

Ayrton Bittencourt Lobo Neto
Fernanda de Santana Villa

José Carlos C. Wasconcellos Jr.
Leonardo Sergio Pontes Gaudenzi¹

Orientador: Prof. Oberdan Pinheiro

**INTELIGÊNCIA ARTIFICIAL PARA A PREDIÇÃO DO ÊXITO
DO ESTADO DA BAHIA EM NOVAS AÇÕES JUDICIAIS**

Salvador - Bahia
Novembro de 2024

¹ Procuradores do Estado da Bahia

RESUMO

O trabalho aborda a aplicação de uma sequência de modelos de inteligência artificial (IA), utilizando técnicas de aprendizado de máquina e processamento de linguagem natural, para viabilizar a predição do resultado de processos judiciais desde a sua ciência pela Procuradoria Geral do Estado da Bahia. A partir do cenário verificado no contencioso judicial da PGE-BA, o trabalho apresenta o desenvolvimento da estratégia, desde a concepção, levantamento e pré-processamento dos dados até a construção dos modelos de IA e a execução de uma prova de conceito, avaliando os resultados obtidos.

Palavras-Chave: Processo judicial. Gestão do contencioso. Predição de êxito. Inteligência Artificial. Aprendizado de máquina. Processamento de Linguagem Natural.

ABSTRACT

This paper explores the application of a sequence of artificial intelligence (AI) models, utilizing machine learning and natural language processing techniques, to predict the outcomes of legal proceedings from the point of their registration by the Attorney General's Office of the State of Bahia. Based on the judicial litigation observed at PGEBA, this study details the development of the strategy, from initial design, data collection, and pre-processing through to model construction and the execution of a proof of concept, with an evaluation of the results obtained.

Keywords: Legal case. Litigation management. Success prediction. Artificial intelligence. Machine learning. Natural Language Processing.

Sumário

1. INTRODUÇÃO.....	4
2. DESAFIO E CONTEXTUALIZAÇÃO.....	4
3. A ESTRATÉGIA PENSADA	8
4. A COLETA DE DADOS	11
5. MÉTODOS UTILIZADOS NO PRÉ-PROCESSAMENTO DOS DADOS	13
6. EXECUÇÃO DA PROVA DE CONCEITO DO PROJETO.....	17
6.1 MODELO (C).....	17
A. Representação de Texto com <i>Word Embeddings</i>	18
a) Treinamento do Modelo Word2Vec:.....	18
b) Processo de Treinamento:.....	18
c) Utilização no Modelo de Classificação:.....	19
B. Pré-processamento dos Dados	19
C. Construção do Modelo de Rede Neural	20
6.2 MODELO (D).....	22
A. Geração do <i>dataset</i>	22
B. Pré-processamento dos Dados	23
C. Construção do Modelo de Rede Neural	23
7. RESULTADOS.....	30
8. REFERÊNCIAS	31

1. INTRODUÇÃO

Nos últimos anos, a Procuradoria Geral do Estado da Bahia (PGE) tem enfrentado uma crescente sobrecarga de processos judiciais, especialmente em casos nos quais o Estado figura no polo passivo.

Diante desse cenário, a gestão eficiente do contencioso se tornou um desafio crítico, exigindo soluções tecnológicas que auxiliem no controle e planejamento das ações judiciais, com foco em otimizar o uso dos recursos públicos e maximizar a efetividade de sua atuação.

Este trabalho propõe uma abordagem inovadora para a resolução desse problema, através do desenvolvimento de modelos de inteligência artificial (IA) que predigam, com uma margem de erro conhecida, a probabilidade de êxito do Estado em processos judiciais recém-ingressados.

2. DESAFIO E CONTEXTUALIZAÇÃO

O processo de redemocratização do Brasil culminou com a promulgação da Constituição de 1988; uma Constituição garantista, que colocou o Poder Público na posição de assegurado de uma larga série de direitos, trazendo novos e enormes desafios para a gestão pública na construção de políticas públicas capazes de assegurar a efetivação dos direitos constitucionais.

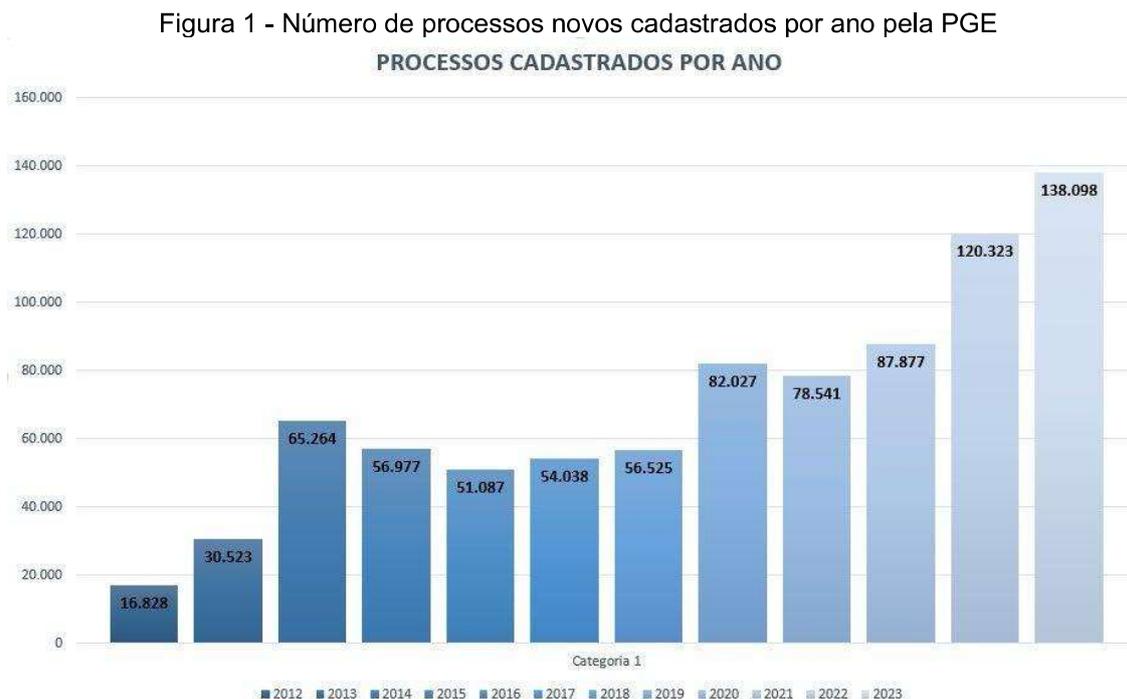
Em acréscimo, a modernização do Processo judicial, tanto no que diz respeito ao arcabouço normativo – com a promulgação da Lei Federal nº. 13.105, de 16 de março de 2015 (Código de Processo Civil) – como no que se refere à incorporação da tecnologia, com o surgimento do processo judicial eletrônico, o peticionamento remoto e a realização de audiências virtuais, por exemplo, ampliaram o acesso ao Poder Judiciário, democratizando o processo judicial e exigindo do Poder Público a adoção de medidas necessárias à gestão do contencioso.

De acordo com o Portal Justiça em Números, do Conselho Nacional de Justiça, no ranking nacional dos 20 (vinte) maiores litigantes, considerando apenas o polo passivo da ação (ou seja, a posição de réu no processo judicial), 10 (dez) são entes públicos, e o estado da Bahia ocupa a 14^a (décima quarta) posição [1].

Conforme dados internos da Procuradoria Geral do Estado da Bahia, o acervo de processos judiciais sob os cuidados dessa instituição ultrapassou a marca de 1.000.000 (um milhão) de processos em 2024.

Considerando que atualmente há 246 (duzentos e quarenta e seis) Procuradores em atividade, significa dizer que se o contencioso fosse a única atribuição da PGE (o que não é verdade), e todos os Procuradores estivessem voltados às lides jurisdicionais (o que não é possível), caberia a cada um, indistintamente, cerca de 4.000 (quatro mil) processos judiciais, o que já seria um absurdo.

O aumento exponencial de novas demandas judiciais segue ilustrado pelo gráfico da Figura 1, que registra o volume de novos processos judiciais recepcionados pela Procuradoria Geral do Estado da Bahia entre 2012 e 2023:



Fonte: base de dados internos da PGE.

Como consequência do aumento exponencial do estoque de processos judiciais, resulta a enorme majoração do volume de demanda, que apresenta crescimento contínuo e alarmante, tendo ultrapassado, no ano de 2023, a marca de

660 mil movimentações processuais tratadas pela Procuradoria Geral do Estado da Bahia, como registra a Figura 2:



Fonte: base de dados internos da PGE.

