

APEB

Associação dos Procuradores
do Estado da Bahia

Ano V - 2025

Revista

APEB

Associação do Procuradores do Estado da Bahia

Prêmio
Joselita Leão

TRABALHOS VENCEDORES

A Revista APEB - é uma publicação oficial da Associação dos Procuradores do Estado da Bahia de periodicidade bianual. Objetiva ser um espaço de disseminação da produção intelectual e acadêmica, com a publicação de artigos, relatos de experiência, textos técnicos e resenhas de temáticas jurídicas de interesse da Associação, dos Procuradores e da comunidade forense.

EQUIPE:

Equipe Editorial

Secretaria APEB
Cannal de Ideias em Comunicação
Gato de AP/Arnoldo Miranda

DIRETORIA

Presidente:

Téssio Rauff de Carvalho Moura

Vice-Presidente:

Alexandre de Souza Araújo

Secretário Geral:

Mariana Cardoso Vaz Santos

Diretor de Inativos e Pensionistas:

Claudio Santos Silva (Claudão)

Diretor Jurídico e de Prerrogativas:

Luís Eduardo Rolin Carneiro de Oliveira

Diretor Social e de Divulgação:

Paloma Teixeira Rey

Diretor Financeiro:

Thales Francisco Amaral Cabral

Representante da APEB no Conselho

Superior:

Eduardo Santos Sales Mesa

COMISSÃO JULGADORA

Presidente:

Marco Aurelio de Castro Junior

Patrícia Werner

Renata Fabiana Santos Silva

Roberto Limas Figueiredo

Thiago Xavier

Viviane Chetto

MESA ASSEMBLEIA GERAL:

1º Efetivo:

André Luiz Peixoto Fernandes

2º Efetivo:

Arley Cavalcante de Oliveira

3º Efetivo:

Antonio César de Carvalho Magaldi

SUPLENTES

Tiago Araujo Costa

Filipe Xavier Ribeiro

CONSELHO FISCAL

1ª Efetivo:

Leonardo Sérgio Pontes Gaudenzi

2º Efetivo:

Clara Meira Costa Sampaio

3º Efetivo:

Vanesca Lopes de Araújo Politano

SUPLENTES

Mariana Machado Nóbrega

Ana Carla Pires Meira Cardoso

CONSELHO DE ÉTICA:

1ª Efetiva:

Ursula da Rocha Viegas

2ª Efetiva:

Rivanda da Costa Santos

3ª Efetiva:

João Luiz Rezende Lamêgo

SUPLENTES

Adriano Oliveira Pessoa

Tatiana Martins Oliveira

SUMÁRIO

INTELIGÊNCIA ARTIFICIAL PARA A PREDIÇÃO DO ÊXITO DO ESTADO DA BAHIA EM NOVAS AÇÕES JUDICIAIS..... 5

1. Introdução.....	6
2. Desafio e Contextualização	6
3. A Estratégia Pensada	10
4. A Coleta de Dados.....	13
5. Métodos a Serem Utilizados no Pré-processamento dos Dados.....	14
6. Execução do Treinamento dos Modelos de IA	16
6.1 Modelo (C).....	16
A) Representação De Texto Com Word Embeddings.....	17
A) Treinamento Do Modelo Word2vec:	17
B) Processo De Treinamento:.....	17
C) Utilização No Modelo De Classificação:	17
B) Pré-Processamento Dos Dados	18
6.2 Modelo (D).....	20
A. Geração Do Dataset.....	20
B. Pré-Processamento Dos Dados.....	20
C. Construção Do Modelo De Rede Neural.....	21
7. Conclusão	21
8. Referências.....	23

Autores: Ayrton Bittencourt Lobo Neto, Fernanda de Santana Villa,
José Carlos C. Wasconcellos Jr e Leonardo Sergio Pontes Gaudenzi.

PARECER JURÍDICO – ATUALIZA E MODERNIZA A POLÍTICA PÚBLICA DE EXPLORAÇÃO MINERAL DO ESTADO DA BAHIA	25
Empresa Estatal.....	26

Autor: Marco Valério Viana Freire.

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

Autores: Ayrton Bittencourt Lobo Neto,
Fernanda de Santana Villa, José Carlos
C. Wasconcellos Jr e Leonardo Sergio
Pontes Gaudenzi.

