

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS  
CURSO DE CIÊNCIAS SOCIAIS

**Aplicação da Inteligência Artificial sobre dados socioeconômicos na previsão de  
Cenários Políticos: reeleição de prefeitos**

Dejalma Mateus Morandin

Porto Alegre  
2024

DEJALMA MATEUS MORANDIN

**Aplicação da Inteligência Artificial sobre dados socioeconômicos na previsão de  
Cenários Políticos: reeleição de prefeitos**

Trabalho de Conclusão de Curso apresentado como requisito parcial para a obtenção do título de Bacharel em Ciências Sociais, pelo Instituto de Filosofia e Ciências Humanas, da Universidade Federal do Rio Grande do Sul.

Orientador: Prof. Dr. Alfredo Alejandro Gugliano.

Porto Alegre  
2024

### CIP - Catalogação na Publicação

Morandin, Dejalma Mateus

Aplicação da Inteligência Artificial sobre dados socioeconômicos na previsão de Cenários Políticos: reeleição de prefeitos / Dejalma Mateus Morandin. -- 2024.

74 f.

Orientador: Alfredo Alejandro Gugliano.

Trabalho de conclusão de curso (Graduação) -- Universidade Federal do Rio Grande do Sul, Instituto de Filosofia e Ciências Humanas, Bacharelado em Ciências Sociais, Porto Alegre, BR-RS, 2024.

1. Inteligência Artificial. 2. Previsão de Cenários Políticos. 3. Reeleição de Prefeitos. 4. Teoria da Escolha Racional. 5. Dados Socioeconômicos. I. Gugliano, Alfredo Alejandro, orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

DEJALMA MATEUS MORANDIN

**Aplicação da Inteligência Artificial sobre dados socioeconômicos na previsão de  
Cenários Políticos: reeleição de prefeitos**

Trabalho de Conclusão de Curso apresentado  
como requisito parcial para a obtenção do título  
de Bacharel em Ciências Sociais, pelo Instituto  
de Filosofia e Ciências Humanas, da  
Universidade Federal do Rio Grande do Sul.

Aprovado em \_\_\_\_ de \_\_\_\_\_ de 2024.

**BANCA EXAMINADORA**

---

Prof. Dr. Alfredo Alejandro Gugliano.  
Orientador  
Universidade Federal do Rio Grande do Sul

---

Luis Gustavo Mello Grohmann  
Universidade Federal do Rio Grande do Sul

---

Mauricio Assumpção Moya  
Universidade Federal do Rio Grande do Sul

## AGRADECIMENTOS

Primeiramente minha gratidão vai para a Universidade Federal do Rio Grande do Sul que tornou possível a realização deste curso de formação. Professores dedicados conduziram minha graduação e realizaram uma verdadeira reconfiguração existencial pela construção e reflexão acerca das realidades sociais, antropológicas e políticas que vivemos. Em seguida, gostaria de agradecer aos professores Luis Gustavo Mello Grohmann e Mauricio Assumpção Moya que gentilmente aceitaram participar da banca avaliadora deste Trabalho de Conclusão de Curso, possibilitando o coroamento dessa caminhada.

Somada a essa gratidão à instituição e à formação, está uma gratidão muito especial ao meu orientador Professor Doutor Alfredo Gugliano que revela uma tamanha relação com a verdadeira Ciência Política que permite a criação de conexão com a diversidade de pensamento e opinião. Sem falar numa característica ímpar de manter autoridade de maneira gentil e carinhosa. Em toda minha vida as pessoas que tiveram autoridade sobre mim fizeram de maneiras diversas, às vezes mais duras outras mais suaves, mas nunca com tamanha gentileza e amorosidade. Isto foi absolutamente marcante e, além de reconfigurar a minha relação com a Ciência Política, marcou minha relação com a família e o mundo.

O mestrando em Engenharia da Computação, Gabriel Bauer de Oliveira, trabalhou na coleta, tratamento de dados e no treinamento da Inteligência Artificial e merece toda minha gratidão, pois caminhamos juntos ao longo de vários meses e o seu trabalho foi essencial para atingirmos os objetivos aqui elencados. Junto conosco trabalhou ainda a bibliotecária Adriana Rossoni, o estudante de Engenharia da Computação Kairan Morandin e a doutoranda em Ciência da Informação e bibliotecária, Janaina Morandin.

Na minha família muitas pessoas precisam ser lembradas, primeiro meus pais. Meu pai jamais abriu mão de ver seu filho formado e minha mãe é a própria expressão da cooperação, do entusiasmo para seguir e vencer e do acolhimento incondicional das mais diversas preocupações e instabilidades. Meus filhos foram absolutamente importantes nos diálogos reflexivos, nas conversas, nos incentivos e na manutenção da perseverança. A Manuela não abre mão de convencer seu pai do seu merecimento e da fortaleza no propósito e o Kairan pega junto, pensa soluções e trabalha na atingimento dos propósitos.

No entanto, sem dúvida a pessoa mais importante nessa jornada foi a Janaina, minha esposa, companheira de vida, mãe dos meus filhos. Serviu sempre de inspiração pela extraordinária dedicação e perseverança. Leu, releu, corrigiu e aconselhou. Permitiu que todo tempo necessário à faculdade estivesse disponível para mim, não mediu esforços em momento

algun para qualquer demanda que, de alguma maneira, pudesse inviabilizar a minha jornada. Jamais aceitou o desânimo e a menor possibilidade da dúvida. Meu amor e gratidão são infinitos e jamais conseguiria expressar o quanto te agradeço por compartilhar comigo essa jornada de vida.

## RESUMO

Este estudo explora a aplicação da Inteligência Artificial na previsão de cenários políticos, com foco na reeleição de prefeitos em municípios brasileiros. A pesquisa utiliza a Teoria da Escolha Racional como base teórica, combinando-a com técnicas de aprendizado de máquina para analisar dados socioeconômicos e identificar padrões que possam prever a reeleição. Foram coletados e tratados dados do Instituto de Pesquisa Econômica Aplicada e da base do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, além de variáveis eleitorais. A metodologia incluiu o uso de redes neurais profundas e técnicas de regularização, além de métodos para lidar com o desbalanceamento de classes. O estudo concluiu que, embora as redes neurais tenham conseguido realizar previsões um pouco acima do acaso, os modelos não alcançaram um nível de precisão que permita sua utilização como previsores confiáveis de reeleição. Os resultados indicam a necessidade de incorporar mais fontes de dados e diferentes abordagens analíticas para melhorar a eficácia dos modelos preditivos. A pesquisa sugere ainda que a integração de dados socioeconômicos com análises de conteúdo jornalístico, manifestações em redes sociais e pesquisas de opinião podem oferecer uma base mais sólida para a previsão de cenários políticos.

**Palavras-chave:** Inteligência Artificial, previsão de cenários políticos, reeleição de prefeitos, Teoria da Escolha Racional, dados socioeconômicos.

## ABSTRACT

This study explores the application of Artificial Intelligence in forecasting political scenarios, focusing on the re-election of mayors in Brazilian municipalities. The research uses Rational Choice Theory as a theoretical foundation, combining it with machine learning techniques to analyze socioeconomic data and identify patterns that may predict re-election. Data was collected and processed from the Institute of Applied Economic Research and the database of the National Institute of Educational Studies and Research Anísio Teixeira, as well as electoral variables. The methodology included the use of deep neural networks and regularization techniques, as well as methods to address class imbalance. The study concluded that, although the neural networks managed to perform slightly better than random predictions, the models did not reach a level of accuracy that would allow their use as reliable predictors of re-election. The results indicate the need to incorporate more data sources and different analytical approaches to improve the effectiveness of predictive models. The research also suggests that integrating socioeconomic data with analyses of journalistic content, social media manifestations, and opinion polls could offer a stronger foundation for forecasting political scenarios.

**Keywords:** Artificial Intelligence, political scenario forecasting, mayor re-election, Rational Choice Theory, socioeconomic data.

## LISTA DE ILUSTRAÇÕES

<b>Figura 1</b> - Arquitetura básica de uma rede neural com seus 4 componentes	32
<b>Figura 2</b> - Elementos da matriz de confusão	35
<b>Figura 3</b> - Classes de um problema de classificação binária	35
<b>Figura 4</b> - Classificação de um conjunto de dados de teste produz quatro resultados – verdadeiro positivo, falso positivo, verdadeiro negativo e falso negativo	35
<b>Figura 5</b> - Exemplo de curva <i>precision-recall</i> obtida no trabalho	36
<b>Figura 6</b> – Primeira Matriz de confusão associada ao treinamento do modelo individual	56
<b>Figura 7</b> – Segunda Matriz de confusão associada ao treinamento do modelo individual	57
<b>Figura 8</b> – <i>Precision-Recall Curve</i>	58
<b>Gráfico 1</b> – Comprovação de Aprovação e Probabilidade de Reeleição por Cidade	59

## LISTA DE QUADROS

**Quadro 1** - Consultas disponíveis através da API do Ipea

40

## LISTA DE ABREVIATURAS E SIGLAS

API	Application Programming Interface
AUC-ROC	Área sob a Curva da Característica de Operação do Receptor
CNN	Redes Neurais Convolucionais
DNN	Deep Neural Networks (Redes Neurais Profundas)
DT	Diferencial Partidário Atual
IA	Inteligência Artificial
IDH	Índice de Desenvolvimento Humano
Ipea	Instituto de Pesquisa Econômica Aplicada
ENEM	Exame Nacional do Ensino Médio
EUA	Estados Unidos da América
L2	Regularização L2 (L2 Regularization)
FN	False Negative (Falso Negativo)
FP	False Positive (Falso Positivo)
$\eta$	Taxa de Aprendizado (Learning Rate)
PIB	Produto Interno Bruto
PNL	Processamento da Linguagem Natural
PR	Precision-Recall
RNN	Redes Neurais Recorrentes
ReLU	Unidade Linear Retificada
RMSprop	Root Mean Square Propagation (Propagação da Média Quadrática)
SMOTE	Synthetic Minority Over-sampling Technique (Técnica de Superamostragem de Minoria Sintética)
TN	True Negative (Verdadeiro Negativo)
TP	True Positive (Verdadeiro Positivo)
UI	Renda de Utilidade Hipotética
UO	Renda de Utilidade
Val_loss	Validation Loss (Perda de Validação)
W	Peso Weight
$\nabla_w L$	Gradiente da Função de Perda em Relação ao Peso $w$ (Gradient of the Loss Function with Respect to Weight $w$ )

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>13</b>
1.1	JUSTIFICATIVA	14
1.2	PROBLEMA	16
1.3	HIPÓTESE	17
1.4	OBJETIVOS	17
<b>1.4.1</b>	<b>Objetivo Geral</b>	<b>17</b>
<b>1.4.2</b>	<b>Objetivos Específicos</b>	<b>17</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>18</b>
2.1	INTELIGÊNCIA ARTIFICIAL	18
2.2	TEORIA DA ESCOLHA RACIONAL	24
<b>3</b>	<b>DEFINIÇÕES DE MÉTRICAS E PROCEDIMENTOS</b>	<b>30</b>
3.1	REDES NEURAS PROFUNDAS	30
<b>3.1.1</b>	<b>Principais Componentes de uma Rede Neural</b>	<b>30</b>
3.2	INICIALIZAÇÃO DE PESOS	31
3.3	PROPAGAÇÃO PARA A FRENTE ( <i>FORWARD PRAPAGATION</i> )	31
<b>3.3.1</b>	<b>Função de Ativação</b>	<b>31</b>
<b>3.3.2</b>	<b>Regra de Aprendizado e <i>Back Propagation</i></b>	<b>31</b>
<b>3.3.3</b>	<b>Cálculo da Função da Perda</b>	<b>32</b>
<b>3.3.4</b>	<b>Atualização dos Pesos</b>	<b>32</b>
<b>3.3.5</b>	<b>Iteração</b>	<b>33</b>
<b>3.3.6</b>	<b>Avaliação e Ajuste</b>	<b>33</b>
3.4	MÉTRICAS DE AVALIAÇÃO	33
<b>3.4.1</b>	<b>Acurácia</b>	<b>33</b>
<b>3.4.2</b>	<b>Precisão</b>	<b>33</b>
<b>3.4.3</b>	<b>Sensibilidade (<i>Recall</i>)</b>	<b>34</b>
<b>3.4.4</b>	<b><i>F1 Score</i></b>	<b>34</b>
<b>3.4.5</b>	<b><i>AUC-ROC</i></b>	<b>34</b>
<b>3.4.6</b>	<b>Matriz de Confusão</b>	<b>34</b>
<b>3.4.7</b>	<b>Curva Precision-Recall</b>	<b>36</b>
3.5	REGULARIZAÇÃO L2	37
3.6	<i>DROPOUT</i>	37
3.7	<i>ESSEMBLE LEARNING</i>	37

3.8	DESBALANCEAMENTO DE CLASSES	38
3.9	<i>RANDOM UNDERSAMPLER</i>	38
<b>4</b>	<b>PROCEDIMENTOS METODOLÓGICOS</b>	<b>39</b>
4.1	COLETA DOS DADOS	39
<b>4.1.1</b>	<b>Instituto de Pesquisa Econômica Aplicada (Ipea)</b>	<b>39</b>
4.1.1.1	PIB Municipal	44
4.1.1.2	Variáveis das Séries Eleitorais	45
<b>4.1.2</b>	<b>Variáveis das Séries Educacionais</b>	<b>48</b>
4.2	TRATAMENTO DOS DADOS	49
<b>4.2.1</b>	<b>Tratamento das Séries Socioeconômicas</b>	<b>50</b>
<b>4.2.2</b>	<b>Tratamento das Séries Eleitorais</b>	<b>51</b>
<b>4.2.3</b>	<b>Tratamento das Séries Educacionais</b>	<b>52</b>
4.3	AGRUPAMENTO DOS DADOS PARA TREINAMENTO	53
4.4	ORGANIZAÇÃO PARA TREINAMENTO DA INTELIGÊNCIA ARTIFICIAL	54
<b>4.4.1</b>	<b>Treinamento</b>	<b>55</b>
<b>5</b>	<b>RESULTADOS</b>	<b>56</b>
<b>6</b>	<b>CONSIDERAÇÕES FINAIS</b>	<b>61</b>
	<b>REFERÊNCIAS</b>	<b>64</b>
	<b>APÊNDICE A</b>	<b>66</b>
	<b>APÊNDICE B</b>	<b>67</b>

## 1 INTRODUÇÃO

Em um mundo marcado por complexas interações políticas, sociais e econômicas, o estudo da Ciência Política desempenha um papel fundamental na compreensão das dinâmicas que moldam a vida em sociedade. Dentro desse cenário, o comportamento do eleitor se destaca como um elemento central para a análise política. Aperfeiçoar a compreensão dos elementos que são relevantes no momento das escolhas dos eleitores é uma tarefa contínua da Ciência Política.

Analisando as perspectivas sobre a escolha do voto pode-se identificar uma série de fatores para essas decisões, desde identificação partidária e posicionamento ideológico até as condições socioeconômicas em que os indivíduos estão inseridos.

Neste contexto, a teoria da escolha racional emerge como uma lente analítica para entender como os indivíduos tomam decisões políticas. Baseada na premissa de que os atores políticos são racionais e buscam maximizar suas próprias preferências ou minimizar os prejuízos, a teoria da escolha racional explora como as decisões são tomadas em um contexto de custos e benefícios percebidos. Ao aplicar essa abordagem à análise do comportamento eleitoral, emerge a pergunta: como as características socioeconômicas podem ajudar a entender os cenários políticos?

A possibilidade de utilização e a disponibilidade de dados socioeconômicos viáveis e relevantes para o propósito do projeto foi uma questão que permeou (e ainda permeia) a viabilidade de desenvolvimento da hipótese fundamental. A forma como tais dados públicos são tratados por diferentes esferas públicas (nacionais, estaduais, regionais e municipais), tanto no âmbito de coleta quanto no âmbito de disponibilização das informações, é um ponto de discussão que deve ser levado em consideração não apenas para o desenvolvimento de ferramentas tecnológicas, mas fundamentalmente para a adoção de medidas públicas efetivas.

A utilização de Inteligência Artificial (IA) na previsão de cenários políticos representa em alguma medida uma inovação na Ciência Política. Com suas capacidades avançadas de processamento de dados e aprendizado de máquina, a IA possibilita a análise de grandes volumes de dados socioeconômicos e políticos. Essa tecnologia tem o potencial de identificar padrões e tendências que podem não ser evidentes através de métodos tradicionais, oferecendo previsões sobre comportamentos e eventos políticos. Além disso, a IA pode simular diferentes cenários, fornecendo uma ferramenta para a formulação de estratégias políticas e a tomada de decisões informadas. A incorporação da IA no campo da Ciência Política abre novas perspectivas para a compreensão e análise da dinâmica política contemporânea.

## 1.1 JUSTIFICATIVA

A IA tem a capacidade de processar grandes volumes de dados de forma eficiente e em tempo real. Isso se mostra crucial para previsões políticas, que dependem da análise de dados de diversas fontes, como redes sociais, pesquisas de opinião, notícias e registros históricos. Neste trabalho, a opção foi por analisar dados socioeconômicos de bases de dados periodicamente atualizadas.

Modelos de IA, como aprendizado de máquina e redes neurais, são especialmente eficazes na identificação de padrões complexos que podem não ser imediatamente aparentes para analistas humanos. Essa habilidade é valiosa para detectar tendências e correlações em dados políticos que podem influenciar previsões.

A utilização da IA pode ser alinhada teoricamente com a Teoria da Escolha Racional, já que a IA pode modelar decisões eleitorais como processos de otimização com base em preferências individuais e contextos socioeconômicos. Figueiredo (2022), argumenta que muitos fenômenos políticos podem ser melhor compreendidos ao reconhecer padrões complexos em dados históricos e atuais. A IA é extremamente eficaz nessa tarefa, pois pode identificar relações e tendências ocultas nos dados socioeconômicos que podem não ser visíveis por meio de métodos tradicionais. Uma das contribuições de Figueiredo (2022), é a integração de múltiplas variáveis para entender dinâmicas políticas. A IA pode processar e analisar simultaneamente um grande número de variáveis socioeconômicas, como renda, educação, emprego, e dados demográficos, permitindo uma visão holística e detalhada dos fatores que influenciam cenários políticos.

Segundo Downs (2013), os eleitores tomam decisões que maximizam sua utilidade esperada. A IA pode utilizar algoritmos de aprendizado de máquina para prever quais políticas ou candidatos são mais prováveis de maximizar a utilidade dos eleitores em diferentes contextos socioeconômicos. A abordagem de Downs (2013), pode ser complementada com a simulação de cenários políticos utilizando IA, que pode criar modelos preditivos que simulam diferentes resultados eleitorais com base em variáveis socioeconômicas e políticas, permitindo uma análise prospectiva e informada.

