

# RESSALVA

Atendendo solicitação do autor,  
o texto completo desta dissertação  
será disponibilizado somente a partir  
de 27/08/2027.



Programa de Pós-Graduação em Ciência da Computação

José Cláudio Guedes das Neves

**BupFlow: Um Pipeline de Mineração de Dados  
para Alocação Orçamentária Pública — Do  
Mapeamento Sistemático à Aplicação no Sistema  
Orçamentário Brasileiro**

Rio Claro – SP

2025

UNIVERSIDADE ESTADUAL PAULISTA  
“Júlio de Mesquita Filho”  
Instituto de Geociências e Ciências Exatas  
Câmpus de Rio Claro

José Cláudio Guedes das Neves

**BupFlow: Um Pipeline de Mineração de Dados para Alocação  
Orçamentária Pública — Do Mapeamento Sistemático à  
Aplicação no Sistema Orçamentário Brasileiro**

Dissertação de Mestrado apresentada ao Instituto de Geociências e Ciências Exatas do Câmpus de Rio Claro, da Universidade Estadual Paulista “Júlio de Mesquita Filho”, como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof<sup>a</sup>. Dr<sup>a</sup>. Veronica Oliveira de Carvalho

Rio Claro – SP  
2025

N518b

Neves, José Cláudio Guedes das

BupFlow: Um pipeline de mineração de dados para alocação  
orçamentária pública : Do mapeamento sistemático à aplicação no  
sistema orçamentário brasileiro / José Cláudio Guedes das Neves. --  
Rio Claro, 2025  
150 p.

Dissertação (mestrado) - Universidade Estadual Paulista (UNESP),  
Instituto de Geociências e Ciências Exatas, Rio Claro

Orientadora: Veronica Oliveira de Carvalho

1. Mineração de dados (Computação). 2. Orçamento. 3. Alocação  
de recursos. 4. Otimização matemática. I. Título.

# Impacto potencial desta pesquisa

Essa pesquisa procurou gerar impacto positivo ao contribuir diretamente para diversos Objetivos de Desenvolvimento Sustentável (ODS) como a erradicação da pobreza (1), saúde e bem-estar (3), trabalho decente e crescimento econômico (8), indústria, inovação e infraestrutura (9), redução das desigualdades (10), entre outros. Buscou ainda promover novos conhecimentos aplicáveis ao avanço científico por meio do mapeamento sistemático da literatura do uso de técnicas de inteligência artificial no contexto dos processos de planejamento e orçamento públicos. Contribuiu também com a análise crítica do estado da arte nesse segmento, além de disponibilizar um *pipeline* aplicável com o objetivo de promover a cultura de valorização do uso estratégico de mineração de dados para a melhoria da governança orçamentária, modernização do ciclo orçamentário e otimização da alocação de recursos públicos.

# Potential impact of this research

This research sought to generate a positive impact by directly contributing to several Sustainable Development Goals (SDGs), such as no poverty (1), good health and well-being (3), decent work and economic growth (8), industry, innovation, and infrastructure (9), and reduced inequalities (10), among others. It also sought to promote new knowledge applicable to scientific advancement through systematically mapping the literature on the use of artificial intelligence techniques in the context of public planning and budgeting processes. It also contributed to a critical analysis of the state of the art in this segment, in addition to providing an applicable pipeline aimed at promoting a culture of valuing the strategic use of data mining to improve budgetary governance, modernize the budget cycle, and optimize the allocation of public resources.

UNIVERSIDADE ESTADUAL PAULISTA  
“Júlio de Mesquita Filho”  
Instituto de Geociências e Ciências Exatas  
Câmpus de Rio Claro

José Cláudio Guedes das Neves

**BupFlow: Um Pipeline de Mineração de Dados para Alocação  
Orçamentária Pública — Do Mapeamento Sistemático à  
Aplicação no Sistema Orçamentário Brasileiro**

Dissertação de Mestrado apresentada ao Instituto de Geociências e Ciências Exatas do Câmpus de Rio Claro, da Universidade Estadual Paulista “Júlio de Mesquita Filho”, como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação.

Comissão Examinadora

- Prof<sup>ª</sup>. Dr<sup>ª</sup>. Veronica Oliveira de Carvalho (Orientadora)  
Departamento de Estatística, Matemática Aplicada e Computação (DEMAC)  
Universidade Estadual Paulista "Júlio De Mesquita Filho" (UNESP) – Câmpus de Rio Claro
- Prof. Dr. Diego Furtado Silva  
Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (ICMC)  
Universidade de São Paulo (USP) – Câmpus de São Carlos
- Prof. Dr. André Luis Debiaso Rossi  
Departamento de Computação - Faculdade de Ciências  
Universidade Estadual Paulista "Júlio De Mesquita Filho" (UNESP) – Câmpus de Bauru

Conceito: Aprovado

Rio Claro (SP), 27 de agosto de 2025

*Este trabalho é dedicado em honra e glória de Nosso Senhor Jesus Cristo e sua mãe Maria  
Santíssima.*

*Non nobis, Dómine, non nobis. Sed nómini tuo da glóriam.*

# AGRADECIMENTOS

Gostaria primeiramente de agradecer a Deus pelo dom da vida através de meus amados pais Jenny e Antônio (*in memoriam*); por ter me concedido fortaleza e sabedoria para atravessar mais este desafio.

Não poderia deixar de agradecer à Secretaria de Orçamento Federal nas pessoas dos colegas Alexis Braga Sotto Maior e Ramon Gomes Brandão pelo incentivo e apoio dado ao longo deste período. Às vezes as atitudes mais simples são as que geram maior impacto em nossas vidas e estas podem ser consideradas como exemplos.

Após um longo tempo fora da academia, retorno à Unesp, casa na qual graduei-me, agradecendo aos professores que ministraram disciplinas neste mestrado. Agradeço especialmente a Profa. Dra. Verônica Oliveira de Carvalho que, corajosa e diligentemente, aceitou o duplo desafio de orientar um aluno experiente da área de planejamento e orçamento, até então muito pouco explorada dentro das ciências da computação. Certamente não foram poucos os desafios, mas os incentivos e aprendizados adquiridos ao longo do período foram mais do que suficientes para suplantá-los, experiências que levo pelo resto de minha vida.

Por fim, à minha querida esposa Lucimara, companheira de uma vida, que além de ter aberto mão de grande parte do tempo de convívio ao longo dos últimos anos, ainda, como bônus, me ensinou os conceitos elementares de genética, importantes para este curso. Deste convívio também incluo meus amados filhos Thiago e Lucas que bem compreenderam minha ausência temporária.

# RESUMO

Esta dissertação teve como objetivo investigar o uso de técnicas de mineração de dados como instrumento de apoio à tomada de decisão na alocação de recursos públicos, com foco na função alocativa do orçamento. Para isso, realizou-se inicialmente um mapeamento sistemático da literatura, a fim de identificar aplicações existentes de mineração de dados no contexto orçamentário. O estudo constatou que, embora a área de orçamento público represente um campo estratégico para a aplicação de soluções inteligentes, sua exploração ainda é incipiente, especialmente no que diz respeito à função alocativa. Dentre as soluções encontradas, destaca-se o Pipe-VC, um pipeline que relaciona gastos públicos a variáveis macroeconômicas como PIB, inflação e índice de Gini. No entanto, o Pipe-VC apresenta limitações, entre elas o tratamento inadequado dos dados temporais, a ausência de técnicas de *feature engineering* e a baixa capacidade de generalização. Diante dessas e outras limitações, este trabalho apresenta o BupFlow (**B**udget **p**ublic **F**low), um pipeline alternativo, mais flexível e interpretável, que incorpora pré-processamento adequado, seleção de atributos e modelos mais simples, porém com maior capacidade de generalização, visando um domínio caracterizado por conjuntos de dados de pequena escala e alta dimensão. O BupFlow foi avaliado empiricamente por meio de sua aplicação a dados do orçamento federal brasileiro, sendo comparado ao Pipe-VC. Os resultados indicam que o BupFlow se apresenta como uma alternativa viável ao Pipe-VC, constituindo uma contribuição relevante ao estado da arte. Conclui-se que a proposta oferece uma abordagem metodológica promissora para apoiar decisões públicas orientadas por dados, contribuindo para uma alocação mais eficiente, transparente e socialmente orientada dos recursos públicos.