O problema é alarmante e chama a atenção dos Entes Públicos para a necessidade urgente de adoção de medidas voltadas à gestão do contencioso, ou seja, desse estoque de processos judiciais e demandas dele derivadas.

Não apenas isso. É preciso reconhecer que, enquanto garantidor de direitos, compete ao Estado melhor estudar as postulações que lhe são judicialmente dirigidas, a fim de adotar soluções que assegurem adequadamente os direitos dos jurisdicionados.

Nesse universo de alarmante crescimento de demandas para serem atendidas e tratadas pela Procuradoria Geral do Estado da Bahia, com seus limitados recursos financeiros e de pessoal, é imperativo o desenvolvimento de ferramentas tecnológicas que possam contribuir com o gestor na busca por soluções sistêmicas, com adoção de medidas alternativas para a solução de conflitos, inclusive no intuito de reduzir a litigiosidade.

Até porque o custo do litígio é muito alto.

De acordo com o referido Painel Justiça em Números, do Conselho Nacional de Justiça, só com despesa de pessoal, o Poder Judiciário gastou, no ano de 2023,

R\$ 132.750.000.000,00 (cento e trinta e dois bilhões, setecentos e cinquenta milhões de reais) no Brasil [2];

Só no Tribunal de Justiça da Bahia, de acordo com o Painel em referência, a despesa de pessoal alcançou R\$ 5.220.000.000,00 (cinco bilhões, duzentos e vinte milhões de reais) no ano de 2023 [2].

Se essas despesas forem somadas às de pessoal do Tribunal Regional Eleitoral, do Tribunal Regional Federal da 1ª Região, e do Tribunal Regional do Trabalho da 5ª Região, totalizam, apenas no ano de 2023, R\$ 10.070.000.000,00 (dez bilhões e setenta milhões de reais) [2].

Cumprе recordar que o sistema de Justiça envolve não apenas o Poder Judiciário, mas também o Ministério Público, a Advocacia Pública e a Defensoria Pública, o que decerto implica em expressiva majoração dessas e outras despesas, para a sua manutenção e desenvolvimento, que repercutem no alto custo do litígio no Brasil.

É esse o contexto em que o presente trabalho foi desenvolvido.

A consciência de que os Estados precisam gerenciar o seu contencioso para melhor se desincumbir de seu papel constitucional de assegurador de direitos provocam as seguintes reflexões:

E se, diante do recebimento da citação em uma nova ação judicial, o oferecimento de defesa (contestação) não fosse a única alternativa possível? E se, dependendo da situação jurídica posta, a conduta mais adequada fosse chamar a parte contrária para uma mesa de negociação? Ou, ainda, e se a melhor conduta para o Ente Público fosse reconhecer o pedido judicialmente feito pelo cidadão, e adotar providências internas, para que o problema apontado por ele naquela ação judicial não voltasse a ocorrer?

Para a gestão, inclusive do contencioso, a análise de dados é essencial.

E se, para além da análise de dados, fosse possível, por meio de ferramentas tecnológicas com Inteligência Artificial incorporada, a previsão do resultado final do processo?

Desenvolver modelo de inteligência artificial que infira, com margem de erro conhecida e razoável, a probabilidade de êxito do Estado em uma ação ajuizada contra ele é o grande propósito do trabalho aqui desenvolvido.

A partir dessa predição, a Procuradoria Geral do Estado da Bahia poderá, de forma efetiva, gerir o contencioso, definindo qual a conduta mais adequada deve ser adotada pelo Estado da Bahia na condução dos processos judiciais.

3. A ESTRATÉGIA PENSADA

Conforme exposto, a Procuradoria Geral do Estado da Bahia não possui atualmente meios para estimar os resultados dos processos judiciais movidos contra os entes públicos que representa, o que termina limitando a atuação em novos processos ao desenvolvimento de novas teses ou à repetição de defesas anteriormente produzidas, sem considerar a probabilidade de ganho ou perda, o que lhe permitiria a melhor definição de estratégias e dimensionamento de esforços.

Há, contudo e como dito acima, um vasto acervo de processos judiciais em que a Procuradoria Geral do Estado já atuou e que poderia, pela análise de seus dados e por técnicas de reconhecimento de padrões, contribuir para a inferência dos resultados de novos processos. Dessa forma, é possível estabelecer estratégias mais informadas para a condução das ações judiciais.

Afinal, se o processo judicial é um conjunto de atos ordenados praticados pelos sujeitos envolvidos a partir de regras pré-estabelecidas, a combinação das características do próprio processo, de seus elementos e de seus sujeitos podem ser determinantes para o resultado da ação judicial.

Seguindo essa linha de raciocínio, e diante da impossibilidade de centenas de milhares de processos judiciais serem analisados por uma atuação humana, os autores deste estudo propõem uma solução inovadora: a aplicação de uma sequência de modelos de inteligência artificial (IA), utilizando técnicas de aprendizado de máquina e processamento de linguagem natural [3], para viabilizar a predição do resultado de processos judiciais desde a sua ciência pela Procuradoria Geral do Estado da Bahia.

A aplicação sequenciada de mais de um modelo decorre da necessidade de montar uma base de dados rica e com variáveis que possam ser determinantes, visando ao treinamento do modelo de predição do resultado do processo judicial.

Em síntese, o trabalho envolve a utilização de quatro modelos de inteligência artificial:

- (A) **Modelo de classificação por Área:** responsável por identificar a área jurídica (Matéria) envolvida em cada processo;
- (B) **Modelo de classificação por Assunto:** destinado à categorização dos processos segundo o tema específico em discussão;
- (C) **Modelo de classificação de sentenças:** visa à análise das sentenças com foco em determinar o teor favorável ou desfavorável ao ente público;
- (D) **Modelo de Predição de Resultados:** aplica-se na avaliação da probabilidade de êxito da PGE em novos processos.

Importante destacar que este trabalho se relaciona diretamente com as entregas da primeira Encomenda Tecnológica² contratada pela Procuradoria Geral do Estado da Bahia junto ao SENAI CIMATEC, na medida em que os modelos supervisionados utilizados para a classificação do processo por Área e Assunto (A e B) são os mesmos já desenvolvidos, treinados e entregues na execução desse contrato.

Partindo, portanto, da premissa de que esses dois primeiros modelos já existem e estão disponíveis à Procuradoria Geral do Estado da Bahia, a estratégia consiste em utilizá-los para agregar novos e determinantes atributos para a análise dos processos judiciais novos. Juntamente com os dados estruturados do processo e das partes, tais atributos serão submetidos ao modelo de predição (D), que, ao final, indicará uma escala de probabilidade entre 0 e 1 para o êxito da PGE nesse novo processo.

Note-se que o último Modelo (D) será aplicado exclusivamente sobre dados estruturados, conforme ilustra a Figura 3:

Figura 3 – Atributos do Modelo (D)

	Dado cadastral do processo 1	Dado cadastral do processo 2	Dado cadastral do processo 3	Dado cadastral do processo ...	Dados das partes 1	Dados das partes 2	Dados das partes 3	Dados das partes ...	Área	Assunto	Probabilidade de Êxito da PGE
Fonte	Attus/PJE	Attus/PJE	Attus/PJE	Attus/PJE	RHBahia	RHBahia	RHBahia	RHBahia	Modelo IA (A)	Modelo IA (B)	Modelo IA (D)

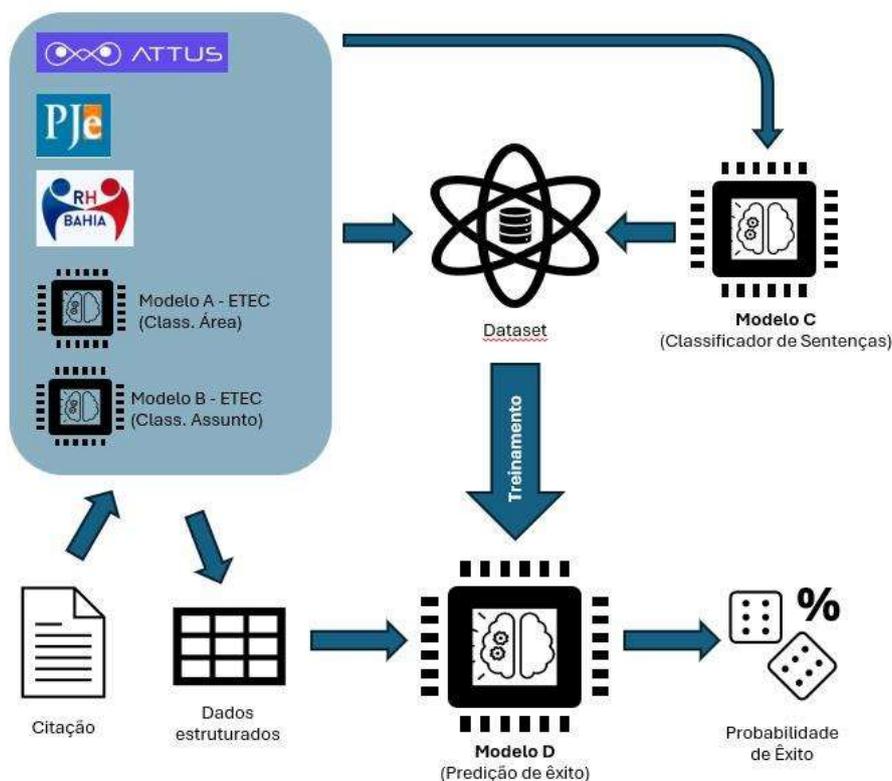
Fonte: Autores.