Shepsle (2016), argumenta que dados empíricos são fundamentais para a análise política. A IA pode processar grandes volumes de dados socioeconômicos, identificando tendências e padrões que influenciam o comportamento político. Isso pode ajudar a prever como variáveis como renda, educação, emprego e desigualdade podem impactar a política. Shepsle

(2016), é um proponente da Teoria da Escolha Racional e da modelagem quantitativa. Grandes volumes de dados permitem a criação de modelos mais complexos e precisos, que podem levar em conta uma maior variedade de variáveis e interações. Isso é alinhado com sua visão de que a política pode ser analisada como um conjunto de decisões racionais dentro de um contexto institucional

A Teoria da Escolha Racional tem sido reconhecida por sua capacidade de explicar o comportamento político por meio da análise de decisões individuais baseadas em benefícios e custos percebidos. Contudo, a aplicação prática desta teoria em contextos políticos dinâmicos e multifacetados pode ser desafiadora. Integrar técnicas de IA com a Teoria da Escolha Racional oferece uma oportunidade para testar e expandir os limites dessa teoria, permitindo uma análise das interações políticas.

O rápido avanço das tecnologias de IA, especialmente em áreas como aprendizado de máquina e análise de *big data*, proporciona ferramentas para a análise de grandes volumes de dados socioeconômicos e políticos. Estas ferramentas podem identificar padrões e prever comportamentos que seriam inacessíveis por meio de métodos tradicionais. Ao explorar essas capacidades, este estudo pode contribuir para a inovação metodológica na Ciência Política.

Prever cenários políticos de forma acurada pode ter impactos significativos em diversos campos, desde a formulação de políticas públicas até a estratégia eleitoral e a gestão de crises. Governos, instituições e organizações podem se beneficiar de previsões mais precisas para tomar decisões informadas e proativas. Além disso, a previsão de cenários políticos pode ajudar a manter a estabilidade social e econômica, ao antecipar conflitos e identificar oportunidades para intervenções preventivas.

Assim, esta pesquisa se posiciona na interseção entre Ciência Política e Ciência de Dados, pretendendo oferecer uma contribuição ao corpo de conhecimento existente. Ao aplicar e avaliar modelos de IA no contexto da Teoria da Escolha Racional, este estudo não apenas reforça a relevância dessa teoria, mas também abre novas possibilidades para sua aplicação prática.

Em suma, esta pesquisa justifica-se pela sua relevância teórica e pelo potencial para a inovação tecnológica, o que pode implicar em um impacto prático significativo. Justifica-se também, pela possível necessidade de uma reflexão ética sobre o uso da IA em cenários políticos. Ao explorar a integração entre a Teoria da Escolha Racional e a IA, este estudo pretende fornecer uma base para futuras pesquisas e aplicações neste campo.

## 1.2 PROBLEMA

Considerando que os indicadores econômicos e sociais dos municípios sofrem variações ao longo do tempo devido às decisões e ações políticas, e que diferentes propostas políticas são interpretadas de maneiras variadas entre as diferentes camadas sociais, pode-se, possivelmente, utilizar técnicas de análise de dados para reconhecer padrões nesses dados. Parte-se do pressuposto de que esses padrões podem fornecer conclusões sobre o cenário político, especialmente do ponto de vista da reeleição de políticos e partidos, por meio de indicadores socioeconômicos e sociais. Investiga-se se indicadores como o PIB per capita, taxa de desemprego e índices de desenvolvimento humano (IDH), juntamente com indicadores como níveis de educação, saúde e segurança são influenciados pelas políticas públicas implementadas pelos governos locais. Essas variações podem refletir a eficácia das políticas e o impacto que elas têm nas condições de vida da população.

Considerando que os indicadores econômicos e sociais dos municípios sofrem variações com o passar do tempo pelas decisões e ações políticas e que diferentes propostas políticas são interpretadas de maneira diferente entre as camadas sociais, é possível reconhecer padrões nesses dados para permitir conclusões sobre o cenário político, principalmente do ponto de vista da reeleição?

## 1.3 HIPÓTESE

A hipótese é que as variações nos dados socioeconômicos de um município permitem extrair conclusões sobre o comportamento eleitoral. A partir da aplicação de análises desses dados com a utilização de algoritmos de Inteligência Artificial é possível reconhecer com alguma precisão esses panoramas eleitorais.

Hipótese fundamental: com a utilização de dados socioeconômicos relevantes, disponibilizados em quantidade suficiente, é possível criar ferramentas baseadas em Inteligência Artificial que permitam realizar a predição de eventos eleitorais com grau de acerto satisfatório.

H1 - Dessa maneira, quando variações positivas acontecem nesses índices, o candidato da situação potencializa sua votação.

H2 - Nas variações negativas dos índices, o candidato da situação tende a recrudescer a sua votação.

#### 1.4 OBJETIVOS

A seguir são descritos o objetivo específico e os objetivos gerais deste estudo.

##### 1.4.1 Objetivo Geral

Desenvolver um modelo preditivo utilizando Inteligência Artificial para analisar dados socioeconômicos visando identificar padrões relevantes nas reeleições municipais.

##### 1.4.2 Objetivos Específicos

São objetivos específicos da pesquisa:

- a) Coletar e preparar um conjunto de dados relevante para o contexto político estudado;
- b) treinar e testar modelos de Inteligência Artificial para prever cenários políticos.

## 2 REFERENCIAL TEÓRICO

O referencial teórico deste trabalho baseia-se na Teoria da Escolha Racional e nas ideias de teóricos sobre comportamento eleitoral e as abordagens contemporâneas da aplicação da Inteligência Artificial.

### 2.1 INTELIGÊNCIA ARTIFICIAL

Na década de 1940, testemunhou-se um salto monumental na maneira como a informação começou a ser processada, graças ao advento dos computadores digitais. Esses computadores emergiram como potentes ferramentas capazes de armazenar vastas quantidades de dados e processá-los com eficiência inédita. De acordo com Luger (2013), esse desenvolvimento estabeleceu um novo paradigma em que a inteligência era vista como uma forma específica de processamento da informação. Segundo Doneda *et al.* (2018), foi em 1945 que Vannevar Bush, antecipando o potencial dessa nova era, sugeriu que o tratamento automatizado da informação poderia eventualmente servir como uma extensão da memória humana, marcando o início da era do processamento automatizado de dados, com seus fundamentos relevantes até hoje.

Em um momento significativo dos anos de 1950, Alan Turing lançou uma questão pioneira com o seu *jogo da imitação*, interrogando se as máquinas poderiam efetivamente pensar, dadas as vantagens proporcionadas pelo progresso da computação e engenharia (Turing, 1950). O Teste de Turing é uma proposta para um critério de inteligência em uma máquina. No teste, um interrogador humano se comunica com dois participantes em outra sala: um humano e uma máquina projetada para gerar respostas semelhantes às humanas. Se o interrogador não consegue determinar consistentemente qual dos dois é a máquina, a máquina é considerada, para todos os efeitos práticos, inteligente (Silva, 2021).

Esse questionamento abriu caminho para a primeira conferência dedicada aos profissionais da área de IA, realizada em 1956 no Dartmouth College, nos Estados Unidos da América (EUA). Este encontro histórico reuniu pesquisadores interessados em combinar os campos da computação e da inteligência, momento em que o termo *Inteligência Artificial* foi formalmente adotado. Os assuntos discutidos na conferência, como esclarece Luger (2013), abrangiam desde a teoria da complexidade e metodologias de abstração até o design de

linguagens de programação e aprendizado de máquina, delineando o que viria a ser o núcleo da Ciência da Computação contemporânea.

Contudo, ao procurar construir uma historicidade profunda da IA, pode-se buscar em Aristóteles os primeiros elementos, quando este define a Física como a ação de estudar as coisas que mudam, e distingue, entre as coisas, a matéria e a forma, criando os fundamentos para o desenvolvimento da computação que é feita da manipulação nas formas de materiais eletromagnéticos.

Fazer um paralelo entre Aristóteles, uma figura central na filosofia antiga, e a IA, uma das fronteiras mais avançadas da tecnologia moderna, pode parecer desafiador à primeira vista. No entanto, existem conceitos fundamentais em seu trabalho que têm ressonância direta ou indireta com os princípios e desafios da IA.

Sobre as questões de lógica e racionalidade, Aristóteles é considerado o pai da lógica formal. Seu trabalho, sobre os princípios da dedução e inferência lógica, permanece fundamental para a Filosofia e a ciência, pois acreditava na capacidade de raciocínio e lógica como as principais características da mente humana. A lógica formal é uma base crucial para muitas áreas da IA, especialmente na programação lógica, sistemas baseados em regras e algoritmos de inferência. A capacidade de processar informações logicamente é essencial para que as máquinas realizem tarefas complexas, como aprender, raciocinar e tomar decisões (Luger, 2013).

A trajetória da ciência desde Copérnico até a era moderna não apenas transformou a compreensão das pessoas sobre o universo, mas também pavimentou o caminho para questionar e redefinir a natureza da inteligência, tanto humana quanto artificial. Em sua obra, Kuhn (2013), oferece uma análise profunda de como as revoluções científicas transformam o entendimento do mundo, abordando mudanças paradigmáticas como as iniciadas por Copérnico.

A separação entre mente e matéria, ideias e coisas, influenciou o desenvolvimento da IA, pois esta se esforça para replicar processos mentais — tradicionalmente vistos como exclusivamente humanos — em sistemas físicos, como computadores. De acordo com Luger (2013), a premissa de que processos cognitivos complexos podem ser alcançados por sistemas físicos é fundamental para a IA, refletindo a visão de que a inteligência, ou pelo menos aspectos dela, pode ser emulada por meio de algoritmos e *hardware*.

De acordo com Onody (2021), Turing estava menos interessado em definir o que seria pensar de maneira abstrata e mais em criar um critério objetivo e prático que pudesse ser testado. Ele previu que, em algum momento, as máquinas poderiam ser construídas para imitar o

pensamento humano tão bem que seria desafiador distinguir suas respostas daquelas de um humano. Esta questão não apenas abriu caminho para a pesquisa em IA, mas também provocou debates filosóficos profundos sobre a natureza da mente, da consciência e da inteligência, tanto humana quanto artificial.

Para Silva (2021), Turing também especulou sobre o futuro da inteligência das máquinas, sugerindo que, eventualmente, elas poderiam aprender e se adaptar ao ponto de exibir comportamentos que poderiam ser considerados inteligentes. Ainda sobre este aspecto, Onody (2021), afirma que o trabalho de Turing colocou as bases para a exploração do conceito de IA como o conhecemos hoje, visto os desdobramentos que vieram do seu trabalho como a competição anual a partir de 1966 após um programa chamado ELISA ter passado por inúmeros testes e inspirar a adoção do Prêmio Loebner que avalia anualmente programas que utilizam o teste de Turing como referência de desempenho.

Um conceito genérico de IA foi proposto por Luger (2013, p. 1): “A Inteligência Artificial pode ser definida como o ramo da ciência da computação que se ocupa da automação do comportamento inteligente”, no entanto pontua que engloba uma vasta gama de subcampos, técnicas e aplicações, desde o processamento de linguagem natural e aprendizado de máquina até robótica e sistemas de recomendação. Uma definição muito genérica pode falhar em capturar a especificidade e a diversidade das abordagens dentro do campo. Ainda na visão de Luger (2013), os objetivos da IA mudaram ao longo do tempo. Inicialmente, o foco estava em replicar o raciocínio humano. Com o tempo, isso se expandiu para incluir a capacidade de lidar com tarefas específicas de maneiras que podem ou não imitar diretamente a cognição humana. Uma definição genérica pode não refletir adequadamente essas mudanças de foco.

Conforme apresentado por Russell e Norvig (2009), a IA é definida como o ramo da Ciência da Computação que se dedica ao desenvolvimento de agentes autônomos. Esses agentes são sistemas capazes de observar o ambiente ao seu redor e tomar decisões que maximizem suas chances de sucesso em um determinado objetivo ou conjunto de objetivos. A IA abrange uma diversidade de subcampos, incluindo aprendizado de máquina, visão computacional, processamento de linguagem natural e robótica, cada um com o propósito de criar sistemas que possam realizar, de maneira autônoma, tarefas que normalmente exigiriam inteligência humana.

A IA é categorizada por Russel e Norvig (2009), em quatro abordagens principais: sistemas que pensam como humanos, sistemas que agem como humanos, sistemas que pensam racionalmente e sistemas que agem racionalmente. Esta classificação reflete não apenas uma ampla gama de tecnologias e métodos empregados no campo da IA, mas também encapsula a filosofia subjacente ao desenvolvimento de sistemas inteligentes: a criação de máquinas

capazes de emular e, eventualmente, superar as capacidades cognitivas humanas em tarefas específicas.

O conceito amplamente adotado como IA corresponde, na realidade, ao campo do aprendizado de máquina (Jordan, 2019). Este último, emergente nas últimas décadas, representa a interseção de disciplinas como Ciência da Computação e Estatística, com o propósito de desenvolver algoritmos capazes de analisar dados, efetuar previsões e executar tomadas de decisões. Por sua vez, Géron (2019, p. 4, grifo do autor) completa essa visão dizendo que “Aprendizado de Máquina é a ciência (e a arte) da programação de computadores para que eles possam *aprender com os dados*”.

O avanço tecnológico e a emergência do conceito de *big data* têm sido cruciais para o desenvolvimento de modelos analíticos avançados. Esses modelos, nutridos por extensos conjuntos de dados, são capazes de gerar hipóteses, embasar suas escolhas e estabelecer padrões de comportamento (Doneda *et al.*, 2018). Conforme apontado por Campos (2020), a complexidade de um modelo está diretamente relacionada à quantidade de dados necessária para seu treinamento eficaz. Além disso, uma infraestrutura capaz de processar rapidamente grandes volumes de dados, otimizando tempo e recursos, é essencial para a operacionalização desses modelos.

Dentro desse contexto, identificam-se três modalidades principais de aprendizado de máquina: aprendizado supervisionado, não supervisionado e por reforço (Campos, 2020; Gerón, 2019; Lurdermir, 2021). O aprendizado supervisionado opera com a rede neural sendo treinada utilizando exemplos específicos de entrada e saída. Para que o treinamento seja efetivo, é preciso que haja uma compreensão antecipada dos resultados esperados, permitindo que um analista determine a adequação das respostas da máquina pela comparação entre os resultados alcançados e os objetivos predeterminados. Caso erros sejam identificados, a máquina é ajustada de forma a aprimorar seu desempenho em interações futuras (Ferneda, 2006).

Para se conceber uma perspectiva mais precisa, é necessário conceituar cada uma das técnicas ou campos da IA. Russell e Norvig (2009) apresentam o aprendizado de máquina (*machine learning*) como um pilar fundamental dessas aplicações, destacando sua capacidade de habilitar máquinas a aprenderem a partir de dados. Os autores categorizam o aprendizado de máquina em três abordagens principais: a primeira é o aprendizado supervisionado onde o modelo é treinado em um conjunto de dados rotulados, ou seja, cada exemplo de treinamento é associado a um rótulo ou saída específica. O objetivo é que, após o treinamento, o modelo seja capaz de prever o rótulo correto para novos dados nunca antes vistos, baseando-se nas

características aprendidas. Esse é o método amplamente utilizado em tarefas como classificação e regressão.

A segunda abordagem de Russell e Norvig (2009), é o aprendizado não supervisionado, que diferentemente do aprendizado supervisionado, lida com dados que não estão rotulados. O modelo procura identificar padrões, agrupamentos ou estruturas subjacentes nos dados sem qualquer orientação específica sobre o resultado desejado. Esta abordagem é útil para explorar dados e identificar relações significativas ou agrupamentos naturais dentro dos conjuntos de dados.

Por fim, os autores apresentam o aprendizado por reforço, este é um paradigma onde o modelo, ou agente, aprende a tomar decisões através de tentativa e erro, recebendo recompensas ou punições por suas ações. O objetivo é desenvolver uma política de ação que maximize a soma das recompensas ao longo do tempo. Russell e Norvig (2009) destacam que essa abordagem é particularmente aplicável em contextos onde um agente deve tomar uma sequência de decisões, como em jogos ou na navegação de robôs e enfatizam que o sucesso das aplicações de aprendizado de máquina depende da qualidade e da quantidade dos dados disponíveis, além da escolha do algoritmo correto para a tarefa em questão.

Alguns subcampos são importantes para o entendimento da IA como as redes neurais e o aprendizado profundo, robótica, processamento da linguagem natural (PLN), sistemas baseados em conhecimento e raciocínio automatizado. Luger (2013) descreve as Redes Neurais como sistemas computacionais inspirados na estrutura e funcionamento do cérebro humano, particularmente na maneira como os neurônios processam e transmitem informações através de sinapses. Esta inspiração biológica confere às redes neurais uma capacidade notável para o processamento de dados sensoriais complexos, como imagens, sons e linguagem natural, de acordo com (Russell e Norvig, 2009). Do ponto de vista de Luger (2013), diversas são as aplicações para as redes neurais, incluindo visão computacional, processamento de linguagem natural e previsão de séries temporais. Ele discute como esses sistemas são capazes de lidar com informações imprecisas ou fragmentadas e ainda assim produzir resultados úteis, uma característica particularmente valiosa em ambientes ruidosos ou onde os dados são incompletos.

Russell e Norvig (2009), exploram o conceito de aprendizado profundo (*deep learning*) como uma evolução significativa dentro do campo do aprendizado de máquina, uma subárea da IA. O aprendizado profundo é apresentado como uma técnica que emprega redes neurais com muitas camadas (ou profundas) para aprender representações de dados em níveis progressivamente mais altos de abstração. Essa abordagem permite que sistemas

computacionais modelem e resolvam problemas complexos com uma eficiência e precisão notáveis, particularmente em áreas como reconhecimento de imagem, processamento de linguagem natural e análise de séries temporais.

Os autores mencionados discutem como as redes neurais, compostas por unidades de processamento inspiradas pelos neurônios biológicos, podem ser empilhadas em várias camadas para formar estruturas profundas. Cada camada transforma as entradas recebidas de maneira não linear, permitindo que o sistema aprenda características complexas dos dados. Uma das principais contribuições é a capacidade de aprender automaticamente as características necessárias para a realização de tarefas, ao invés de depender de características manualmente projetadas por humanos. Isso simplifica o processo de modelagem para muitos problemas difíceis.

Outro apontamento que Russell e Norvig (2009) apresentam, são as numerosas aplicações do aprendizado profundo que revolucionaram campos específicos, como o uso de Redes Neurais Convolucionais (CNNs) para tarefas de visão computacional e Redes Neurais Recorrentes (RNNs) para processar sequências de dados, como texto ou séries temporais. Apesar dos avanços, eles também discutem os desafios atuais do aprendizado profundo, incluindo a necessidade de interpretabilidade dos modelos, a questão da transferência de aprendizado e a busca por métodos que possam funcionar com conjuntos de dados menores ou mais ruidosos. Em síntese o aprendizado profundo representa uma fronteira avançada no desenvolvimento de sistemas de IA, oferecendo uma abordagem poderosa para extrair conhecimento de grandes volumes de dados e resolver problemas que eram anteriormente considerados inacessíveis para as máquinas.