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

RESUMO

A partir do cenário verificado no levantamento e pré-processamento contencioso judicial da Procuradoria dos dados até a construção dos Geral do Estado da Bahia, o trabalho modelos de IA para viabilizar a apresenta o desenvolvimento da predição do resultado de processos estratégia, desde a concepção, judiciais desde a sua ciência pela PGE.

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, desde o recebimento da citação pela PGE e 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

assegurador 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, como 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

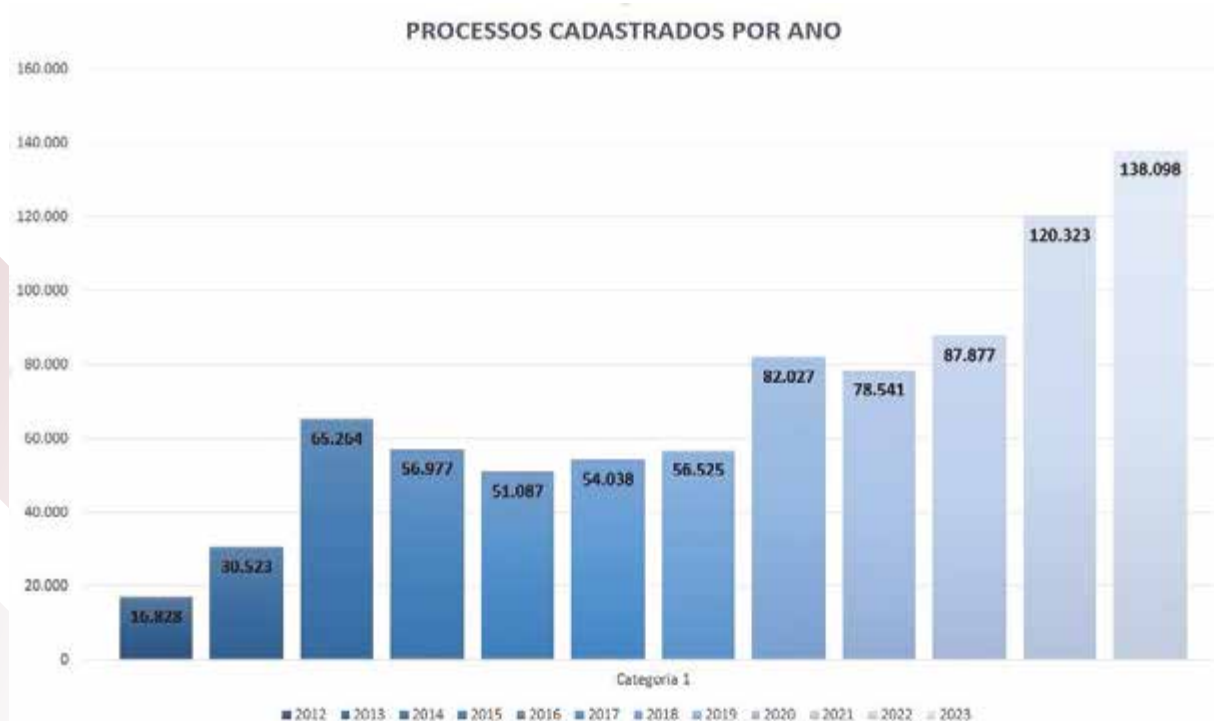
sob os cuidados dessa instituição ultrapassou a marca de 1.000.000 (um milhão) de processos em 2024.

Ao final do ano de 2024, quando havia cerca de 246 (duzentos e quarenta e seis) Procuradores em atividade, significaria 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:

Figura 1 - Número de processos novos cadastrados por ano pela PGE



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:

Figura 2 - Número de movimentações cadastradas por ano pela PGE



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 lhes 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 provoca as seguintes reflexões: diante do recebimento da citação em uma nova ação judicial, o oferecimento de defesa (contestação) seria 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 esta proposta se relaciona diretamente com as entregas da Encomenda Tecnológica¹ 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:

¹ Contrato nº. PGE 032/2022, celebrado em 01/07/2022.

