UNIVERSIDADE FEDERAL DO TOCANTINS
CÂMPUS UNIVERSITÁRIO DE PALMAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GOVERNANÇA E TRANSFORMAÇÃO DIGITAL

LUÍS CLÁUDIO FERNANDES

PREDIÇÃO DE RESTOS A PAGAR NO TRE-GO
APLICAÇÃO DE APRENDIZADO DE MÁQUINA PARA OTIMIZAÇÃO ORÇAMENTÁRIA

PALMAS (TO)
2026

Luís Cláudio Fernandes

Predição de Restos a Pagar no TRE-GO
Aplicação de Aprendizado de Máquina para Otimização Orçamentária

Dissertação apresentada ao Programa de Pós-Graduação em Governança e Transformação Digital da Universidade Federal do Tocantins como requisito para obtenção do título de Mestre(a) em Governança e Transformação Digital.

Orientador: Dr. Rafael Lima de Carvalho
Coorientador: Dr. Flávio Roldão de Carvalho Lelis

PALMAS (TO)

2026

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da Universidade Federal do Tocantins

F363p Fernandes, Luís Cláudio.

Predição de Restos a Pagar no TRE-GO: Aplicação de Aprendizado de Máquina para Otimização Orçamentária. / Luís Cláudio Fernandes. – Palmas, TO, 2026.

101 f.

Dissertação (Mestrado Profissional) - Universidade Federal do Tocantins – Câmpus Universitário de Palmas - Curso de Pós-Graduação (Mestrado Profissional) em Governança e Transformação Digital - PPGGTD, 2026.

Orientador: Rafael Lima de Carvalho

Coorientador: Flávio Roldão de Carvalho Lélis

1. Execução orçamentária. 2. Restos a pagar. 3. Otimização. 4. Modelagem preditiva. I. Título

CDD 004

Elaborado pelo sistema de geração automática de ficha catalográfica da UFT com os dados fornecidos pelo(a) autor(a).

Luís Cláudio Fernandes

Predição de Restos a Pagar no TRE-GO

Aplicação de Aprendizado de Máquina para Otimização Orçamentária

Dissertação apresentada ao Programa de Pós-Graduação em Governança e Transformação Digital, foi avaliada para a obtenção do título de Mestre(a) em Governança e Transformação Digital e aprovada em sua forma final pelo Orientador e pela Banca Examinadora.

Data da Defesa: 22 / 6 / 2026

Banca Examinadora:

Prof. Dr. Celso Gonçalves Camilo Junior (UFG)

Prof. Dr. Douglas de Oliveira Cardoso (UC)

Prof. Dr. George Lauro Ribeiro de Brito ((PPGGTD-UFT))

Profa. Dra. Suzana Gilioli da Costa Nunes (PPGGTD-UFT)

À minha esposa Karine, cujo
amor e suporte irrestrito me
impulsiona todos os dias a me
tornar uma pessoa melhor, e às
minhas filhas Louise e Isis, por
quem tudo faz sentido.

AGRADECIMENTOS

Aos meus pais Jaci e Velma, pela minha vida e por sempre acreditarem em mim.

Ao meu orientador prof. Dr. Rafael Lima, que nos momentos de maior escuridão iluminou meu caminho com sua admirável humanidade.

Ao meu coorientador prof. Dr. Flávio Roldão, que, com sua gentileza e sabedoria, sempre me motivou a continuar.

Ao meu compadre e chefe Leonardo Antônio, que viabilizou o desenvolvimento deste trabalho com todo o seu apoio e compreensão.

Aos amigos Humberto Vilani, Paulo Kliemann e Christine Helman, que me ajudaram a abrir a porta e a trilhar o caminho da investigação do orçamento público.

RESUMO

A execução orçamentária no setor público brasileiro enfrenta o desafio persistente da inscrição de despesas em Restos a Pagar (RP), fenômeno que compromete a credibilidade e a transparência do orçamento público. Este estudo de caso tem por objetivo investigar a aplicação de técnicas de aprendizado de máquina para desenvolver um modelo preditivo capaz de identificar notas de empenho a serem inscritas em Restos a Pagar ao final do exercício financeiro. A pesquisa utiliza dados extraídos do Sistema Integrado de Administração Financeira (SIAFI) e do Sistema Eletrônico de Informações (SEI) do Tribunal Regional Eleitoral de Goiás (TRE-GO), abrangendo o período de 2022 a 2025. A metodologia emprega sete algoritmos de classificação binária supervisionada: Regressão Logística, Árvore de Decisão, *Random Forest*, XGBoost, KNN, SVC e *Multilayer Perceptron* (MLP). Para tratar o desbalanceamento natural entre notas de empenho inscritas e não inscritas em RP, aplica-se a técnica de Sobreamostragem Sintética da Minoria (SMOTE). O conjunto estruturado de dados contempla variáveis relacionadas às características da despesa, ao processo de tramitação administrativa e ao contexto temporal de emissão. Os resultados demonstram que o modelo *Random Forest* alcançou o melhor desempenho global, obtendo um F1-Score de 66%, uma Área sob a Curva Precisão-Recall (AUC-PR) de 69% e uma precisão de 84% na identificação da classe minoritária. A interpretabilidade do modelo, conduzida via metodologia SHAP, revelou que o mês de emissão da nota de empenho, a interação com o ano eleitoral, o valor da despesa, a taxa de reenvios processuais e a modalidade de licitação constituem os principais fatores preditivos. O estudo conclui que é possível desenvolver uma ferramenta de apoio à decisão gerencial baseada em inteligência artificial para otimização da execução orçamentária, permitindo o monitoramento proativo e ações preventivas na gestão de despesas com maior risco de inscrição em RP. A pesquisa pode contribuir para o avanço do conhecimento na intersecção entre gestão orçamentária, eficiência administrativa e ciência de dados aplicada ao setor público.

Palavra-chave: Execução orçamentária. Restos a pagar. Otimização. Modelagem preditiva. Aprendizado de máquina.

ABSTRACT

Budget execution in the Brazilian public sector faces the persistent challenge of registering expenditures as Unpaid Obligations (Restos a Pagar - RP), a phenomenon that compromises the credibility and transparency of public budgets. This case study aims to investigate the application of machine learning techniques to develop a predictive model capable of identifying commitment notes to be registered as RP at the end of the fiscal year. The research utilizes data extracted from the Integrated Financial Administration System (SIAFI) and the Electronic Information System (SEI) of the Regional Electoral Court of Goiás (TRE-GO), covering the period from 2022 to 2025. The methodology employs seven supervised binary classification algorithms: Logistic Regression, Decision Tree, Random Forest, XGBoost, KNN, SVC, and Multilayer Perceptron (MLP). To address the natural imbalance between commitment notes registered and not registered as RP, the Synthetic Minority Oversampling Technique (SMOTE) is applied. The structured dataset encompasses variables related to expenditure characteristics, administrative processing procedures, and the temporal context of issuance. The results demonstrate that the Random Forest model achieved the best overall performance, obtaining an F1-Score of 66%, an Area Under the Precision-Recall Curve (AUC-PR) of 69%, and a precision of 84% in identifying the minority class. The model's interpretability, conducted via the SHAP methodology, revealed that the month of commitment note issuance, the interaction with the electoral year, the expenditure value, the rate of procedural resubmissions, and the procurement modality constitute the main predictive factors. The study concludes that it is possible to develop an artificial intelligence-based managerial decision support tool for budget execution optimization, enabling proactive monitoring and preventive actions in managing expenditures with a higher risk of RP registration. The research may contribute to advancing knowledge at the intersection of budget management, administrative efficiency, and data science applied to the public sector.