**Palavras-chave:** mineração de dados; orçamento público; alocação; otimização; pipeline.

# ABSTRACT

This dissertation sought to explore the application of data mining techniques as a means to facilitate decision-making in the allocation of public resources, with a particular focus on the budget's allocative function. To achieve this, we initially conducted a systematic literature review to identify existing applications of data mining within the budgetary context. The findings revealed that, while the public budget represents a strategic area for the implementation of intelligent solutions, its exploration remains nascent, particularly concerning the allocative function. Among the solutions identified, Pipe-VC stands out as a pipeline that correlates public spending with macroeconomic variables such as GDP, inflation, and the Gini index. Nonetheless, Pipe-VC exhibits certain limitations, including inadequate handling of temporal data, a lack of feature engineering techniques, and limited generalizability. In light of these issues, this paper introduces BupFlow (**B**udget **p**ublic **F**low), a more flexible and interpretable pipeline that integrates appropriate preprocessing, feature selection, and simpler models, all while exhibiting enhanced generalization capacity, particularly suited for small-scale, high-dimensional datasets. BupFlow was empirically assessed through its application to Brazilian federal budget data and was compared with Pipe-VC. The results indicate that BupFlow serves as a viable alternative to Pipe-VC, making a significant contribution to the current state of the art. This proposal offers a promising methodological framework to support data-driven public decision-making, ultimately contributing to a more efficient, transparent, and socially responsible allocation of public resources.

**Keywords:** data mining; public budget; allocation; optimization; pipeline.

# LISTA DE ILUSTRAÇÕES

Figura 1 – PIB a preços de mercado - Taxa acumulada em 4 trimestres (%), 1º trimestre 1996 - 1º trimestre 2023. . . . .	16
Figura 2 – Processo de mineração de dados temporais. . . . .	16
Figura 3 – Decomposição da série apresentada na Figura 1 via composição aditiva. . .	19
Figura 4 – SVM linear de margem rígida. . . . .	22
Figura 5 – SVR linear de margem rígida. . . . .	24
Figura 6 – Redes neurais de única camada (Perceptron) e múltiplas camadas (MLP). . .	25
Figura 7 – Fluxo básico de um algoritmo genético. . . . .	27
Figura 8 – Fluxo geral do algoritmo NSGA-II. . . . .	29
Figura 9 – Etapas de treinamento e predição do método SST. . . . .	31
Figura 10 – Etapas de treinamento e predição do método ERC. . . . .	32
Figura 11 – Correlograma da função de autocorrelação e autocorrelação parcial da série apresentada na Figura 1. . . . .	33
Figura 12 – Métodos de imputação de dados sugeridos pela literatura em função das características da série. . . . .	39
Figura 13 – Estratégias de avaliação apresentadas em <a href="#">Cerqueira, Torgo e Mozetic (2020)</a> no contexto de séries temporais. . . . .	43
Figura 14 – Dinâmica das entradas e saídas do orçamento público no trabalho de <a href="#">Valle-Cruz, Fernandez-Cortez e Gil-Garcia (2022)</a> . . . . .	65
Figura 15 – Pipeline utilizado no trabalho de <a href="#">Valle-Cruz, Fernandez-Cortez e Gil-Garcia (2022)</a> . . . . .	66
Figura 16 – Arquitetura da MLP utilizada no trabalho de <a href="#">Valle-Cruz, Fernandez-Cortez e Gil-Garcia (2022)</a> . . . . .	68
Figura 17 – Desvios percentuais obtidos em relação as diferenças nos cálculos apresentados no artigo. . . . .	70
Figura 18 – Séries temporais de entrada e saída utilizadas no trabalho de <a href="#">Valle-Cruz, Fernandez-Cortez e Gil-Garcia (2022)</a> . . . . .	71
Figura 19 – Resultados obtidos na partição de teste referente a reprodução do trabalho. . .	73
Figura 20 – Comparação entre os resultados originais e os reproduzidos. . . . .	74
Figura 21 – Resultados obtidos na partição de teste referente a reprodução modificada do trabalho. . . . .	75
Figura 22 – Comparação entre os resultados originais e os reproduzidos na versão modificada na partição de teste. . . . .	76
Figura 23 – BupFlow. . . . .	78

Figura 24 – BupFlow - etapa de “Seleção e Importação de Dados”.	88
Figura 25 – Dinâmica das entradas e saídas para o sistema orçamentário federal brasileiro.	89
Figura 26 – Comportamento das variáveis para o sistema orçamentário federal brasileiro.	89
Figura 27 – Modelo ontológico da classificação das despesas orçamentárias federais brasileiras.	90
Figura 28 – Diagnóstico de variáveis: PIB e inflação.	94
Figura 29 – Diagnóstico de variáveis: índice de Gini e índice de desenvolvimento humano (IDH).	95
Figura 30 – Identificação de anomalias nas séries utilizadas.	96
Figura 31 – BupFlow - etapa de “Transformação, Limpeza e Normalização de Dados”.	97
Figura 32 – Resultados da MLP na partição de teste, via <i>holdout</i> , para os dados do Brasil.	98
Figura 33 – Resultados da MLP na partição de teste, via WFV, para os dados do Brasil.	99
Figura 34 – Desempenho na partição de teste tanto em relação a média quanto ao valor predito - modelo treinado via WFV.	100
Figura 35 – Resultados da partição de teste em versão modificada do Pipe-VC: modelagens.	103
Figura 36 – Sobreposição dos resultados obtidos na partição de teste em uma versão modificada do Pipe-VC: modelagens.	104
Figura 37 – Resultados da partição de teste em versão modificada do Pipe-VC: modelagens + expansão.	105
Figura 38 – Sobreposição de resultados obtidos na partição de teste em uma versão modificada do Pipe-VC: modelagens + expansão.	106
Figura 39 – Comparativo dos modelos em termos de desempenho nas duas versões modificadas do Pipe-VC: modelagens; modelagens+expansão.	106
Figura 40 – Resultados obtidos na partição de teste para o modelo Global-RF após execução das etapas descritas.	108
Figura 41 – Resultados obtidos na partição de teste para o modelo Local-RF após execução das etapas descritas.	109
Figura 42 – Resultados obtidos na partição de teste para o modelo Local-SVR após execução das etapas descritas.	110
Figura 43 – Resultados obtidos na partição de teste para todos os modelos após execução das etapas descritas.	112
Figura 44 – Comparação dos resultados obtidos na partição de teste, para todos os modelos após execução das etapas descritas, em relação a predição média.	113
Figura 45 – Desempenho do modelo MLP (Pipe-VC) incorporando-se <i>features</i> temporais.	115
Figura 46 – Resultados da partição de teste após expansão de dados.	117
Figura 47 – Desvios obtidos entre o valor computado (predito) e de referência (real).	120
Figura 48 – Visualização das soluções da otimização por pares de variáveis de saída.	122

Figura 49 – Visualização das soluções da otimização em relação a todas as variáveis de saída. . . . .	122
Figura 50 – Visualização para conferência das restrições especificadas por escolha de alocação: individual (local) ou combinada (global). . . . .	123
Figura 51 – Visualização em bolha para comparação dos valores a serem alocados em cada variável de entrada por tipo de orçamento. . . . .	126
Figura 52 – Visualização em barra para comparação dos valores a serem alocados em cada variável de entrada por tipo de orçamento. . . . .	127
Figura 53 – Top-10 <i>features</i> relacionadas ao modelo Local-RF. . . . .	129
Figura 54 – Top-10 <i>features</i> relacionadas ao modelo Global-RF. . . . .	131
Figura 55 – Matriz de correlação de Pearson entre as variáveis de saída utilizadas. . . . .	133