Figura 3 – Atributos do Modelo (D)

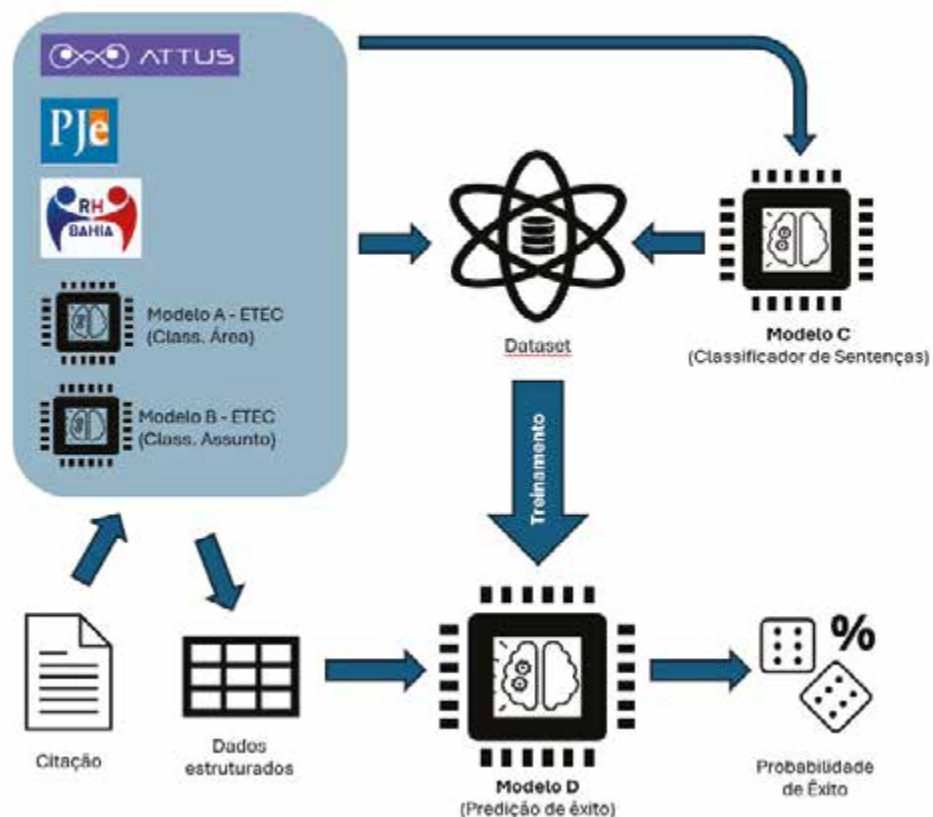
	Dado cadastral do processo 1	Dado cadastral do processo 2	Dado cadastral do processo 3	Dado cadastral do processo ...	Dados das partes 1	Dados das partes 2	Dados das partes 3	Dados das partes ...	Área	Assunto	Probabilidade de Êxito da PGE
Fonte	Attus/PJE	Attus/PJE	Attus/PJE	Attus/PJE	RHBahia	RHBahia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

Fonte: Autores.

A estratégia pensada, portanto, permitirá utilizar apenas dados estruturados para o consumo diário do modelo de predição, reduzindo o esforço computacional necessário. A entrada e a saída do modelo de predição (D):

A Figura 4, abaixo, demonstra os fluxos e os relacionamentos entre as

Figura 4 – ilustração dos fluxos e relacionamentos na estratégia pensada



Fonte: Autores.

Note-se 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 está disponível à Procuradoria Geral do Estado, propõe-se 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 primeiros dados devem ser coletados a partir do sistema *Attus*², que se integra ao sistema *PJE – Processo Judicial Eletrônico*³ por meio de *WebService* padronizado por *MNI – Modelo Nacional de Interoperabilidade*.

Os dados serão 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.

Diretamente do sistema *Attus*, é possível realizar a coleta dos seguintes dados e variáveis:

- Data de ajuizamento

² Sistema de gerenciamento e acompanhamento de processos judiciais desenvolvido pela empresa *Attus* e implantado na Procuradoria Geral do Estado, em janeiro de 2024, no qual estão cadastrados todos os processos em que a PGE atua.

³ Sistema de tramitação de processos judiciais desenvolvido pelo Conselho Nacional de Justiça e implantado pelo Tribunal de Justiça do Estado da Bahia desde 2014.