Keywords: Budget execution. Unpaid obligations. Optimization. Predictive modeling. Machine learning.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

O orçamento público constitui instrumento fundamental da administração pública, materializando o planejamento governamental através da previsão de receitas e fixação de despesas para um período determinado. No Brasil, o sistema orçamentário estrutura-se em torno de três instrumentos integrados - Plano Plurianual, Lei de Diretrizes Orçamentárias e Lei Orçamentária Anual. Este modelo integrado compõe o Sistema de Planejamento e Orçamento Federal (Abraham, 2025, p. 20), que instrumentaliza a "presença e atuação do Estado perante a sociedade, nas áreas econômica e social..." (Abraham, 2025, p. 5).

A execução orçamentária, regida pelo princípio da anualidade, estabelece que as despesas autorizadas devem ser realizadas dentro do exercício financeiro correspondente. Contudo, a prática administrativa brasileira revela um fenômeno persistente que desafia este princípio: a inscrição de despesas empenhadas mas não pagas até o final do ano de exercício, conhecida como Restos a Pagar (Giacomoni, 2012, p. 336).

Os Restos a Pagar, embora constituam instrumento necessário para garantir a continuidade da ação administrativa, têm apresentado crescimento significativo nas últimas décadas. Dados do Tesouro Nacional indicam que o volume de Restos a Pagar de 2020 a 2024 é de aproximadamente 45% no Poder Legislativo, 25% no Judiciário e 15% no Executivo. Este crescimento concentra-se principalmente nas despesas discricionárias, em especial investimentos e outras despesas correntes, enquanto as despesas obrigatórias, como pessoal e encargos, mantêm baixos índices de inscrição.

A problemática dos Restos a Pagar transcende a mera questão contábil, impactando a credibilidade do orçamento como instrumento de planejamento, reduzindo a transparência das contas públicas e comprometendo a programação financeira dos exercícios subsequentes (Aquino; Azevedo, 2017, p. 591)[1]. Valle-Cruz et al. (2020) destacam que a aplicação de técnicas de inteligência artificial no setor público, em particular no orçamento público, pode contribuir para melhorar a tomada de decisão governamental, oferecendo novas perspectivas para problemas complexos da administração pública.

Dentro deste contexto, este trabalho procura fazer uma contribuição na área de gestão orçamentária através da aplicação de técnicas de aprendizado de máquina para desenvolver modelos preditivos capazes de identificar despesas a serem inscritas em Restos a Pagar.

[1] Optou-se pela inclusão deste artigo de 2017 no presente trabalho por tratar especificamente da transparência e perda de credibilidade orçamentária causadas pelos Restos a Pagar.

1.1 Problema de pesquisa

A inscrição excessiva de despesas em Restos a Pagar representa um dos principais desafios da execução orçamentária brasileira, afetando todos os níveis de governo. No acórdão do TCU nº 2.823/2015, o ministro José Múcio Monteiro afirma:

> O uso desmesurado de inscrições de obrigações financeiras na rubrica Restos a Pagar configura desvirtuamento do princípio da anualidade. [...] a rubrica Restos a Pagar apresentou um aumento expressivo de seu montante nos últimos cinco exercícios; apesar de, em princípio, não violar o princípio da legalidade, [...] sua utilização tem sido desvirtuada, pois deixou ela de ser uma ferramenta de exceção para tornar-se uma modalidade amplamente utilizada de execução da despesa, [...].

Abraham (2025, p. 192) afirma que as despesas inscritas em Restos a Pagar "devem sempre apresentar crédito financeiro suficiente e adequado para seu adimplemento", e que não se pode conceber o comprometimento de recursos financeiros arrecadados nos anos subsequentes para sua quitação, sob o risco de se violarem as boas práticas orçamentárias e a gestão fiscal responsável.

A complexidade do problema decorre da interação entre múltiplas variáveis que influenciam a execução da despesa pública. Fatores como deficiências no planejamento da execução orçamentária, a complexidade intrínseca de determinadas naturezas de despesa e de modalidades de licitação, dificuldade em se antecipar ações relacionadas às contratações e aquisições, e deficiências ou inexistência de modelagem de processos de negócio relacionadas ao planejamento e execução orçamentária concorrem de forma não-linear, dificultando a identificação de padrões através de métodos tradicionais de análise.

O mapeamento de literatura realizado para o desenvolvimento deste trabalho, abordado no Capítulo 3, identificou o interesse dos pesquisadores pela problemática dos Restos a Pagar. Abordagens descritivas e preditivas para a caracterização desta problemática e proposições foram identificadas.

Diante deste cenário, o presente trabalho busca responder ao seguinte problema de pesquisa: Modelos preditivos de aprendizado de máquina são capazes de identificar quais notas de empenho serão inscritas em Restos a Pagar no TRE-GO, e quais fatores orçamentários e processuais mais influenciam nessa identificação?

1.1.1 Hipótese

A hipótese central desta pesquisa fundamenta-se na premissa de que as características da despesa e os padrões temporais e de tramitação processual contêm informações suficientes para predizer as Notas de Empenho que serão inscritas em Restos a Pagar no exercício orçamentário corrente.

Pode-se enunciar a hipótese principal da seguinte forma: “É possível desenvolver e avaliar a aplicabilidade de um modelo preditivo para identificar notas de empenho a serem inscritas em Restos a Pagar, utilizando-se variáveis relacionadas às características da despesa (valor, natureza, modalidade de licitação), aspectos temporais (mês de empenho, ano eleitoral) e métricas de tramitação processual (quantidade de tramitações, tempo de paralisação, número de reenvios)”.

Esta hipótese baseia-se na observação de que determinados tipos de despesa, como por exemplo aquelas de maior complexidade processual, apresentam maior propensão à inscrição em Restos a Pagar, e que algoritmos de aprendizado de máquina podem capturar esses padrões de forma mais eficaz que métodos tradicionais de análise.

1.1.2 Delimitação de Escopo

O presente estudo delimita-se ao desenvolvimento e avaliação comparativa de modelos preditivos para identificação de notas de empenho a serem inscritas em Restos a Pagar, utilizando dados do Tribunal Regional Eleitoral de Goiás.

O trabalho abrange as notas de empenho emitidas pelo TRE-GO no período de 2022 a 2025, sobre as quais são desenvolvidos e comparados sete algoritmos de aprendizado de máquina com diferentes abordagens (KNN, Regressão Logística, Árvore de Decisão, Random Forest, XGBoost, SVC e MLP), com aplicação da técnica de balanceamento SMOTE e otimização de hiperparâmetros via *GridSearchCV*.

O desempenho dos modelos é avaliado por métricas apropriadas à classificação binária desbalanceada, e o modelo de melhor resultado é submetido à análise de interpretabilidade por meio de valores SHAP, com vistas a identificar os fatores determinantes da inscrição em Restos a Pagar.

O trabalho não abrange outros órgãos ou esferas de governo, tampouco a generalização dos resultados para contextos institucionais distintos do TRE-GO. Estão igualmente fora do escopo o desenvolvimento de sistema informatizado para implementação operacional e a análise de impactos financeiros ou fiscais decorrentes dos Restos a Pagar.