# LISTA DE TABELAS

Tabela 1 – Transposição de parte da sequência de dados apresentada na Figura 1 por meio de uma janela deslizante de tamanho três ( $l=3$ ). . . . .	20
Tabela 2 – Taxa de preenchimento das variáveis de entrada e saída utilizadas. . . . .	72
Tabela 3 – Ranqueamento das 15 melhores arquiteturas MLP pelo menor RMSE. . . . .	98
Tabela 4 – Comparação ponto a ponto, para cada variável de saída, entre os modelos tendo como base a Figura 43. . . . .	114
Tabela 5 – Comparação entre modelos e variáveis de saída em relação aos pontos de predição problemáticos referentes a Figura 43. . . . .	114
Tabela 6 – Valores a serem alocados em cada variável de entrada, por tipo de orçamento, obtidos via otimização. . . . .	124

# LISTA DE ABREVIATURAS E SIGLAS

ACF	<i><u>A</u>uto<u>C</u>orrelation <u>F</u>unction</i>
AG	<i><u>A</u>lgoritmo <u>G</u>enético</i>
AM	<i><u>A</u>prendizado de <u>M</u>áquina</i>
COFOG	<i><u>C</u>lassification <u>O</u>f <u>F</u>unctions <u>O</u>f <u>G</u>overnment</i>
CRISP-DM	<i><u>C</u>Ross-<u>I</u>ndustry <u>S</u>tandard <u>P</u>rocess for <u>D</u>ata <u>A</u>alysis</i>
DT	<i><u>D</u>ecision <u>T</u>ree</i>
ERC	<i><u>E</u>nsemble of <u>R</u>egressor <u>C</u>hains</i>
GFSM	<i><u>G</u>overnment <u>F</u>inance <u>S</u>tatistics <u>M</u>anual</i>
IA	<i><u>I</u>nteligência <u>A</u>rtificial</i>
IDH	<i><u>Í</u>ndice de <u>D</u>esenvolvimento <u>H</u>umano</i>
KDD	<i><u>K</u>nowledge <u>D</u>iscovery in <u>D</u>atabases</i>
LOA	<i><u>L</u>ei <u>O</u>rçamentária <u>A</u>nual</i>
LSTM	<i><u>L</u>ong <u>S</u>hort-<u>T</u>erm <u>M</u>emory</i>
MD	<i><u>M</u>ineração de <u>D</u>ados</i>
MDT	<i><u>M</u>ineração de <u>D</u>ados <u>T</u>emporais</i>
MLP	<i><u>M</u>ulti <u>L</u>ayer <u>P</u>erceptron</i>
MSL	<i><u>M</u>apeamento <u>S</u>istemático da <u>L</u>iteratura</i>
MTR	<i><u>M</u>ulti <u>T</u>arget <u>R</u>egression</i>
OCDE	<i><u>O</u>rganização para a <u>C</u>ooperacão e o <u>D</u>esenvolvimento <u>E</u>conômico</i>
ODS	<i><u>O</u>bjtivos de <u>D</u>esenvolvimento <u>S</u>ustentável</i>
ONU	<i><u>O</u>rganização das <u>N</u>ações <u>U</u>nidas</i>
PACF	<i><u>P</u>artial <u>A</u>uto<u>C</u>orrelation <u>F</u>unction</i>
PIB	<i><u>P</u>roduto <u>I</u>nterno <u>B</u>ruto</i>
PPA	<i><u>P</u>lano <u>P</u>luri<u>A</u>nual</i>
RC	<i><u>R</u>egressor <u>C</u>hains</i>
RNA	<i><u>R</u>ede <u>N</u>eural <u>A</u>rtificial</i>
RF	<i><u>R</u>andom <u>F</u>orest</i>
SEMMA	<i><u>S</u>ample, <u>E</u>xplore, <u>M</u>odify, <u>M</u>odel and <u>A</u>ssess</i>
SOF	<i><u>S</u>ecretaria de <u>O</u>rçamento <u>F</u>ederal</i>
SST	<i><u>S</u>tacked <u>S</u>ingle <u>T</u>arget</i>
ST	<i><u>S</u>ingle <u>T</u>arget</i>
STR	<i><u>S</u>ingle <u>T</u>arget <u>R</u>egression</i>
SVM	<i><u>S</u>upport <u>V</u>ector <u>M</u>achine</i>
SVR	<i><u>S</u>upport <u>V</u>ector <u>R</u>egression</i>

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>11</b>
<b>2</b>	<b>FUNDAMENTOS</b>	<b>15</b>
2.1	Mineração de Dados Temporais	15
2.2	Séries Temporais	17
2.3	Métodos de Aprendizado de Máquina	19
2.3.1	Árvore de Decisão e Random Forest	20
2.3.2	SVR	21
2.3.3	Redes Neurais Artificiais	23
2.3.4	Algoritmos Genéticos	26
2.3.5	Regressão com Múltiplas Saídas	29
2.4	Pré-processamento em Séries Temporais	31
2.5	Séries Temporais Curtas	35
2.5.1	Imputação de Dados	38
2.6	Avaliação de Modelos	41
2.6.1	Medidas de Avaliação	41
2.6.2	Estratégias de Avaliação	42
<b>3</b>	<b>MAPEAMENTO SISTEMÁTICO DA LITERATURA</b>	<b>46</b>
<b>4</b>	<b>PIPE-VC: PIPELINE DE ALOCAÇÃO ORÇAMENTÁRIA DE VALLE-CRUZ ET AL. (2022)</b>	<b>64</b>
4.1	Pipe-VC	65
4.2	Limitações	69
4.3	Reprodução do Trabalho	70
4.3.1	Descrição do Experimento	71
4.3.2	Resultados Iniciais	73
4.3.3	Resultados Adicionais	74
4.4	Considerações Finais	75
<b>5</b>	<b>BUPFLOW: UM PIPELINE PARA APOIO A ALOCAÇÃO ORÇAMENTÁRIA PÚBLICA</b>	<b>77</b>
5.1	Seleção e Importação de Dados	79
5.2	Transformação, Limpeza e Normalização de Dados	79
5.3	Tratamento Temporal (TT): Feature Engineering	79

<b>5.4</b>	<b>Indução de Modelo</b>	<b>80</b>
<b>5.5</b>	<b>Feature Selection: “TT: Feature Engineering” + “Indução de Modelo”</b>	<b>83</b>
<b>5.6</b>	<b>Expansão (E): Imputação</b>	<b>84</b>
<b>5.7</b>	<b>Extração de Funções</b>	<b>84</b>
<b>5.8</b>	<b>Otimização das Funções via AG</b>	<b>85</b>
<b>5.9</b>	<b>Análise e Interpretação</b>	<b>86</b>
<b>6</b>	<b>BUPFLOW NA ALOCAÇÃO ORÇAMENTÁRIA DO GOVERNO FEDERAL BRASILEIRO</b>	<b>87</b>
<b>6.1</b>	<b>BupFlow: Etapa de Seleção e Importação de Dados</b>	<b>87</b>
<b>6.2</b>	<b>BupFlow: Etapa de Transformação, Limpeza e Normalização de Dados</b>	<b>93</b>
<b>6.3</b>	<b>Pipe-VC: Indução de Modelo via MLP</b>	<b>97</b>
<b>6.4</b>	<b>Pipe-VC Modificado: Indução de Modelo e Expansão</b>	<b>100</b>
6.4.1	Adaptação via Modelos	101
6.4.2	Adaptação via Modelos e Expansão	102
<b>6.5</b>	<b>BupFlow: TT:Feature Engineering + Indução de Modelo + Feature Selection</b>	<b>104</b>
<b>6.6</b>	<b>BupFlow: E: Imputação</b>	<b>115</b>
<b>6.7</b>	<b>BupFlow: Extração de Funções</b>	<b>116</b>
<b>6.8</b>	<b>BupFlow: Otimização das Funções via AG</b>	<b>120</b>
<b>6.9</b>	<b>BupFlow: Análise e Interpretação</b>	<b>123</b>
6.9.1	Local-SVR	123
6.9.2	Local-RF e Global-RF	125
<b>6.10</b>	<b>Considerações Finais</b>	<b>130</b>
<b>7</b>	<b>CONCLUSÃO</b>	<b>134</b>
	<b>REFERÊNCIAS</b>	<b>138</b>

# 1 Introdução

A era digital tem promovido a geração e o armazenamento contínuo de grandes volumes de dados, disponíveis em formatos variados, como texto, vídeo e áudio. Para extrair conhecimento útil a partir desses dados e aplicá-lo em diferentes tarefas, faz-se necessário um processo sistemático conhecido como mineração de dados. Trata-se de um processo iterativo e interativo que permite a extração de padrões implícitos nos dados (TAN *et al.*, 2020). Dentre as metodologias que viabilizam esse processo, destacam-se a *Knowledge Discovery in Databases* (KDD), *CRoss-Industry Standard Process for Data Mining* (CRISP-DM) e *Sample, Explore, Modify, Model and Assess* (SEMMA) (PLOTNIKOVA; DUMAS; MILANI, 2020).