A robótica, para Luger (2013), é vista como a aplicação de sistemas de IA em entidades físicas ou robôs. Estes sistemas são projetados para perceber o ambiente ao seu redor, processar essas informações e agir de maneira autônoma ou semi-autônoma para alcançar objetivos específicos. É, portanto, a incorporação de uma intersecção de *hardware* (corpos robóticos, sensores, atuadores) e *software* (algoritmos de IA, processamento de dados, decisão autônoma), representando um campo prático onde as teorias e modelos de IA são testados e implementados.

Finalmente, pode-se incluir o PLN. Luger (2013) apresenta o PLN como uma subárea crítica da IA dedicada à compreensão, interpretação e geração de linguagem humana por máquinas, destacando que o PLN envolve a capacidade das máquinas de entender a linguagem natural humana, abrangendo desde a análise sintática (estrutura da frase) e semântica (significado) até a pragmática (uso da linguagem em contextos específicos).

Além de compreender a linguagem, o PLN também se estende à geração de textos coerentes e contextualmente apropriados, permitindo que as máquinas comuniquem ideias complexas de forma natural. Dispõe ainda de aplicações em sistemas de diálogo e *chatbots*, que são interfaces conversacionais que podem interagir com humanos em linguagem natural, conversão de texto ou fala de um idioma para outro, ainda na análise de sentimentos identificando opiniões e emoções em textos, útil para análises de mercado e redes sociais e também a geração de resumos concisos de documentos extensos. Modelos como o ChatGPT, citado por Larroyed (2023), apresentam ampla aplicabilidade em tarefas de PLN, incluindo a criação de conteúdo textual, tradução de idiomas e o fornecimento de respostas a consultas específicas.

## 2.2 TEORIA DA ESCOLHA RACIONAL

Ao se fazer uma reflexão sobre a evolução da Ciência Política desde o fim da Segunda Guerra Mundial até o século XXI, pode-se destacar a transição da simples descrição e julgamento para a explicação e análise sistematizada dos fenômenos políticos como acentua Shepsle (2016). No pós-guerra, a Ciência Política era dominada por descrições detalhadas e análises históricas das práticas políticas, como as do Congresso dos Estados Unidos e outras instituições. Essas descrições eram minuciosas, mas frequentemente careciam de uma organização que transcendesse a mera acumulação de fatos. Além disso, esses estudos incluíam críticas ao *status quo* e sugeriam reformas sem, no entanto, analisar profundamente as causas dos problemas ou o impacto potencial das reformas.

Shelpse (2016) prossegue no raciocínio pontuando que a partir dos anos 1960, houve uma mudança significativa com um foco crescente nas questões do *porquê*, marcando um movimento em direção à análise científica que busca explicações sistemáticas para os padrões observados. Essa nova abordagem científica pretende entender regularidades empíricas e suas causas, o que é crucial para transformar a coleta de dados e a descrição detalhada em verdadeira compreensão científica. Além disso, a importância dos modelos e teorias na Ciência Política, como simplificações estilizadas da realidade podem ajudar a entender fenômenos complexos. Esses modelos permitem intuições sobre como indivíduos e grupos tomam decisões, que são fundamentais para analisar comportamentos políticos em diversos contextos. Ainda servem de ferramenta para ajudar leitores e estudiosos a aplicar conceitos teóricos para entender e analisar

eventos políticos reais, melhorando assim sua capacidade de prever e explicar tais eventos por meio de uma abordagem científica mais madura e analítica.

Em *Analyzing Politics: Rationality, Behavior, and Institutions* (traduzido como "Analisar a Política: Racionalidade, Comportamento e Instituições"), Shepsle (2016), explora a evolução da Ciência Política ao longo do tempo. O autor enfatiza como a disciplina se desenvolveu por meio da incorporação de métodos e abordagens analíticas mais rigorosas. Um ponto principal que Shepsle (2016) aborda sobre a evolução da Ciência Política é a tradição clássica e normativa onde inicialmente, a Ciência Política estava fortemente enraizada na tradição clássica, com foco em questões normativas e filosóficas sobre o melhor tipo de governo e a justiça e o bem comum. Grandes filósofos como Platão e Aristóteles contribuíram para essa tradição.

Outro ponto é o Empirismo e Comportamentalismo, onde, com o tempo, a Ciência Política começou a se afastar das questões puramente normativas e a adotar uma abordagem mais empírica. O comportamento dos indivíduos e grupos tornou-se um foco central, levando ao surgimento do comportamentalismo na primeira metade do século XX. Os pesquisadores começaram a utilizar métodos quantitativos e qualitativos para estudar o comportamento político. Em seguida, a Teoria da Escolha Racional, que se desenvolve a partir da segunda metade do século XX e ganha destaque. Essa abordagem utiliza modelos econômicos para entender como os indivíduos fazem escolhas políticas baseadas na maximização de sua utilidade pessoal. Shepsle (2016) é um dos principais defensores dessa abordagem e argumenta que ela proporciona uma estrutura rigorosa para analisar o comportamento político.

Outro ponto se refere às Instituições e Regras do Jogo, pois, além do comportamento individual, Shepsle (2016) destaca a importância das instituições políticas e das regras do jogo. O autor argumenta que as instituições moldam e constroem o comportamento dos indivíduos, criando incentivos e desincentivos específicos. A análise institucional é, portanto, crucial para entender a política. Destacam-se também os Modelos Formais e Análise Matemática, em que Shepsle (2016) promove o uso de modelos formais e análise matemática na Ciência Política. Para o autor, esses métodos ajudam a clarificar teorias, fornecer previsões precisas e testar hipóteses de maneira rigorosa. O uso de jogos, teoria dos jogos e modelos espaciais são exemplos de ferramentas analíticas empregadas.

Por fim destaca-se a interdisciplinaridade e expansão de métodos, onde a Ciência Política contemporânea é caracterizada por uma abordagem interdisciplinar, incorporando *insights* de Economia, Sociologia, Psicologia e outras disciplinas. Além disso, há uma expansão contínua dos métodos de pesquisa, incluindo análise de dados, experimentos de campo e laboratoriais, e simulações computacionais. Shepsle (2016) argumenta que a evolução da Ciência Política

reflete um movimento em direção a uma disciplina mais científica e analítica, com ênfase em rigor metodológico, teoria formal e compreensão das instituições e do comportamento racional dos atores políticos.

Em *A Decisão do Voto: Democracia e Racionalidade*, Figueiredo (2022), faz uma leitura do contexto político parecido com Shepsle (2016) e retrata a evolução da Ciência Política com um foco especial na interação entre democracia e racionalidade na tomada de decisão dos eleitores. Figueiredo (2022) considera várias etapas e influências na evolução da Ciência Política, destacando alguns pontos principais: primeiro, a integração da Teoria da Escolha Racional enfatizando a importância dessa teoria na Ciência Política, onde os eleitores são vistos como agentes racionais que tomam decisões com base em suas preferências ou para evitar maiores prejuízos e na informação disponível. Isso trouxe uma perspectiva mais analítica e quantitativa para o estudo da política, influenciando fortemente os métodos de pesquisa e a interpretação dos comportamentos eleitorais. Segundo, a influência do contexto histórico e social que molda a evolução da Ciência Política. A democratização e a ampliação do sufrágio, por exemplo, trouxeram novos desafios e questões para os cientistas políticos, exigindo uma compreensão mais profunda dos comportamentos eleitorais e das dinâmicas políticas. Em terceiro lugar, destaca o papel das instituições e das regras eleitorais argumentando que as instituições políticas e as regras eleitorais têm um papel crucial na formação das preferências dos eleitores e na estruturação da competição política. A evolução da Ciência Política inclui a análise de como diferentes sistemas eleitorais e arranjos institucionais influenciam o comportamento dos eleitores e os resultados das eleições. Um quarto ponto apontado por Figueiredo (2022) são os métodos de pesquisa e análise de dados manifestando a evolução da Ciência Política, o que também se reflete na sofisticação dos métodos de pesquisa e análise de dados. O uso de modelos estatísticos avançados, experimentos e *big data* permite uma compreensão mais detalhada e precisa dos fenômenos políticos. Por fim, entende que a interdisciplinaridade é fundamental na evolução da Ciência Política, com a incorporação de *insights* da Sociologia, Economia, Psicologia e outras disciplinas para enriquecer a análise política.

Os principais pontos da Teoria da Escolha Racional na Ciência Política dão um panorama do alcance dessa teoria para as compreensões das ações dos atores políticos, principalmente os eleitores, onde a análise começa com o indivíduo. Mesmo ao estudar grupos ou instituições, os teóricos buscam entender o comportamento individual subjacente. Os atores são considerados autônomos e suas ações são derivadas de cálculos individuais (Downs, 2013). Isto é, entendendo as motivações, preferências e ações individuais, pode-se depois compreender

fenômenos coletivos, geralmente operando sob a suposição de que os indivíduos são racionais, ou seja, eles fazem escolhas para maximizar sua utilidade ou satisfação.

No entanto, os indivíduos não necessariamente têm informações completas e podem enfrentar restrições ao fazer suas escolhas. Ao entender o comportamento de muitos indivíduos e como eles interagem, é possível fazer previsões ou compreender fenômenos sociais em uma escala maior. Downs (2013) argumenta que os cidadãos votam de forma a maximizar seus benefícios pessoais, um exemplo clássico de individualismo metodológico aplicado ao comportamento eleitoral.

A maximização da utilidade propõe que os atores tomam decisões para maximizar sua utilidade ou benefício. Eles fazem escolhas que acreditam trazer-lhes a maior recompensa possível, dadas as opções disponíveis (Becker, 1976). Por outro lado, as informações são restritas e os custos de transação calculados, enquanto os atores buscam maximizar sua utilidade, eles operam sob certas restrições. Eles podem não ter informações completas ou enfrentar custos para obter essa informação (North, 1990). Além disso, a decisão de participar da política, como votar, também tem custos associados (Downs, 2013).

A interdependência estratégica é o cenário onde as escolhas feitas por um ator frequentemente dependem das escolhas feitas por outros. Em muitos casos, como jogos de dilema de prisioneiro ou tragédia dos comuns, os atores tomam decisões sabendo que outros estão fazendo seus próprios cálculos (Axelrod, 1984). Por fim, as instituições e as regras do jogo moldam o comportamento dos atores ao estabelecer incentivos e penalidades. As instituições podem estabilizar ou alterar os resultados da interação estratégica (North, 1990; Ostrom, 1990).

Aprofundando a compreensão da tomada de decisão propriamente dita, aborda-se a decisão do voto a partir de uma perspectiva que combina elementos da Teoria da Escolha Racional com fatores sociopsicológicos. Figueiredo (2022) argumenta que a decisão do voto é um processo complexo que envolve a avaliação racional dos eleitores sobre diferentes candidatos e partidos, mas também é influenciada por fatores como identidade partidária, contexto social e influência dos meios de comunicação.

Neste contexto, segundo Downs (2013), o eleitor decide o voto pela lógica do voto racional e se baseia no axioma de que os cidadãos agem racionalmente na política, votando no partido que acreditam que lhes proporcionará maiores benefícios. Os principais pontos na decisão do voto incluem o fato de que o eleitor vota no partido que acredita que proporcionará mais benefícios, ou maior "renda de utilidade", em comparação aos outros partidos. A Renda de Utilidade são os benefícios recebidos das atividades governamentais e considerados como

fluxos de utilidade. É uma medida dos benefícios que um cidadão usa para decidir entre alternativas e pode incluir benefícios não diretamente percebidos pelo cidadão, como serviços públicos essenciais. Avaliando os diferenciais partidários o eleitor compara a Renda de Utilidade que acredita que receberá de cada partido se este estiver no poder votando no partido que oferece um diferencial partidário positivo (maior renda de utilidade esperada). Essa avaliação é feita considerando tanto o desempenho atual quanto as expectativas futuras dos partidos. O diferencial partidário atual ( $D_t$ ) é a diferença entre a Renda de Utilidade que o eleitor realmente recebeu ( $U_o$ ) do partido no poder e a Renda de Utilidade Hipotética ( $U_i$ ) que ele acredita que teria recebido se o partido de oposição estivesse no poder:  $D_t = U_o - U_i$ .

Downs (2013) segue assumindo ainda que o eleitor concebe um fator tendencial ajustando sua avaliação com base nas tendências observadas no desempenho do governo atual e melhorias ou deteriorações nesse governo são consideradas ao projetar o desempenho futuro. Em caso de empate entre partidos com políticas idênticas, o eleitor usa avaliações de desempenho comparando o governo atual com governos anteriores.

Em sistemas multipartidários o eleitor pode votar estrategicamente em partidos que não são sua primeira preferência para evitar a vitória de um partido menos desejado. Essa decisão é baseada em previsões sobre as preferências de outros eleitores e nas chances de vitória dos partidos. Nesse processo o eleitor ainda lida com incertezas e falta de informação ao fazer estimativas de rendas de utilidade futuras e novas informações podem mudar a decisão de voto se afetarem as percepções sobre políticas partidárias ou a concepção de uma *boa sociedade*.

Resumindo, o eleitor racional compara os benefícios recebidos e esperados dos partidos, vota no partido que espera proporcionar maior Renda de Utilidade, ajusta a decisão com base em tendências e desempenho passado e em sistemas multipartidários e pode votar estrategicamente para influenciar o resultado desejado.

Por sua vez, o governante segue uma lógica básica de decisão governamental que é a perspectiva da maximização de votos, afirma Downs (2013), em vez da maximização da utilidade social ou bem-estar. O governo, em seu modelo, busca maximizar o apoio político realizando atos que ganhem a maior quantidade de votos com o menor custo em termos de perda de votos. Os principais pontos sobre como os governos tomam decisões para maximizar seus resultados políticos são as operações marginais onde o governo realiza gastos que proporcionam o maior ganho de votos e financia esses gastos de maneiras que resultem na menor perda de votos. Isso é feito até o ponto em que o ganho marginal em votos iguale a perda marginal em votos. Além do princípio da maioria, onde as decisões do governo são baseadas em seguir o que a maioria dos eleitores prefere. Isso se deve ao fato de que, se o governo não seguir a

vontade da maioria em todas as questões, a oposição pode se posicionar ao lado da maioria em questões específicas, ganhando assim mais votos.

### 3 DEFINIÇÕES DE MÉTRICAS E PROCEDIMENTOS

Na sequência são explicitadas algumas métricas e procedimentos essenciais ao entendimento metodológico desta pesquisa.

#### 3.1 REDES NEURAI PROFUNDAS

Redes Neurais Profundas (Deep Neural Networks, DNNs) representam uma classe poderosa de modelos em aprendizado de máquina, especialmente notáveis por suas capacidades em tarefas complexas tanto para dados estruturados (tabulares) quanto para dados não estruturados (como imagens).

As redes neurais são inspiradas no sistema nervoso biológico, mais especificamente no funcionamento dos neurônios humanos. Uma rede neural é composta por uma ou mais camadas ocultas de neurônios entre as camadas de entrada de dados e a camada de saída. Cada neurônio das camadas ocultas realiza uma soma ponderada de suas entradas seguida pela aplicação de uma função de ativação não-linear. O que distingue uma rede neural profunda de outras arquiteturas é a presença de múltiplas camadas ocultas entre a entrada e a saída, permitindo a composição das entradas de cada camada de maneira não-linear.

##### 3.1.1 Principais componentes de uma rede neural

###### a) Entradas (*Inputs*)

A camada de entrada de uma rede neural constitui a primeira camada do modelo e é composta por neurônios que recebem diretamente os dados externos. Cada neurônio nesta camada está associado a uma entrada específica ( $x_1, x_2, \dots, x_n$ ), que representa uma característica ou atributo dos dados analisados. Essas entradas podem incluir uma ampla gama de tipos de dados, dependendo do contexto de aplicação da rede.

###### b) Somatório (*Net Input*)

O somatório ( $net = \sum_{i=0}^n w_i x_i$ ) representa a soma ponderada das entradas, onde cada entrada  $x_i$  é multiplicada por um peso  $w_i$  correspondente. Este somatório calcula a entrada líquida para o neurônio antes de ser transformada pela função de ativação. Os pesos  $w_i$  são parâmetros ajustáveis da rede que são aprendidos durante o processo de treinamento para minimizar o erro de saída da rede.

### 3.2 INICIALIZAÇÃO DE PESOS

Antes do treinamento ser iniciado, os pesos e bias da rede neural são inicializados. Geralmente, esses pesos são inicializados com pequenos valores aleatórios, o que ajuda a quebrar a simetria e permite que a aprendizagem ocorra. Métodos comuns de inicialização incluem a inicialização *Xavier* e *He*, que consideram o número de entradas e saídas da camada para calibrar a variância dos pesos.

### 3.3 PROPAGAÇÃO PARA FRENTE (*FORWARD PROPAGATION*)

Durante a propagação para frente, os dados de entrada são passados através da rede, camada por camada. Em cada neurônio, o seguinte cálculo é realizado, conforme visto nos principais componentes de uma rede neural:

- a) Cálculo da entrada líquida: A soma ponderada das entradas é calculada, adicionando-se o bias.
- b) Ativação: A entrada líquida é então passada por uma função de ativação (por exemplo, *ReLU*, *Sigmoid*), que introduz não-linearidades ao modelo, permitindo-lhe aprender padrões complexos.

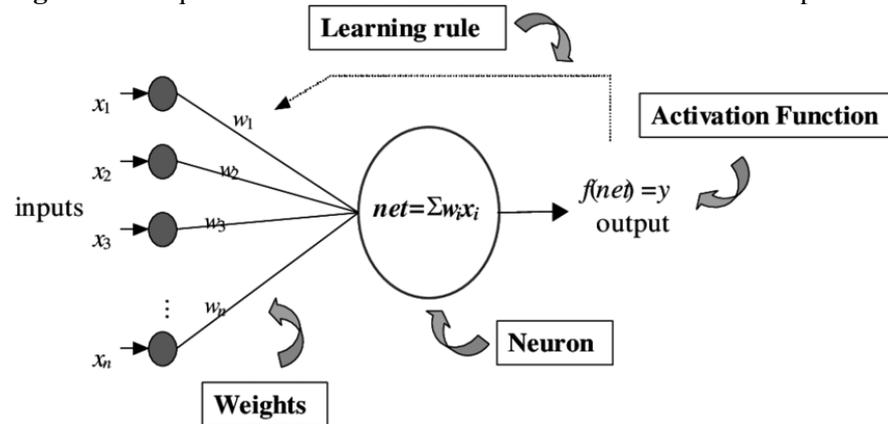
#### 3.3.1 Função de Ativação

Após calcular o somatório das entradas ponderadas, a rede aplica uma função de ativação ao resultado. Esta função de ativação ( $f(\text{net})$ ) pode ser linear ou, mais comumente, não-linear, permitindo que a rede capture relações complexas nos dados. Funções de ativação comuns incluem *ReLU* (Unidade Linear Retificada), *Sigmoid* e *Tanh*. A escolha da função de ativação depende da natureza do problema e da necessidade de não-linearidade no modelo.