1.2 Justificativa

O volume de Restos a Pagar inscritos anualmente pelos órgãos da administração pública federal revela uma tensão estrutural entre o planejamento orçamentário e a capacidade efetiva de execução das despesas dentro do exercício financeiro. No TRE-GO, esse fenômeno se repete de forma recorrente: parcela significativa das notas de empenho emitidas ao longo do exercício não é liquidada até 31 de dezembro, sendo inscrita em Restos a Pagar e carregada para o exercício seguinte. Essa situação compromete a previsibilidade da execução orçamentária, dificulta o planejamento de caixa e pode resultar no cancelamento de despesas que não são liquidadas dentro do prazo legal, afetando a

continuidade de serviços e contratos.

A despeito da recorrência do problema, a gestão orçamentária ainda opera predominantemente de forma reativa: a inscrição em Restos a Pagar é constatada ao final do exercício, quando as possibilidades de intervenção já são limitadas. Não há, no TRE-GO, instrumento sistemático que permita identificar, durante o exercício, quais empenhos apresentam maior risco de não serem liquidados ou pagos a tempo. Essa lacuna operacional é um forte motivador da presente pesquisa.

Do ponto de vista científico, a literatura sobre aprendizado de máquina aplicado à gestão orçamentária pública ainda tem muito a crescer, considerando-se o momento do desenvolvimento deste trabalho. Dos estudos relacionados a RP apresentados no Capítulo 3, apenas Santos (2023) apresenta um estudo voltado especificamente à predição de Restos a Pagar e à interpretação dos fatores que os determinam. Diante do potencial de melhoria da gestão do orçamento público por meio da predição de notas de empenho a serem inscritas em RP, o assunto merece ser explorado.

A disponibilidade de dados históricos estruturados no TRE-GO, aliada ao amadurecimento das técnicas de aprendizado de máquina interpretável, cria condições favoráveis para o desenvolvimento de um modelo preditivo que possa subsidiar decisões de priorização orçamentária durante o exercício. Justifica-se, assim, a realização desta pesquisa pela convergência entre uma necessidade institucional identificada e a viabilidade técnica de abordá-las conjuntamente.

A viabilidade da solução proposta fundamenta-se na disponibilidade de dados estruturados nos sistemas corporativos do órgão estudado, na maturidade das técnicas de aprendizado de máquina aplicadas, e no crescente interesse pela aplicação dessas técnicas em contextos similares. Por exemplo:

- Valle-Cruz et al. (2020) demonstraram o potencial da aplicação de técnicas de Inteligência Artificial, como redes neurais e algoritmos genéticos, para prever o impacto da alocação orçamentária em áreas como desenvolvimento social e economia no México;

- Zarzà et al. (2023) explora o uso de *Large Language Models* para aconselhamento de orçamento doméstico, apontando para a possibilidade de aplicação em um contexto organizacional.

1.3 Objetivos

1.3.1 Objetivo Geral

O objetivo geral do trabalho pode ser assim expresso:

"Desenvolver e avaliar comparativamente modelos preditivos de aprendizado de máquina para a detecção antecipada de notas de empenho a serem inscritas em Restos a

Pagar no TRE-GO".

1.3.2 Objetivos Específicos

1. Verificar se dados orçamentários e processuais disponíveis no TRE-GO permitem antecipar a inscrição de notas de empenho em Restos a Pagar;

2. Identificar o algoritmo de aprendizado de máquina que oferece o melhor equilíbrio entre precisão e sensibilidade na predição de Restos a Pagar;

3. Identificar os fatores que mais influenciam a probabilidade de inscrição de uma nota de empenho em Restos a Pagar no TRE-GO;

4. Avaliar em que medida as predições do modelo são úteis como ferramenta de apoio à tomada de decisão orçamentária.

1.3.3 Estrutura da Dissertação

O trabalho está organizado em 6 capítulos correlacionados. O Capítulo 1 apresentou por meio de sua contextualização o tema proposto neste trabalho. Também foram estabelecidos os resultados esperados por meio da definição de seus objetivos e apresentadas as limitações do trabalho, permitindo uma visão clara do escopo proposto.

O Capítulo 2 apresenta a fundamentação teórica, que aborda conceitos relacionados a orçamento público e conceitos de aprendizado de máquina.

O Capítulo 3 apresenta trabalhos relacionados ao tema desta dissertação. Foram selecionados trabalhos com temática relevante a Restos a Pagar, preferencialmente num horizonte temporal de 5 anos em relação ao momento do desenvolvimento deste trabalho.

O Capítulo 4 apresenta os procedimentos metodológicos utilizados para o desenvolvimento deste trabalho, desde a preparação do conjunto de dados de notas de empenho até a aplicação de técnicas de aprendizado de máquina.

O Capítulo 5 apresenta os resultados da pesquisa e uma discussão sobre a interpretação dos coeficientes encontrados com a criação do modelo preditivo.

No Capítulo 6 são tecidas as conclusões do trabalho, relacionando os objetivos identificados inicialmente com os resultados alcançados. São ainda propostas possibilidades de continuação da pesquisa desenvolvida a partir das experiências adquiridas com a execução do trabalho, e abordadas limitações identificadas.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, apresentam-se os conceitos essenciais relacionados a orçamento público para que se possa compreender o restante deste trabalho, abordando desde fundamentos do orçamento público brasileiro até as técnicas de aprendizado de máquina aplicadas ao problema de pesquisa, passando pela problemática específica dos Restos a Pagar e as soluções tecnológicas propostas.

Na Seção 2.1, explicam-se os conceitos fundamentais do orçamento público. Na Seção 2.2, abordam-se os elementos do ciclo orçamentário brasileiro de forma simplificada. Na seção 2.3, abordam-se as classificações de despesas relevantes à compreensão desta dissertação. Na Seção 2.4, apresenta-se a problemática dos Restos a Pagar, objeto de estudo central do presente trabalho. Na Seção 2.5, abordam-se fundamentos de aprendizado de máquina. Por fim, na Seção 2.6, são apresentados os fundamentos dos modelos preditivos e das métricas utilizadas nessa dissertação.

2.1 Conceitos Fundamentais do Orçamento Público

Nesta subseção, aborda-se a caracterização do orçamento público, os princípios orçamentários essenciais à compreensão deste trabalho e discorre-se sobre as leis orçamentárias.

Natureza e Definição do Orçamento Público

O orçamento público constitui instrumento fundamental da administração pública, diferenciando-se substancialmente do orçamento doméstico ou empresarial por sua natureza política, econômica, jurídica, administrativa, financeira e contábil (Giacomoni, 2012, p. 54).

Segundo Abraham (2025, p. 40), o orçamento público pode ser definido como "o instrumento de planejamento, administração e controle financeiro estatal que possibilita prever receitas e fixar despesas em um determinado lapso temporal, de forma transparente, equilibrada e eficiente".

Esta definição evidencia três características essenciais do orçamento público brasileiro. Primeiro, seu caráter legal, uma vez que constitui lei em sentido formal e material, aprovada pelo Poder Legislativo e sancionada pelo Poder Executivo. Segundo, sua função de previsão e autorização, estabelecendo limites para a arrecadação de receitas e a realização de despesas. Terceiro, sua temporalidade definida, coincidindo com o ano civil e respeitando o princípio da anualidade orçamentária. Os princípios orçamentários mais relevantes ao presente trabalho são descritos na Subseção 2.1.2.

Princípios Orçamentários Fundamentais

Os princípios orçamentários constituem diretrizes que orientam a elaboração, execução e controle do orçamento público, conferindo-lhe racionalidade, transparência e efetividade. Para os propósitos deste trabalho, destacam-se três princípios fundamentais que se relacionam diretamente com a problemática dos Restos a Pagar.