² Contrato n°. PGE 032/2022, celebrado em 01/07/2022.

A estratégia pensada, portanto, permitirá utilizar apenas dados estruturados para o consumo diário do modelo de predição, reduzindo o esforço computacional necessário.

A Figura 4, abaixo, demonstra os fluxos e os relacionamentos entre as fontes dos dados, os quatro modelos de inteligência artificial acima citados e o *dataset*, destacando, também, a entrada e a saída do modelo de predição (D):

Figura 4 – ilustração dos fluxos e relacionamentos na estratégia pensada



Fonte: Autores.

Note-se que a que a saída do modelo de classificação de sentenças (C) não é utilizada para a *aplicação* do modelo de predição (D), mas apenas para o *treinamento* deste, enriquecendo o *dataset*.

De fato, o modelo de classificação de sentenças (C) tem o exclusivo propósito de conferir mais um atributo determinante a cada processo que componha o *dataset* de treinamento do modelo de predição (D), indicando, por análise de sentimentos, se

a decisão foi ou não favorável ao ente público representado pela Procuradoria Geral do Estado. Prevê-se que, anualmente, o modelo de predição (D) precise ser retreinado com novas sentenças que tenham sido proferidas nos 12 (doze) meses anteriores em processos judiciais sob o acompanhamento da PGE, de modo a manter o contexto atualizado.

Como esse modelo de classificação de sentenças (C) não estava disponível à Procuradoria Geral do Estado no momento da concepção da estratégia, tornou-se necessário incluir o seu desenvolvimento e treinamento no escopo da análise experimental tratada neste trabalho, como se verá adiante.

4. A COLETA DE DADOS

Para o treinamento dos modelos de classificação de sentenças (C) e de predição do êxito (D), os autores deste trabalho coletaram os primeiros dados a partir do sistema *Attus*³, que se integra ao sistema *PJE – Processo Judicial Eletrônico*⁴ por meio de *WebService* padronizado por *MNI – Modelo Nacional de Interoperabilidade*.

Os dados foram extraídos a partir da geração de relatório, em formato *xlsx* (Microsoft Excel), de processos judiciais que atendessem concomitantemente aos seguintes parâmetros:

- cadastrados a partir do ano de 2018;
- em trâmite em uma das varas do sistema dos Juizados Especiais da Fazenda Pública da Bahia;
- a PGE represente o polo passivo;
- sentenciados a partir do mês de julho de 2023.

Gerou-se, com base nesses critérios, um relatório contendo 24.528 processos, a partir do qual foi possível obter as chaves necessárias para requisitar,

³ Sistema de gerenciamento e acompanhamento de processos judiciais desenvolvido pela empresa *Attus* e implantado na Procuradoria Geral Estado, em janeiro de 2024, no qual estão cadastrados todos os processos em que a PGE atua.

⁴ Sistema de tramitação de processos judiciais desenvolvido pelo Conselho Nacional de Justiça e implantado pelo Tribunal de Justiça do Estado da Bahia desde 2014.

via *WebService* ao sistema PJE, os documentos do tipo sentença em formatos *html* ou *pdf*. Das 24.528 sentenças requisitadas, por restrições computacionais só retornaram 18.994 sentenças, que passaram a integrar a base de dados utilizada para o treinamento e teste do modelo de classificação de sentenças (C).

Com base no mesmo relatório, agora reduzido para 18.994 processos judiciais (para coincidir com os processos cujas sentenças se conseguiu obter), foi possível a coleta dos seguintes dados e variáveis diretamente do sistema Attus:

- Data de ajuizamento
- Data de cadastro (recebimento da citação pela PGE)
- Classe processual
- Valor da causa
- Matéria (Área)
- Assunto institucional
- Comarca
- Juízo
- Nome da parte Autora
- CPF/CNPJ da parte Autora
- Advogado da parte Autora
- CPF do advogado da parte Autora
- Pessoa jurídica representada pela PGE
- Data de sentença
- Tipo de andamento (classificação da sentença)⁵.

A partir de outro trabalho desenvolvido pelos mesmos autores durante o *Módulo 4 – Introdução à Estatística Descritiva* do curso de Especialização em Ciências de Dados e Inteligência Artificial Aplicados às Ciências Jurídicas⁶, intitulado *Panorama da Litigiosidade dos Servidores Públicos Estaduais contra o Estado da Bahia*, observou-se que, em 62% (sessenta e dois por cento) dos processos judiciais

⁵ Dado utilizado apenas para o treinamento do modelo de classificação de sentenças (C).

⁶ Ministrado pelo SENAI CIMATEC, turma 85584.

movidos por pessoas físicas contra o Estado da Bahia até 31/05/2023, a parte Autora é composta por servidores públicos do Poder Executivo estadual.

Diante disso, recorreu-se ao banco de dados do sistema *RH Bahia*⁷ para coletar novos dados que pudessem enriquecer o *dataset* de treinamento dos modelos de inteligência artificial. Utilizando, como chave primária, o CPF dos autores de cada um dos processos constantes do relatório extraído do sistema Attus, foi possível coletar os seguintes dados do *RH Bahia*:

- Cargo público
- Tipo de cargo (vínculo)
- Lotação
- Valor da última remuneração
- Órgão de atuação
- Situação funcional

Cumprido registrar que os autores deste projeto realizam atividades finalísticas e gerenciais com o mesmo conjunto de dados na Procuradoria Geral do Estado da Bahia, legitimando o seu acesso.

5. MÉTODOS UTILIZADOS NO PRÉ-PROCESSAMENTO DOS DADOS

Concluída a fase de coleta e agrupamento dos dados, a pesquisa evoluiu para a etapa seguinte, qual seja o pré-processamento desses dados, visando a execução de uma prova de conceito (*Proof of Concept - POC*) dos Modelos (C) e (D) [4].

Nesse ponto uma explicação se impõe em relação a opção dos autores por concluir este estudo com uma prova de conceito dos algoritmos propostos. Essa escolha está baseada em dois fundamentos.

⁷ Sistema informatizado de gestão de Recursos Humanos do Estado da Bahia. Implantado pela Secretaria de Administração do Estado da Bahia (SAEB) em parceria com a Companhia de Processamento de Dados do Estado (PRODEB) e a empresa contratada *Resource*, e disponível em todo o Estado desde janeiro de 2019. Mensalmente, a SAEB disponibiliza uma base de dados atualizada do RH Bahia à PGE, para alimentação e hidratação das suas soluções tecnológicas.

O primeiro deles diz respeito à realização de provas de conceito ser uma prática comum no mercado de tecnologia. Isso porque a complexidade dos desafios a serem superados pelos produtos faz com que a mera avaliação de descritivos técnicos não garanta a um contratante se um produto ou serviço de fato será a solução ideal para o seu problema. Assim, a prova de conceito se apresenta como uma forma de testar um produto não sobre um conjunto de dados fictícios, mas sobre um recorte dos efetivos dados e cenários que serão utilizados pelo contratante se vier a efetivar a aquisição do produto.

O segundo, e mais relevante, seria a inviabilidade do trabalho de conclusão de curso diante da complexidade da tarefa que se apresentaria para os autores, se pretendessem eles realizar coleta, pré-processamento e treinamento de inteligência artificial sobre o escopo completo das situações que são alvo de tratamento judicial no âmbito da Procuradoria Geral do Estado da Bahia.

Feito esses esclarecimentos e partindo dos dados reunidos no item anterior, foram constituídos dois *datasets* visando o desenvolvimento dos Modelos (C) e (D). Tais dados, em maior ou menor grau, não se apresentavam prontos para o processamento pelos modelos de inteligência artificial, pelo que se fez necessário o tratamento desses conjuntos de dados.