- 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)⁴.

Diante disso, recorrer-se-á ao banco de dados do sistema *RH Bahia*⁵ para coletar novos dados que possam 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, sendo possível coletar os seguintes dados do *RH Bahia*:

- Cargo público

⁴ Dado a ser utilizado apenas para o treinamento do modelo de classificação de sentenças (C).

⁵ 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.

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

5. MÉTODOS A SEREM UTILIZADOS NO PRÉ-PROCESSAMENTO DOS DADOS

Concluída a fase de coleta e agrupamento dos dados, avança-se para o seu pré-processamento, visando à execução do treinamento dos Modelos (C) e (D) [4].

A partir dos dados reunidos no item anterior, serão constituídos dois *datasets*, visando ao desenvolvimento dos Modelos (C) e (D). Tais dados, em maior ou menor grau, não se apresentam prontos para o processamento pelos modelos de inteligência artificial, pelo que se faz 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, pode ser montado inicialmente a partir dos dados extraídos do sistema Attus, conforme já indicado.

Três situações se apresentam norteadoras para a seleção dos dados: a) a saída desejada para esse modelo seria a avaliação se a sentença de um processo judicial é favorável ou não para o Estado; b) tal resultado deve ser apresentado numa grandeza de 0 (integralmente desfavorável) a 1 (integralmente favorável); c) o resultado do modelo deve ser obtido através de processamento da linguagem natural contida nas sentenças a serem avaliadas.

A partir dessas condicionantes, para a finalidade de treinamento do Modelo (C), serão necessários apenas os dados de N° Processo (CNJ), a Data do Andamento e o Tipo de Andamento. Dessa forma, processados esses ajustes e feita a referida extração das sentenças que seriam classificadas, estará pronto o primeiro conjunto de dados, que irá 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), deverá ser montado a partir de duas planilhas: uma extraída do sistema Attus, contendo os dados estruturados eminentemente relacionados ao processo judicial, e outra extraída do sistema RH Bahia, com os dados funcionais dos servidores do Poder

Executivo estadual, autores de expressiva parcela dessas ações.

O primeiro processamento busca trazer os dados da segunda planilha para a primeira. Para tanto, identifica-se o CPF como o dado que serve como chave em ambas as planilhas, visando proceder à unificação.

A partir dessa planilha unificada, passa-se à fase de pré-processamento dos dados, utilizando-se o Microsoft Power Query. Nessa etapa – que contempla desde operações mais simples, como remoção de colunas, até situações mais complexas, como divisão de colunas a partir de *strings* pré-determinados –, será possível corrigir inconsistências de formatação contidas atualmente nos dados.

Uma vez concluída essa etapa de pré-processamento, terá sido gerada uma nova planilha de Excel, agora com os dados prontos para serem submetidos ao Modelo (D). Para tal finalidade, apenas se acrescentaria 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* separado para treinamento do modelo, pelo resultado da predição para cada um desses processos realizados pelo Modelo (C).

Dessa forma, alcançar-se-á o ponto para a execução da Prova de Conceito (PoC), cujo roteiro ideal consiste 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.

Os dados a serem obtidos são de acesso assegurado à PGE em função de sua atuação institucional. 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, proceder-se-á à 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 seriam 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, devem ser alvo da técnica de anonimização.

6. EXECUÇÃO DO TREINAMENTO DOS MODELOS DE IA

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 será 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

Essa abordagem permite 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 serão 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, utilizar-se-á 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, apenas palavras com uma frequência mínima de 5 (cinco) ocorrências devem ser incluídas no vocabulário para filtrar termos muito raros que podem introduzir ruído; os vetores de *embeddings* devem ser configurados com 500 dimensões para capturar relações semânticas complexas entre as palavras; deve-

se utilizar uma janela de contexto de tamanho 5, significando que o modelo considerará as 5 palavras anteriores e posteriores a uma palavra-alvo durante o treinamento para capturar relações de proximidade; o modelo será treinado por 10 iterações (épocas) para assegurar a convergência e estabilidade das representações vetoriais, e uma semente fixa (42) a ser utilizada para garantir a reprodutibilidade dos resultados.

b) Processo de Treinamento:

O treinamento deve ser conduzido utilizando a biblioteca *DeepLearning4j* (DL4J) [6], especificamente a classe *Word2Vec*. O modelo iterará 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 pretende resultar 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 serão utilizados para converter as sentenças em vetores de características. Isso será

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 servirão como entrada para a rede neural do modelo de classificação, permitindo que a rede aprenda 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 consiste em textos de sentenças judiciais e seus

respectivos rótulos 1 (favorável) ou 0 (desfavorável); o pré-processamento envolve os seguintes passos: as sentenças serão *tokenizadas*, dividindo-as em *tokens* (palavras) utilizando espaços em branco como delimitadores; em seguida, o texto deverá ser tratado e limpo, com a remoção de caracteres especiais e normalização para padronização; para otimizar o processamento e focar nas partes mais relevantes, textos com mais de 500 caracteres serão 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 será 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 deve ser balanceado para conter quantidades iguais de sentenças favoráveis e desfavoráveis ao Estado. Os dados devem ser embaralhados aleatoriamente e divididos em conjuntos de treinamento (70%) e teste (30%); o conjunto de treinamento será utilizado para ajustar os pesos do modelo durante o pro-

cesso de aprendizagem, enquanto o conjunto de teste servirá 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

Será desenvolvida uma Rede Neural Artificial do tipo *Multilayer Perceptron* (MLP) com múltiplas camadas densas, utilizando a biblioteca DL4J.

O modelo deve ser treinado por 500 (quinhentas) épocas, utilizando o conjunto de dados de treinamento; durante o processo, o conjunto de treinamento será fornecido ao modelo em lotes (*batches*) para otimizar o processo de aprendizagem e utilizar eficientemente os recursos computacionais; além disso, será necessário utilizar 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 é crucial para evitar que o modelo seja tendencioso em favor da classe majoritária, melhorando a equidade na classificação; a técnica de média dos vetores de palavras

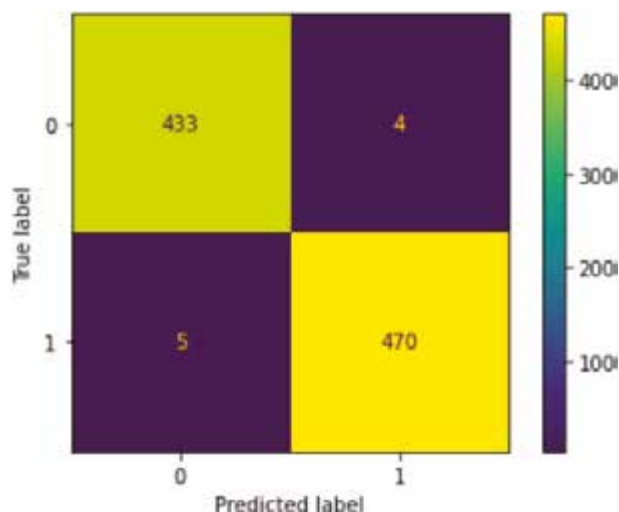
permitirá representar sentenças de comprimento variável em um espaço de dimensão fixa, facilitando a entrada na rede neural e reduzindo a complexidade computacional; e o truncamento dos textos aos últimos 500 (quinhentos) caracteres se baseia na hipótese de que o resultado das sentenças geralmente está presente no final do documento, permitindo focar nas informações mais relevantes para a classificação.

Após o treinamento, o modelo será avaliado utilizando o conjunto de teste para verificar sua eficácia; para isso, serão utilizadas métricas de avaliação como a matriz de confusão (Figura 5), 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 consiste 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.

Figura 5 – 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, devem ser utilizadas técnicas de aprendizado de máquina com ênfase em redes neurais profundas. Uma característica fundamental deste modelo é que o *dataset* a ser utilizado será 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 associa-

da ao êxito da sentença por meio do neurônio de saída da rede neural. Os valores dessas probabilidades serão 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 (cf. tópicos 3 e 4), contendo atributos processuais e funcionais dos processos judiciais e seus Autores, será complementado com a probabilidade de êxito fornecida pelo Modelo (C). Essa integração permite que o Modelo (D) utilize 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 será carregado a partir de um arquivo de extensão CSV, contendo as *features* originais e a probabilidade de êxito das sentenças. A normalização dos dados deve ser 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

do para a eficiência e estabilidade do treinamento do modelo. A divisão do conjunto de dados a ser 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 será realizada utilizando a *API Sequential* do Keras para criar a rede de forma linear. A arquitetura da rede incluirá 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 consistirão 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 su-

gerida é 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), a ser incorporada como uma das *features* de entrada na rede neural, permitindo que o Modelo (D) se beneficie diretamente das análises anteriores, enriquecendo o aprendizado e potencializando a capacidade preditiva.