O princípio da **anualidade** estabelece que o orçamento deve ser elaborado e executado para um período determinado, coincidente com o ano civil. Giacomoni (2012, p. 73) explica que este princípio encontra aceitação praticamente unânime entre as nações modernas. A anualidade orçamentária constitui mecanismo fundamental de controle democrático, permitindo que o Legislativo exerça sua função fiscalizadora de forma regular e sistemática (Giacomoni, 2012, p. 343).

O princípio da **universalidade** determina que o orçamento deve conter todas as receitas e despesas do Estado, sem exceções. Abraham (2025, p. 116) destaca que o cumprimento deste princípio faz com que todas as receitas e despesas do Estado passem pelo processo orçamentário de elaboração, aprovação, execução e controle. A universalidade facilita o controle e a fiscalização dos recursos públicos, contribuindo para a transparência da gestão fiscal.

O princípio da **legalidade** estabelece que a administração pública se obriga a realizar suas atividades de acordo com o previsto nas leis orçamentárias (Abraham, 2025, p. 111). O arcabouço jurídico em matéria orçamentária prevê que a competência para a elaboração dos projetos de lei do plano plurianual, de diretrizes orçamentárias e do orçamento anual é do Presidente da República.

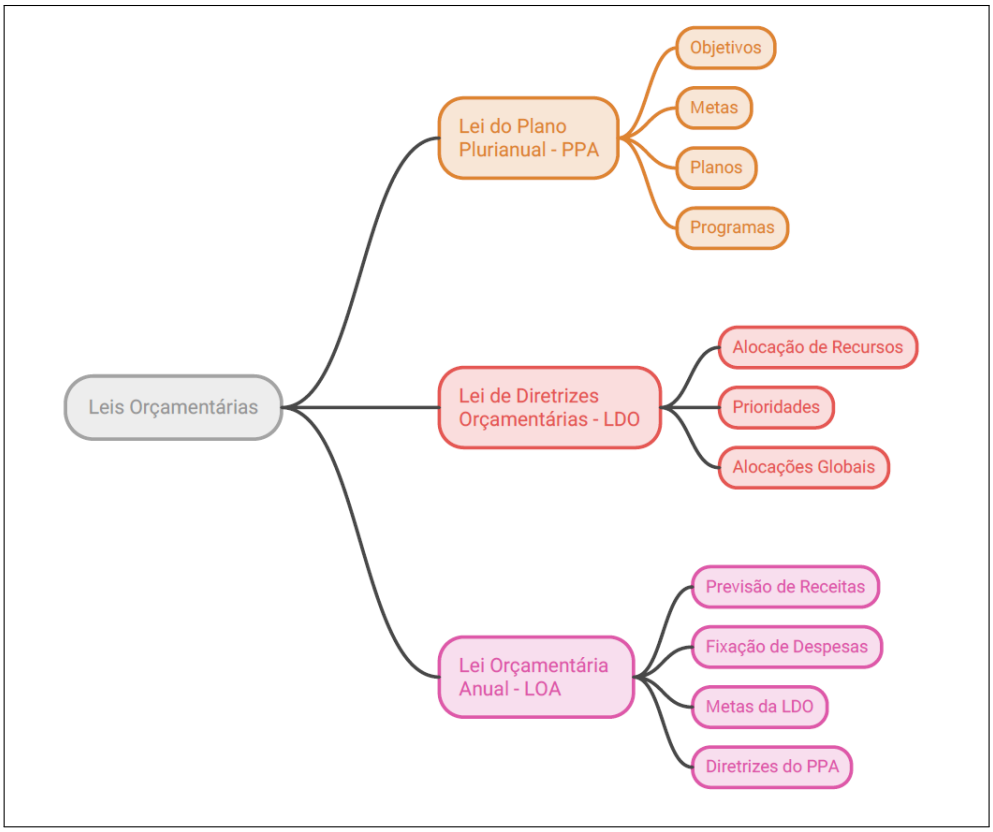
As Leis Orçamentárias

O sistema brasileiro de planejamento e orçamento público é estruturado por três leis interdependentes, previstas no art. 165 da Constituição Federal: o PPA, a LDO e a LOA. Em conjunto, elas organizam a atuação estatal em diferentes níveis de planejamento, articulando objetivos de médio prazo, prioridades anuais e execução concreta das receitas e despesas públicas. Na Figura 1, as três leis orçamentárias e suas principais características são apresentadas.

A **Lei do Plano Plurianual - PPA** encontra-se no nível estratégico. Contém os objetivos, metas, planos e programas da Administração Pública em um horizonte temporal de quatro anos. Deve ser encaminhado pelo Executivo ao Legislativo até 31 de agosto do primeiro ano de cada governo. Tal mecanismo permite a continuidade administrativa em caso de mudança de governo.

A **Lei de Diretrizes Orçamentárias - LDO** tem por finalidade constituir o planejamento operacional do governo. Desta forma, posiciona-se no nível tático. Possui duração de um ano e contém “definições fundamentais de alocação de recursos, grandes

Figura 1 – Leis Orçamentárias



Fonte: Elaborado pelo autor, com o apoio de Napkin AI

prioridades e alocações globais, por áreas ou setores" (Abraham, 2025, p. 89). Sua elaboração deve seguir as diretrizes do PPA.

Já a **Lei Orçamentária Anual - LOA** pode ser considerada o orçamento público propriamente dito (*Ibidem*, p. 89). Contém a previsão de receitas e fixação das despesas para a concretização das metas e prioridades da LDO e das diretrizes, objetivos e metas do PPA. A Constituição brasileira, em seu art. 165, § 8º, prevê que não se pode fazer constar na LOA matéria estranha à previsão de receitas e fixação de despesas, para evitar a tentativa de se aprovar conteúdos não relacionados ao orçamento público.

2.2 O Processo Orçamentário e a Execução da Despesa

Nesta seção, abordam-se as fases do processo orçamentário e as fases obrigatórias da execução da despesa pública.

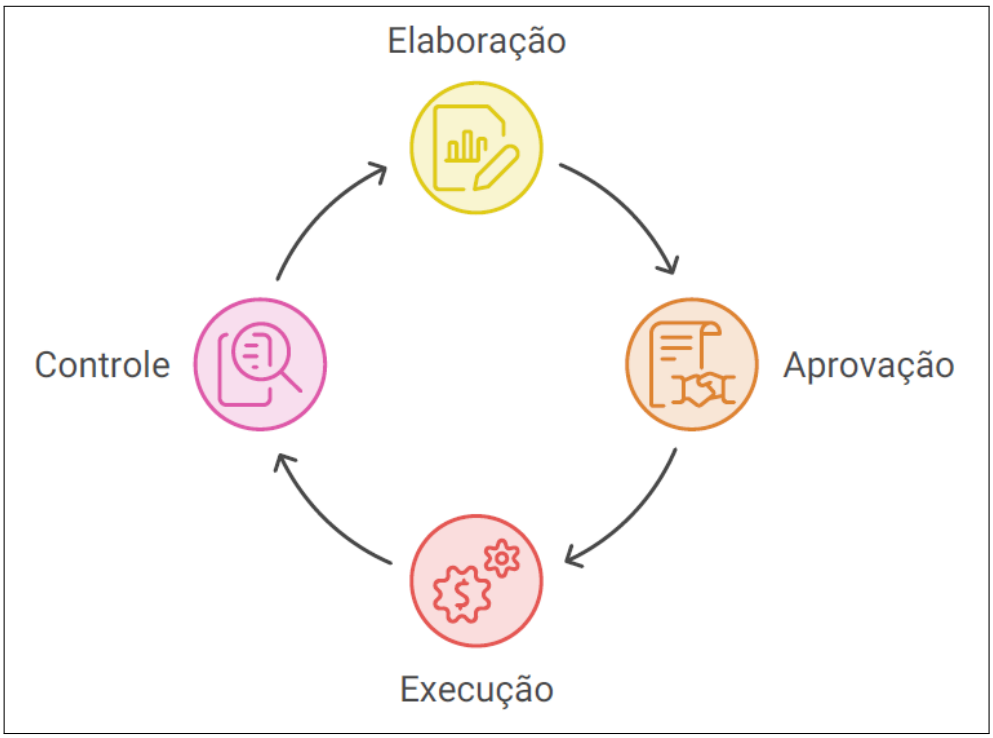
Fases do Processo Orçamentário Brasileiro