O primeiro *dataset*, a ser utilizado como matéria prima do Modelo (C), que visa a classificação das sentenças judiciais, foi montado inicialmente a partir dos dados extraídos do sistema Attus, conforme já indicado.

A planilha obtida trazia os seguintes atributos: Nº da Pasta, Nº Processo, UF, Município, Matéria, Assunto do Processo no sistema da PGE, Tipo de Andamento, Polo, Agrupados Tipo de Andamento, Data do Andamento, Instância e Situação do Processo.

Avaliando tais dados, três situações foram norteadoras do primeiro ajuste dos dados: a) a saída desejada para esse modelo era a avaliação se a sentença de um processo judicial era favorável ou não para o Estado; b) tal resultado deveria ser apresentado numa grandeza de 0 (integralmente desfavorável) a 1 (integralmente favorável); c) o resultado do modelo deveria ser obtido através de processamento da linguagem natural contida nas sentenças a serem avaliadas.

A partir dessas condicionantes, decidiu-se pela desnecessidade, para a finalidade de treinamento do Modelo (C), de praticamente todos os atributos que

foram extraídos do sistema Attus. Permaneceram no grupo de dados apenas o N° Processo (CNJ), a Data do Andamento e o Tipo de Andamento.

Os dois primeiros atributos foram mantidos tão somente para que fosse possível, através de uma API⁸, a busca da decisão judicial que seria classificada.

Já o terceiro foi preservado por se tratar de rótulo para confirmação do resultado do modelo, configurando tal informação como marcador para avaliação da classificação.

Dessa forma, processados esses ajustes e feita a referida extração das sentenças que seriam classificadas, estava pronto o primeiro conjunto de dados, que viria a compor o *dataset* para a aplicação do modelo de classificação das sentenças judiciais.

Já o segundo *dataset*, que serviria de insumo para o Modelo (D), reservou um maior trabalho para que fossem feitos os ajustes necessários para que seu processamento pelo algoritmo escolhido fosse possível.

Duas planilhas serviram de base para a montagem desse grupo de dados, uma extraída do sistema Attus, que continha dados estruturados eminentemente relacionados ao processo judicial, e uma segunda planilha, extraída do sistema RH Bahia, que trazia dados funcionais dos servidores do Poder Executivo estadual que eram autores de uma boa parte dessas ações.

O primeiro processamento feito foi buscar trazer os dados da segunda planilha para a primeira. Para tanto, identificou-se o CPF como o dado que servia como chave em ambas as planilhas e se procedeu a unificação.

A partir dessa planilha unificada, passou-se à fase de pré-processamento dos dados, utilizando-se o Microsoft Power Query. Nessa etapa, mais de 50 operações de processamento nas colunas foram realizadas, contemplando desde operações mais simples, como remoção de colunas, até situações mais complexas, como divisão de colunas a partir de *strings* pré-determinados, visando corrigir inconsistências de formatação contidas nos dados.

Numa situação particularmente trabalhosa, foi necessário remover do campo “Lotação” os dados sobre o “Órgão de Atuação”, visto que há no *dataset* coluna específica para essa última informação, que se repetia no primeiro campo.

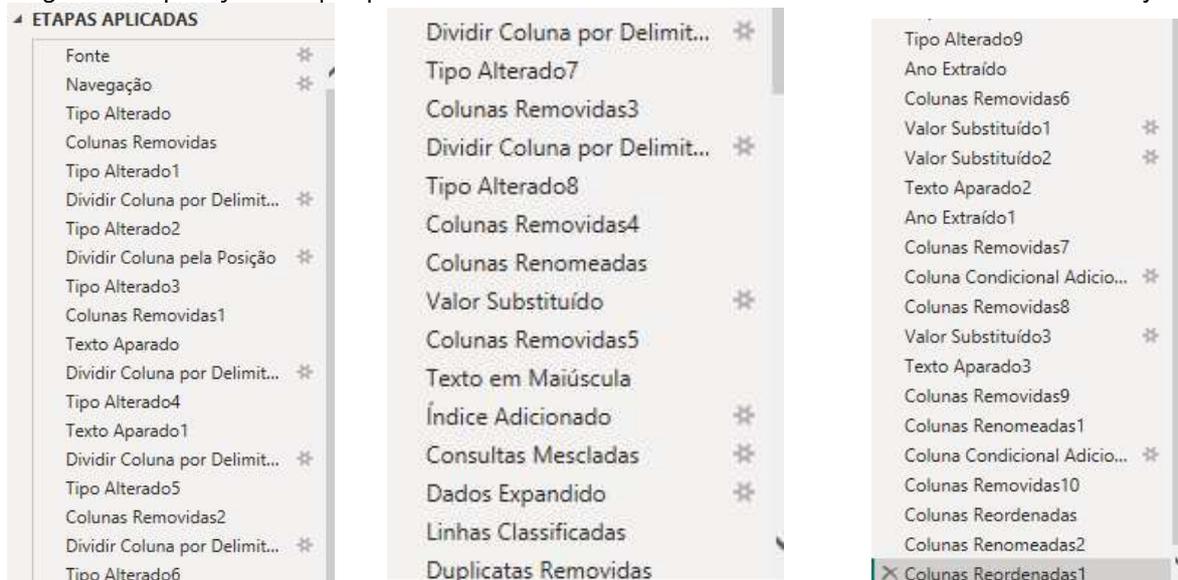
⁸ Application Programming Interface. Ferramenta que possibilita que dois sistemas troquem dados entre si por meio da internet sem intervenção do usuário.

Mais ainda, nessa oportunidade algumas decisões de negócio, ligadas aos dados, foram tomadas. Por exemplo, a conversão de todos os campos de data para o formato ano, em razão do julgamento feito pela equipe de que a relevância da informação se perderia se considerado o dia exato de cada marco, ao invés do ano, que funcionaria como um agrupador de dados, preservando a relevância desse atributo para o processamento das informações.

Nesse pré-processamento também foram afastadas algumas instâncias que apresentavam dados incompletos ou que possuíam dados muito divergentes do conjunto do *dataset*.

A Figura 5 expõe a lista integral dessas operações:

Figura 5 - Operações de pré-processamento e tratamento realizadas no Microsoft Power Query



Fonte: Autores.

Uma vez concluída essa etapa de pré-processamento, foi gerada nova planilha de Excel, agora com os dados prontos para serem submetidos ao Modelo (D). Para tal finalidade, apenas se acrescentou uma coluna *class*, justamente a que representa o dado que se pretende encontrar, qual seja, a análise de sentimento em relação a um processo ainda em fase inicial sobre o seu desfecho.

Tal atributo *class*, como já destacado, deverá ser complementado, na porção deste *dataset* separada para treinamento do modelo, pelo resultado da predição para cada um desses processos realizados pelo Modelo (C).

Desta forma, o projeto chegou ao ponto em que estava pronto para a execução da prova de conceito, cujo roteiro ideal consistia em rodar o Modelo (C), transladar os resultados obtidos para o *dataset* do Modelo (D), para poder treinar este modelo, e obter como resultado um algoritmo capaz de realizar a predição do resultado de uma ação judicial aforada contra a PGE no momento de seu ajuizamento.

Finalmente, necessário trazer algumas notas sobre os dados tratados neste processo e o alinhamento desta manipulação com a Lei Geral de Proteção de Dados.

Conforme indicado anteriormente, os dados obtidos são de acesso assegurado aos autores deste trabalho em função de sua atuação funcional. Ainda assim, cumpre anotar que se trata de dados que podem ser obtidos por qualquer cidadão, visto serem dados de natureza pública.

Mais ainda, mesmo diante da natureza pública desses dados, em alinhamento com a supracitada norma, procedeu-se a anonimização das informações sensíveis, notadamente no que tange a identificação das pessoas físicas que litigam contra o Estado nos processos usados para o treinamento dos modelos.

Dessa forma, os nomes dos autores foram descartados antes mesmo da realização do pré-tratamento no Microsoft Power Query. Por seu turno, os números dos CPF desses autores, necessários para que se procedesse o relacionamento das planilhas obtidas, foi alvo de processo de anonimização.

6. EXECUÇÃO DA PROVA DE CONCEITO DO PROJETO

A partir dos dados coletados e pré-processados nas etapas anteriores, foram construídos dois conjuntos de dados (*datasets*) para serem utilizados nos treinamentos dos Modelos (C) e (D).

