metrics_names[1], metrics[1]))

3/3 [=====] - 0s 96ms/step - loss: 0.3947 - auc_3:
0.8224
auc_3: 0.8223943710327148

[26]: results.loc[0, 'Bi-LSTM'] = metrics[1]
results

[26]: DNN CNN LSTM Bi-LSTM
0 0.79818 0.82687 0.781545 0.822394

[27]: df_acordaos = pd.read_csv('dados/df_acordaos_proc.csv', sep='|')

[28]: start_time = datetime.now()
df_acordaos.DECISAO = hero.remove_diacritics(df_acordaos.DECISAO)
print('Duration: {}'.format(datetime.now() - start_time))

Duration: 0:06:36.879381

[29]: start_time = datetime.now()
tokenizer.fit_on_texts(df_acordaos.DECISAO)
print('Duration: {}'.format(datetime.now() - start_time))

start_time = datetime.now()
sequences = tokenizer.texts_to_sequences(df_acordaos.DECISAO)
print('Duration: {}'.format(datetime.now() - start_time))

start_time = datetime.now()
x = pad_sequences(sequences, maxlen=maxlen)
print('Duration: {}'.format(datetime.now() - start_time))

Duration: 0:02:26.951062
Duration: 0:01:31.191726
Duration: 0:00:04.030195

[38]: saida = cnn_model.predict(x)
df_predict = pd.DataFrame(data=saida, columns=list(df_rotulados.columns[2:]))
pd.concat([df_acordaos.drop(['DECISAO'], axis=1), df_predict], axis=1).
 to_csv('dados/df_predict.csv', sep='|', index=False)

[]: %load_ext tensorboard
%tensorboard --logdir logs

[]: tf.keras.utils.plot_model(dnn_model)
tf.keras.utils.plot_model(cnn_model)
tf.keras.utils.plot_model(lstm_model)
tf.keras.utils.plot_model(bilstm_model)
```

## APÊNDICE H – LINHA JURISPRUDENCIAL DE ENTENDIMENTO DA CLASSE 7