O processo orçamentário brasileiro, também conhecido como ciclo orçamentário, compreende quatro fases distintas e inter-relacionadas: elaboração, aprovação, execução e controle. Trata-se de um processo contínuo e sistemático que se renova anualmente,

garantindo a continuidade da ação governamental e fornecendo instrumentos de controle dos recursos públicos (Abraham, 2025, p. 236).

A Figura 2 representa de forma esquemática a inter-relação cíclica entre as quatro fases do processo orçamentário do Brasil.

Figura 2 – Processo Orçamentário



Fonte: Adaptado de Giacomoni (2012, p. 217), com o apoio de Napkin AI

A fase de elaboração inicia-se com a definição de diretrizes macroeconômicas e culmina com o envio da proposta orçamentária ao Poder Legislativo. Esta fase envolve a participação de múltiplos atores, incluindo órgãos centrais de planejamento, órgãos setoriais e unidades orçamentárias, em processo que busca conciliar demandas setoriais com restrições fiscais e prioridades governamentais. O TRE-GO enquadra-se na categoria de unidade orçamentária.

A fase de aprovação compreende a análise, discussão e votação da proposta orçamentária pelo Poder Legislativo. Giacomoni (2012, p. 270-276) explica que esta fase contém subdivisões. Primeiramente, apresenta-se o projeto da LOA no Congresso. Em seguida, os parlamentares discutem e votam emendas, que são alterações ao projeto de lei original.

Após esta etapa, o projeto segue para a sanção presidencial, que tem 15 dias para vetar total ou parcialmente o projeto da LOA. Em caso de veto total ou parcial, o Congresso tem até 30 dias para rejeitar o veto. Por fim, o projeto de lei é encaminhado à Presidência para promulgação. Destaca-se que a rejeição é um evento raro no Brasil.

A fase de execução é a responsável por concretizar os programas e as ações go-

vernamentais, por meio da realização das despesas fixadas (Abraham, 2025, p. 189). Os recursos orçamentários são distribuídos às unidades orçamentárias ou a órgãos centrais de administração geral, com a devida previsão legal.

A fase de controle provê mecanismos de acompanhamento e fiscalização necessários a qualquer atividade humana para prevenir falhas e erros, sejam eles involuntários ou intencionais (*Ibidem*, p. 209). O acompanhamento da execução orçamentária pode ser feito por qualquer interessado (os cidadãos, as instituições e órgãos de fiscalização e controle), que podem acessar relatórios periódicos de divulgação obrigatória.

Já a fiscalização é um procedimento formal realizado principalmente pelo controle interno e por Tribunais e Conselhos de Contas. Caso se identifiquem desvios, o controle é aplicado para retificação de tais equívocos e/ou irregularidades.

Execução da Despesa Pública

A execução da despesa pública constitui fase crucial do ciclo orçamentário, materializando o planejamento governamental através da realização efetiva dos gastos autorizados. No Brasil, a execução da despesa segue as etapas clássicas estabelecidas pela Lei nº 4.320/1964: empenho, liquidação e pagamento.

O **empenho** constitui o primeiro estágio da execução da despesa, representando, conforme o art. 58 da referida lei, "o ato emanado de autoridade competente que cria para o Estado obrigação de pagamento pendente ou não de implemento de condição". O empenho significa a reserva de dotação orçamentária para fazer face a compromisso assumido. Esta reserva é formalizada através da nota de empenho, documento que especifica o credor, a importância, a dotação orçamentária e o fundamento legal da despesa.

A **liquidação** representa o segundo estágio, consistindo na verificação do direito adquirido pelo credor com base nos títulos e documentos comprobatórios do respectivo crédito. Em outras palavras, se houve a devida entrega de produto ou serviço, o fornecedor ou prestador tem o direito de receber o pagamento. Abraham (2025, p. 191) explica que a liquidação apura a origem e o objeto do que se deve pagar, a importância exata a pagar e a quem se deve pagar, com o objetivo de dar fim à obrigação e assegurar a regularidade da despesa.

O **pagamento** constitui o estágio final, efetuando-se através da transferência de recursos financeiros ao credor, extinguindo a obrigação (*Ibidem*, p. 191). O pagamento só pode ser efetuado após a regular liquidação da despesa, garantindo que o Estado cumpra suas obrigações de forma legal e transparente.

2.3 Classificações da Despesa Pública

Nesta subseção, apresentam-se a classificação por natureza de despesa e as modalidades de licitação relevantes ao desenvolvimento deste trabalho.

Classificação por Natureza da Despesa

A classificação da despesa por natureza constitui instrumento fundamental para a compreensão dos diferentes tipos de gastos públicos e suas características específicas. Esta classificação, estabelecida pela Portaria Interministerial STN/SOF nº 163/2001, organiza as despesas em categorias que refletem sua natureza econômica e finalidade.

As **despesas de pessoal e encargos sociais** compreendem gastos com remuneração de servidores ativos e inativos, bem como encargos sociais e benefícios. Trata-se de despesas de natureza obrigatória e fluxo contínuo, apresentando alta previsibilidade e baixa discricionariedade gerencial. Esta característica torna as despesas de pessoal menos propensas à inscrição em Restos a Pagar, uma vez que seguem cronograma regular de pagamento.

As **despesas correntes** abrangem gastos com material de consumo, serviços de terceiros, diárias, auxílios e outras despesas de custeio. Estas despesas apresentam maior variabilidade e permitem maior poder de escolha aos gestores públicos, dependendo de sua natureza e do contexto da execução orçamentária. Por essa razão, são mais suscetíveis à inscrição em Restos a Pagar.

Já as **despesas de investimento** referem-se a gastos com obras, instalações, equipamentos e aquisição de bens de capital. Caracterizam-se por projetos de maior porte e complexidade, com prazos de execução mais longos e, consequentemente, maior risco de inscrição em Restos a Pagar. A natureza plurianual de muitos investimentos contribui para que parte da despesa empenhada em um exercício seja liquidada e paga em exercícios subsequentes.

Modalidades de Licitação

As modalidades de licitação constituem procedimentos específicos para a seleção de fornecedores, influenciando diretamente os prazos e a complexidade da execução da despesa. A Lei nº 14.133/2021 (Nova Lei de Licitações) estabelece as modalidades atualmente vigentes.

O **pregão** constitui modalidade destinada à aquisição de bens e serviços comuns, caracterizando-se pela celeridade e simplicidade procedimental. Permite execução mais rápida da despesa, reduzindo a probabilidade de inscrição em Restos a Pagar devido aos prazos reduzidos entre a licitação e a entrega.