6.1 MODELO (C)

O *dataset* elaborado conforme descrito no tópico anterior foi então utilizado para treinar o modelo de classificação, ensinando a inteligência artificial a correlacionar o conteúdo textual das sentenças ao seu respectivo desfecho em relação ao Estado da Bahia, conforme detalhado a seguir:

A. Representação de Texto com *Word Embeddings*

Ao invés de utilizar um modelo pré-treinado de Word2Vec, optou-se por treinar um modelo Word2Vec [5] a partir dos próprios textos das sentenças judiciais disponíveis no conjunto de dados. Essa abordagem permitiu capturar as nuances e especificidades do vocabulário jurídico presente nas sentenças, proporcionando representações vetoriais mais adequadas para a tarefa.

a) Treinamento do Modelo Word2Vec:

As sentenças judiciais foram coletadas e pré-processadas, incluindo a conversão dos textos para letras minúsculas e a remoção de caracteres não alfabéticos como pontuações e símbolos especiais, visando padronizar e limpar os dados textuais; em seguida, utilizou-se um *tokenizador* padrão com um *CommonPreprocessor* para segmentar o texto em tokens (palavras) e realizar limpeza básica adicional, como remoção de espaços extras e normalização; na configuração do modelo, definiu-se que apenas palavras com uma frequência mínima de 5 ocorrências seriam incluídas no vocabulário para filtrar termos muito raros que poderiam introduzir ruído; os vetores de *embeddings* foram configurados com 500 dimensões para capturar relações semânticas complexas entre as palavras; utilizou-se uma janela de contexto de tamanho 5, significando que o modelo considerou as 5 palavras anteriores e posteriores a uma palavra-alvo durante o treinamento para capturar relações de proximidade; o modelo foi treinado por 10 iterações (épocas) para assegurar a convergência e estabilidade das representações vetoriais, e uma semente fixa (42) foi utilizada para garantir a reprodutibilidade dos resultados.

b) Processo de Treinamento:

O treinamento foi conduzido utilizando a biblioteca *DeepLearning4j* (DL4J) [6], especificamente a classe Word2Vec. O modelo iterou sobre a coleção de sentenças pré-processadas, ajustando os vetores de palavras para minimizar a diferença entre palavras que aparecem em contextos semelhantes. O treinamento resultou em um modelo que encapsula as relações semânticas e sintáticas presentes nas sentenças judiciais do conjunto de dados, refletindo o uso específico da linguagem jurídica.

c) Utilização no Modelo de Classificação:

Os vetores de palavras gerados pelo modelo Word2Vec treinado foram utilizados para converter as sentenças em vetores de características. Isso foi feito calculando a média dos vetores das palavras presentes em cada sentença, resultando em um vetor representativo da sentença como um todo. Esses vetores de sentenças serviram como entrada para a rede neural do modelo de classificação, permitindo que a rede aprendesse a mapear características semânticas das sentenças aos rótulos de saída (favorável ou desfavorável ao Estado da Bahia).

Ao treinar o modelo com os próprios textos das sentenças judiciais, as representações vetoriais capturam termos técnicos e expressões específicas do contexto legal, melhorando a eficácia na tarefa de classificação. Palavras que aparecem em contextos similares nas sentenças terão vetores próximos no espaço vetorial, auxiliando o modelo de classificação a distinguir nuances entre sentenças favoráveis e desfavoráveis. A criação de um modelo de *embeddings* interno elimina possíveis incompatibilidades ou defasagens que poderiam ocorrer ao utilizar modelos pré-treinados em domínios diferentes.

B. Pré-processamento dos Dados

O conjunto de dados consistia em textos de sentenças judiciais e seus respectivos rótulos 1 (favorável) ou 0 (desfavorável); o pré-processamento envolveu os seguintes passos: as sentenças foram *tokenizadas*, dividindo-as em *tokens* (palavras) utilizando espaços em branco como delimitadores; em seguida, realizou-se a limpeza do texto, removendo caracteres especiais e normalizando o texto para padronização; para otimizar o processamento e focar nas partes mais relevantes, textos com mais de 500 caracteres foram truncados para considerar apenas os últimos 500 caracteres, baseando-se na premissa de que o desfecho da sentença geralmente está presente no final do documento; por fim, cada sentença foi convertida em um vetor de características calculado pela média dos vetores das palavras presentes no texto, obtidos do modelo Word2Vec, considerando apenas as palavras reconhecidas pelo modelo para garantir a relevância semântica.

Para evitar viés no modelo devido a classes desbalanceadas, o conjunto de treinamento foi balanceado para conter quantidades iguais de sentenças favoráveis e

desfavoráveis ao Estado; para isso, as sentenças foram separadas em duas listas, uma para cada classe, identificou-se o tamanho da classe minoritária e, em seguida, ambas as classes foram reduzidas ao tamanho da classe minoritária por meio de subamostragem, garantindo assim um balanceamento adequado no conjunto de treinamento. Os dados foram embaralhados aleatoriamente e divididos em conjuntos de treinamento (70%) e teste (30%); o conjunto de treinamento foi utilizado para ajustar os pesos do modelo durante o processo de aprendizagem, enquanto o conjunto de teste serviu para avaliar o desempenho do modelo em dados não vistos durante o treinamento, permitindo estimar sua capacidade de generalização.

C. Construção do Modelo de Rede Neural

Desenvolveu-se uma Rede Neural Artificial do tipo *Multilayer Perceptron* (MLP) com múltiplas camadas densas, utilizando a biblioteca DL4J. A arquitetura da rede consistiu em uma camada de entrada que recebe vetores de dimensão 500 (resultantes da média dos *embeddings* das palavras); seguida por quatro camadas ocultas densas com quantidades de neurônios variando entre 800 e 200, utilizando a função de ativação ReLU e inicialização de pesos Xavier, onde algumas camadas incluíram regularização como *Dropout* para prevenir *overfitting*; e, por fim, uma camada de saída com dois neurônios (correspondentes às classes favorável e desfavorável), utilizando a função de ativação Sigmoid e a função de perda *Cross-Entropy* [7].

As configurações de treinamento incluíram a utilização do algoritmo de otimização Nadam (uma variante do Adam com *momentum* de Nesterov), visando uma convergência mais rápida e estável; aplicou-se a normalização do gradiente para evitar explosão de gradientes durante o treinamento; e foi adicionada penalização L2 aos pesos para reduzir *overfitting* e melhorar a capacidade de generalização do modelo.

O modelo foi treinado por 500 épocas utilizando o conjunto de dados de treinamento; durante o processo, o conjunto de treinamento foi fornecido ao modelo em lotes (*batches*) para otimizar o processo de aprendizagem e utilizar eficientemente os recursos computacionais; além disso, utilizou-se um *ScoreIterationListener* para

acompanhar o progresso do treinamento e a convergência do modelo, permitindo ajustes se necessário.

O balanceamento das classes no conjunto de treinamento foi crucial para evitar que o modelo fosse tendencioso em favor da classe majoritária, melhorando a equidade na classificação; a técnica de média dos vetores de palavras permitiu representar sentenças de comprimento variável em um espaço de dimensão fixa, facilitando a entrada na rede neural e reduzindo a complexidade computacional; e o truncamento dos textos aos últimos 500 caracteres se baseou na hipótese de que o resultado das sentenças geralmente está presente no final do documento, permitindo focar nas informações mais relevantes para a classificação.

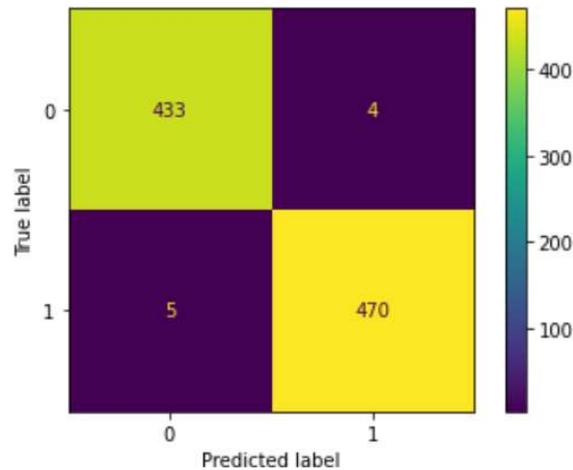
Após o treinamento, o modelo foi avaliado utilizando o conjunto de teste para verificar sua eficácia; para isso, foram utilizadas métricas de avaliação como a matriz de confusão (Figura 6), que visualiza a performance do modelo em termos de verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos, além de acurácia, precisão e *recall* [8], que são cálculos estatísticos para avaliar o desempenho global do modelo; o processo de avaliação consistiu em o modelo realizar previsões sobre o conjunto de teste e compará-las com os rótulos reais, permitindo calcular as métricas de avaliação e medir a capacidade do modelo em classificar corretamente novas sentenças.