#### 3.3.2 Regra de Aprendizado e *Backpropagation*

A regra de aprendizado governa como os pesos da rede são ajustados durante o treinamento. O método mais comum é a *backpropagation*, que envolve o ajuste dos pesos com base no gradiente da função de perda. O processo de *backpropagation* calcula o gradiente da função de perda com respeito a cada peso pelo método da cadeia, propagando o erro de saída de volta através da rede para atualizar os pesos. Isso ajuda a reduzir o erro entre a saída prevista e a saída real, refinando o modelo para fazer previsões mais precisas.

**Figura 1** - Arquitetura básica de uma rede neural com seus 4 componentes



**Fonte:** Jang; Yang; Kang, 2003.

Nas arquiteturas desenvolvidas no projeto, as redes são construídas com várias camadas densas (ou completamente conectadas), onde cada neurônio de uma camada está conectado a todos os neurônios na próxima camada. As funções de ativação não-lineares são um elemento fundamental das redes neurais, pois são capazes de permitir que a rede aprenda e generalize os dados em dimensões acima dos dados de entrada.

### 3.3.3 Cálculo da Função de Perda

Após a propagação para frente, a saída da rede é comparada com o valor real usando uma função de perda (também conhecida como função de custo). Esta função quantifica o erro entre a previsão da rede e o valor real. Funções de perda comuns incluem a entropia cruzada para classificação e o erro quadrático médio para regressão. Para o caso do problema abordado neste estudo, tendo em vista a existência de duas possíveis classificações para uma instância (reeleito ou não-reeleito), a função de perda utilizada nas redes neurais criadas foi a entropia cruzada binária.

### 3.3.4 Atualização dos Pesos

Com o gradiente da função de perda conhecido, os pesos são atualizados para reduzir o erro. Geralmente, isso é feito usando um método de otimização como o Gradiente Descendente ou suas variantes (*Adam*, *RMSprop*). A fórmula geral para a atualização é:

$$w = w - \eta \cdot \nabla_w L$$

onde  $\eta$  é a taxa de aprendizado e  $\nabla wL$  é o gradiente da função de perda em relação ao peso  $w$ .

### 3.3.5 Iteração

O processo de propagação para frente, cálculo da função de perda, retropropagação e atualização dos pesos é repetido para um número definido de iterações ou até que a função de perda seja minimizada ao ponto desejado. Este processo é comumente dividido em batches, que permite que a rede seja treinada com mais eficiência em grandes conjuntos de dados.

### 3.3.6 Avaliação e Ajuste

Após o treinamento, a rede é avaliada em um conjunto de dados de teste para verificar seu desempenho. Dependendo dos resultados, ajustes podem ser necessários, seja no design da rede, nos parâmetros de treinamento ou na escolha da função de ativação e de perda.

## 3.4 MÉTRICAS DE AVALIAÇÃO

Métricas de avaliação são essenciais para entender a eficácia de modelos de aprendizado de máquina, especialmente em tarefas de classificação. Cada métrica oferece uma perspectiva sobre aspectos diferentes do desempenho de um modelo.

### 3.4.1 Acurácia

Responsável por medir a proporção de previsões corretas (tanto positivas quanto negativas) em relação ao total de casos. É útil quando as classes estão equilibradas, mas deve ser utilizada com cautela em dados desbalanceados, como o caso dos dados de reeleição municipal, onde a tendência é a não reeleição.

$$\text{Acurácia} = (\text{Verdadeiros Positivos} + \text{Verdadeiros Negativos}) / (\text{Verdadeiros Positivos} + \text{Falsos Positivos} + \text{Verdadeiros Negativos} + \text{Falsos Negativos})$$

### 3.4.2 Precisão

Indica a proporção de previsões positivas corretas em relação ao total de previsões positivas feitas pelo modelo. Alta precisão significa que um modelo gera poucos falsos positivos, ideal para situações onde os falsos positivos têm consequências severas.

$$\text{Precisão} = \text{Verdadeiros Positivos} / (\text{Verdadeiros Positivos} + \text{Falsos Positivos})$$

### 3.4.3 Sensibilidade (*Recall*)

Referente à proporção de casos positivos reais que foram corretamente identificados pelo modelo. É uma métrica a ser considerada em contextos onde é crítico capturar todos os casos positivos, como no diagnóstico de doenças.

$$\text{Sensibilidade} = \text{Verdadeiros Positivos} / (\text{Verdadeiros Positivos} + \text{Falsos Negativos})$$

### 3.4.4 *F1-Score*

Combina precisão e recall em uma única métrica através de sua média harmônica. É útil quando você precisa de um equilíbrio entre precisão e recall, e é particularmente valioso em cenários com classes desequilibradas.

$$\text{F1 Score} = 2 \times (\text{Precisão} \times \text{Sensibilidade}) / (\text{Precisão} + \text{Sensibilidade})$$

### 3.4.5 *AUC-ROC* (Área sob a Curva da Característica de Operação do Receptor)

Esta métrica avalia a capacidade do modelo de discriminar entre as classes em todos os limiares de classificação. AUC é a área sob a curva ROC, que é um gráfico de taxa de verdadeiros positivos contra taxa de falsos positivos. Um valor de AUC próximo de 1 indica um modelo muito bom, enquanto um valor próximo de 0.5 sugere não haver capacidade discriminativa melhor que o acaso.

### 3.4.6 *Matriz de Confusão*

Fornecer uma visão detalhada do desempenho do modelo, mostrando os verdadeiros positivos, falsos positivos, verdadeiros negativos e falsos negativos. Esta ferramenta é extremamente útil para entender o comportamento do modelo em diferentes classes e ajustar o limiar de decisão conforme necessário.

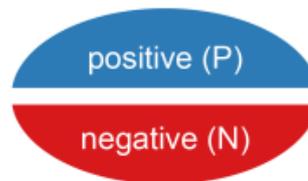
**Figura 2** - Elementos da matriz de confusão

		Valor Predito	
		Sim	Não
Real	Sim	Verdadeiro Positivo (TP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (TN)

**Fonte:** Disponível em: <https://diegonogare.net/wp-content/uploads/2020/04/matrizConfusao-600x381.png>

**Figura 3** - Classes de um problema de classificação binária

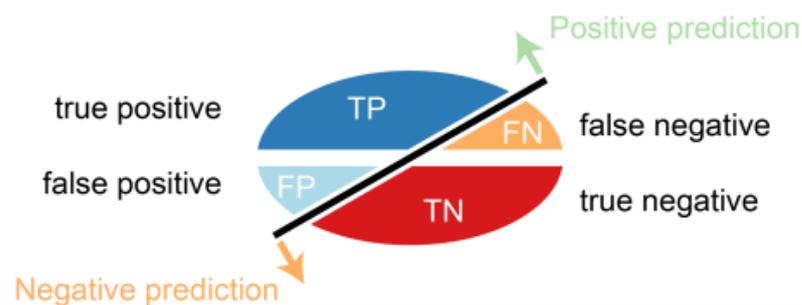
**Two actual classes or observed labels**



**Fonte:** Disponível em: <https://classeval.wordpress.com/wp-content/uploads/2015/06/observed-labels.png>

**Figura 4** - Classificação de um conjunto de dados de teste produz quatro resultados – verdadeiro positivo, falso positivo, verdadeiro negativo e falso negativo.

**Four outcomes of a classifier**

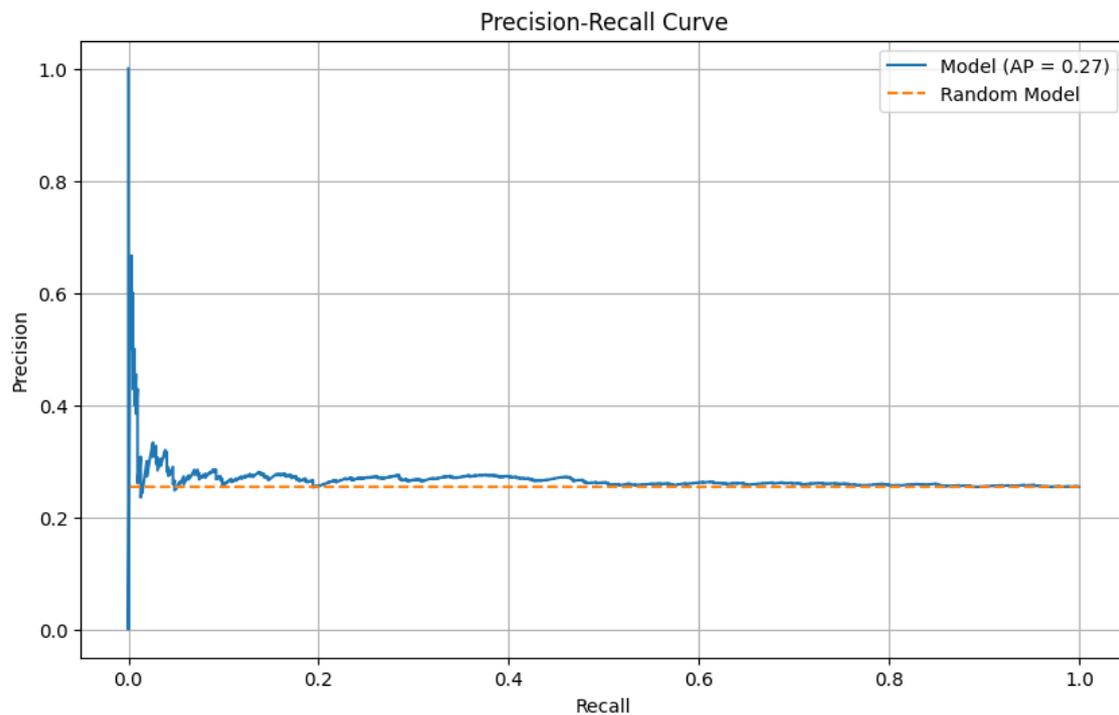


**Fonte:** Disponível em <https://classeval.wordpress.com/wp-content/uploads/2015/06/four-outcomes-of-classifier.png>

### 3.4.7 Curva Precision-Recall

A curva *Precision-Recall* (PR) é uma ferramenta essencial para a avaliação de modelos de classificação, especialmente em cenários de desequilíbrio de classes. Ela traça a relação entre a precisão (*precision*) e o *recall* (também conhecido como sensibilidade ou taxa de verdadeiros positivos) para diferentes *thresholds* de decisão. Precisão é definida como a proporção de verdadeiros positivos (TP) sobre o total de previsões positivas, incluindo verdadeiros e falsos positivos (TP + FP), enquanto *recall* é a proporção de TP sobre o total de casos positivos reais (TP + FN). A curva PR é particularmente informativa quando se lida com datasets onde a classe positiva é rara, pois destaca o trade-off entre a precisão e a capacidade de recuperação de exemplos positivos pelo modelo. A área sob a curva *Precision-Recall* (AUC-PR) é uma métrica crítica que quantifica a performance global do modelo; uma AUC-PR mais alta indica uma melhor capacidade do modelo em identificar a classe positiva. Diferentemente da curva ROC, que pode ser otimista em cenários com classes desbalanceadas devido à ênfase nos verdadeiros negativos, a curva PR é mais sensível à performance em prever a classe positiva, tornando-a mais adequada para tais contextos.

**Figura 5** - Exemplo de curva *precision-recall* obtida no trabalho



**Fonte:** elaborada pelo autor, 2024.

### 3.5 REGULARIZAÇÃO L2

A regularização L2 é uma técnica utilizada para prevenir o *overfitting* em modelos de aprendizado de máquina, especialmente em redes neurais profundas que possuem um grande número de parâmetros. Esta técnica funciona adicionando um termo de penalidade ao custo total do modelo que é proporcional ao quadrado dos valores dos pesos.

O impacto dessa penalidade é que ela encoraja a rede a manter os pesos pequenos, o que tende a resultar em modelos menos complexos e mais generalizáveis. Ao penalizar os pesos grandes, a regularização L2 efetivamente reduz a variância do modelo sem aumentar substancialmente o viés, tornando o modelo menos suscetível a aprender o ruído nos dados de treinamento como se fossem características verdadeiras.

### 3.6 DROPOUT

*Dropout* é uma técnica de regularização durante o treinamento de redes neurais que envolve desativar aleatoriamente uma proporção de neurônios em cada iteração do treinamento, ou seja, seus pesos são temporariamente reduzidos a zero. Esta desativação é aleatória e independente entre os neurônios, e ocorre com uma probabilidade  $p$  (por exemplo,  $p = 0.5$ ) a cada atualização do modelo durante o treinamento.

A ideia principal por trás do Dropout é que ao desativar aleatoriamente diferentes partes da rede neural durante o treinamento, o modelo não pode contar com a presença de nenhum sinal específico (ou qualquer configuração específica de sinais) para gerar previsões. Isso força a rede a aprender caminhos redundantes para representar os dados, aumentando sua robustez e diminuindo a chance de *overfitting*. A rede torna-se, portanto, menos sensível à perda específica de informações de qualquer um de seus neurônios, dado que qualquer neurônio pode ser "desligado" a qualquer momento durante o treinamento.

### 3.7 ENSEMBLE LEARNING

*Ensemble learning* é uma técnica utilizada em aprendizado de máquina que envolve a combinação de múltiplos modelos de aprendizado para obter uma previsão ou classificação mais robusta e precisa do que qualquer modelo individual poderia alcançar. Este método se baseia no princípio de que diferentes modelos capturam diferentes padrões e erros nos dados, e que a agregação dessas previsões pode minimizar a variância, o viés e reduzir o risco de

*overfitting*. *Ensemble learning* é particularmente poderoso em contextos onde os dados são complexos e os modelos simples não conseguem capturar toda a variabilidade, fornecendo soluções que são mais generalizáveis e resilientes a ruídos nos dados.

### 3.8 DESBALANCEAMENTO DE CLASSES

O problema de desbalanceamento de classes ocorre quando as classes em um conjunto de dados de classificação não são representadas de forma equitativa. Isso significa que uma ou mais classes possuem significativamente mais exemplos do que outras. Este desbalanceamento pode levar a problemas no treinamento de modelos de aprendizado de máquina, pois o modelo pode se tornar tendencioso em favor das classes majoritárias, ignorando ou classificando incorretamente as classes minoritárias. Outros problemas que podem ocorrer durante a etapa de treinamento e validação é a obtenção de métricas que não refletem de maneira fiel a capacidade de previsão do modelo.

### 3.9 *RANDOM UNDERSAMPLER*

O *Random Undersampler* é uma técnica usada para lidar com o desbalanceamento de classes ao reduzir o número de exemplos da classe majoritária. Em vez de aumentar o número de exemplos da classe minoritária, o *undersampling* reduz o número de exemplos da classe majoritária (selecionando aleatoriamente uma amostra de exemplos da classe majoritária) para igualar o número de exemplos na classe minoritária.

## 4 PROCEDIMENTOS METODOLÓGICOS

Os procedimentos metodológicos aplicados a esta pesquisa se referem à natureza, abordagem, objetivo, procedimento, reunião, apresentação e interpretação dos dados da mesma.

De natureza básica e abordagem quantitativa, esta pesquisa caracteriza-se como exploratória, pois, segundo Gil (2002), proporciona o aprimoramento de ideias e uma nova visão da questão proposta, permitindo flexibilidade no seu planejamento.

### 4.1 COLETA DOS DADOS

As duas principais características elencadas como desejadas nas fontes a serem utilizadas para o estudo foram:

- a) Captação dos dados da fonte demandar um esforço viável e de acordo com o escopo do projeto;
- b) quantidade satisfatória de dados disponíveis com relação às granularidades pretendidas, sendo dados municipais e estaduais.

Assim, as bases utilizadas foram o Ipeadata<sup>1</sup>, base do Instituto de Pesquisa Econômica Aplicada (Ipea) e a base do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira<sup>2</sup>, visando coletar os dados referentes ao desempenho dos estudantes nas edições do Exame Nacional do Ensino Médio (ENEM).

#### 4.1.1 Instituto de Pesquisa Econômica Aplicada (Ipea)

O Ipeadata possui bases de dados Macroeconômicos, Regionais e Sociais. Uma série de dados está associada apenas a uma dessas bases de dados e é representada por seus metadados e valores ao longo do tempo e pode estar associada a um território nacional ou a um país. Assim como no site do Ipeadata, cada série também pode estar associada a um tema. As seguintes entidades estão disponíveis para consulta através de uma API (*Application Programming Interface*): Metadados, Valores, Países, Territórios e Temas. As informações contidas na base são agrupadas em três principais categorias:

---

<sup>1</sup> <http://www.ipeadata.gov.br/Default.aspx>

<sup>2</sup> [www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem](http://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem)

- a) **Macroeconômica:** Base de dados econômicos e financeiros mantida pelo Ipea incluindo séries estatísticas da economia brasileira e dos aspectos que lhe são mais pertinentes na economia internacional. Os dados são atualizados e documentados de forma sistemática e apresentados na mesma unidade monetária. Recursos disponíveis permitem a manipulação matemática e a extração dos resultados em planilhas ou gráficos.
- b) **Regional:** Base de dados demográficos, econômicos e geográficos para as regiões, estados e municípios brasileiros que se iniciam no Censo Demográfico de 1970. A busca das séries estatísticas pode ser feita por *Palavras-chave* na sua descrição, *Temas* ou assuntos de interesse, *Fontes* de publicação ou pelo *Nível geográfico* para o qual é possível agregar os dados. Os níveis geográficos disponíveis no Ipeadata são município, área metropolitana, micro e mesorregião geográfica, estado, região administrativa (como Amazônia Legal, Sudene, Fome Zero, entre outras), bacia e sub-bacia hidrográfica, grandes regiões.
- c) **Social:** Base de dados e indicadores sociais abrangendo temas diversos, como nível de renda per capita, desigualdade na distribuição de renda dos indivíduos e domicílios, desempenho educacional, condições de saúde e habitação, inserção no mercado de trabalho, situação dos direitos humanos da população, entre outros. Sempre que possível, apresentam-se séries históricas de periodicidade mensal, anual e decenal para o país, grandes regiões, estados, regiões metropolitanas, desagregadas segundo a localização rural ou urbana dos domicílios, cortes etários, níveis de desempenho educacional, anos de escolaridade, sexo e cor dos indivíduos.

Conforme consta na definição da estruturação da base, o Ipea possui uma API que permite a realização da consulta dos dados armazenados através de requisições HTTP, em que cada requisição é realizada a partir de um endereço URL específico da base de dados, permitindo, em diferentes casos, a passagem de parâmetros de pesquisa para a requisição. Os principais endereços URL que a API disponibiliza e que foram considerados relevantes para o contexto do projeto, bem como as informações obtidas por meio de requisições para cada endereço, são apresentados no quadro 1.