A **concorrência** destina-se a contratações de maior valor e complexidade, especialmente obras e serviços de engenharia. Envolve procedimentos mais longos e complexos, aumentando a probabilidade de que a execução da despesa se estenda além do exercício de empenho.

A **dispensa** e **inexigibilidade de licitação** aplicam-se a casos específicos previstos em lei, permitindo contratação direta. Estas modalidades podem acelerar a execução

da despesa em situações específicas, mas requerem justificação detalhada e controle rigoroso.

O **suprimento de fundos**, por sua vez, constitui um adiantamento de numerário a servidor (chamado de suprido) para despesas que, por sua excepcionalidade, caráter urgente ou pequeno vulto, não podem aguardar o processo normal de aplicação (licitação/empenho prévio ao fornecedor).

2.4 Restos a Pagar: Conceito e Problemática

Nesta subseção, apresenta-se a natureza jurídica dos Restos a Pagar, um pouco da sua evolução histórica, a magnitude do problema e seus impactos na gestão do orçamento público.

Definição e Natureza Jurídica

Por vezes, não é viável à Administração Pública pagar todas as despesas até a data legalmente prevista de 31 de dezembro do ano de exercício, especialmente por questões operacionais (Abraham, 2025, p. 192). Quando isto ocorre, as despesas são inscritas em Restos a Pagar, ou seja, são reconhecidas em exercícios posteriores aos originários. Os Restos a Pagar constituem despesas empenhadas mas não pagas até a data limite, distinguindo-se em processados e não processados conforme o estágio de execução alcançado.

Os **Restos a Pagar Processados** referem-se a despesas empenhadas e liquidadas, mas não pagas até o encerramento do exercício. Estas despesas representam obrigações líquidas e certas do Estado, uma vez que o bem foi entregue ou o serviço foi prestado, restando apenas o pagamento.

Os **Restos a Pagar Não Processados** compreendem despesas empenhadas mas não liquidadas até 31 de dezembro. Em outras palavras, existe o reconhecimento do adimplemento da Administração Pública, mas ainda não houve a entrega do bem ou serviço, representando compromissos assumidos mas não efetivados.

Para o contexto deste trabalho, tanto os Restos a Pagar processados como os não processados foram incluídos no conjunto de dados de criação dos modelos preditivos, por se considerar que o foco é o **não pagamento no ano de exercício**, independentemente da entrega do bem ou serviço por parte do contratado.

Evolução Histórica e Magnitude do Problema

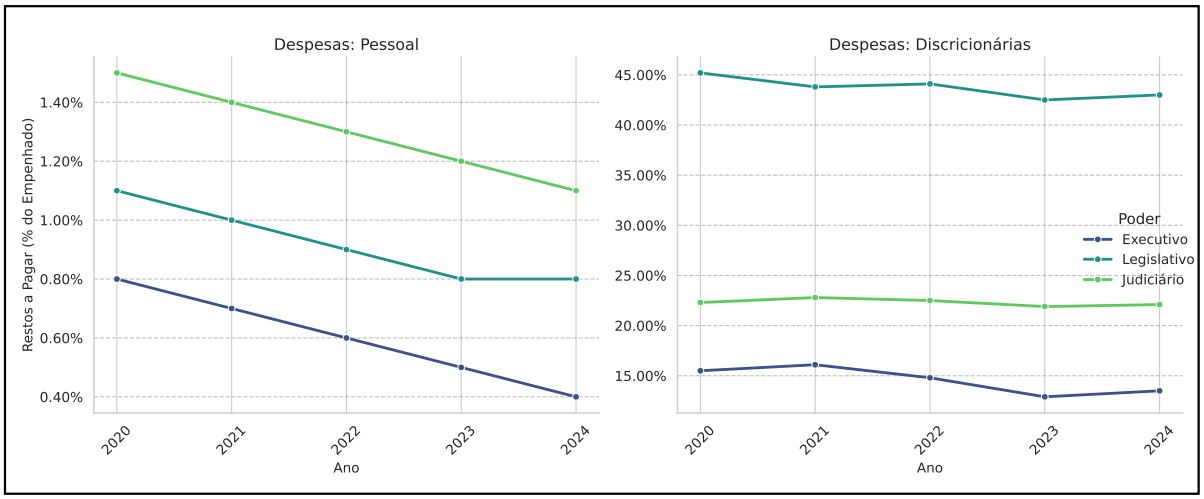
A problemática dos Restos a Pagar tem se intensificado nas últimas décadas, afetando todos os níveis de governo. Esta evolução concentra-se principalmente nas despesas discricionárias, especialmente investimentos e outras despesas correntes, enquanto as despesas obrigatórias, como pessoal e encargos sociais, mantêm baixos índices de inscrição.

Pode-se observar, pela análise de dados do Tesouro Transparente (Tesouro Nacional, 2025), que o volume percentual de despesas discricionárias inscritas em Restos a Pagar de 2020 a 2024, nos três Poderes da União, continua altíssimo em relação às despesas de pessoal, como mostra a Figura 3.

A figura exibe dois gráficos comparativos que evidenciam a disparidade na inscrição de Restos a Pagar (RP) entre diferentes tipos de despesa. Enquanto as despesas de pessoal mantêm taxas residuais (abaixo de 1,5%), as despesas discricionárias apresentam percentuais alarmantes, variando de 15% a 45%, com o Poder Executivo liderando os índices. Apesar da tendência geral de queda no período, os dados confirmam que o problema dos RP é estrutural e afeta majoritariamente o orçamento de custeio e investimento.

Abraham (2025, p. 192) destaca que tal artifício pode estar sendo usado como instrumento indireto de rolagem de dívida, e deve-se a fatores como receitas desconectadas da realidade e infladas, bem como à prática de não se realizar o acompanhamento adequado da execução orçamentária ao longo do ano.

Figura 3 – Evolução do Percentual de Restos a Pagar



Fonte: Elaborado pelo autor (2025), com dados de Tesouro Nacional (2025)

Impactos na Gestão Orçamentária

A inscrição excessiva de despesas em Restos a Pagar gera múltiplos impactos negativos na gestão orçamentária. Termos como "orçamento de restos a pagar" e "orçamento paralelo" são indicativos de que o problema é crônico no Brasil. Aquino e Azevedo (2017) identificam três dimensões principais destes impactos: endividamento, credibilidade e transparência.

Os autores sugerem que o uso excessivo deste artifício é indicativo da baixa saúde financeira dos governos das três esferas (federal, estadual e municipal). Os mecanismos de controle pelos Tribunais e Conselhos de Contas dependem da atuação do Legislativo no

sentido de referendar a reprovação das contas do Presidente da República, dos governadores e dos prefeitos, o que "dependeria da dinâmica política de coalização no momento" (Melo et al., 2009 apud Aquino; Azevedo, 2017, p. 591) .

Quanto à credibilidade, a inscrição excessiva em Restos a Pagar ocasiona a não execução completa do orçamento do exercício, pois parte da receita já está comprometida, tornando o orçamento uma "peça de ficção" para a opinião pública. Além disso, como os Restos a Pagar têm prioridade sobre as despesas executadas no ano (Lei no 8.666/1993, art. 5º), ocorre um efeito cascata que gera atraso em pagamentos de fornecedores e empurra despesas para o ano seguinte, o que estabelece um círculo vicioso no orçamento público (*Ibidem*, p. 592).