Após o treinamento, o modelo realizará previsões sobre o conjunto de teste. Para quantificar o desempenho do modelo, devem ser calculadas as métricas *Mean Squared Error* (MSE), *Mean Absolute Error* (MAE) e *R² Score* [8], as quais fornecerão *insights* sobre a precisão e a capacidade de generalização das previsões. Além disso, visualizações dos resultados devem ser 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.

7. CONCLUSÃO

Desenvolver modelos de inteligência artificial que, concatenados, sejam capazes de inferir, com margem

de erro conhecida e razoável, a probabilidade de êxito do Estado em uma ação ajuizada contra ele é medida viável do ponto de vista tecnológico e importante à gestão eficiente dos processos judiciais.

A estratégia apresentada de criação dos dois Modelos (C) e (D), de maneira que os dados de classificação obtidos pelo Modelo (C) venham a ser incorporados como mais um atributo no *dataset* de treinamento do Modelo (D), mostra-se adequada ao alcance dos melhores resultados na performance preditiva do Modelo (D), que operará 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 vantajosa é a estratégia sugerida de segmentação dos dados. Os treinos do Modelo (D) para incluir apenas dois assuntos específicos por treino reduzem a variabilidade e complexidade dos dados de entrada, o que permitirá ao modelo aprender de forma mais precisa.

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 significa verdadeiro tesouro de dados, pois carregam valiosos atributos processuais e funcionais dos processos judiciais e seus autores, que, complementado com a probabilidade de êxito fornecida pelas análises de sentimento de sentenças - Modelo (C), proporcionará uma qualificada predição do êxito do Estado em Juízo através de técnicas de aprendizado de máquina e de inteligência artificial - Modelo (D), permitindo uma melhor gestão do contencioso estadual.

Há que se registrar, finalmente, 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.

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.

PARECER JURÍDICO –
ATUALIZA E MODERNIZA A
POLÍTICA PÚBLICA
DE EXPLORAÇÃO MINERAL
DO ESTADO DA BAHIA

Autor: Marco Valério Viana Freire.

PARECER JURÍDICO –
ATUALIZA E MODERNIZA A POLÍTICA PÚBLICA
DE EXPLORAÇÃO MINERAL DO ESTADO DA BAHIA



EMPRESA ESTATAL.

Autorização legal para criação de subsidiária. Desnecessidade. Competências pretendidas para a subsidiária proposta que já se acham atribuídas à Empresa Estatal Consulente mediante fórmula normativa de grande amplitude semântica. Ajustes internos que se mostram suficientes para que a Companhia requerente opere segundo todos os moldes pretendidos. Lei alteradora do diploma autorizativo atual. Possibilidade. Minuta de projeto de lei alteradora da atual, da lavra da [REDACTED], que já segue compondo a instrução do expediente examinado. Exame das minutas de instrumentos relativos à empresa subsidiária proposta.

Inviabilidade momentânea. Ausência de evidências objetivas, de natureza contábil, relativas à viabilidade econômica da estrutura proposta. Necessidade de reapresentação do tema às D.D. Autoridades do Centro do Governo, para que, sob a orientação do Exm.º Sr. Governador do Estado, seja firmada diretriz objetiva acerca de prosseguir a Consulente operando com base na lei atual,

com base em lei alteradora da atualmente vigente ou com base em lei criadora da empresa subsidiária pretendida. Duas primeiras alternativas que, firmadas pelas esferas decisórias governamentais competentes, devem ensejar providências imediatas na Estatal consulente. Última alternativa que, uma vez definida, deve ensejar o retorno do expediente respectivo a esta [REDACTED] para o exame das minutas relativas à empresa subsidiária apresentada.