**Quadro 1** - Consultas disponíveis através da API do Ipea

URL (à partir da URL base <a href="http://www.ipeadata.gov.br/api/odata4/">http://www.ipeadata.gov.br/api/odata4/</a> )	Descrição

Metadados	Registros de metadados de todas as séries disponíveis para consulta.
Metadados('{SERCODIGO}')	Metadados de apenas uma série, de código igual a {SERCODIGO}.
Metadados('{SERCODIGO}')/Valores ou ValoresSerie(SERCODIGO='{SERCODIGO}')	Todos os valores de uma série numérica.
Metadados('{SERCODIGO}')/ValoresStr ou ValoresStrSerie(SERCODIGO='{SERCODIGO}')	Todos os valores de uma série alfanumérica.
Territorios	Registros de todos os territórios cadastrados.
Territorios(TERCODIGO='{TERCODIGO}',NIVNOME='{NIVNOME}')	Um único registro de país, de código igual a {TERCODIGO} e nível territorial igual a {NIVNOME}. Os níveis territoriais possíveis são: <ul style="list-style-type: none"> <li>• Brasil</li> <li>• Regiões</li> <li>• Estados</li> <li>• Microrregiões</li> <li>• Mesorregiões</li> <li>• Municípios</li> <li>• Municípios por bacia <ul style="list-style-type: none"> <li>• Área metropolitana</li> </ul> </li> <li>• Estado/RM</li> <li>• AMC 20-00</li> <li>• AMC 40-00</li> <li>• AMC 60-00</li> </ul>

	<ul style="list-style-type: none"> <li>• AMC 1872-00</li> <li>• AMC 91-00</li> <li>• AMC 70-00</li> <li>• Outros Países</li> </ul>
--	--

**Fonte:** elaborado pelo autor, 2024.

Utilizando a API, o primeiro passo do processo foi a busca de todos os agrupamentos de dados existentes na base, doravante denominados *séries*. Uma série, no contexto da base do Ipea, é o agrupamento de dados de mesma natureza, registrados através de séries históricas em diferentes níveis territoriais, e identificados através de um código alfanumérico. Por meio do processo de busca inicial, foi constatada a existência de mais de 1.300 séries, onde todos os códigos existentes foram armazenados.

Após o processo inicial de busca inicial dos códigos das séries existentes, o próximo passo realizado foi obter os metadados associados a cada uma das séries. Os metadados são diferentes informações sobre os detalhes de uma série. Alguns dos metadados existentes, cuja descrição é mais relevante, são:

- SERCODIGO: Código único de identificação da série.
- SERCOMENTARIO: Comentários relativos à série, em português.
- SERSTATUS: Indica se uma série macroeconômica ainda é atualizada. Valores: ‘A’ (Ativa) para séries atualizadas ou ‘I’ (Inativa) para séries que não são atualizadas.
- BASNOME: Nome da base de dados da série.
- FNTSIGLA: Sigla ou nome abreviado da fonte da série, em português.

- FNTURL: URL para o site da fonte da série.
- PERNOME: Nome da periodicidade, em português.
- UNINOME: Nome da unidade dos valores da série.

Com os metadados coletados, foi realizada uma análise sobre os comentários e status de manutenção e atualização das séries, identificando por meio dos seus códigos aquelas cuja utilização no projeto poderia ser viável. Com base nos seus códigos, houve um processo de busca sistematizada das informações contidas na fonte de dados do Ipea, onde os dados coletados foram tratados e inseridos em uma base de dados criada especialmente para o projeto. Cada uma das séries existentes na base é formada pelos seguintes campos:

- NIVNOME: Nome do nível territorial (Brasil, Regiões, Estados, Municípios etc.).
- SERCODIGO: Código único de identificação da série.
- TERCODIGO: Código identificador de território, do IBGE, para territórios dos níveis Brasil, Região, Estado, Mesorregião, Microrregião e Município. Para os demais territórios possui o mesmo valor de UFMUNDV.
- VALDATA: Data do valor.
- VALVALOR: Valor para série na data e no território.

Para cada uma das séries buscadas, foi realizada uma análise exploratória dos valores dos dados existentes em níveis territoriais de municípios e estados, alvo da utilização no projeto. O processo de busca dos dados, análise dos níveis territoriais existentes e da viabilidade de uso no projeto foi repetido durante todo o processo entre novembro de 2023 e março de 2024.

Ao final da análise dos dados, as séries escolhidas para compor o projeto foram (consultar apêndice para a definição completa de cada):

- DESPCUTM: Despesa de custeio - municipal
- DESPCORRM: Despesas correntes - municipal
- RECTRIBM: Receita corrente - receita bruta - impostos, taxas e contribuições de melhoria - municipal
- REKM: Receita de capital - receita bruta - municipal
- RIPTUM: Receita corrente - receita bruta - impostos, taxas e contribuições de melhoria - IPTU - municipal
- RTRCORTOM: Receita corrente - receita bruta - transferências correntes - municipal
- DESPKM: Despesa de capital - municipal
- DESPORM: Despesa orçamentária - municipal
- DFEDUCM: Despesa por função - educação e cultura - empenhada - municipal
- DFHABM: Despesa por função - urbanismo e habitação - empenhada - municipal

- RTRKTOM: Receita de capital - receita bruta - transferências de capital - municipal
- TACIDT: Taxa de vítimas de acidentes de trânsito a óbito (100.000 Habitantes)
- THOMIC: Taxa de homicídios (100.000 Habitantes)
- FAMILIAS\_BOLSA\_FAMILIA: Programa Bolsa Família (PBF) - número de benefícios em dezembro
- VALORES\_BOLSA\_FAMILIA: Programa Bolsa Família (PBF) - valor total dos benefícios em dezembro
- PIB: Produto Interno Bruto (PIB) municipal

#### 4.1.1.1 PIB municipal

Para a utilização de informações sobre o PIB dos municípios, foi necessária a aplicação de um cálculo baseado em diferentes séries existentes na base. Isso se deve ao fato de, durante a análise das séries, ser constatada ausência dos dados em alguns anos.

Para a realização do cálculo, foram utilizadas as seguintes séries:

- PIB\_IBGE\_5938\_37: PIB Municipal - preços de mercado (preços de 2010)
- IMPPIB: PIB Municipal - impostos, líquidos de subsídios, sobre produtos (preços de 2010)
- PIBAG: PIB Municipal (valor adicionado a preços básicos) - agropecuária (preços de 2010)
- PIBG: PIB Municipal (valor adicionado a preços básicos) - serviços - administração, defesa, educação e saúde públicas e seguridade social (preços de 2010)
- PIBI: PIB Municipal (valor adicionado a preços básicos) - indústria (preços de 2010)
- PIBSE: PIB Municipal (valor adicionado a preços básicos) - serviços, exclusive administração, defesa, educação e saúde públicas e seguridade social (preços de 2010).

Pelo fato dos dados da série PIB\_IBGE\_5938\_37 de cada município não estarem completos para todos os anos do intervalo de 2000 a 2020, foi necessário a utilização de um cálculo com base em diferentes séries para o preenchimento dos dados faltantes. A expressão apresentada a seguir é verdadeira para todos os anos em que os dados da série PIB\_IBGE\_5938\_37 não estão presentes:

$$\text{PIB\_IBGE\_5938\_37} = \text{IMPPIB} + \text{PIBAG} + \text{PIBG} + \text{PIBI} + \text{PIBSE}$$

#### 4.1.1.2 Variáveis das séries eleitorais

Para a utilização das variáveis eleitorais contidas na base, o primeiro passo foi a análise da base com relação às séries existentes nesse contexto. Para o âmbito de eleições municipais, foram identificadas as seguintes séries relevantes:

- VDEMC: Total de vereadores eleitos do DEM
- VNOVOC: Total de vereadores eleitos do NOVO
- VPANC: Total de vereadores eleitos do PAN
- VPCBC: Total de vereadores eleitos do PCB
- VPCDOBC: Total de vereadores eleitos do PCDOB
- VPCOC: Total de vereadores eleitos do PCO
- VPDSC: Total de vereadores eleitos do PDS
- VPDTC: Total de vereadores eleitos do PDT
- VPENC: Total de vereadores eleitos do PEN
- VPFLC: Total de vereadores eleitos do PFL
- VPGTC: Total de vereadores eleitos do PGT
- VPHSC: Total de vereadores eleitos do PHS
- VPLC: Total de vereadores eleitos do PL
- VPMBC: Total de vereadores eleitos do PMB
- VPMDBC: Total de vereadores eleitos do PMDB
- VPMNC: Total de vereadores eleitos do PMN
- VPPBC: Total de vereadores eleitos do PPB
- VPPC: Total de vereadores eleitos do PP
- VPPLC: Total de vereadores eleitos do PPL
- VPPSC: Total de vereadores eleitos do PPS
- VPRBC: Total de vereadores eleitos do PRB
- VPRC: Total de vereadores eleitos do PR
- VPRNC: Total de vereadores eleitos do PRN
- VPRONAC: Total de vereadores eleitos do PRONA
- VPROSC: Total de vereadores eleitos do PROS
- VPRPC: Total de vereadores eleitos do PRP
- VPRTBC: Total de vereadores eleitos do PRTB
- VPSBC: Total de vereadores eleitos do PSB

- VPSCC: Total de vereadores eleitos do PSC
- VPSDBC: Total de vereadores eleitos do PSDB
- VPSDC: Total de vereadores eleitos do PSD
- VPSDCC: Total de vereadores eleitos do PSDC
- VPSLC: Total de vereadores eleitos do PSL
- VPSNC: Total de vereadores eleitos do PSN
- VPSOLC: Total de vereadores eleitos do PSOL
- VPSTC: Total de vereadores eleitos do PST
- VPSTUC: Total de vereadores eleitos do PSTU
- VPTBC: Total de vereadores eleitos do PTB
- VPTC: Total de vereadores eleitos do PT
- VPTCC: Total de vereadores eleitos do PTC
- VPTDOBC: Total de vereadores eleitos do PTDOB
- VPTNC: Total de vereadores eleitos do PTN
- VPVC: Total de vereadores eleitos do PV
- VREDEC: Total de vereadores eleitos do REDE
- VSOLIDARIEDADEC: Total de vereadores eleitos do SOLIDARIEDADE
- FDEM: Votação do DEM - prefeito no primeiro turno
- FDEM2: Votação do DEM - prefeito no segundo turno
- FNOVO: Votação do NOVO - prefeito no primeiro turno
- FNOVO2: Votação do NOVO - prefeito no segundo turno
- FNULO: Votos nulos - prefeito no primeiro turno
- FNULO2: Votos nulos - prefeito no segundo turno
- FPAN: Votação do PAN - prefeito no primeiro turno - INATIVA
- FPCB: Votação do PCB - prefeito no primeiro turno
- FPCDOB: Votação do PCdoB - prefeito no primeiro turno
- FPCDOB2: Votação do PC do B - prefeito no segundo turno
- FPCO: Votação do PCO - prefeito no primeiro turno
- FPDS: Votação do PDS - prefeito no primeiro turno - INATIVA
- FPDT: Votação do PDT - prefeito no primeiro turno
- FPDT2: Votação do PDT - prefeito no segundo turno
- FPEN: Votação do PATRI - prefeito no primeiro turno
- FPFL: Votação do DEM - prefeito no primeiro turno - INATIVA
- FPFL2: Votação do DEM - prefeito no segundo turno - INATIVA

- FPGT: Votação do PGT - prefeito no primeiro turno - INATIVA
- FPHS: Votação do PODE - prefeito no primeiro turno
- FPL: Votação do PL - prefeito no primeiro turno - INATIVA
- FPL2: Votação do PL - prefeito no segundo turno - INATIVA
- FPMB: Votação do PMB - prefeito no primeiro turno
- FPMB2: Votação do PMB - prefeito no segundo turno
- FPMDB: Votação do MDB - prefeito no primeiro turno
- FPMDB2: Votação do MDB - prefeito no segundo turno
- FPMN: Votação do PMN - prefeito no primeiro turno
- FPP: Votação do PP - prefeito no primeiro turno
- FPP2: Votação do PP - prefeito no segundo turno
- FPPB: Votação do PPB - prefeito no primeiro turno - INATIVA
- FPPB2: Votação do PPB - prefeito no segundo turno - INATIVA
- FPPL: Votação do PCdoB - prefeito no primeiro turno
- FPPS: Votação do Cidadania - prefeito no primeiro turno
- FPPS2: Votação do Cidadania - prefeito no segundo turno
- FPR: Votação do PL - prefeito no primeiro turno - INATIVA
- FPR2: Votação do PL - prefeito no segundo turno
- FPRB: Votação do PRB - prefeito no primeiro turno
- FPRN: Votação do PRN - prefeito no primeiro turno - INATIVA
- FPRONA: Votação do PRONA - prefeito no primeiro turno - INATIVA
- FPROS: Votação do PROS - prefeito no primeiro turno
- FPROS2: Votação do PROS - prefeito no segundo turno
- FPRP: Votação do PATRI - prefeito no primeiro turno
- FPRTB: Votação do PRTB - prefeito no primeiro turno
- FPSB: Votação do PSB - prefeito no primeiro turno
- FPSB2: Votação do PSB - prefeito no segundo turno
- FPSC: Votação do PSC - prefeito no primeiro turno
- FPSC2: Votação do PSC - prefeito no segundo turno
- FPSD: Votação do PSD - prefeito no primeiro turno
- FPSD2: Votação do PSD - prefeito no segundo turno
- FPSDB: Votação do PSDB - prefeito no primeiro turno
- FPSDB2: Votação do PSDB - prefeito no segundo turno
- FPSDC: Votação do DC - prefeito no primeiro turno

- FPSDC2: Votação do DC - prefeito no segundo turno - INATIVA
- FPSL: Votação do PSL - prefeito no primeiro turno
- FPSL2: Votação do PSL - prefeito no segundo turno - INATIVA
- FPSN: Votação do PSN - prefeito no primeiro turno - INATIVA
- FPSOL: Votação do PSOL - prefeito no primeiro turno
- FPST: Votação do PST - prefeito no primeiro turno - INATIVA
- FPSTU: Votação do PSTU - prefeito no primeiro turno
- FPT: Votação do PT - prefeito no primeiro turno
- FPT2: Votação do PT - prefeito no segundo turno
- FPTB: Votação do PTB - prefeito no primeiro turno
- FPTB2: Votação do PTB - prefeito no segundo turno
- FPTC: Votação do PTC - prefeito no primeiro turno
- FPTDOB: Votação do Avante - prefeito no primeiro turno
- FPTN: Votação do PTN - prefeito no primeiro turno
- FPV: Votação do PV - prefeito no primeiro turno
- FPV2: Votação do PV - prefeito no segundo turno
- FREDE: Votação do REDE - prefeito no primeiro turno
- FREDE2: Votação do REDE - prefeito no segundo turno
- FSD: Votação do SD - prefeito no primeiro turno
- FUP: Votação do UP - prefeito no primeiro turno

#### **4.2.2 Variáveis das séries educacionais**

A segunda fonte escolhida para fazer parte do estudo foi a base dos dados referentes ao desempenho dos estudantes nas edições do ENEM. A base realiza a separação dos dados por ano de realização do exame. Para cada ano de realização do exame, uma planilha reúne dados de todos os participantes através de variáveis, sendo cada participante identificado por um número de inscrição único. Para cada ano, há também uma planilha de metadados das variáveis, provendo informações destinadas à manipulação das informações contidas nos dados dos estudantes.

Para cada participante, as principais variáveis existentes, e que foram utilizadas no estudo, foram:

- NU\_INSCRICAO: Número de inscrição
- TP\_ST\_CONCLUSAO: Situação de conclusão do Ensino Médio

- NU\_INSCRICAO: Número de inscrição
- TP\_ST\_CONCLUSAO: Situação de conclusão do Ensino Médio
- SG\_UF\_ESC: Sigla da Unidade da Federação da escola
- CO\_MUNICIPIO\_ESC: Código do município da escola
- TP\_PRESENCA\_CN: Presença na prova objetiva de Ciências da Natureza
- TP\_PRESENCA\_CH: Presença na prova objetiva de Ciências Humanas
- TP\_PRESENCA\_LC: Presença na prova objetiva de Linguagens e Códigos
- TP\_PRESENCA\_MT: Presença na prova objetiva de Matemática
- NU\_NOTA\_CN: Nota da prova de Ciências da Natureza
- NU\_NOTA\_CH: Nota da prova de Ciências Humanas
- NU\_NOTA\_LC: Nota da prova de Linguagens e Códigos
- NU\_NOTA\_MT: Nota da prova de Matemática
- TP\_STATUS\_REDACAO: Situação da redação do participante
- NU\_NOTA\_REDACAO: Nota da prova de redação

O código do município da escola de cada participante (CO\_MUNICIPIO\_ESC) é detalhado em dígitos que representam região, Unidade da Federação (UF) e município. O padrão de formação do código é adotado por diferentes bases, como a base de dados do Ipea. Assim, é possível identificar e agregar todos os dados de um mesmo município através do mesmo código, permitindo a coerência dos dados. O padrão de formação dos códigos é dado por:

- 1º dígito: Região
- 1º e 2º dígitos: UF
- 3º, 4º, 5º e 6º dígitos: Município
- 7º dígito: dígito verificador

O processo de coleta dos dados teve início em janeiro de 2024 e se estendeu até maio de 2024. Para cada ano de realização do exame, o primeiro passo realizado foi a extração da planilha contendo os dados de todos os estudantes que participaram da edição do exame, bem como o arquivo de metadados relativos às variáveis da planilha.

## 4.2 TRATAMENTO DOS DADOS

Para o tratamento dos dados foi escrito um *script* com o intuito de detectar e corrigir inconsistências nos dados como valores faltantes, valores absurdos ou valores desformatados.

Assim, variáveis derivadas são utilizadas para que o modelo de IA reconheça padrões mais facilmente.

#### 4.2.1 Tratamento das séries socioeconômicas

A partir das séries escolhidas, foram criadas variáveis derivadas com relação ao PIB de cada município. As variáveis a seguir representam a razão entre o valor da série (especificada anteriormente pelo seu código) em uma faixa de anos com relação à média do PIB calculado no mesmo intervalo de anos (e.g. média dos valores de RECTRIBM entre os anos de 2016 a 2020 sobre a média dos valores do PIB entre os anos de 2016 a 2020).