Apesar das diferenças entre essas metodologias, todas compartilham três etapas fundamentais: pré-processamento, extração de padrões e pós-processamento (TAN *et al.*, 2020; PARMEZAN; BATISTA, 2016). No pré-processamento, os dados são preparados para análise, considerando tarefas como tratamento de valores ausentes, discretização, normalização e padronização. Na extração, aplicam-se algoritmos de aprendizado de máquina para identificar padrões relevantes. Por fim, no pós-processamento, os resultados obtidos são avaliados com base em medidas de desempenho.

Embora amplamente adotadas em setores privados e industriais, observa-se uma utilização ainda incipiente dessas técnicas no setor público, especialmente no planejamento e no orçamento governamentais (ver Capítulo 3). O orçamento público constitui um instrumento central de planejamento estatal, por meio do qual o governo exerce influência política e administrativa sobre a alocação de recursos, a gestão e o controle de serviços e políticas públicas. Sua elaboração envolve um conjunto de tarefas complexas e dinâmicas, exigindo articulação entre diferentes níveis da administração pública (BENNETT; LEMOINE, 2014). Assim, as decisões de financiamento tomadas nesse processo são de extrema relevância, pois impactam diretamente a qualidade dos serviços prestados à sociedade e os resultados das políticas públicas (TASCHE; DHUNGEL; HEINE, 2022; FERNANDEZ-CORTEZ; VALLE-CRUZ; GIL-GARCIA, 2020).

A escassez de recursos públicos impõe ao Estado o desafio constante de definir quanto arrecadar e como distribuir tais recursos de forma eficiente e transparente. As leis orçamentárias, como principal instrumento da política fiscal, orientam essas decisões e buscam assegurar a continuidade e a qualidade dos serviços públicos, refletindo escolhas estratégicas dos formuladores de políticas voltadas ao bem-estar social.

Nesse contexto, o uso de soluções baseadas em Inteligência Artificial (IA) tem se mostrado promissor para apoiar os gestores na tomada de decisão orçamentária. A mineração de

dados, em particular, tem se destacado como uma alternativa viável para modelar e compreender questões orçamentárias. Sendo a antecipação estratégica uma das práticas recomendadas pela OCDE (Organização para a Cooperação e Desenvolvimento Econômico), a criação de grupos de trabalho focados no uso de técnicas de IA no setor público tornou-se imperativa. De fato, no ano de 2024, na “46ª Reunião Anual do Comitê de Altos Funcionários do Orçamento”, o tema “Uso de Inteligência Artificial na Gestão Financeira Pública” foi abordado<sup>1</sup>. A própria OCDE disponibiliza em sua biblioteca digital diversos artigos recentes sobre o poder da IA e suas implicações para economias e sociedades<sup>2</sup>. Portanto, é essencial entender como as soluções são estruturadas.

Considerando o exposto, como um **primeiro objetivo** deste trabalho (**Obj.-1**) realizou-se um mapeamento sistemático da literatura (MSL) visando (i) fornecer uma visão geral dos aspectos relacionados às etapas de mineração de dados no contexto apresentado, assim como (ii) identificar lacunas que possam ser abordadas e/ou exploradas. Tal mapeamento é apresentado no Capítulo 3 e encontra-se disponível em [Neves e Carvalho \(2025\)](#). Assim, uma primeira **contribuição** desta dissertação se refere a tal estudo, o qual tem o potencial de impactar significativamente futuras pesquisas e práticas em planejamento público e mineração de dados.

Uma das vertentes abordadas no MSL foi a identificação, de forma genérica, considerando a divisão lógica proposta em [Giacomoni \(2021\)](#), em quais assuntos, temas e subtemas orçamentários, técnicas de mineração de dados vêm sendo aplicadas. Os assuntos identificados foram categorizados em “Processo orçamentário” e “O Estado na economia”. O assunto “O Estado na economia” foi dividido em: (i) “Responsabilidades econômicas do Estado” e (ii) “O crescimento da despesa pública”. Em relação ao tema “Responsabilidades econômicas do Estado”, encontram-se os subtemas relativos às funções fiscais (funções orçamentárias), os quais dividem-se em ([MUSGRAVE; SIMÕES, 1976](#)): (i) “Função Alocativa”, para promover ajustes na alocação de recursos; (ii) “Função Distributiva”, para promover ajustes na distribuição de renda; (iii) “Função Estabilizadora”, para manter a estabilidade econômica, tendo entre seus principais objetivos macroeconômicos a manutenção de um alto nível de emprego e a estabilidade dos níveis de preços ([GIACOMONI, 2021](#)).

Destaca-se neste cenário a “Função Alocativa” do Estado, fundamental para garantir que os recursos públicos sejam aplicados em áreas que gerem maior retorno social, especialmente diante da escassez orçamentária e da alta demanda por serviços essenciais, como saúde, educação e infraestrutura. Seu papel é corrigir falhas de mercado, promover equidade e impulsionar o desenvolvimento. Neste contexto, destaca-se a atuação da Secretaria de Orçamento Federal (SOF)<sup>3</sup>, responsável por formular e coordenar a proposta orçamentária da União. A alocação dos

<sup>1</sup> O documento está disponível em [https://one.oecd.org/document/GOV/SBO\(2024\)14/en/pdf](https://one.oecd.org/document/GOV/SBO(2024)14/en/pdf).

<sup>2</sup> [https://www.oecd-ilibrary.org/science-and-technology/oecd-artificial-intelligence-papers\\_dee339a8-en](https://www.oecd-ilibrary.org/science-and-technology/oecd-artificial-intelligence-papers_dee339a8-en).

<sup>3</sup> <https://www.gov.br/planejamento/pt-br/composicao/orgaos/secretaria-de-orcamento-federal>.

recursos públicos é estruturada a partir do Plano Plurianual (PPA), orientada pela Lei de Diretrizes Orçamentárias (LDO) e concretizada na Lei Orçamentária Anual (LOA), instrumentos que juntos garantem coerência entre o planejamento de médio prazo, as prioridades do governo e a execução orçamentária. Deste modo, investigar a função alocativa com o apoio de técnicas de mineração de dados permite identificar padrões, gargalos e oportunidades de melhoria, contribuindo para uma gestão pública mais eficiente, transparente e alinhada às reais necessidades da sociedade.

No MSL realizado, apenas quatro trabalhos encontram-se relacionados à função alocativa, a saber: [Abdelsalam \*et al.\* \(2014\)](#), [Guariso, Castañeda e Guerrero \(2023\)](#), [Ivaşcu e Ştefoni \(2023\)](#) e [Valle-Cruz, Fernandez-Cortez e Gil-Garcia \(2022\)](#). Contudo, apenas [Valle-Cruz, Fernandez-Cortez e Gil-Garcia \(2022\)](#) procuram responder em que medida as técnicas de IA podem ajudar a distribuir os gastos públicos para aumentar o PIB, diminuir a inflação e reduzir o índice de Gini através de um *pipeline*, denominado aqui de Pipe-VC, que processa entradas orçamentárias (despesas específicas) para gerar resultados econômicos, políticos e sociais. O *pipeline* de apoio à decisão na alocação de gastos públicos contribui, a partir de uma visão prospectiva, sobre como a função alocativa pode ser modernizada e otimizada no contexto da transformação digital do setor público. Vale mencionar que o trabalho dos autores foi citado na “46ª Reunião” acima referida.