A análise da matriz de confusão mostra o desempenho do modelo em termos de classificação, com 470 verdadeiros positivos e 433 verdadeiros negativos, indicando que o modelo foi capaz de prever corretamente a maioria das amostras em ambas as classes. O número de falsos positivos e falsos negativos é baixo, com apenas 4 e 5 casos, respectivamente.

Com base nesses valores, a acurácia do modelo foi de 99.01%, o que reflete uma alta proporção de previsões corretas em relação ao total de previsões realizadas. A precisão, que mede a proporção de previsões positivas corretas em relação ao total de previsões positivas feitas pelo modelo, foi de 99.16%, indicando que quase todas as previsões positivas foram verdadeiras. O *recall*, que avalia a capacidade do modelo de identificar corretamente todas as amostras positivas, foi de 98.95%, mostrando que o modelo conseguiu detectar quase todos os casos positivos. Esses resultados sugerem que o modelo tem um desempenho bom em termos de classificação, com alta capacidade de generalização e mínima ocorrência de erros de classificação. A

combinação de alta acurácia, precisão e recall reflete um modelo bem calibrado e confiável para a tarefa em questão, sendo capaz de classificar corretamente a maioria das amostras enquanto mantém a ocorrência de erros em um nível muito baixo.

Figura 6 – Matriz de confusão do Modelo (C)



Fonte: Autores.

6.2 MODELO (D)

Para o desenvolvimento do Modelo (D), cujo objetivo é prever a probabilidade de êxito da Procuradoria Geral do Estado (PGE) em novos processos judiciais, foram utilizadas técnicas de aprendizado de máquina com ênfase em redes neurais profundas. Uma característica fundamental deste modelo é que o *dataset* utilizado foi gerado com base nas inferências do Modelo (C), reutilizando os valores do neurônio que calcula a probabilidade de êxito da sentença, conforme detalhado a seguir:

A. Geração do *dataset*

O Modelo (C), responsável por classificar sentenças judiciais como favoráveis ou desfavoráveis ao Estado, produz uma probabilidade associada ao êxito da sentença por meio do neurônio de saída da rede neural. Os valores dessas probabilidades foram extraídos e incorporados ao novo *dataset* do Modelo (D) como uma variável adicional, enriquecendo o conjunto de dados com *insights* preditivos já estabelecidos.

O *dataset* original, criado antes da prova de conceito (cf. tópicos 3 e 4), contendo atributos processuais e funcionais dos processos judiciais, foi complementado com a probabilidade de êxito fornecida pelo Modelo (C). Essa integração permitiu que o Modelo (D) utilizasse tanto as informações estruturadas dos processos quanto as inferências obtidas da análise das sentenças, proporcionando uma visão mais abrangente para a predição.

B. Pré-processamento dos Dados

O conjunto de dados foi carregado a partir de um arquivo de extensão CSV, contendo as *features* originais e a probabilidade de êxito das sentenças. Os valores ausentes foram removidos para assegurar a qualidade dos dados. A normalização dos dados foi realizada utilizando o *MinMaxScaler* para escalonar as *features* e a variável alvo no intervalo de 0 a 1; esse procedimento é essencial para redes neurais, pois padroniza a escala dos dados, facilitando o processo de aprendizado e contribuindo para a eficiência e estabilidade do treinamento do modelo. A divisão do conjunto de dados foi realizada utilizando a função *train_test_split*, alocando 80% dos dados para o conjunto de treinamento e 20% para o conjunto de teste, com o parâmetro *random_state=42* para garantir a reprodutibilidade dos resultados; essa divisão é essencial para avaliar o desempenho do modelo em dados não vistos durante o treinamento, permitindo estimar sua capacidade de generalização e validar sua eficácia em cenários reais [9].

C. Construção do Modelo de Rede Neural

A construção do modelo de rede neural foi realizada utilizando a *API Sequential* do Keras para criar a rede de forma linear. A arquitetura da rede incluiu uma primeira camada oculta composta por uma camada *Dense* com 256 neurônios, função de ativação ReLU, regularização L2 [*kernel_regularizer=l2(0.0001)*] e *Dropout* de 20%. As camadas subsequentes consistiram em quatro camadas *Dense* com números decrescentes de neurônios (128, 128, 64, 32), todas utilizando a função de ativação ReLU, regularização L2 e *Dropout* de 10% após cada camada. A camada de saída foi uma *Dense* com 1 neurônio e função de ativação linear, adequada para

problemas de regressão. A probabilidade de êxito da sentença, fornecida pelo Modelo (C), foi incorporada como uma das *features* de entrada na rede neural, permitindo que o Modelo (D) se beneficiasse diretamente das análises anteriores, enriquecendo o aprendizado e potencializando a capacidade preditiva. Nas configurações de treinamento, o modelo foi compilado utilizando o otimizador Adam com uma taxa de aprendizado reduzida (*learning_rate*=0.0001) para melhorar a convergência e estabilidade; a função de perda escolhida foi a *Mean Absolute Error* (MAE), adequada para medir a diferença média entre as previsões e os valores reais, e também foi utilizada como métrica de avaliação durante o treinamento. O modelo foi treinado por 300 épocas com um *batch_size* de 32, utilizando um *validation_split* de 20% para monitorar o desempenho em um subconjunto de validação durante o treinamento [10].

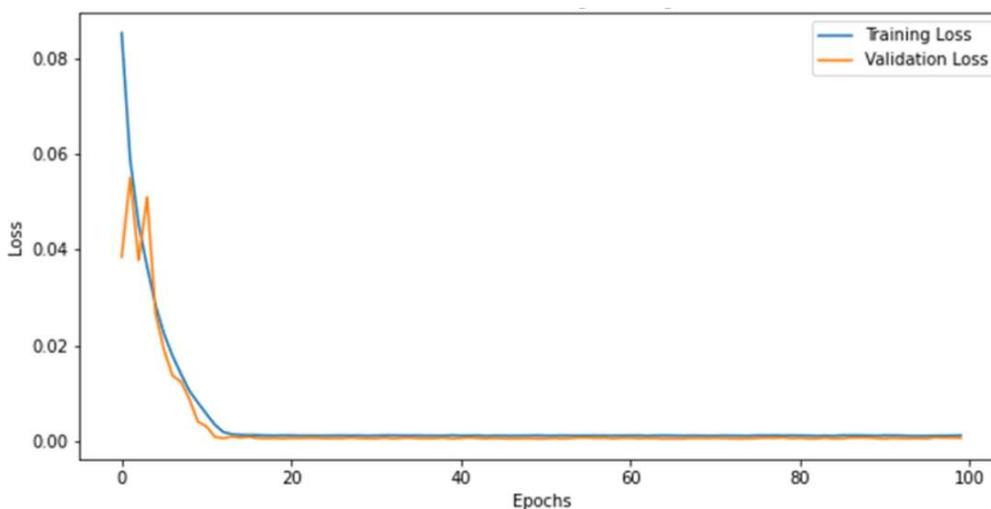
Após o treinamento, o modelo realizou previsões sobre o conjunto de teste. Para quantificar o desempenho do modelo, foram calculadas as métricas *Mean Squared Error* (MSE), *Mean Absolute Error* (MAE) e *R² Score* [8], as quais forneceram *insights* sobre a precisão e a capacidade de generalização das previsões. Além disso, visualizações dos resultados foram geradas, incluindo a curva de perda que mostra a evolução da perda durante o treinamento nos conjuntos de treino e validação, um gráfico comparativo entre os valores reais e previstos indicando a proximidade das previsões em relação aos valores observados.

Os valores previstos e reais foram desnormalizados para a escala original, facilitando a interpretação dos resultados no contexto real; a inclusão da probabilidade de êxito da sentença, proveniente do Modelo (C), como *feature* no Modelo (D) mostrou-se significativa, observando-se uma melhoria no desempenho preditivo, o que indica que as inferências anteriores adicionaram valor ao modelo de regressão; a construção do Modelo (D) foi significativamente enriquecida pela integração das inferências do Modelo (C), pois ao reutilizar os valores do neurônio que calcula a probabilidade de êxito da sentença, o Modelo (D) pôde incorporar conhecimento prévio diretamente em seu processo de aprendizado; essa abordagem permitiu capturar de forma mais efetiva as nuances dos dados judiciais e melhorou a capacidade preditiva em relação ao êxito do Estado da Bahia em novas ações judiciais; a utilização combinada dos dois modelos demonstra a eficácia de *pipelines* de aprendizado, onde modelos distintos contribuem sequencialmente para um objetivo comum, e, no contexto deste projeto, essa metodologia potencializa a

estratégia jurídica da PGE, oferecendo previsões mais precisas e informadas para a tomada de decisões.