1. Os autos do expediente em epígrafe emergem da CBPM - Companhia Baiana de Pesquisa Mineral demandando opinativo sobre a viabilidade jurídica de criação de uma empresa subsidiária voltada à exploração econômico-financeira de ativos de origem mineral, especialmente aqueles para os quais a Entidade já *“detém autorizações de pesquisa para cerca de 700 (setecentas) áreas com potencial de aproveitamento econômico mineral no estado da Bahia”* (ID 00090420519), tendo a Consulente instruído o feito com as minutas de estatuto, acordo de acionistas, projeto de lei (PL), justificativa do PL, convite/convocação empresarial, termo de

concordância e compromisso, bem como com as razões de apresentação da pretensão criadora da subsidiária aludida ao r. Conselho de Administração da Companhia (IDs 00090424130, 00090424305, 00090424406, 00090424509, 00090424746, 00090424867 e 00090425031).

2. Com isso posto, avanço imediatamente à constatação no sentido de que a Lei Estadual nº 3.093, de 18.12.1972, ao autorizar a criação da CBPM, a esta atribuiu, em seu art. 1º, competência para desempenhar “*qualquer forma de aproveitamento econômico de minérios*”, consagrando tal dicção normativa um grau de amplitude semântica que comporta e autoriza o desempenho de todas as atividades operacionais ou financeiras pelas quais a Entidade possa, de forma direta ou indireta, explorar economicamente ativos resultantes de atividades ou projetos da cadeia produtiva da mineração.

3. Se já não fosse apenas por isso, é também certo que a Lei das Empresas Estatais (LEE) - Lei Federal nº 13.303, de 30.06.2016 -, em seu art. 28, incisos I, II e § 4º, exponencializa a amplitude semântica da fórmula normativa

“*qualquer forma de aproveitamento econômico de minérios*” vertida no suso referido art. 1º da Lei Estadual nº 3.093, de 18.12.1972, ao estabelecer que as empresas estatais estão inclusive dispensas de realizar procedimento licitatório **(a)** para a comercialização, prestação ou execução, de forma direta, de produtos, serviços ou obras especificamente relacionados com seus respectivos objetos sociais, bem como **(b)** para os casos em que a escolha do parceiro esteja associada a suas características particulares, vinculada a oportunidades de negócio definidas e específicas, justificada a inviabilidade de procedimento competitivo, definindo-se ali como oportunidades de negócio a formação e a extinção de parcerias e outras formas associativas, societárias ou contratuais, a aquisição e a alienação de participação em sociedades e outras formas associativas, societárias ou contratuais e as operações realizadas no âmbito do mercado de capitais, respeitada a regulação pelas esferas públicas competentes.

4. Assim, à luz de uma interpretação sistemática e integrada dos arts. 1º da Lei Estadual nº 3.093, de 18.12.1972, e 28, incisos I, II e § 4º, da LEE, assumo que a CBPM já se encontra autorizada a desempenhar *toda e qualquer*

competência que pudesse vir a ser atribuída à empresa subsidiária cuja viabilidade jurídica para criação ora se examina, sendo certo que, nessa perspectiva, é bastante que a Companhia (**a**) promova as adaptações estatutárias e normativas internas na direção de ampliar seu leque de atuação operacional e financeira, (**b**) qualifique os dirigentes e os empregados que irão lidar diretamente com as atividades de sua nova carta de serviços, (**c**) defina, em bases estratégicas, as linhas em que prioritariamente atuará e os projetos que abraçará em razão das distintas capacidades necessárias para tocá-los, notadamente quanto ao volume de investimentos a ser mobilizado e, por fim, ainda que não exclusivamente, (**d**) dê início à prospecção e à captação de parceiros técnica, econômica e financeiramente qualificados para atuar nas distintas frentes de investimentos e de projetos definidas como prioritárias, com eles celebrando os negócios jurídicos formalizadores de cada parceria, sob a consultoria da Área Jurídica da Empresa e, quando for o caso, da [REDACTED]

5. Não sendo assim, se por razões associadas ao valor segurança jurídica, p.ex., persistir como decisão

de Centro de Governo, o propósito de promover a reconfiguração da carta de serviços da CBPM mediante lei, assumo que, ao menos por hora, a lei a ser editada deve ter por objeto a alteração da suso referida Lei Estadual nº 3.093, de 18.12.1972, nos moldes [REDACTED] cados nas minutas anexas que seguem em aplicativos WORD, para permitir sugestões de ajustes, e PDF, para estabilizar a última formatação atribuída ao PL da lavra deste (IDs 00101366447 e 00101366582).