Embora o *pipeline* proposto pelos autores seja interessante por tentar objetivamente relacionar variáveis macroeconômicas clássicas de saída (PIB, inflação e Gini) a um conjunto de variáveis de entrada na busca de uma melhor alocação orçamentária, alguns pontos apresentam limitações do ponto de vista computacional e orçamentário, como, por exemplo, o não tratamento dos dados em relação aos seus aspectos temporais, desprezando importantes aspectos de *feature engineering*. Ademais, o modelo proposto por eles, de acordo com os experimentos realizados, não generaliza, uma vez que baseia-se em uma rede neural e a quantidade de dados disponível é pequena. O Capítulo 4 apresenta uma análise detalhada do referido trabalho, a qual se configura como uma **contribuição** adicional desta dissertação.

Diante do exposto, considerando as limitações brevemente mencionadas, apresenta-se neste trabalho como um **segundo objetivo (Obj.-2)** um *pipeline* para apoiar os tomadores de decisão no que se refere a alocação de recursos públicos. O *pipeline*, denominado BupFlow (*Budget public Flow*), baseia-se no trabalho de [Valle-Cruz, Fernandez-Cortez e Gil-Garcia \(2022\)](#). Contudo, o BupFlow visa avançar o estado da arte, uma vez que foi projetado para ser mais flexível que o Pipe-VC, incorporando aspectos temporais, seleção de atributos, modelos mais simples, porém com melhor capacidade de generalização, visando um domínio caracterizado por conjuntos de dados de pequena escala e alta dimensão. Pode-se pensar no BupFlow como um *guideline* metodológico a ser seguido no contexto aqui apresentado, o qual é desenvolvido em detalhes no Capítulo 5. A fim de avaliar sua aplicabilidade, assim como verificar seu desempenho em relação ao Pipe-VC, o BupFlow foi instanciado com os dados do governo federal brasileiro. Os resultados indicam que o BupFlow se apresenta como uma alternativa viável ao Pipe-VC,

configurando-se como uma **contribuição** para o estado da arte, conforme será visto no Capítulo 6.

A fim de apresentar o trabalho desenvolvido, esta dissertação encontra-se estruturada da seguinte maneira:

- Capítulo 2: esclarece os conceitos que embasam esta pesquisa, a saber: mineração de dados temporais, séries temporais, métodos de aprendizado de máquina, pré-processamento em séries temporais, séries temporais curtas e avaliação de modelos;
- Capítulo 3: divulga o MSL realizado, tendo sido publicado e disponível neste [link](#);
- Capítulo 4: relata e discute o trabalho de [Valle-Cruz, Fernandez-Cortez e Gil-Garcia \(2022\)](#);
- Capítulo 5: explica o BupFlow como processo de apoio a tomada de decisão no escopo orçamentário;
- Capítulo 6: desenvolve a instanciação do BupFlow na alocação orçamentária do governo federal brasileiro;
- Capítulo 7: apresenta as conclusões e trabalhos futuros.

## 7 Conclusão

Este trabalho investigou o uso de técnicas de mineração de dados como instrumento de apoio à tomada de decisão na alocação de recursos públicos, com foco na função alocativa do orçamento. Para tal, um MSL foi realizado. Foi possível constatar que a utilização dessas técnicas no contexto do planejamento público ainda está em sua fase inicial, tendo apresentado leve crescimento a partir de 2019. Notou-se também que apenas 4 trabalhos estavam relacionados com a função alocativa do Estado. Outras informações relevantes também foram levantadas e apresentadas em [Neves e Carvalho \(2025\)](#).

Com base no mapeamento realizado, analisou-se criticamente o trabalho de [Valle-Cruz, Fernandez-Cortez e Gil-Garcia \(2022\)](#), o único entre os 4 que apresenta maior relacionamento com a tarefa de alocação orçamentaria a partir de critérios objetivos, utilizado aqui como *baseline*. O mesmo apresenta um *pipeline* referenciado como Pipe-VC. Na medida do possível, o trabalho foi reproduzido, sendo possível verificar seus pontos fortes e limitações. Entre as limitações identificadas pode-se mencionar o não tratamento dos dados em relação aos seus aspectos temporais, desprezando importantes aspectos de *feature engineering*.

Com base na análise do trabalho de [Valle-Cruz, Fernandez-Cortez e Gil-Garcia \(2022\)](#), um *pipeline* de recomendação de alocação do orçamento público, denominado BupFlow, foi desenvolvido. O BupFlow incorpora aspectos temporais, seleção de atributos, modelos menos complexos, porém com melhor capacidade de generalização, visando um domínio caracterizado por conjuntos de dados de pequena escala e alta dimensão.

A fim de avaliar sua aplicabilidade, assim como verificar seu desempenho em relação ao Pipe-VC, o BupFlow foi instanciado com os dados do governo federal brasileiro visando maximizar o crescimento econômico (PIB) e o índice de desenvolvimento humano (IDH), assim como reduzir a inflação e a desigualdade de renda (índice de Gini). A partir da extração das equações aprendidas via utilização do algoritmo SVR (Local-SVR), posteriormente otimizadas via NSGA-III, foi possível avaliar as sugestões de alocação propostas. Os resultados apontaram para o aprimoramento da alocação balanceada do orçamento público a partir do aumento dos gastos nas rubricas em assistência social e urbanismo, financiadas pela redução dos gastos em encargos especiais e previdência social. Considerando-se exclusivamente o crescimento econômico como objetivo, os resultados indicam que os gastos prioritários deveriam concentrar-se nas áreas de assistência social, ciência e tecnologia, e indústria, mantendo-se o padrão de financiamento previamente estabelecido. Por sua vez, caso a meta fosse unicamente a redução da inflação, sugere-se a alocação mais expressiva de recursos no setor industrial.

Sabe-se que o processo orçamentário caracteriza-se como prioritariamente político, sendo essa componente de difícil análise e parametrização em qualquer sistema computacional; porém, acredita-se que a metodologia proposta possa ser utilizada como um ponto de partida para aprimorar decisões do ponto de vista técnico a partir de uma análise baseada em dados.

Em um fluxo alternativo do BupFlow, tendo como base as variações do algoritmo Random Forest (Global-RF e Local-RF), verifica-se a importância dos gastos em energia, comunicações, agricultura, cultura e principalmente indústria para se alocar o orçamento de forma equilibrada a fim de maximizar resultados econômicos e sociais simultaneamente. Os resultados foram consistentes com estudos orçamentários e econômicos anteriores, demonstrando o potencial do BupFlow em auxiliar os tomadores de decisões sobre otimização de alocações orçamentárias.

Foi possível verificar, durante o desenvolvimento do trabalho, em função da análise de anomalias, a complexidade inerente a política fiscal, causada por fatores como crises econômicas e a pandemia da COVID-19. A alocação dos gastos públicos deve atender as necessidades da população para gerar bem-estar e crescimento econômico, variáveis destacadas nos modelos. Assim, um dos principais benefícios do uso de técnicas de mineração de dados na alocação de gastos públicos baseia-se no fornecimento de critérios objetivos de alocação para maximizar objetivos múltiplos, buscando trazer algum grau de racionalidade técnica ao processo orçamentário, a partir de ferramentas de apoio para analisar e compreender os dados utilizados nos programas e políticas governamentais.

Entende-se que o *pipeline* proposto pode ser útil para fornecer informações sobre alocações orçamentárias alternativas e seus potenciais efeitos na sociedade para alocadores orçamentários, tendo em vista que essas informações podem ser usadas em um processo de tomada de decisão mais complexo e com forte viés político.

Em relação às questões metodológicas do BupFlow, entende-se que a combinação de algoritmos que sejam capazes de aprender, generalizar e extrair equações, que reflitam as relações lineares e não lineares entre as despesas públicas, sob o total controle do macro alocador orçamentário, com saídas econômicas e sociais, tem o potencial de serem maximizadas a partir de algoritmos genéticos multiobjetivos para gerar cenários úteis para a alocação do orçamento público.