As variáveis derivadas obtidas a partir do processo foram:

- DESPCORRM\_OVER\_PIB
- DESPCUTM\_OVER\_PIB
- DESPKM\_OVER\_PIB
- DESPORM\_OVER\_PIB
- DFEDUCM\_OVER\_PIB
- DFHABM\_OVER\_PIB
- RECTRIBM\_OVER\_PIB
- REKM\_OVER\_PIB
- RIPTUM\_OVER\_PIB
- RTRCORTOM\_OVER\_PIB
- RTRKTOM\_OVER\_PIB

Para cada uma das variáveis acima, foi calculada uma nova variável derivada, cujo propósito é indicar a razão entre o valor da variável derivada original no intervalo de anos e o valor da variável no intervalo de anos imediatamente anterior, indicando o aumento ou diminuição relativa do valor entre os dois intervalos de anos consecutivos. O processo foi aplicado tanto às variáveis diretamente relacionadas às séries originais da base de dados quanto às variáveis derivadas descritas anteriormente. As variáveis obtidas foram:

- DESPCORRM\_OVER\_PIB\_RATIO
- DESPCORRM\_RATIO
- DESPCUTM\_OVER\_PIB\_RATIO
- DESPCUTM\_RATIO
- DESPKM\_OVER\_PIB\_RATIO
- DESPKM\_RATIO

- DESPORM\_OVER\_PIB\_RATIO
- DESPORM\_RATIO
- DFEDUCM\_OVER\_PIB\_RATIO
- DFEDUCM\_RATIO
- DFHABM\_OVER\_PIB\_RATIO
- DFHABM\_RATIO
- RECTRIBM\_OVER\_PIB\_RATIO
- RECTRIBM\_RATIO
- REKM\_OVER\_PIB\_RATIO
- REKM\_RATIO
- RIPTUM\_OVER\_PIB\_RATIO
- RIPTUM\_RATIO
- RTRCORTOM\_OVER\_PIB\_RATIO
- RTRCORTOM\_RATIO
- RTRKTOM\_OVER\_PIB\_RATIO
- RTRKTOM\_RATIO
- FAMILIAS\_BOLSA\_FAMILIA\_RATIO
- VALORES\_BOLSA\_FAMILIA\_RATIO
- PIB\_RATIO
- THOMIC\_RATIO
- TACIDT\_RATIO

#### 4.2.2 Tratamento das séries eleitorais

Com as séries identificadas, o primeiro passo foi a criação de uma variável derivada indicando a reeleição de um partido em um município. Para o contexto do estudo, e para fins de simplificação dado o escopo, foi adotado o seguinte critério de reeleição:

- A reeleição é considerada quando o mesmo partido assume a prefeitura em um período de quatro anos subsequente a um período em que esteve à frente do comando municipal. Com base no critério de reeleição, foi criada a variável REELEITO, cujas características

são:

- REELEITO: Variável indicando se o partido do prefeito eleito é o mesmo (REELEITO = 1) ou distinto (REELEITO = 0) com relação ao mandato anterior

Assim, para cada período de quatro anos de mandato, é anexada a variável indicando se o conjunto de dados representa os dados de uma situação de reeleição. No contexto de ferramentas de IA, dada a hipótese inicial do estudo, a coluna REELEITO é denominada coluna alvo, necessária para a utilização de diferentes abordagens de aprendizado de máquina. Com base na coluna alvo, diferentes algoritmos podem realizar inferências sobre os dados, identificando de diferentes formas os padrões existentes nos dados.

Após o processo da identificação de partidos reeleitos, foi identificada a necessidade da criação de uma variável derivada indicando a composição da câmara de vereadores do município no mandato, indicando a porcentagem da câmara de vereadores pertencente ao mesmo partido do prefeito. Com isso, foi criada a variável derivada PROPORCAO\_CAMARA, cuja definição é:

- PROPORCAO\_CAMARA: Razão entre o número de vereadores eleitos do mesmo partido do prefeito eleito e o total de vereadores que compõem a câmara do município

Ao final do processo de tratamento e agrupamento dos dados relativos às eleições municipais, foi obtida uma planilha contendo as seguintes informações:

- TERCODIGO: Código do território
- INICIO\_INTERVALO: Data do início do intervalo do mandato, no formato 01/01/AAAA
- FIM\_INTERVALO: Data do fim do intervalo do mandato, no formato 31/12/AAAA
- PROPORCAO\_CAMARA: Razão entre o número de vereadores eleitos do mesmo partido do prefeito eleito e o total de vereadores que compõem a câmara do município
- REELEITO: Variável indicando se o partido do prefeito eleito é o mesmo ou distinto com relação ao mandato anterior

#### **4.2.3 Tratamento das séries educacionais**

Após a extração dos dados dos estudantes, foi realizado o tratamento dos dados, mantendo apenas os dados de estudantes que estiveram presentes em todas as provas objetivas e que cuja redação tenha sido aceita e tido uma nota atribuída. Os dados válidos foram organizados por UF através da variável SG\_UF\_ESC. Com isso, cada estado passou a possuir uma planilha própria de todos os estudantes que prestaram o exame no ano.

Com os dados dos estudantes organizados por estado, o próximo passo realizado foi a organização dos estudantes por meio da variável CO\_MUNICIPIO\_ESC, onde foi realizado o

cálculo da média nota de cada município, tanto para as provas objetivas quanto para as redações.

Foi realizado também o levantamento dos dados da situação escolar informada pelos estudantes, obtendo os dados de evasão escolar do ensino médio com base nas informações do ENEM. Assim, realizado o cálculo e a organização, o resultado obtido foi uma nova planilha contendo as seguintes informações por município:

- CO\_MUNICIPIO\_ESC: Código do município da escola
- NU\_NOTA\_OBJETIVA: Média das notas das provas objetivas
- NU\_NOTA\_REDACAO: Média das notas das redações
- CONCLUSAO: Número de estudantes que concluíram o ensino médio
- EVASAO: Número de estudantes cuja situação escolar é de evasão
- NAO\_EVASAO: Número de estudantes cuja situação escolar é regular

Ao fim, os dados municipais de todos os estados foram agregados, formando uma única planilha, contendo as mesmas colunas acima, mas resumindo a edição do ENEM em nível municipal. Essa planilha passou a ser utilizada para a realização dos agrupamentos em diferentes intervalos de anos.

#### 4.3 AGRUPAMENTO DOS DADOS PARA TREINAMENTO

Após a coleta, tratamento, organização, cálculo e criação das variáveis necessárias para a utilização de algoritmos relacionados à IA, houve a necessidade de agrupar todos os dados em uma única planilha, cujo propósito é servir como entrada para as diferentes abordagens adotadas utilizando IA.

Cada registro (linha da planilha) possui a seguinte definição:

“Resumo de todas variáveis citadas ao longo do texto com relação ao período do mandato municipal entre as datas INICIO\_INTERVALO e FIM\_INTERVALO.”

Ao final, a relação completa de variáveis compondo cada um dos registros final da tabela é:

TERCODIGO, INICIO\_INTERVALO, FIM\_INTERVALO, CONCLUSAO,  
 CONCLUSAO\_RATIO, DESPCORRM, DESPCORRM\_OVER\_PIB,  
 DESPCORRM\_OVER\_PIB\_RATIO, DESPCORRM\_RATIO, DESPCUTM,  
 DESPCUTM\_OVER\_PIB, DESPCUTM\_OVER\_PIB\_RATIO, DESPCUTM\_RATIO,  
 DESPKM, DESPKM\_OVER\_PIB, DESPKM\_OVER\_PIB\_RATIO,  
 DESPKM\_RATIO, DESPORM, DESPORM\_OVER\_PIB,  
 DESPORM\_OVER\_PIB\_RATIO, DESPORM\_RATIO, DFEDUCM,

DFEDUCM\_OVER\_PIB, DFEDUCM\_OVER\_PIB\_RATIO, DFEDUCM\_RATIO,  
 DFHABM, DFHABM\_OVER\_PIB, DFHABM\_OVER\_PIB\_RATIO,  
 DFHABM\_RATIO, EVASAO, EVASAO\_RATIO, FAMILIAS\_BOLSA\_FAMILIA,  
 FAMILIAS\_BOLSA\_FAMILIA\_RATIO, NAO\_EVASAO,  
 NAO\_EVASAO\_RATIO, NU\_NOTA\_OBJETIVA,  
 NU\_NOTA\_OBJETIVA\_RATIO, NU\_NOTA\_REDACAO,  
 NU\_NOTA\_REDACAO\_RATIO, PIB, PIB\_RATIO, POPULACAO,  
 POPULACAO\_RATIO, PROPORCAO\_CAMARA, RECTRIBM,  
 RECTRIBM\_OVER\_PIB, RECTRIBM\_OVER\_PIB\_RATIO, RECTRIBM\_RATIO,  
 REKM, REKM\_OVER\_PIB, REKM\_OVER\_PIB\_RATIO, REKM\_RATIO,  
 RIPTUM, RIPTUM\_OVER\_PIB, RIPTUM\_OVER\_PIB\_RATIO, RIPTUM\_RATIO,  
 RTRCORTOM, RTRCORTOM\_OVER\_PIB, RTRCORTOM\_OVER\_PIB\_RATIO,  
 RTRCORTOM\_RATIO, RTRKTOM, RTRKTOM\_OVER\_PIB,  
 RTRKTOM\_OVER\_PIB\_RATIO, RTRKTOM\_RATIO, TACIDT, TACIDT\_RATIO,  
 THOMIC, THOMIC\_RATIO, VALORES\_BOLSA\_FAMILIA,  
 VALORES\_BOLSA\_FAMILIA\_RATIO, REELEITO

#### 4.4 ORGANIZAÇÃO PARA TREINAMENTO DA IA

O conjunto de dados inicial foi dividido em subconjuntos de treinamento, validação e teste. O conjunto completo de dados extraídos do processo de obtenção de dados possui um total de 16.163 entradas. Desse conjunto, 12.143 entradas são referentes a mandatos até 2020, que podem ser utilizados na análise, pois são os últimos dados com possibilidade de validação. Cada entrada corresponde a um mandato em um município em uma eleição entre os anos de 2000 e 2020 que possuíam todos os atributos presentes e consistentes nos dados coletados.

Os dados referentes a mandatos até 2020 foram divididos em 3 conjuntos: de treino, teste e validação para a utilização. A proporção de divisão utilizada foi:

- Treino: 60% dos dados
- Validação: 20% dos dados
- Teste: 20% dos dados

Para o conjunto de dados de treinamento, o desbalanceamento identificado foi de aproximadamente 1:3, onde para cada dado com a coluna alvo REELEITO = 1, existem 3 dados com a coluna alvo REELEITO = 0.

Tendo isso em vista, as melhores possibilidades de resultado são observadas em amostras balanceadas. No treinamento de cada rede neural foi utilizado um subconjunto balanceado das amostras existentes. Para isso, foi utilizada a técnica do Random Undersampling, onde o treinamento de cada rede neural, foi utilizada uma subamostra dos dados que ajuste a proporção para 1:1. A adoção dessa técnica teve como objetivo o de minimizar a distorção gerada pelos desbalanceamentos durante a etapa de treinamento.

#### 4.4.1 Treinamento

Para o treinamento das redes, devido às limitações de *hardware*, as redes criadas tiveram suas características de número de neurônios por camada e número de camadas reduzidas para que o treinamento fosse possível.

Após o teste com a criação de dezenas de arquiteturas, a que apresentou os melhores resultados durante a etapa de treinamento foi uma rede neural profunda, que recebe 65 entradas, processa ao longo de 6 camadas, de 10240, 6144, 4096, 2048, 1024 e 1 neurônio, respectivamente. Com todos os neurônios de cada camada conectados a todos os neurônios da camada subsequente. A última camada, possui um único neurônio que retorna valores entre 0 e 1, indicando a confiança da rede na resposta, valores mais próximos de 1, indicam maior certeza da rede em relação a reeleição e valores mais próximos de 0 indicam maior certeza na não reeleição.

Para fins de comparação com o conjunto de dados de teste. Valores acima de 0,5 são considerados previsões de reeleição e valores abaixo como não reeleição.

A partir da melhor rede neural encontrada, foi replicada arquitetura para formar um ensemble composto por 3 modelos da mesma rede neural treinados de maneira independente, tendo como dados de treinamento subconjuntos distintos de amostras.

Os parâmetros utilizados para a criação da melhor configuração encontrada de redes foram:

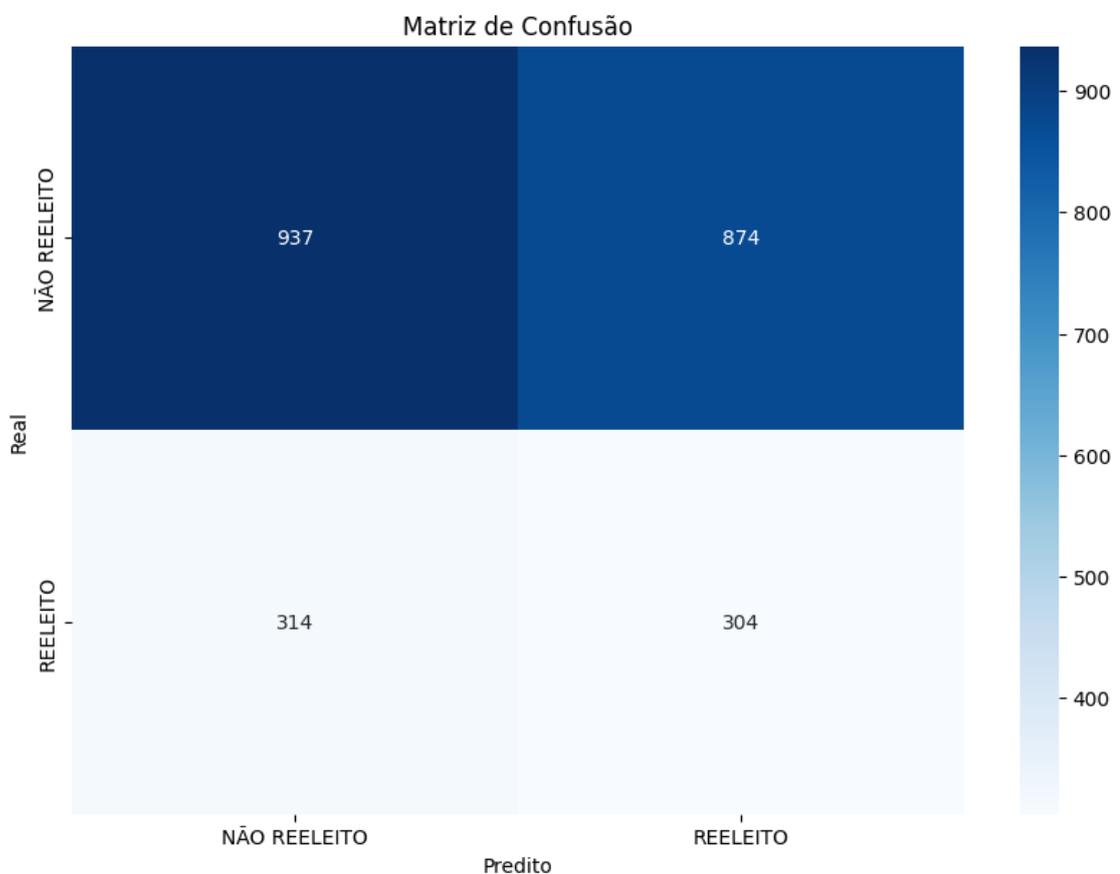
- Neurônios: 2048
- *Dropout\_rate*: 0.25
- Épocas: 250

## 5 RESULTADOS

Resultados das métricas de treinamento de um dos modelos de rede neural:

- Acurácia: 0.5109098394400988
- Precisão: 0.25806451612903225
- Recall: 0.4919093851132686
- F1-Score: 0.33853006681514475
- AUC-ROC: 0.5046515451242765
- Matriz de Confusão:
  - $\begin{bmatrix} 937 & 874 \end{bmatrix}$
  - $\begin{bmatrix} 314 & 304 \end{bmatrix}$

**Figura 6** – Primeira Matriz de confusão associada ao treinamento do modelo individual

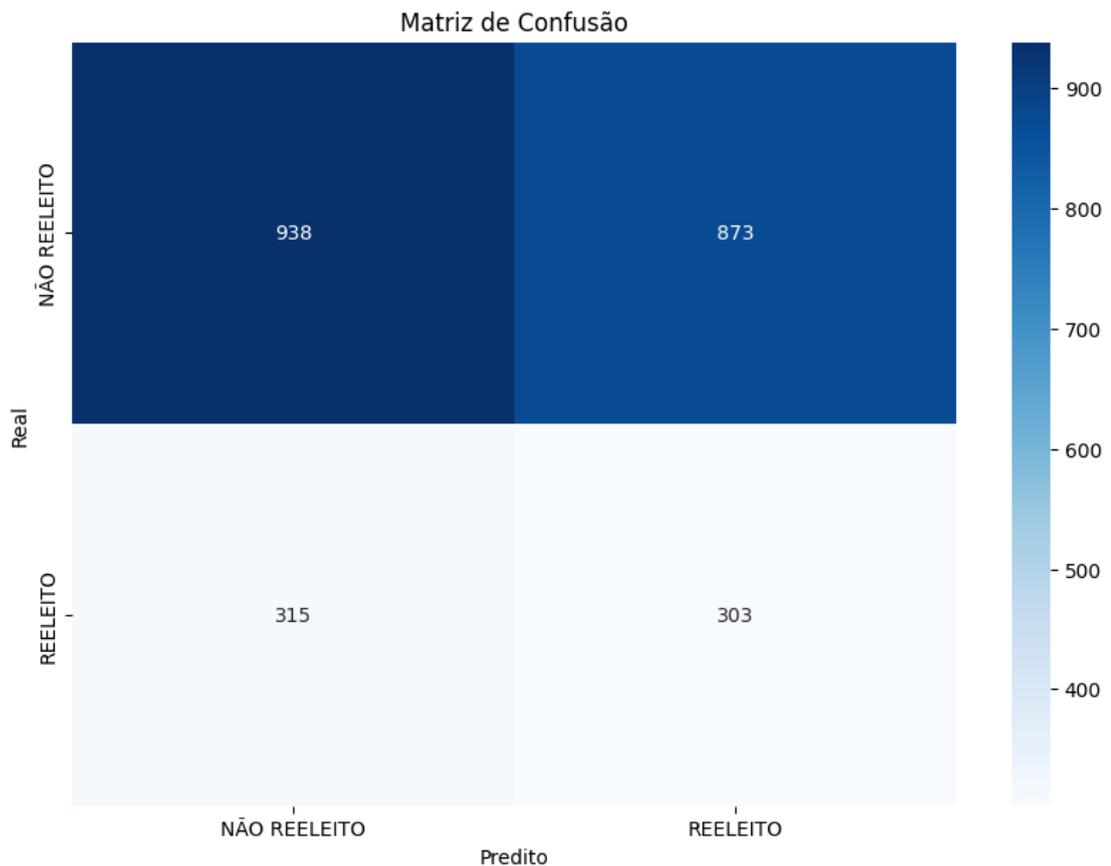


**Fonte:** elaborada pelo autor, 2024.

Para o treinamento do ensemble, foi utilizada a mesma abordagem da criação de subconjuntos dos dados de teste. Após o treinamento dos modelos que passariam a compor o ensemble, as métricas obtidas foram:

- Acurácia: 0.5109098394400988
- Precisão: 0.2576530612244898
- Recall: 0.49029126213592233
- F1-Score: 0.33779264214046817
- AUC-ROC: 0.5041185741933063
- Matriz de Confusão:
  - $\begin{bmatrix} 938 & 873 \end{bmatrix}$
  - $\begin{bmatrix} 315 & 303 \end{bmatrix}$

**Figura 7** – Segunda Matriz de confusão associada ao treinamento do modelo individual



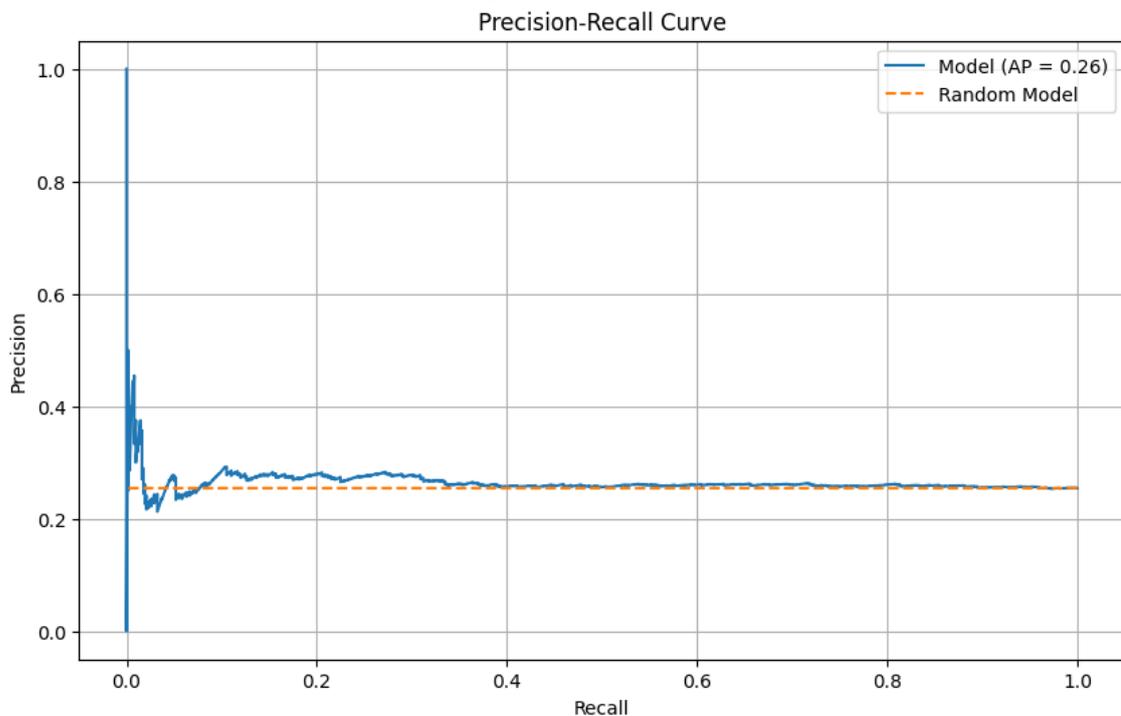
**Fonte:** elaborada pelo autor, 2024.

Os valores das métricas apresentadas na seção 4 do treinamento das redes, tanto para um modelo individual quanto para os modelos que compunham o ensemble de predições, indicam que mesmo uma rede neural com diversas camadas e utilizando diferentes subconjuntos dos dados de treinamento não foi capaz de captar de maneira satisfatória as relações existentes entre os dados a fim de ser utilizada como um modelo preditivo.

O seguinte gráfico apresenta a curva *Precision-Recall* para o ensemble dos modelos criados. Com base no desbalanceamento 1:3 indicado anteriormente, um modelo que gera previsões ao acaso possui uma taxa de acertos fixa em 25% (linha tracejada amarela).

Conforme o Recall tende a 1, indicando que todos os casos positivos verdadeiros foram capturados, o modelo tende a aproximar a Precisão com o resultado de uma previsão aleatória. Para valores de Recall entre 0,15 até 0,4, o modelo apresenta uma leve superioridade nas Precisão entre 0,25 e 0,30, indicando que a sua performance quanto à previsão é um pouco melhor que a de uma escolha aleatória.

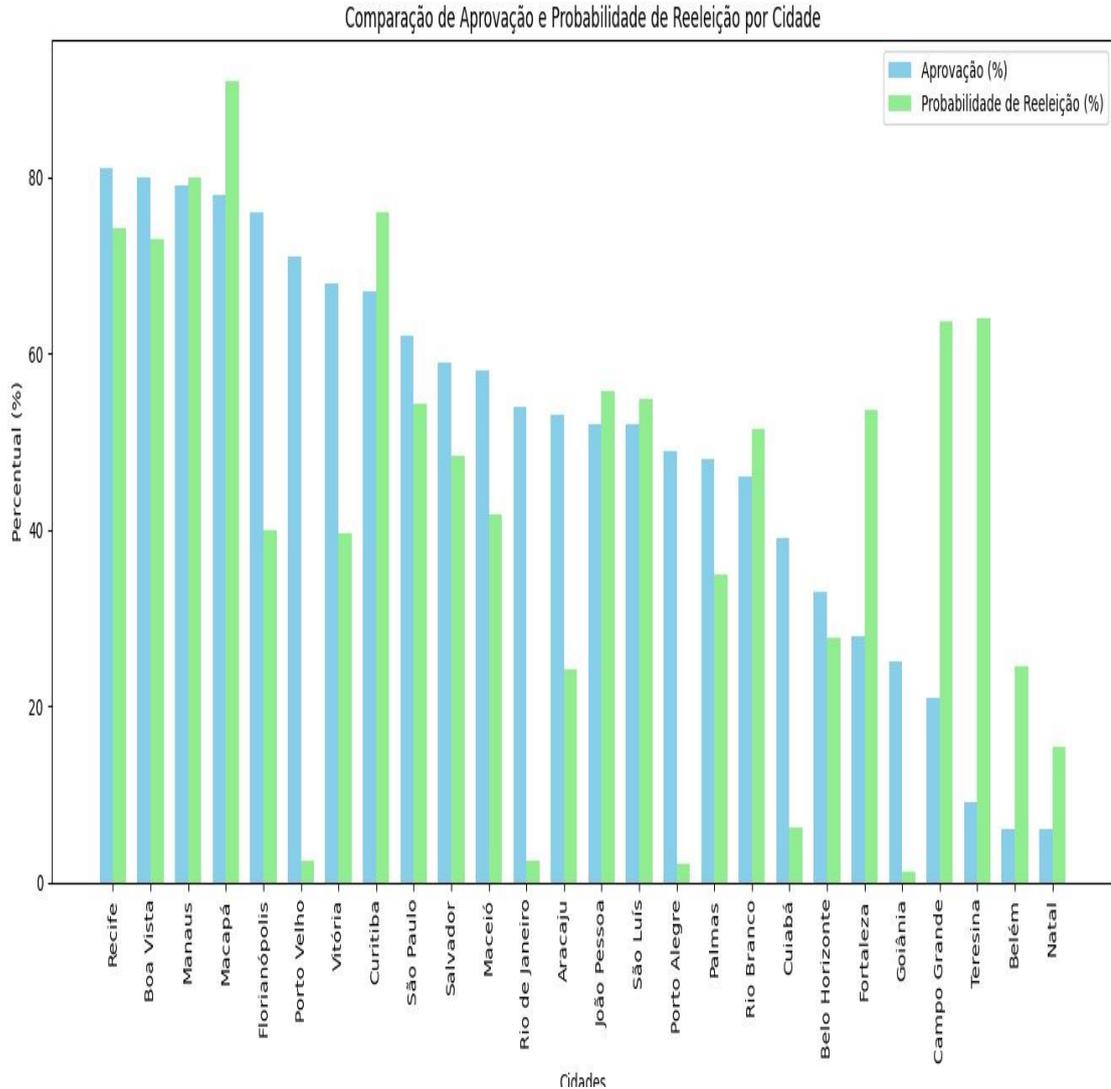
**Figura 8 – Precision-Recall Curve**



**Fonte:** elaborada pelo autor, 2024.

Com base nos algoritmos e no que foi construído pela presente pesquisa foi realizada uma estimativa da probabilidade de reeleição nas capitais brasileiras comparada com a aprovação dos prefeitos e pode-se verificar que a análise algorítmica dos dados revela o gráfico a seguir, numa comparação com a aprovação dos atuais prefeitos.

**Gráfico 1** – Comprovação de Aprovação em dezembro de 2023 e Probabilidade de Reeleição por Cidade



**Fonte:** elaborado pelo autor, 2024.

Legenda: O eixo horizontal representa as capitais dos estados brasileiros, são elas: Recife, Boa Vista, Manaus, Macapá, Florianópolis, Porto Velho, Vitória, Curitiba, São Paulo, Salvador, Maceió, Rio de Janeiro, Aracaju, João Pessoa, São Luís, Porto Alegre, Palmas, Rio Branco, Cuiabá, Belo Horizonte, Fortaleza, Goiânia, Campo Grande, Teresina, Belém e Natal. O eixo vertical representa o percentual (%), com valores que variam de 0 a 100%. As barras azuis representam a aprovação (%) do prefeito em cada cidade e as barras verdes representam a probabilidade de reeleição (%) em cada cidade segundo o algoritmo de IA.

As informações sobre a aprovação dos prefeitos foram retiradas de uma pesquisa da empresa Atlas Intel publicada em 31 de dezembro de 2023<sup>3</sup>. A pesquisa foi realizada entre os dias 18 e 31 de dezembro em todo território nacional e coletou mais de 14 mil questionários.

<sup>3</sup> <https://x.com/atlaspolitico/status/1741563235357069713>

Em algumas cidades, a aprovação e a probabilidade de reeleição são muito próximas, como em Recife, Boa Vista e Manaus. No entanto, em outras cidades, como Porto Velho e Rio Branco, há uma diferença significativa entre aprovação e probabilidade de reeleição. Cidades como Boa Vista e Manaus mostram altos valores tanto para aprovação quanto para a probabilidade de reeleição, sugerindo uma correlação positiva entre esses dois indicadores. Em cidades como Natal e Macapá, tanto a aprovação quanto a probabilidade de reeleição são relativamente baixas, indicando possíveis desafios para a reeleição nesses locais. Algumas cidades, como Florianópolis, Porto Velho e Porto Alegre mostram uma alta aprovação, mas uma probabilidade de reeleição consideravelmente menor, sugerindo que os dados produzidos no período do atual mandato e analisados pela IA denotam decréscimos e não sustentam a reeleição desses prefeitos. Além de cidades como Campo Grande e Teresina, onde a aprovação é baixa, mas a probabilidade de reeleição segundo a IA é alta mostrando que os dados revelam um desempenho maior nos dados socioeconômicos que a aprovação desses prefeitos.

O gráfico permite uma comparação direta entre a percepção pública de líderes (medida pela aprovação) e a viabilidade de sua reeleição medida pela IA a partir dos dados eleitos nesse trabalho. Em geral, a expectativa seria que uma alta aprovação correspondesse a uma alta probabilidade de reeleição, mas as discrepâncias observadas sugerem que os dados revelam em alguns casos probabilidades muito diferentes das aprovações desses prefeitos.

## 6 CONSIDERAÇÕES FINAIS

A análise dos modelos de redes neurais para a previsão de reeleições municipais possibilitou o estudo e sobre a aplicabilidade e eficácia de técnicas avançadas de aprendizado de máquina em dados socioeconômicos. Os modelos evoluíram significativamente ao longo do processo, desde a implementação de redes neurais básicas até modelos mais complexos com técnicas de *ensemble* e *boosting*.

Cada modelo sucessivo incorporou melhorias em relação ao anterior, seja na complexidade da arquitetura, no uso de técnicas de regularização como *Dropout* e *L2*, ou na otimização dos hiperparâmetros. Isso resultou em melhorias graduais nas métricas de desempenho, como acurácia, precisão, recall, *F1-Score* e *AUC-ROC*. A introdução de regularização *L2* e *Dropout* ajudou significativamente a reduzir o *overfitting*, permitindo que os modelos generalizassem melhor para dados não vistos. Essas técnicas são cruciais para manter o equilíbrio entre a capacidade de aprendizado e a generalização em modelos complexos.

As métricas revelaram que, embora a acurácia tenha melhorado consistentemente, outras métricas como precisão e recall variaram mais, refletindo os desafios em equilibrar a sensibilidade e especificidade dos modelos. O *F1-Score* e o *AUC-ROC* foram particularmente úteis para avaliar o equilíbrio entre essas métricas.

Os modelos de *ensemble* demonstraram a eficácia de combinar diferentes modelos para melhorar a robustez e a precisão das previsões. Eles mostraram melhorias notáveis em termos de *recall* e *AUC-ROC*, indicando uma capacidade aprimorada de identificar corretamente os casos de reeleição.

A coleta de dados aconteceu de maneira bastante satisfatória e relevante conforme o objetivo específico, mais de 3.000 séries de dados foram incorporadas à base de dados utilizada para a pesquisa, o que representou um conjunto muito significativo, entretanto havia muitas inconsistências nessas séries e o tempo não permitiu tratar todas elas de maneira satisfatória para utilização no treinamento dos algoritmos de IA, assim a opção foi pelas séries que estão descritas na metodologia.

Com base nos resultados obtidos durante os treinamentos dos modelos, ficou evidente a necessidade da utilização de novas e diferentes fontes de dados socioeconômicos para que o objetivo da proposta do trabalho possa ter resultados mais efetivos.

A capacidade preditiva dos modelos treinados inspira a necessidade de mais acúmulo de competência com dados e com os próprios algoritmos. Conforme pode-se observar no gráfico que relaciona a aprovação dos prefeitos com a probabilidade de sua reeleição calculada pelo modelo preditivo para 2024. Na maioria das capitais se vê que o que foi conquistado por esse processo de IA se aproxima da aprovação, mas noutros está muito distante. Certamente apenas a eleição em si poderá revelar a eficiência dessa previsão, mas a comparação com a aprovação dos prefeitos revela algum tipo de necessidade de aperfeiçoamento.

Com a utilização dos dados captados, as arquiteturas de redes neurais criadas realizaram de maneira um pouco acima da escolha aleatória a previsão de resultados de reeleições em capitais, não podendo imputar aos modelos gerados o rótulo de previsores confiáveis dos resultados nas eleições. Além disso, características inerentes à forma como a política ocorre demandam mais que a utilização apenas de fatores econômicos para a previsão de resultados.

Com isso, é de bom tom classificar que a pesquisa e a utilização mais a fundo de ferramentas de inteligência artificial baseadas em dados socioeconômicos pode gerar resultados que podem ser incorporados a outros métodos de pesquisas sobre reeleições municipais que abordam diferentes aspectos, fornecendo novas visões que agreguem à avaliação final.

O estudo destacou a importância de continuar explorando e refinando técnicas de aprendizado de máquina para questões políticas e sociais. Sugere-se como estudos futuros explorar a integração de mais fontes de dados e a aplicação de técnicas de aprendizado ainda mais avançadas, como redes neurais profundas adaptativas e métodos de aprendizado federado para abordar preocupações de privacidade e descentralização de dados.

O estudo indica ainda que outras dimensões podem ser analisadas para atingir melhores previsões. Os dados socioeconômicos revelam o resultado das decisões políticas vividas no passado, o que permite um moderado padrão de assertividade. No entanto, análises sobre o presente como o conteúdo jornalístico e manifestações nas redes sociais podem ajudar significativamente para afinar esse processo preditivo. Somam-se aí as pesquisas de opinião e pesquisas eleitorais, além das Análises de Sentimento sobre a comunicação política presente e passada.

O desenvolvimento de ferramenta preditiva muito assertiva exigiria a incorporação desses elementos e isso sugere a necessidade de estudos nessas áreas, que seriam a IA aplicada ao reconhecimento dos elementos midiáticos relativos ao cenário político que se pretende prever, bem como a IA aplicada no reconhecimento das manifestações pessoais nas redes sociais. Somado aos estudos dos dados e das manifestações presentes, se mostra imperativo o

estudo da IA aplicada às pesquisas de opinião e pesquisas eleitorais, além de pesquisas de Análise de Sentimentos aplicadas à comunicação política.

## REFERÊNCIAS

- AXELROD, R. **The Evolution of Cooperation**. New York: Basic Books, 1984.
- BECKER, G. S. **The Economic Approach to Human Behavior**. Chicago: University of Chicago Press, 1976.
- CAMPOS, R. S. Desmistificando a inteligência artificial: uma breve introdução conceitual ao aprendizado de máquina. **International Journal of Phenomenology, Hermeneutics and Metaphysics**, Toledo, v. 3, n. 1, p. 106-123, 2020. DOI: <https://doi.org/10.48075/aoristo.v3i1.24880> Disponível em: <https://saber.unioeste.br/index.php/aoristo/article/view/24880> Acesso em: 06 fev. 2022.
- DONEDA, D. C. M. *et al.*  
Considerações iniciais sobre inteligência artificial, ética e autonomia pessoal. **Pensar-Revista de Ciências Jurídicas**, Fortaleza, v. 23, n. 4, p. 1-17, out./dez. 2018. DOI: <http://dx.doi.org/10.5020/2317-2150.2018.8257>. Acesso em: 06 fev. 2022.
- DOWNS, A. **Uma teoria econômica da Democracia**. São Paulo: Editora da Universidade de São Paulo, 2013.
- FERNEDA, E. Redes neurais e sua aplicação em sistemas de recuperação de informação. **Ciência da Informação**, Brasília, v. 35, n. 1, p. 25-30, jan./abr. 2006. DOI: <http://dx.doi.org/10.1590/S0100-19652006000100003> Acesso em: 08 jun. 2022.
- FIGUEIREDO, M. **A decisão do voto: democracia e racionalidade**. São Paulo: DOXA-IESP, 2022.
- GÉRON, A. **Mãos à obra aprendizado de máquina com Scikit-Learn & TensorFlow: conceitos, ferramentas e técnicas para construções inteligentes**. Rio de Janeiro: Alta Books, 2019. Disponível em: <https://altabooks.com.br/produto/maos-a-obra-aprendizado-de-maquina-com-scikit-learn-.keras-tensorflow/> . Acesso em: 07 abr. 2024.
- GIL, A. C. **Como elaborar projetos de pesquisa**. São Paulo: Atlas, 2002.
- JORDAN, M. I. Artificial Intelligence: the revolution hasn't happened yet. **Harvard Data Science Review**, Boston, v. 1, n. 1, p. 1-8, jul. 2019. DOI: <http://dx.doi.org/10.1162/99608f92.f06c6e61> Acesso em: 10 fev. 2022.
- KUHN, T. S. **A estrutura das revoluções científicas**. Tradução: Beatriz Vianna Boeira e Nelson Boeira. 12. ed. São Paulo: Perspectiva, 2013. Disponível em: <https://ppec.ufms.br/files/2020/10/A-estrutura-das-revolu%C3%A7%C3%B5es-cient%C3%ADficas-Kuhn.pdf>. Acesso em: 03 de abr. 2024.
- LARROYED, A. Redefining Patent Translation: The Influence of ChatGPT and the Urgency to Align Patent Language Regimes in Europe with Progress in Translation Technology. **GRUR International**, Munich, v. 72, n. 11, p. 1009–1017, 2023. DOI: <https://doi.org/10.1093/grurint/ikad099> Acesso em: 06 nov. 2023.