6. Considerando que as decisões de Centro de Governo relativas a (**a**) prosseguir a CBPM com a ampliação desua cartadeserviçoscom base na lei já existente e demais ajustes internos sugeridos acima **ou** a (**b**) viabilizar a CBPM a ampliação de sua carta de serviços com base na edição de lei alteradora do diploma já em vigor constituem premissas indispensáveis à adoção de qualquer outra diretriz a ser firmada e a seus respectivos desdobramentos, descabe, por enquanto, na visão desta [REDACTED] [REDACTED], trabalhar sobre o conteúdo de todas as minutas referidas no **item 1.** supra deste opinativo, tanto mais porque, não estando evidenciada na instrução processual, em bases numéricas e contábeis objetivas, a viabilidade econômico-financeira

motivadora da criação de uma nova entidade pública com todas os impactos de finanças públicas disso decorrentes - pessoal, patrimônio, estrutura, integralização do capital inicial, estimativas associadas a ativos, passivos, patrimônio líquido, custos, despesas, fluxo de caixa e demais indicadores aptos a permitirem análises das estimativas consignadas nas demonstrações contábeis projetadas para a estrutura que se pretende criar -, seria *bastante prematuro*, da parte da [REDACTED] avançar na direção de cancelar a proposta de criação de uma nova empresa estatal.

7. Nada obstante, tal como assinalado acima, as pretensões que se almeja atribuir à empresa subsidiária indicada podem ser - *todas elas, sem qualquer exceção* - já desempenhadas pela CBPM mediante a ampliação de sua carta de serviços com base na lei já existente e demais ajustes sugeridos **ou** mediante a ampliação do seu portfólio de atividades com base na edição de lei alteradora do diploma em vigor, revelando-se - repita-se - prematura qualquer chancela à criação de um novo componente empresarial para a Administração Estadual à míngua de dados objetivos relativos à viabilidade econômica do

empreendimento suscitado.

8. Em tais condições, **concluo por orientar o retorno do expediente à CBPM**, a fim de que sejam ali adotadas as seguintes **providências**:

1ª) reapresentação do tema às D.D. Autoridades do Centro do Governo, sob inafastável orientação do Exm.º Sr. Governador do Estado, a fim de que, em tal esfera decisória, seja firmada diretriz no sentido de **(a)** prosseguir a CBPM com a ampliação de sua carta de serviços com base na lei já existente e demais ajustes sugeridos acima **ou** a **(b)** viabilizar a CBPM a ampliação de sua carta de serviços com base na edição de lei alteradora do diploma já em vigor;

2ª) firmada diretriz governamental no sentido da **providência 1ª), letra (a) supra** - prosseguir a CBPM com a ampliação de sua carta de serviços com base na *lei já existente e demais ajustes sugeridos acima* -, cabe à Companhia proceder conforme orientado no **item 4., letras (a) a (d)**, deste opinativo,

3ª) firmada diretriz governamental no sentido da **providência 1ª), letra (b)**

supra - viabilizar a CBPM a ampliação de sua carta de serviços com base na *edição de lei alteradora do diploma já em vigor* -, cabe à Companhia proceder conforme orientado no **item 5.** *supra*, deste opinativo, enviando à Casa Civil a minuta de PL retratada nos instrumentos concernentes aos IDs 00101366447 e 00101366582, e procedendo, em seguida, aos ajustes internos necessários;

4ª) firmada diretriz governamental no sentido da criação de uma empresa subsidiária da CBPM, este expediente deverá retornar a esta PGE / PCT / NAAE, para exame das

minutas referidas [REDACTED] [REDACTED] *supra* deste opinativo.

9. É como me cabe orientar. que considerar pertinentes.

10. [REDACTED] para os anexos, destaques e indicações que considerar pertinentes.

Salvador / BA, 25 de outubro de 2024.