A interpretação dos resultados coloca-se também como fator crítico para o sucesso da utilização de tais soluções, tendo em vista que as características temporais criadas devem estar associadas a conceitos orçamentários sedimentados. O fato da solução estar limitada por desconsiderar modelos incrementais orçamentários estabelecidos ao longo dos anos não impede que o motor de regras orçamentárias constitucionais e legais seja incluído gradativamente ao processo, trazendo maior realidade aos novos cenários projetados, auxiliando assim a tomada de decisão.

Não se pode descartar ainda a utilização do *pipeline* no âmbito micro orçamentário, i.e., adequar-se entradas e saídas para refletir realidades específicas de cada unidade orçamentaria, como, por exemplo, a instanciação adaptada do BupFlow para determinado ministério baseado em classificação de subfunções relativas a essa determinada unidade. Procura-se assim fornecer alternativas de alocação ao invés de prescrever uma solução única definitiva.

As limitações desta pesquisa estão relacionadas ao fato de não se incluir no *pipeline* toda a gama de complexidades relacionadas ao processo orçamentário em seus aspectos econômicos, políticos e sociais. Como mencionado anteriormente, o processo orçamentário não é um exercício puramente técnico, mas prioritariamente político, sujeito a interesses de diversos atores envolvidos. O BupFlow analisa a alocação orçamentaria com base em uma perspectiva nacional brasileira caracterizada em um orçamento base zero ideal. A inclusão do motor de regras constitucionais e legais poderia identificar quais despesas orçamentarias tiveram o melhor impacto no PIB, na inflação, desigualdade e desenvolvimento humano em um orçamento incremental mais realista.

Em contrapartida, manter todas as restrições orçamentarias em um orçamento extremamente rígido tende a reduzir as opções de uma visão mais abrangente da dinâmica das relações entre as variáveis econômicas, sociais e orçamentarias, restringindo assim em demasia as possibilidades de reformas orçamentarias tão necessárias ao atual contexto brasileiro.

Entende-se que os resultados deste estudo são encorajadores quanto ao potencial da realização de análises baseadas em mineração de dados para apoiar a tomada de decisões governamentais. Acredita-se que a mineração de dados pode aprimorar o processo orçamentário, reconhecendo que há diversos fatores não técnicos a serem considerados. Entende-se também que abordagens algorítmicas e suas respectivas técnicas podem resultar em viés, omissões e erros. Ademais, a análise foi conduzida utilizando os dados abertos do Banco Mundial, das Nações Unidas e da Secretaria de Orçamento Federal, podendo também conter vieses nos conjuntos de dados. Assim sendo, os resultados sugeridos devem ser vistos com cautela, não podendo ser generalizados. Pesquisas futuras podem incluir regras orçamentarias legais e constitucionais para avaliar a adequação do modelo aos contextos específicos, suas necessárias adaptações e restrições a diferentes cenários.

Por fim, comparações das técnicas de inteligência artificial com modelos econométricos tradicionais nas tarefas de projeção e otimização poderia servir para novas recomendações de alocação orçamentaria.

Em resumo, como contribuições computacionais e orçamentarias do trabalho pode-se ressaltar:

- o MSL realizado visando (i) fornecer uma visão geral dos aspectos relacionados às etapas de mineração de dados no contexto apresentado, assim como (ii) identificar lacunas que

possam ser abordadas e/ou exploradas. O mesmo encontra-se disponível em [Neves e Carvalho \(2025\)](#);

- a análise crítica do trabalho de [Valle-Cruz, Fernandez-Cortez e Gil-Garcia \(2022\)](#);
- a disponibilização de um *pipeline* genérico (BupFlow), aplicável a qualquer país e ente governamental a partir do ajuste e adequação das variáveis de entrada e saída;
- que, tendo como base os resultados do Pipe-VC, quando as cinco modelagens do BupFlow foram exploradas em tal *pipeline*, a escolha de uma modelagem mais adequada leva a resultados mais satisfatórios;
- a melhora nos resultados referentes aos dados do governo federal brasileiro após aplicação efetiva do BupFlow. Pôde-se observar o impacto positivo das *features* temporais, em conjunto com a etapa de *feature selection*, nas modelagens apresentadas. Ademais, realizando uma análise tanto gráfica quanto numérica, observou-se um bom desempenho do modelo Local-SVR em relação aos demais (Global-RF e Local-RF), os quais são usados como uma referência de desempenho não linear ao problema. Contudo, exceções foram encontradas nos anos referentes a pandemia da COVID-19;
- que o uso das técnicas de expansão de dados no BupFlow não se mostrou efetivo;
- a otimização da alocação orçamentaria brasileira a partir do modelo Local-SVR especificado até 2022, assim como a identificação das variáveis orçamentarias para cada solução de otimização orçamentária proposta.

Como trabalhos futuros, sugere-se, como já mencionado ao longo deste capítulo: (i) a possibilidade de incorporação de regras fiscais e legais ao *pipeline*; (ii) a adaptação do modelo para contextos micro orçamentários, como ministérios específicos; (iii) a comparação entre modelos de inteligência artificial e técnicas econométricas tradicionais; e (iv) a ampliação da análise para outros países ou entes subnacionais. Ademais, sugere-se também as melhorias relacionadas ao BupFlow mencionadas na Seção 6.10.

# REFERÊNCIAS

- ABDELSALAM, H. M. *et al.* Prioritization of public expenditure for a better return on social development: A data mining approach. In: SPRINGER. *Advanced Machine Learning Technologies and Applications: Second International Conference, AMLTA 2014, Cairo, Egypt, November 28-30, 2014. Proceedings 2.* [S.l.], 2014. p. 523–530. Citado 3 vezes na(s) página(s) 13, 64 e 92.
- AGGARWAL, C. C. *Data Mining - The Textbook.* Springer, 2015. ISBN 978-3-319-14141-1. Disponível em: <<https://doi.org/10.1007/978-3-319-14142-8>>. Citado 2 vezes na(s) página(s) 23 e 25.
- ALMEIDA, M. F. d.; GASPARINI, C. E. Plurianualidade orçamentária no brasil: Diagnóstico, rumos e desafios. Escola Nacional de Administração Pública (Enap), 2017. Citado na página 130.
- ARAÚJO, G. S.; GAGLIANONE, W. P. Machine learning methods for inflation forecasting in brazil: New contenders versus classical models. *Latin American Journal of Central Banking*, v. 4, n. 2, p. 100087, 2023. Citado na página 36.
- ARAÚJO, L. S. de O.; SANTOS, M. T.; SILVA, D. A. The brazilian federal budget ontology: a semantic web case of public open data. In: *Proceedings of the 7th International Conference on Management of computational and collective intelligence in Digital EcoSystems.* [S.l.: s.n.], 2015. p. 85–89. Citado na página 90.
- AWAD, M.; KHANNA, R. *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers.* [S.l.]: Apress Berkeley, CA, 2015. Citado na página 23.
- BALTAGI, B. H.; BALTAGI, B. H. *Econometric analysis of panel data.* [S.l.]: Springer, 2008. v. 4. Citado na página 39.
- BECK, M. W. Neuralnettools: Visualization and analysis tools for neural networks. *Journal of statistical software*, NIH Public Access, v. 85, n. 11, p. 1, 2018. Citado 2 vezes na(s) página(s) 67 e 69.
- BENNETT, N.; LEMOINE, G. J. *What a difference a word makes: Understanding threats to performance in a VUCA world.* 2014. Citado na página 11.
- BERTUSSI, G. L.; JUNIOR, R. E. Infraestrutura de transporte e crescimento econômico no brasil. *Journal of Transport Literature*, SciELO Brasil, v. 6, p. 101–132, 2012. Citado na página 128.
- BLANK, J.; DEB, K. Pymoo: Multi-objective optimization in python. *IEEE Access*, v. 8, p. 89497–89509, 2020. Citado 2 vezes na(s) página(s) 85 e 120.
- BORCHANI, H. *et al.* A survey on multi-output regression. *WIREs Data Mining Knowl. Discov.*, v. 5, n. 5, p. 216–233, 2015. Citado na página 29.