A análise da curva de perda apresentada na Figura 7 revela o comportamento do Modelo (D) durante o processo de treinamento, com a visualização das curvas de perda para os conjuntos de treinamento e validação ao longo de 100 épocas. Observa-se uma diminuição acentuada da perda nas primeiras 20 épocas, indicando que o modelo rapidamente aprendeu a capturar os padrões nos dados. Após esse ponto, as curvas de perda tanto para o conjunto de treinamento quanto para o conjunto de validação se estabilizam em um patamar próximo de zero.

Figura 7 - Perda do Modelo (D) durante o 1º treinamento



Fonte: Autores.

O comportamento semelhante entre as curvas de treinamento e validação ao longo das épocas é um indicativo positivo de que o modelo não sofreu de *overfitting* ou *underfitting*. A proximidade entre as perdas de treinamento e de validação sugere que o modelo conseguiu generalizar bem para os dados de validação, mantendo um desempenho consistente e evitando a adaptação excessiva aos dados de treinamento. Outro ponto a destacar é que, após a estabilização por volta da 20ª época, a diferença entre as perdas de treinamento e de validação é mínima, o que reflete uma boa capacidade de generalização do modelo. Isso demonstra que o ajuste dos hiperparâmetros, como a taxa de aprendizado e as estratégias de regularização empregadas (L2 e *Dropout*), contribuíram para um treinamento eficiente e estável.

Embora as curvas de perda durante o treinamento e a validação mostrem um comportamento satisfatório, com rápida convergência e estabilização das perdas, a análise dos indicadores de desempenho (Tabela 1) aponta para desafios na precisão preditiva do Modelo (D). O *R-squared* (R^2) médio negativo de -0.0316 indica que o modelo não conseguiu explicar a variabilidade da variável alvo em relação aos valores reais e que, na prática, ele performou pior do que um modelo base que simplesmente preveja a média dos valores. Essa métrica sugere que o modelo pode não estar capturando bem as relações entre as variáveis de entrada e a saída, apesar da boa curva de perda observada durante o treinamento. O *Mean Squared Error* (MSE) de 0.000408 é relativamente baixo, o que normalmente indicaria que as diferenças entre os valores previstos e os reais são pequenas. No entanto, combinado com o R^2 negativo, isso pode significar que, apesar de as previsões serem próximas entre si, elas não estão suficientemente alinhadas com os valores reais, resultando em um baixo desempenho preditivo em termos de variabilidade explicada. O *Mean Absolute Error* (MAE) de 0.000961 também sugere que a magnitude média dos erros entre as previsões e os valores reais é baixa, o que indicaria um bom ajuste em termos absolutos. No entanto, assim como o MSE, o MAE deve ser considerado em conjunto com o R^2 para uma compreensão mais completa do desempenho. O baixo MAE pode indicar que o modelo faz previsões com desvios pequenos, mas a falta de correlação com a variável alvo reflete um problema de aprendizado subjacente.

Tabela 1 - Métricas de desempenho do Modelo (D) no 1º treinamento

R-squared	MSE	MAE
-0.03162212328057934	0.0004082173656040227	0.0009610776635329797

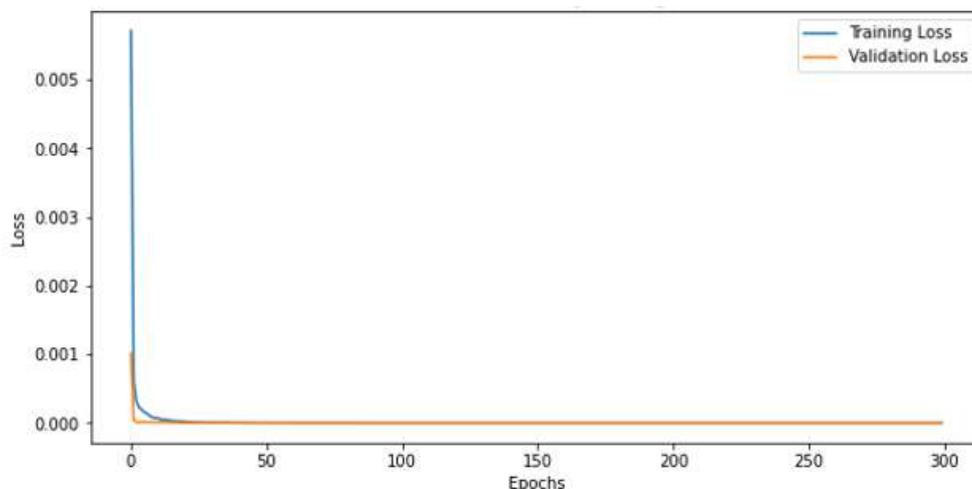
Fonte: Autores.

Em resumo, a análise das curvas de perda e das métricas de desempenho sugere que, embora o Modelo (D) tenha apresentado uma boa curva de treinamento e validação, ele pode estar limitado em capturar de maneira efetiva as relações complexas entre as variáveis de entrada e a variável alvo, resultando em um baixo R^2 . Isso aponta para a necessidade de reavaliar aspectos como a seleção de *features*, complexidade do modelo ou mesmo a qualidade dos dados de entrada para aprimorar o desempenho preditivo.

Para melhorar o desempenho preditivo do Modelo (D) e abordar as limitações identificadas na análise anterior, foram realizados ajustes significativos na arquitetura do modelo e no pré-processamento dos dados. Em vez de uma rede neural densa convencional, optou-se por uma abordagem que incorpora camadas LSTM (Long Short-Term Memory) para capturar dinâmicas temporais e sequenciais nos dados, simulando uma abordagem de Liquid Neural Network. Essa mudança permitiu ao modelo lidar melhor com possíveis dependências temporais ou relacionamentos complexos nas *features*. Adicionalmente, a regularização foi intensificada com camadas Dropout de 30% para evitar *overfitting*.

O pré-processamento também foi aprimorado com o uso do StandardScaler para escalonar as *features* e a variável alvo, garantindo que a distribuição dos dados estivesse mais adequada para o treinamento de modelos complexos. A configuração do treinamento foi ajustada com a introdução da técnica de EarlyStopping, monitorando a perda de validação e interrompendo o treinamento quando a performance não melhorasse por 10 épocas consecutivas, além de restaurar os melhores pesos obtidos. O modelo foi compilado com o otimizador Adam e uma taxa de aprendizado reduzida (0.0001) para assegurar uma convergência mais estável e precisa. Esses ajustes visaram melhorar a capacidade do modelo de capturar relações complexas nos dados e generalizar melhor em novos conjuntos de teste. A Figura 8 apresenta o resultado do modelo durante o 2º treinamento.

Figura 8 - Perda do Modelo (D) durante o 2º treinamento



Fonte: Autores.

A análise da curva de perda e das métricas de desempenho (Tabela 2) para a nova arquitetura do Modelo (D) evidencia alguns pontos críticos. Na Figura 8, observa-se que tanto a perda de treinamento quanto a de validação convergem rapidamente e permanecem próximas de zero ao longo das 300 épocas, sugerindo um treinamento estável e um modelo aparentemente bem ajustado. O comportamento da curva sem flutuações ou divergências entre as perdas de treinamento e validação indica que a regularização com *Dropout* e a técnica de *EarlyStopping* foram eficazes em prevenir *overfitting*, permitindo que o modelo generalizasse bem durante o treinamento.

Tabela 2 - Métricas de desempenho do Modelo (D) no 2º treinamento

R-squared	MSE	MAE
0.000000e+00	0.0024326607062739062	0.008268834548566457

Fonte: Autores.