LUDERMIR, T. B. Inteligência artificial e aprendizado de máquina: estado atual e tendências. **Estudos Avançados**, São Paulo, v. 35, n. 101, p. 85-94, 2021. DOI: <http://dx.doi.org/10.1590/s0103-4014.2021.35101.007> . Acesso em: 07 jun. 2022.

LUGER, G. F. **Inteligência Artificial**. 6. ed. São Paulo: Pearson Education do Brasil, 2013.

NORTH, D. C. **Institutions, Institutional Change, and Economic Performance**. Cambridge: Cambridge University Press, 1990.

ONODY, R. N. **Teste de Turing e Inteligência Artificial**. Instituto de Física de São Carlos-USP, São Paulo, p. 1-3, 2021. Disponível em: [https://repositorio.usp.br/directbitstream/8a6402dc-7cc6-4f66-aae6-93865fbfe05c/PROD032068\\_3044270.pdf](https://repositorio.usp.br/directbitstream/8a6402dc-7cc6-4f66-aae6-93865fbfe05c/PROD032068_3044270.pdf) . Acesso em: 03 abr. 2024.

OSTROM, E. **Governing the Commons: The Evolution of Institutions for Collective Action**. Cambridge: Cambridge University Press, 1990.

RUSSELL, S. J.; NORVIG, P. **Artificial Intelligence: a modern approach**. New Jersey: Pearson Education, 2009. Disponível em: [https://people.engr.tamu.edu/guni/csce421/files/AI\\_Russell\\_Norvig.pdf](https://people.engr.tamu.edu/guni/csce421/files/AI_Russell_Norvig.pdf) . Acesso em: 07 abr. 2024.

SHEPSLE, K. A. **Analizar la política: Comportamiento, instituciones y racionalidad**. México: Centro de Investigación y Docencia Economicas, 2016.

SILVA, R.I V. T. e. **Teste de Turing e sua alusão a consciência de máquina**. 2021. 46 f. Trabalho de Conclusão de Curso (Bacharel em Engenharia da Computação) - Escola de Negócios, Tecnologia e Inovação, Centro Universitário do Estado do Pará (ARGO), Belém, 2021. Disponível em: <http://repositorio.cesupa.br:8080/jspui/bitstream/prefix/222/1/TC%20-%20Rafael%20Veiga.pdf> . Acesso em: 03 abr. 2024.

TURING, A. Computing machine and intelligence. **Mind a quarterly review of psychology and philosophy**, v. 59 n. 236, p. 433-460, out. 1950. DOI: <https://doi.org/10.1093/mind/LIX.236.433> . Acesso em: 03 abr. 2024.

## APÊNDICE A

Código de criação da melhor rede neural utilizada para realizar as predições.

```
def create_model(dropout_rate, neurons, input_shape):
    initializer = HeNormal()
    model = Sequential([
        Dense(5*neurons,          activation='relu',          kernel_initializer=initializer,
kernel_regularizer=l1(0.01), input_shape=(input_shape,)),
        BatchNormalization(),
        Dropout(dropout_rate),
        Dense(3*neurons,          activation='elu',          kernel_initializer=initializer,
kernel_regularizer=l1(0.01)),
        BatchNormalization(),
        Dropout(dropout_rate),
        Dense(2*neurons,          activation='tanh',        kernel_initializer=initializer,
kernel_regularizer=l2(0.02)),
        BatchNormalization(),
        Dropout(dropout_rate),
        Dense(neurons,            activation='sigmoid',    kernel_initializer=initializer,
kernel_regularizer=l1(0.005)),
        BatchNormalization(),
        Dropout(dropout_rate),
        Dense(neurons // 2,       activation='relu',          kernel_initializer=initializer,
kernel_regularizer=l2(0.01)),
        BatchNormalization(),
        Dropout(dropout_rate),
        Dense(1, activation='sigmoid')
    ])
    model.compile(optimizer=Adam(learning_rate=0.00005), loss='binary_crossentropy',
metrics=[Accuracy(), Precision(), Recall(), AUC()])
    return model
```

## APÊNDICE B

### Definição das séries

#### DESPCUTM: Despesa de custeio - municipal

Destina-se ao registro do total das despesas com pessoal, encargos, serviços de terceiros e outros custeios necessários à operação e manutenção dos serviços públicos estaduais anteriormente criados e instalados, inclusive aquelas destinadas a obras de conservação, adaptação e manutenção do seu patrimônio. Nota: a modalidade de aplicação é uma informação gerencial que tem por finalidade indicar se os recursos são aplicados diretamente por órgãos ou entidades no âmbito da mesma esfera de Governo ou por outro ente da Federação e suas respectivas entidades. O universo de municípios da tabela é definido pelo IBGE no levantamento censitário e não necessariamente coincide com aquele utilizado pelo STN ou oficialmente existente ou instalado na data de referência. Mais informações: Manual de Contabilidade Aplicado ao Setor Público, 8ª edição, Glossário do Tesouro Nacional e Manual de Demonstrativos Financeiros, 9ª edição.

Atualização: Anual

Unidade: R\$

#### DESPCORRM: Despesas correntes - municipal

Destina-se ao registro do valor de todas as operações destinadas à manutenção e funcionamento de serviços públicos, bem como as relacionadas com obras de conservação, adaptação e manutenção de bens móveis e imóveis, tais como pagamento de pessoal, aquisição de material de consumo, pagamento de serviços prestados por terceiros, operação de escolas e de centros de saúde, dentre outras. Este item está desdobrado em despesas de custeio, transferências correntes e outras despesas correntes. O universo de municípios da tabela é definido pelo IBGE no levantamento censitário e não necessariamente coincide com aquele utilizado pelo STN ou oficialmente existente ou instalado na data de referência. Para os anos de 2002 a 2004 a diferença entre receita orçamentária e o somatório da receita corrente e receita de capital se deve às deduções relativas ao Fundef. Nota: a modalidade de aplicação é uma informação gerencial que tem por finalidade indicar se os recursos são aplicados diretamente por órgãos ou entidades no âmbito da mesma esfera de Governo ou por outro ente da Federação e suas respectivas entidades. Mais informações: Manual de Contabilidade Aplicado ao Setor Público, 8ª edição, Glossário do Tesouro Nacional e Manual de Demonstrativos Financeiros, 9ª edição.

Atualização: Anual

Unidade: R\$

**RECTRIBM:** Receita corrente - receita bruta - impostos, taxas e contribuições de melhoria - municipal

Engloba os tributos de competência do município: impostos, taxas e contribuição de melhoria.

Nota: o universo de municípios da tabela é definido pelo IBGE no levantamento censitário e não necessariamente coincide com aquele utilizado pelo STN ou oficialmente existente ou instalado na data de referência. Para os anos de 2002 a 2004 a diferença entre receita orçamentária e o somatório da receita corrente e receita de capital se deve às deduções relativas ao Fundef. Assim, receita orçamentária = receita corrente (-) deduções receita corrente (+) receita de capital. Mais informações: Manual de Contabilidade Aplicado ao Setor Público, 8ª edição, Glossário do Tesouro Nacional e Manual de Demonstrativos Financeiros, 9ª edição.

Atualização: Anual

Unidade: R\$

**REKM:** Receita de capital - receita bruta - municipal

As receitas de capital são aquelas provenientes tanto da realização de recursos financeiros oriundos da constituição de dívidas e da conversão, em espécie, de bens e direitos, quanto de recursos recebidos de outras pessoas de direito público ou privado e destinados a atender despesas classificáveis em despesas de capital. Ao contrário da receita corrente, a receita de capital não provoca efeito sobre o patrimônio líquido. Considera-se a receita bruta, sem nenhum tipo de deduções. Nota: para os anos de 2002 a 2004 a diferença entre receita orçamentária e o somatório da receita corrente e receita de capital se deve às deduções relativas ao Fundef. Assim, receita orçamentária = receita corrente (-) deduções receita corrente (+) receita de capital. O universo de municípios da tabela é definido pelo IBGE no levantamento censitário e não necessariamente coincide com aquele utilizado pelo STN ou oficialmente existente ou instalado na data de referência. Mais informações: Manual de Contabilidade Aplicado ao Setor Público, 8ª edição, Glossário do Tesouro Nacional e Manual de Demonstrativos Financeiros, 9ª edição.

Atualização: Anual

Unidade: R\$

**RIPTUM:** Receita corrente - receita bruta - impostos, taxas e contribuições de melhoria - IPTU - municipal

As receitas correntes são aquelas provenientes de tributos, de contribuições, da exploração do patrimônio estatal, da exploração de atividades econômicas, de recursos financeiros recebidos com intuito de atender às despesas correntes e por aquelas receitas não mencionadas acima e que não se enquadram como receitas de capital. Neste caso, trata-se da subconta Impostos, Taxas e Contribuições de Melhoria, que registra as receitas originadas de impostos (fato gerador não depende de qualquer atividade estatal), taxas (fato gerador depende da utilização de serviços públicos) e contribuições de melhoria (arrecadada dos proprietários de imóveis beneficiados por obras públicas). Mais especificamente, versa sobre a arrecadação municipal com o Imposto Predial e Territorial Urbano, que tem como fato gerador a propriedade, o domínio útil ou a posse de bem imóvel por natureza ou por acessão física, como definido na lei civil, localizado na zona urbana do Município. Nota: para os anos de 2002 a 2004 a diferença entre receita orçamentária e o somatório da receita corrente e receita de capital se deve às deduções relativas ao Fundef. Assim, receita orçamentária = receita corrente (-) deduções receita corrente (+) receita de capital. O universo de municípios da tabela é definido pelo IBGE no levantamento censitário e não necessariamente coincide com aquele utilizado pelo STN, ou oficialmente existente ou instalado na data de referência. Mais informações: Manual de Contabilidade Aplicado ao Setor Público, 8ª edição; Glossário do Tesouro Nacional, Manual de Demonstrativos Financeiros, 9ª edição e Lei 5.172 de 1996.

Atualização: Anual

Unidade: R\$

**RTRCORTOM:** Receita corrente - receita bruta - transferências correntes - municipal

As receitas correntes são aquelas provenientes de tributos, de contribuições, da exploração do patrimônio estatal, da exploração de atividades econômicas, de recursos financeiros recebidos com intuito de atender às despesas correntes e por aquelas receitas não mencionadas acima e que não se enquadram como receitas de capital. Neste caso, trata-se da subconta Transferências Correntes, que registra as receitas provenientes de recursos financeiros decorrentes de doações, contratos, convênios, acordos, ajustes, termos de parceria ou outros instrumentos, quando destinados a atender despesas classificáveis como correntes. Nota: para os anos de 2002 a 2004 a diferença entre receita orçamentária e o somatório da receita corrente e receita de capital se deve às deduções relativas ao Fundef. Assim, receita orçamentária = receita corrente (-) deduções receita corrente (+) receita de capital. O universo de municípios da tabela é definido

pelo IBGE no levantamento censitário e não necessariamente coincide com aquele utilizado pelo STN, ou oficialmente existente ou instalado na data de referência. Mais informações: Manual de Contabilidade Aplicado ao Setor Público, 8ª edição; Glossário do Tesouro Nacional e Manual de Demonstrativos Financeiros, 9ª edição.

Atualização: Anual

Unidade: R\$

DESPKM: Despesa de capital - municipal

As despesas de capital são aquelas que contribuem diretamente para a formação ou aquisição de um bem de capital. Seus grupos de natureza (subcontas gerais) são Investimento, Inversões Financeiras e Amortização da Dívida. Aqui trata-se do valor inteiro das despesas de capital, incluindo todas suas subcontas. Corresponde às despesas empenhadas, considerando-se as intra-orçamentárias. Nota: O universo de municípios da tabela é definido pelo IBGE no levantamento censitário e não necessariamente coincide com aquele utilizado pelo STN ou oficialmente existente ou instalado na data de referência. Mais informações: Manual de Contabilidade Aplicado ao Setor Público, 8ª edição; Glossário do Tesouro Nacional e Manual de Demonstrativos Financeiros, 9ª edição.

Atualização: Anual

Unidade: R\$

DESPORM: Despesa orçamentária - municipal

Destina-se ao registro do valor total da soma das Despesas Correntes com as Despesas de Capital. O universo de municípios da tabela é definido pelo IBGE no levantamento censitário e não necessariamente coincide com aquele utilizado pelo STN ou oficialmente existente ou instalado na data de referência. Para os anos de 2002 a 2004 a diferença entre receita orçamentária e o somatório da receita corrente e receita de capital se deve às deduções relativas ao Fundef. Nota: a modalidade de aplicação é uma informação gerencial que tem por finalidade indicar se os recursos são aplicados diretamente por órgãos ou entidades no âmbito da mesma esfera de Governo ou por outro ente da Federação e suas respectivas entidades. Mais informações: Manual de Contabilidade Aplicado ao Setor Público, 8ª edição, Glossário do Tesouro Nacional e Manual de Demonstrativos Financeiros, 9ª edição.

Atualização:

DFEDUCM: Despesa por função - educação e cultura - empenhada - municipal

A função expressa o maior nível de agregação das ações da administração pública, nas diversas áreas de despesa que competem ao setor público. Neste caso, refere-se à soma das despesas com as funções Educação e Cultura. Corresponde às despesas empenhadas, não considerando as intra-orçamentárias. Nota: O universo de municípios da tabela é definido pelo IBGE no levantamento censitário e não necessariamente coincide com aquele utilizado pelo STN ou oficialmente existente ou instalado na data de referência. Mais informações: Manual de Contabilidade Aplicado ao Setor Público, 8ª edição, Glossário do Tesouro Nacional e Manual de Demonstrativos Financeiros, 9ª edição.

Atualização: Anual

Unidade: R\$

**DFHABM:** Despesa por função - urbanismo e habitação - empenhada - municipal

A função expressa o maior nível de agregação das ações da administração pública, nas diversas áreas de despesa que competem ao setor público. Neste caso, refere-se à soma das despesas com as funções Urbanismo e Habitação. Corresponde às despesas empenhadas, não considerando as intra-orçamentárias. Nota: O universo de municípios da tabela é definido pelo IBGE no levantamento censitário e não necessariamente coincide com aquele utilizado pelo STN ou oficialmente existente ou instalado na data de referência. Para os anos de 2002 a 2004 a diferença entre receita orçamentária e o somatório da receita corrente e receita de capital se deve às deduções relativas ao Fundef. Assim, receita orçamentária = receita corrente (-) deduções receita corrente (+) receita de capital. Mais informações: Manual de Contabilidade Aplicado ao Setor Público, 8ª edição, Glossário do Tesouro Nacional e Manual de Demonstrativos Financeiros, 9ª edição.

Atualização: Anual

Unidade: R\$

**RTRKTOM:** Receita de capital - receita bruta - transferências de capital - municipal

As receitas de capital são aquelas provenientes tanto da realização de recursos financeiros oriundos da constituição de dívidas e da conversão, em espécie, de bens e direitos, quanto de recursos recebidos de outras pessoas de direito público ou privado e destinados a atender despesas classificáveis em despesas de capital. Ao contrário da receita corrente, a receita de capital não provoca efeito sobre o patrimônio líquido. Neste caso, trata-se da subconta Transferências de Capital, que são recursos financeiros recebidos de outras pessoas de direito público ou privado e destinados para atender despesas em investimentos ou inversões

financeiras, a fim de satisfazer finalidade pública específica; sem corresponder, entretanto, a contraprestação direta ao ente transferidor. Nota: para os anos de 2002 a 2004 a diferença entre receita orçamentária e o somatório da receita corrente e receita de capital se deve às deduções relativas ao Fundef. Assim, receita orçamentária = receita corrente (-) deduções receita corrente (+) receita de capital. O universo de municípios da tabela é definido pelo IBGE no levantamento censitário e não necessariamente coincide com aquele utilizado pelo STN, ou oficialmente existente ou instalado na data de referência. Mais informações: Manual de Contabilidade Aplicado ao Setor Público, 8ª edição; Glossário do Tesouro Nacional e Manual de Demonstrativos Financeiros, 9ª edição.

Atualização: Anual

Unidade: R\$

TACIDT: Taxa de vítimas de acidentes de trânsito a óbito (100.000 Habitantes)

Óbito por causa externa ou não-natural, indiferente do tempo entre o evento lesivo e a morte propriamente, é categorizado como consequente de lesão provocada por violência (acidentes, homicídios, suicídios ou morte suspeita). Neste caso, a taxa por 100.000 habitantes é calculada através da divisão do indicador principal (número de óbitos por acidente de trânsito) pelo total da população em questão, sendo este resultado multiplicado por 100.000. Fonte: Os dados originais são provenientes do SIM-DATASUS. Até 1995 as informações são do CID9. A partir de 1996 são do CID10. Elaboração Ipeadata: Cálculo da Taxa, divisão do grupo populacional multiplicado por 100.000 pela população de referência. Mais informações: Texto para Discussão IPEA 1848 e Declaração de Óbito.

Atualização: Anual

Unidade: -

THOMIC: Taxa de homicídios (100.000 Habitantes)

Óbito por causa externa ou não-natural, indiferente do tempo entre o evento lesivo e a morte propriamente, é categorizado como consequente de lesão provocada por violência (acidentes, homicídios, suicídios ou morte suspeita). Neste caso, a taxa por 100.000 habitantes é calculada através da divisão do indicador principal (número de homicídios) pelo total da população em questão, sendo este resultado multiplicado por 100.000. Fonte: Os dados originais são provenientes do SIM-DATASUS. Até 1995 as informações são do CID9. A partir de 1996 são do CID10. Elaboração Ipeadata: Cálculo da Taxa, divisão do grupo populacional multiplicado

por 100.000 pela população de referência. Mais informações: Texto para Discussão IPEA 1848 e Declaração de Óbito.

Atualização: Anual

Unidade: -

FAMILIAS\_BOLSA\_FAMILIA: Programa Bolsa Família (PBF) - número de benefícios em dezembro

Esta série apresenta o número de famílias beneficiadas pelo programa em dezembro de cada ano.

Atualização: Anual

Unidade: Pessoa

VALORES\_BOLSA\_FAMILIA: Programa Bolsa Família (PBF) - valor total dos benefícios em dezembro

Esta série apresenta o valor nominal total das transferências do programa em dezembro de cada ano.

Atualização: Anual

Unidade: R\$

PIB: PIB municipal