- BRASIL. *PORTARIA SOF/ME Nº 2.520, DE 21 DE MARÇO DE 2022*. Ministério da Economia/Secretaria Especial do Tesouro e Orçamento/Secretaria de Orçamento Federal, 2022. Disponível em: <<https://www.in.gov.br/en/web/dou/-/portaria-sof/me-n-2.520-de-21-de-marco-de-2022-387374808>>. Citado na página 91.
- BRASIL. *SCNT - Sistema de Contas Nacionais Trimestrais*. 2023. Disponível em: <<https://www.ibge.gov.br/estatisticas/economicas/contas-nacionais/9300-contas-nacionais-trimestrais.html?edicao=37025&t=destaques>>. Citado na página 15.
- BRASIL, G. F. do. *Lei de Diretrizes Orçamentárias para 2024*. 2024. Anexo II da Lei de Diretrizes Orçamentárias de 2024. Disponível em: <[https://www.gov.br/planejamento/pt-br/assuntos/orcamento/orcamento/orcamentos-anuais/2024/ploa/projeto-de-lei-orcamentaria-anual-ploa-2024/4\\_volume-4\\_grade-de-parametros.pdf](https://www.gov.br/planejamento/pt-br/assuntos/orcamento/orcamento/orcamentos-anuais/2024/ploa/projeto-de-lei-orcamentaria-anual-ploa-2024/4_volume-4_grade-de-parametros.pdf)>. Citado na página 92.
- BRASIL, S. de O. F. *CLASSIFICAÇÃO DA DESPESA*. 2023. Disponível em: <[https://www1.siof.planejamento.gov.br/mto/doku.php/mto2023:cap4#classificacao\\_funcional\\_da\\_despesa](https://www1.siof.planejamento.gov.br/mto/doku.php/mto2023:cap4#classificacao_funcional_da_despesa)>. Citado na página 91.
- BROWNLEE, J. *Introduction to Time Series Forecasting With Python: How to Prepare Data and Develop Models to Predict the Future*. Machine Learning Mastery, 2017. Disponível em: <<https://books.google.com.br/books?id=-AiqDwAAQBAJ>>. Citado 2 vezes na(s) página(s) 38 e 44.
- BURKHEAD, J. V. Budget classification and fiscal planning. *Public Administration Review*, JSTOR, v. 7, n. 4, p. 228–235, 1947. Citado na página 91.
- CARVALHO, L.; RUAS, J. X. Housing prices in brazil and their impacts on interest rate and income in the inflation targeting regime between 2007 and 2017. *Revista de Estudos Sociais*, Faculdade de economia, v. 23, n. 47, p. 51–74, 2021. Citado na página 128.
- CERQUEIRA, V.; TORGO, L.; MOZETIC, I. Evaluating time series forecasting models: an empirical study on performance estimation methods. *Mach. Learn.*, v. 109, n. 11, p. 1997–2028, 2020. Citado 3 vezes na(s) página(s) ix, 43 e 44.
- CORMEN, T. H. *et al.* Algoritmos: teoria e prática. *Editora Campus*, v. 2, p. 296, 2002. Citado na página 35.
- DAMA, F.; SINOQUET, C. *Time Series Analysis and Modeling to Forecast: a Survey*. 2021. Citado na página 19.
- DEB, K. *et al.* A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.*, v. 6, n. 2, p. 182–197, 2002. Disponível em: <<https://doi.org/10.1109/4235.996017>>. Citado 2 vezes na(s) página(s) 28 e 29.
- FACELI, K. *et al.* *Inteligência artificial: uma abordagem de aprendizado de máquina*. [S.l.]: LTC, 2011. Citado 6 vezes na(s) página(s) 20, 22, 23, 24, 25 e 26.
- FERNANDEZ-CORTEZ, V.; VALLE-CRUZ, D.; GIL-GARCIA, J. R. Can artificial intelligence help optimize the public budgeting process? lessons about smartness and public value from the mexican federal government. In: IEEE. *2020 Seventh International Conference on EDemocracy & EGovernment (ICEDEG)*. [S.l.], 2020. p. 312–315. Citado 2 vezes na(s) página(s) 11 e 28.

FEU, A. *et al.* A relação entre avaliação e orçamento e o papel do conselho de monitoramento e avaliação de políticas públicas. Instituto de Pesquisa Econômica Aplicada (Ipea), 2022. Citado na página 91.

FIZAINÉ, F.; COURT, V. Energy expenditure, economic growth, and the minimum eroi of society. *Energy Policy*, Elsevier, v. 95, p. 172–186, 2016. Citado na página 127.

GIACOMONI, J. *Orçamento Público*. [S.l.: s.n.], 2021. ISBN 9788597026122. Citado na página 12.

GOH, A. Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering*, v. 9, n. 3, p. 143–151, 1995. Citado 2 vezes na(s) página(s) 67 e 69.

GONZÁLEZ-CRESPO, R. *et al.* Finding an accurate early forecasting model from small dataset: A case of 2019-ncov novel coronavirus outbreak. *International Journal of Interactive Multimedia and Artificial Intelligence*, v. 6, n. 1, p. 132–140, 2020. Citado na página 36.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. MIT Press, 2016. ISBN 9780262035613. Disponível em: <<https://www.deeplearningbook.org/>>. Citado na página 97.

GUARISO, D.; CASTAÑEDA, G.; GUERRERO, O. A. Budgeting for sdgs: Quantitative methods to assess the potential impacts of public expenditure. *Development Engineering*, Elsevier, v. 8, p. 100113, 2023. Citado 2 vezes na(s) página(s) 13 e 64.

HAN, J.; KAMBER, M.; PEI, J. *Data Mining. Concepts and Techniques*. Third. [S.l.]: Elsevier, 2012. Citado na página 23.

HARRELL, F. E. *et al.* *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. [S.l.]: Springer, 2001. v. 608. Citado na página 40.

HAYASHI, S.; TANIMOTO, A.; KASHIMA, H. Long-term prediction of small time-series data using generalized distillation. In: *2019 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2019. p. 1–8. Citado na página 36.

HEATON, J. *Introduction to Neural Networks with Java*. 2nd. ed. [S.l.]: Heaton Research, Inc., 2008. ISBN 9781604390087. Citado na página 97.

HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: Principles and Practice*. 2nd. ed. Melbourne, Australia: OTexts, 2018. Disponível em: <<https://otexts.com/fpp2/>>. Citado 2 vezes na(s) página(s) 82 e 111.

HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: Principles and Practice*. 3rd. ed. Australia: OTexts, 2021. Disponível em: <[OTexts.com/fpp3](https://otexts.com/fpp3/)>. Citado na página 44.

IVAȘCU, C.; ȘTEFONI, S. Modelling the non-linear dependencies between government expenditures and shadow economy using data-driven approaches. *Scientific Annals of Economics and Business*, v. 70, n. 1, p. 97–114, 2023. Citado 2 vezes na(s) página(s) 13 e 64.

KALTON, G.; KISH, L. Some efficient random imputation methods. *Communications in Statistics-Theory and Methods*, Taylor & Francis, v. 13, n. 16, p. 1919–1939, 1984. Citado na página 39.

KOCHENDERFER, M. J.; WHEELER, T. A. *Algorithms for Optimization*. [S.l.]: The MIT Press, 2019. ISBN 0262039427. Citado 2 vezes na(s) página(s) 26 e 28.