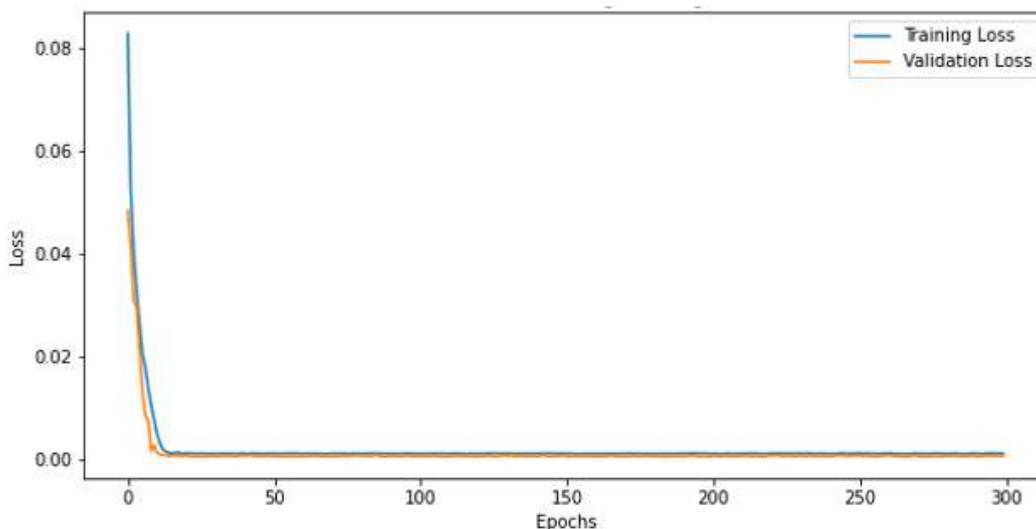
No entanto, ao correlacionar essa observação com as métricas de desempenho, percebe-se um paradoxo. O *R-squared* (R^2) é zero, o que indica que o modelo não conseguiu explicar a variabilidade dos dados de teste em relação aos valores reais, revelando um desempenho preditivo limitado. Esse resultado sugere que, embora a perda durante o treinamento tenha sido mínima, o modelo pode ter aprendido padrões que não se traduzem em uma capacidade robusta de generalização para novos dados. O *Mean Squared Error* (MSE) de 0.0024 e o *Mean Absolute Error* (MAE) de 0.0083 são relativamente baixos, indicando pequenas diferenças médias absolutas e quadráticas entre as previsões e os valores reais, mas, ao serem considerados juntamente com o R^2 , eles sugerem que o modelo está fazendo previsões que não estão alinhadas com a variabilidade esperada da variável alvo.

Para aprofundar a análise e entender melhor o desempenho do Modelo (D), adotou-se a estratégia de selecionar apenas dois assuntos com maior amostragem de dados do conjunto original. Essa abordagem visou reduzir a complexidade e a variabilidade dos dados, permitindo que o modelo se concentrasse em padrões mais consistentes e bem representados. A escolha desses dois assuntos buscou otimizar o processo de aprendizado, possibilitando uma avaliação mais precisa da capacidade

do modelo de fazer previsões eficazes quando exposto a um conjunto de dados menos heterogêneo. Essa estratégia também ajudou a identificar se a redução da variabilidade influenciaria significativamente a precisão preditiva e a estabilidade do modelo, proporcionando *insights* sobre sua aplicabilidade em contextos mais específicos.

A análise da curva de perda apresentada na Figura 9, juntamente com as métricas de desempenho na Tabela 3, revela um resultado positivo para o Modelo (D) após a aplicação do filtro nos dados, considerando apenas dois assuntos do conjunto de dados. Observa-se na imagem que tanto a perda de treinamento quanto a de validação convergem rapidamente e se mantêm próximas de zero ao longo de 300 épocas, indicando que o modelo conseguiu aprender os padrões dos dados de forma eficiente e estabilizou-se sem sinais de *overfitting*.

Figura 9 - Perda do Modelo (D) durante o 3º treinamento



Fonte: Autores.

O desempenho do modelo é corroborado pelas métricas de avaliação, que apresentam resultados excelentes. O *R-squared* (R^2) de 1.0 indica que o modelo conseguiu explicar perfeitamente a variabilidade da variável alvo, sugerindo um ajuste completo aos dados de teste e um alto nível de precisão preditiva. O *Mean Squared Error* (MSE) de 0.000181 e o *Mean Absolute Error* (MAE) de 0.000559 reforçam essa avaliação, mostrando que a diferença média entre as previsões e os valores reais é extremamente baixa. Esses resultados indicam que a decisão de filtrar os dados para

incluir apenas dois assuntos específicos reduziu a variabilidade e complexidade dos dados de entrada, o que permitiu ao modelo aprender de forma mais precisa.

Tabela 3 - Métricas de desempenho do Modelo (D) no 3º treinamento

R-squared	MSE	MAE
1.0	0.00018149667364891406	0.0005591318234788372

Fonte: Autores.

No entanto, é importante considerar que o ajuste perfeito do R^2 pode ser um indicativo de que o modelo está ajustado especificamente para este subconjunto dos dados. Isso pode limitar a generalização do modelo para outros tipos de assuntos ou conjuntos de dados com maior variabilidade.

Outra consideração necessária, diz respeito a avaliação, na amostra, da possibilidade da presença de dados que possam induzir o modelo a repetir padrões que, ainda que refletindo a realidade passada, não são considerados desejáveis do ponto de vista ético e moral. Ainda que a avaliação da presença desses vieses seja indispensável quando da operação efetiva em produção da solução desenvolvida, entendeu-se que, por se tratar de uma prova de conceito, mais relevante seria preservar o tamanho da amostra, visto que os resultados seriam avaliados tão somente do ponto de vista da técnica.

Impõe, contudo, deixar desde logo registrado que na hipótese de se avançar para o desenvolvimento de modelo pleno, para uso pela Procuradoria Geral do Estado, a questão da avaliação de vieses precisará ser contemplada no *pipeline* de construção da solução.

Em resumo, embora os resultados atuais sejam impressionantes, estudos adicionais com diferentes subconjuntos de dados e avaliações de robustez são recomendados para confirmar a capacidade do modelo de manter essa performance em cenários mais amplos e variados.

7. RESULTADOS

A prova de conceito realizada apresentou resultados positivamente surpreendentes, demonstrando a viabilidade prática das ideias planejadas.

A estratégia de criação dos dois Modelos (C) e (D), de maneira que os dados de classificação obtidos pelo Modelo (C) viessem a ser incorporados como mais um atributo no *dataset* de treinamento do Modelo (D), mostrou-se acertada. Isto resultou em ótimos resultados na performance preditiva do Modelo (D), que opera apenas com dados tratados e pré-processados, de maneira enxuta, sem necessidade de processamento de linguagem natural, o que reduz abissalmente os recursos computacionais necessários para rodar o modelo.

Igualmente se mostrou vantajosa a estratégia de segmentação dos dados. Os treinos do Modelo (D) para incluir apenas dois assuntos específicos reduziu a variabilidade e complexidade dos dados de entrada, o que permitiu ao modelo aprender de forma mais precisa, o que deve ser replicado.

Conclui-se que a Procuradoria Geral do Estado da Bahia possui um acervo de centenas de milhares de dados, históricos e em produção e captura diários, que significam verdadeiros atributos processuais e funcionais dos processos judiciais, que, complementado com a probabilidade de êxito fornecida pelas análises de sentimento de sentenças - Modelo (C), proporcionam uma qualificada predição do êxito do Estado em Juízo através de técnicas de aprendizado de máquina e de inteligência artificial - Modelo (D). CQD.

8. REFERÊNCIAS

1. Disponível em: <<https://justica-em-numeros.cnj.jus.br/painel-litigantes/>>. Acesso em: 28 Nov. 2024.
2. Disponível em: <<https://justica-em-numeros.cnj.jus.br/painel-despesas/>>. Acesso em: 28 Nov. 2024.
3. Ivano Lauriola, Alberto Lavelli, Fabio Aiolli (2022). *An introduction to Deep Learning in Natural Language Processing: Models, techniques, and tools*. ISSN 0925-2312, Neurocomputing.
4. Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras and TensorFlow: concepts, tools, and techniques to build intelligent systems* (2nd ed.). O'Reilly.

5. Hauschild, J., & Eskridge, K. (2024). *Word embedding and classification methods and their effects on fake news detection. Machine Learning with Applications, 17*, 100566. doi:10.1016/j.mlwa.2024.100566.
6. Konduit. (n.d.). *Deeplearning4j: Open-source, distributed deep learning for the JVM*. <https://deeplearning4j.org>.
7. Heidari, E., Sobati, M. A., & Movahedirad, S. (2016). Accurate prediction of nanofluid viscosity using a multilayer perceptron artificial neural network (MLP-ANN). *Chemometrics and Intelligent Laboratory Systems, 155*, 73–85. doi:10.1016/j.chemolab.2016.03.031.
8. Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*. 2nd ed. New York, Springer.
9. H. W. Herwanto, A. N. Handayani, A. P. Wibawa, K. L. Chandrika and K. Arai, "Comparison of Min-Max, Z-Score and Decimal Scaling Normalization for Zoning Feature Extraction on Javanese Character Recognition," *2021 7th International Conference on Electrical, Electronics and Information Engineering (ICEEIE)*, Malang, Indonesia, 2021, pp. 1-3, doi: 10.1109/ICEEIE52663.2021.9616665.
10. J. Si, S. L. Harris and E. Yfantis, "A Dynamic ReLU on Neural Network," *2018 IEEE 13th Dallas Circuits and Systems Conference (DCAS)*, Dallas, TX, USA, 2018, pp. 1-6, doi: 10.1109/DCAS.2018.8620116.