Na dimensão da transparência, os Restos a Pagar obscurecem a real situação dos orçamentos dos entes federativos. Como a regulação atual (art. 48 da Lei de Responsabilidade Fiscal) não alcança as despesas executadas pelo "orçamento paralelo", a consulta a informações orçamentárias mascara o montante observado. Tal fato reduz a possibilidade de controle por parte do Legislativo e permite o uso da cifra do orçamento para gerenciar limites fiscais de forma a agravar a situação do orçamento público (*Ibidem*, p. 592).

2.5 Aprendizado de Máquina: Conceitos e Aplicações

Esta subseção apresenta elementos de aprendizado de máquina essenciais à compreensão do modelo preditivo desenvolvido no presente trabalho.

Fundamentos do Aprendizado de Máquina

O aprendizado de máquina constitui subcampo da inteligência artificial que se concentra no desenvolvimento de algoritmos capazes de aprender padrões a partir de dados, sem serem explicitamente programados para cada tarefa específica. Segundo Raschka e Mirjalili (2019, p. 1), aprendizado de máquina é uma subárea da inteligência artificial que envolve "algoritmos capazes de aprender por si mesmos que produzem conhecimento a partir de dados com o objetivo de fazer previsões".

Esta definição evidencia três características fundamentais do aprendizado de máquina. Primeiro, sua capacidade de generalização, permitindo que algoritmos façam predições sobre dados não observados anteriormente. Segundo, sua natureza adaptativa, possibilitando melhoria de performance com a exposição a novos dados. Terceiro, sua independência de programação explícita para cada situação específica.

Paradigmas de Aprendizado

Embora haja diferentes formas de classificação a respeito dos paradigmas de aprendizado de máquina, optou-se neste trabalho pela utilização de uma organização em três

paradigmas principais, cada um adequado a diferentes tipos de problemas e disponibilidade de dados. Raschka e Mirjalili (2019, p. 2) classificam estes paradigmas conforme a natureza dos dados de treinamento e o tipo de resposta disponível.

O aprendizado supervisionado utiliza conjuntos de dados de treinamento rotulados, por exemplo, sexo do paciente e seus sintomas. A partir deste conjunto de objetos (pacientes) conhecidos, pode-se prever para um novo objeto o tipo de doença a partir do seu conjunto de atributos (Faceli et al., 2021, p. 3). Este paradigma subdivide-se em problemas de classificação (predição de categorias) e regressão (predição de valores contínuos, como o montante de uma nota de empenho).

O aprendizado não supervisionado trabalha com dados não rotulados, buscando descobrir padrões ocultos ou estruturas subjacentes nos dados. Este paradigma inclui técnicas como agrupamento (dados de entradas são divididos em grupos segundo sua similaridade), sumarização (busca-se uma descrição simples e compacta para um conjunto de dados) e associação (procuram-se padrões frequentes de associações entre os atributos de um conjunto de dados (Faceli et al., 2021, p. 3)).

O aprendizado por reforço baseia-se na interação com um ambiente, onde um agente aprende através de tentativa e erro, recebendo recompensas ou punições por suas ações. Este paradigma é particularmente adequado para problemas de tomada de decisão sequencial (Raschka; Mirjalili, 2019, p. 6).

Dados e sua Caracterização

Técnicas de aprendizado de máquina são aplicadas a conjuntos de dados que representam entidades do mundo real e suas características. Por exemplo, em um contexto hospitalar, um conjunto de dados *Pacientes* pode ser representado de maneira similar à Tabela 1, em que cada paciente possui um conjunto de características: *Id.* (identificação do paciente),*Nome*, *Idade*, *Sexo*, *Peso*, *Manchas* (presença e distribuição de manchas pelo corpo), *Temp.* (temperatura corporal), *#Int* (número de internações) e *Diagnóstico*, que indica o diagnóstico do paciente e corresponde ao *atributo alvo*, ou seja, aquilo que se deseja prever dadas as características do paciente. Tem-se, na Tabela 1, um exemplo de conjunto de dados de pacientes.

Tabela 1 – Conjunto de dados *Pacientes*

Id	Nome	Idade	Sexo	Peso	Manchas	Temp.	#Int	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2	Doente
3217	Maria	18	F	79	Inexistentes	39,5	4	Doente
4039	Luiz	49	M	79	Espalhadas	38,0	2	Saudável
1920	José	18	M	79	Inexistentes	38,5	8	Doente
4340	Cláudia	21	F	79	Uniformes	37,6	1	Saudável

Fonte: Adaptado de Faceli et al. (2021)

Para os propósitos do presente trabalho, adota-se a classificação por **tipo** e **escala**. O **tipo** define se o atributo representa **quantidades** (e neste caso é denominado quantitativo ou numérico) ou **qualidades** (caso em que é denominado qualitativo, simbólico ou categórico (Faceli et al., 2021)).

Valores quantitativos podem ser contínuos (como a temperatura) ou discretos (como a idade do paciente, quando expressa em anos, por exemplo). Atributos quantitativos podem ser ordenados e podem ser utilizados em operações aritméticas. Já os atributos qualitativos podem ser ordenados em alguns casos, como no conjunto de atributos {pequeno, médio e grande}, mas operações aritméticas não podem ser aplicadas em seus valores. Na Tabela 2, apresentam-se os tipos dos atributos da Tabela 1.

Tabela 2 – Tipo dos atributos do conjunto de dados *Pacientes*

Atributo	Classificação
Id.	Qualitativo
Nome.	Qualitativo
Idade	Quantitativo discreto
Sexo	Qualitativo
Peso	Quantitativo contínuo
Manchas	Qualitativo
Temp.	Quantitativo contínuo
# Int.	Quantitativo discreto
Diagnóstico	Qualitativo

Fonte: Adaptado de Faceli et al. (2021)

Já a **escala** define quais operações podem ser realizadas sobre os valores do atributo. Em relação a essa classificação, os atributos podem ser **nominais**, **ordinais**, **intervalares** e **racionais**. Os dois primeiros são para tipos qualitativos e os dois últimos, para tipos quantitativos.

Na escala **nominal**, os valores são representados por nomes distintos, que carregam a menor quantidade de informação possível, e não existe uma relação de ordem entre seus valores. Tipicamente, utilizam-se as operações $=$ (igual a) e $\neq$ (diferente de) para valores nominais, por exemplo, Sexo = "M".

Os valores em uma escala **ordinal**, por sua vez, refletem também uma ordem das categorias representadas. Assim, além da igualdade e desigualdade, operadores como $<$, $>$, $\leq$, $\geq$ também podem ser utilizados. Por exemplo, estes operadores podem ser utilizados no conjunto de valores {pequeno, médio, grande}.

Na escala **intervalar**, os valores numéricos variam dentro de um intervalo. Desta forma, é possível definir tanto a ordem quanto a diferença em magnitude entre dois valores. A diferença em magnitude representa a distância entre dois valores no intervalo de valores possíveis, e o valor zero não possui o mesmo significado do zero utilizado em operações aritméticas. Por exemplo, para a escala de temperatura em graus Celsius, a variação

entre 26 e 34 graus em um dia de verão e entre 13 e 21 graus em um dia de inverno é de 8 graus. No entanto, 90 graus Celsius é totalmente diferente de 90 graus Fahrenheit, pois a referência zero não é a mesma.