- KONE, L. A. *et al.* Short time series forecasting method based on genetic programming and kalman filter. In: COLLET, P. *et al.* (Ed.). *Complex Computational Ecosystems*. Cham: Springer Nature Switzerland, 2023. p. 77–89. Citado 2 vezes na(s) página(s) 36 e 81.
- KORSTANJE, J. *Advanced Forecasting with Python: With State-of-the-Art-Models Including LSTMs, Facebook's Prophet, and Amazon's DeepAR*. [S.l.]: Apress, 2021. Citado 3 vezes na(s) página(s) 41, 42 e 44.
- LAZZERI, F. *Machine Learning for Time Series Forecasting with Python*. [S.l.]: Wiley, 2020. Citado 6 vezes na(s) página(s) 16, 32, 33, 34, 35 e 38.
- LEPOT, M.; AUBIN, J.-B.; CLEMENS, F. H. Interpolation in time series: An introductory overview of existing methods, their performance criteria and uncertainty assessment. *Water*, v. 9, n. 10, 2017. Citado 2 vezes na(s) página(s) 38 e 40.
- LIU, Z. *et al.* Forecast methods for time series data: A survey. *IEEE Access*, v. 9, p. 91896–91912, 2021. Citado na página 81.
- MUSGRAVE, R. A.; SIMÕES, A. B. Teoria das finanças públicas: um estudo de economia governamental. (*No Title*), 1976. Citado 2 vezes na(s) página(s) 12 e 64.
- NEUSSER, K. *Time series econometrics*. [S.l.]: Springer, 2016. Citado na página 35.
- NEVES, J. C. G. D.; CARVALHO, V. O. D. The use of data mining in public budgeting: A systematic literature mapping. *IEEE Access*, IEEE, 2025. Citado 3 vezes na(s) página(s) 12, 134 e 137.
- PARMEZAN, A. R. S.; BATISTA, G. E. A. P. A. *Descrição de Modelos Estatísticos e de Aprendizagem de Máquina para Predição de Séries Temporais*. 2016. Relatório Técnico da Universidade de São Paulo, São Carlos. Disponível em: <<https://repositorio.usp.br/item/002772986>>. Citado 5 vezes na(s) página(s) 11, 16, 17, 18 e 20.
- PARMEZAN, A. R. S.; SOUZA, V. M. A. de; BATISTA, G. E. A. P. A. Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. *Inf. Sci.*, v. 484, p. 302–337, 2019. Citado 2 vezes na(s) página(s) 16 e 17.
- PLOTNIKOVA, V.; DUMAS, M.; MILANI, F. Adaptations of data mining methodologies: A systematic literature review. *PeerJ Computer Science*, v. 6, p. e267, 2020. Citado 2 vezes na(s) página(s) 11 e 15.
- PRATAMA, I. *et al.* A review of missing values handling methods on time-series data. In: IEEE. *2016 international conference on information technology systems and innovation (ICITSI)*. [S.l.], 2016. p. 1–6. Citado na página 39.
- RAJWAR, K.; DEEP, K.; DAS, S. An exhaustive review of the metaheuristic algorithms for search and optimization: taxonomy, applications, and open challenges. *Artif. Intell. Rev.*, v. 56, n. 11, p. 13187–13257, 2023. Disponível em: <<https://doi.org/10.1007/s10462-023-10470-y>>. Citado na página 85.
- REZENDE, S. O. *Sistemas Inteligentes: Fundamentos e Aplicações*. [S.l.]: Editora Manole Ltda, 2003. Citado na página 16.

RIBEIRO, S. M. *Imputation by Decomposition and by Time Series Nature: Novel Imputation Methods for Missing Data in Time Series*. Dissertação (Mestrado) — Universidade Federal de Minas Gerais (UFMG), 2022. Disponível em: <<http://hdl.handle.net/1843/46099>>. Citado 2 vezes na(s) página(s) 38 e 39.

RIBEIRO, S. M.; CASTRO, C. L. Missing data in time series: A review of imputation methods and case study. *Learning & Nonlinear Models, SBIC*, v. 20, n. 1, p. 31–46, 2022. Citado 2 vezes na(s) página(s) 38 e 40.

ROJAS, A. L. de. Data augmentation in economic time series: Behavior and improvements in predictions. *AIMS Mathematics*, v. 8, n. 10, p. 24528–24544, 2023. Citado 3 vezes na(s) página(s) 37, 38 e 84.

SANTOS, R. d. C. L. F. d. Agenda formal e agenda substantiva na adesão do Brasil às recomendações de governança orçamentária da OCDE. Instituto de Pesquisa Econômica Aplicada (Ipea), 2022. Citado na página 91.

SPYROMITROS-XIOUFIS, E. *et al.* Multi-target regression via input space expansion: treating targets as inputs. *Machine Learning*, Springer, v. 104, p. 55–98, 2016. Citado 4 vezes na(s) página(s) 29, 30, 31 e 32.

TAN, P.-N. *et al.* *Introduction to Data Mining*. Second. [S.l.]: Pearson Education Limited, 2020. Citado 5 vezes na(s) página(s) 11, 20, 21, 22 e 83.

TASCHE, N.; DHUNGEL, A.-K.; HEINE, M. How to design user-centered decision support systems in public budgeting? guidelines and a web-based prototype with first insights from a mixed-methods study. In: *Proceedings of Mensch und Computer 2022*. [S.l.: s.n.], 2022. p. 425–430. Citado na página 11.

THIRLWALL, A. P. The nature of economic growth: an alternative framework for understanding the performance of nations. In: *The Nature of Economic Growth*. [S.l.]: Edward Elgar Publishing, 2002. Citado na página 124.

TRAN, N. K.; KÜHLE, L. C.; KLAU, G. W. A critical review of multi-output support vector regression. *Pattern Recognit. Lett.*, v. 178, p. 69–75, 2024. Disponível em: <<https://doi.org/10.1016/j.patrec.2023.12.007>>. Citado na página 81.

VALLE-CRUZ, D.; FERNANDEZ-CORTEZ, V.; GIL-GARCIA, J. R. From e-budgeting to smart budgeting: Exploring the potential of artificial intelligence in government decision-making for resource allocation. *Government Information Quarterly*, v. 39, n. 2, p. 101644, 2022. ISSN 0740-624X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0740624X21000800>>. Citado 15 vezes na(s) página(s) ix, 13, 14, 64, 65, 66, 68, 71, 75, 76, 77, 91, 92, 134 e 137.

VALLE-CRUZ, D. *et al.* Public budget simulations with machine learning and synthetic data: Some challenges and lessons from the Mexican case. *Communications in Computer and Information Science*, v. 1666 CCIS, p. 141 – 160, 2022. Cited by: 0. Disponível em: <[https://www.scopus.com/inward/record.uri?eid=2-s2.0-85148032201&doi=10.1007%2f978-3-031-22950-3\\_12&partnerID=40&md5=89f7cbdfc5f277ee55b762f4db1e8511](https://www.scopus.com/inward/record.uri?eid=2-s2.0-85148032201&doi=10.1007%2f978-3-031-22950-3_12&partnerID=40&md5=89f7cbdfc5f277ee55b762f4db1e8511)>. Citado 3 vezes na(s) página(s) 36, 38 e 84.

- VIEIRA, F. V.; AVELLAR, A. P.; VERÍSSIMO, M. P. Indústria e crescimento econômico: evidências para países desenvolvidos e em desenvolvimento. *Brazilian Journal of Political Economy*, SciELO Brasil, v. 34, p. 485–502, 2014. Citado na página 127.
- VIEIRA, M. A.; CERETTA, P. S. Impacto das tecnologias da informação e comunicação sobre o crescimento econômico em escala global. *Economia e Sociedade*, SciELO Brasil, v. 33, n. 2, p. e238408, 2024. Citado na página 127.
- WIRSANSKY, E. *Hands-on genetic algorithms with Python: applying genetic algorithms to solve real-world deep learning and artificial intelligence problems*. [S.l.]: Packt Publishing Ltd, 2020. Citado 2 vezes na(s) página(s) 27 e 28.
- XIOUFIS, E. S. *et al.* Multi-label classification methods for multi-target regression. *CoRR*, abs/1211.6581, 2012. Disponível em: <<http://arxiv.org/abs/1211.6581>>. Citado na página 29.

# DADOS CURRICULARES

Nome: José Cláudio Guedes das Neves

Data de nascimento: 17/09/1971

Naturalidade: Vitória-ES

Nacionalidade: Brasileiro

Filiação: Antônio das Neves e Jenny Guedes das Neves

Nome em citações bibliográfica:

- J. C. G. D. Neves
- Neves, José Cláudio Guedes das

Lattes: <<http://lattes.cnpq.br/4212965309582606>>

Orcid: <<https://orcid.org/0000-0003-3966-8297>>

Formação acadêmica:

- Escola Nacional de Administração Pública (ENAP): Pós graduação em planejamento e orçamento
- *Coventry University: Master of communications management*
- Universidade Federal Fluminense (UFF): MBA em serviços de telecomunicações
- Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp): Graduação em Engenharia Elétrica