Por fim, os atributos com escala racional são os que carregam mais informações. Os números possuem um significado absoluto, isto é, existe um zero absoluto junto com uma unidade de medida, de modo que a razão tenha significado. Por exemplo, considerando-se o número de vezes que um paciente foi internado, o ponto zero corresponde a nenhuma internação. Além disso, é possível afirmar que 8 internações correspondem a 4 vezes mais internações do que 2. Os atributos do conjunto de dados da Tabela 1 podem ser classificados da forma como se apresenta na Tabela 3.

Tabela 3 – Escala dos atributos do conjunto de dados *Pacientes*

Atributo	Classificação
Id.	Nominal
Nome.	Nominal
Idade	Racional
Sexo	Nominal
Peso	Racional
Manchas	Nominal
Temp.	Intervalar
#Int.	Racional
Diagnóstico	Nominal

Fonte: Adaptado de Faceli et al. (2021)

2.6 Modelos Preditivos de Aprendizado de Máquina

O aprendizado de máquina supervisionado para classificação binária dispõe de uma vasta gama de algoritmos, cada qual fundamentado em diferentes pressupostos matemáticos e heurísticos. Para a predição da inscrição de notas de empenho em Restos a Pagar, optou-se pela avaliação comparativa de múltiplas abordagens, desde modelos lineares interpretáveis até métodos de conjunto (*ensemble*) de alta complexidade e Redes Neurais Artificiais (RNA). A seguir, apresentam-se os fundamentos teóricos dos 7 algoritmos selecionados para este estudo.

2.6.1 Modelos Lineares e Probabilísticos

Apresentaremos o algoritmo Regressão Logística, um exemplo da classe dos modelos lineares e probabilísticos, que foi escolhido para o desenvolvimento deste trabalho.

Regressão Logística

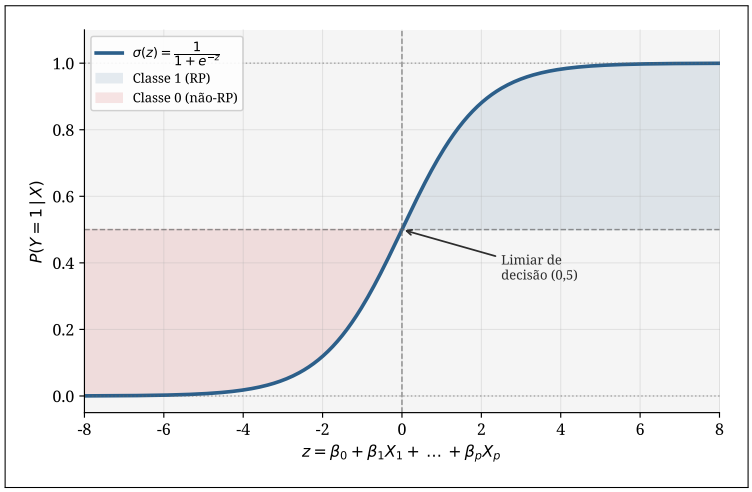
A Regressão Logística, formalizada por Cox (1958), é um modelo estatístico fundamental para problemas de classificação binária. Diferentemente da regressão linear clássica, que prevê valores contínuos no conjunto dos números reais, a regressão logística estima a probabilidade de uma observação pertencer a uma classe específica (neste estudo, a probabilidade de um empenho ser inscrito em Restos a Pagar).

Para garantir que as previsões permaneçam no intervalo $[0, 1]$, o modelo utiliza a função logística (ou sigmoide), definida como:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0+\beta_1 X_1+\cdots+\beta_p X_p)}} \tag{1}$$

onde $P(Y = 1|X)$ é a probabilidade da classe positiva, β_0 é o intercepto e $\beta_1, \ldots, \beta_p$ são os coeficientes associados às variáveis preditoras $X_1, \ldots, X_p$. A decisão final de classificação é tomada com base em um limiar de probabilidade (frequentemente $0,5$). Devido à sua interpretabilidade direta, na qual a razão de chances (*odds ratio*) indica a direção e a magnitude do impacto de cada variável, a Regressão Logística é amplamente utilizada como modelo *baseline* em estudos preditivos. Na Figura 4, tem-se a representação da função sigmoide.

Figura 4 – Representação da função sigmoide



Fonte: Elaborado pelo autor (2026), com o apoio da ferramenta Manus AI

A curva em formato de "S" demonstra como a função é capaz de mapear qualquer valor contínuo do eixo X (combinação linear das variáveis) para um intervalo estritamente contido entre 0 e 1 no eixo Y. Essa propriedade é fundamental para a modelagem preditiva, pois permite interpretar a saída do modelo, por exemplo, a probabilidade de um empenho ser inscrito em Restos a Pagar.

Exemplo Didático: Regressão Logística

Considere um modelo simplificado com duas variáveis: X_1 (Valor da NE em R$ mil) e X_2 (Dias até o fim do ano). Suponha que, após o treinamento, o modelo tenha ajustado os seguintes coeficientes:

Tabela 4 – Coeficientes hipotéticos do modelo de Regressão Logística

Parâmetro	**Valor**	**Interpretação**
Intercepto (β_0)	$-1{,}0$	Tendência base de não inscrição em RP
Coef. Valor (β_1)	$+0{,}08$	Valores maiores aumentam a chance de RP
Coef. Dias (β_2)	$-0{,}10$	Mais dias até o fim do ano reduzem a chance de RP

Fonte: Elaborado pelo autor (2026).

Para uma nova nota de empenho com **Valor = R$ 50 mil** e emitida a **10 dias** do fim do ano, calcula-se primeiro o logito (z):

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 = -1{,}0 + (0{,}08 \times 50) + (-0{,}10 \times 10) = -1{,}0 + 4{,}0 - 1{,}0 = 2{,}0$$

Em seguida, aplica-se a função sigmoide para obter a probabilidade $P(Y = 1)$:

$$P = \frac{1}{1 + e^{-z}} = \frac{1}{1 + e^{-2{,}0}} \approx \frac{1}{1 + 0{,}135} \approx 0{,}881 \;\; (88{,}1\%)$$

Como a probabilidade calculada (88,1%) supera o limiar de decisão padrão de 50%, o modelo classifica esta nota de empenho como **RP**. O resultado é intuitivo: uma nota de alto valor emitida a apenas 10 dias do encerramento do exercício financeiro apresenta elevada probabilidade de não ser liquidada dentro do prazo, sendo inscrita em Restos a Pagar.

2.6.2 Modelos Baseados em Árvores e Métodos de Conjunto (*Ensemble*)

Os algoritmos Árvore de Decisão, Random Forest e XGBoost, apresentados a seguir, foram selecionados para o desenvolvimento do presente trabalho.

Árvore de Decisão (*Decision Tree*)

As Árvores de Decisão são modelos não paramétricos que particionam o espaço de características em regiões retangulares, criando regras de decisão interpretáveis em formato de fluxograma. O algoritmo seminal CART (*Classification and Regression Trees*), desenvolvido por Breiman et al. (1984), constrói a árvore dividindo recursivamente os dados de treinamento.

Em cada nó, o algoritmo seleciona a variável e o ponto de corte que maximizam a pureza dos nós filhos resultantes. Para problemas de classificação, a impureza é frequentemente medida pelo Índice de Gini:

$$Gini(p) = 1 - \sum_{i=1}^{C} p_i^2 \tag{2}$$

onde C é o número de classes e p_i é a proporção de observações da classe i no nó. Embora sejam altamente interpretáveis e capazes de capturar relações não lineares complexas entre as características das despesas, as árvores de decisão individuais são propensas ao sobreajuste (*overfitting*[1]), tendendo a memorizar ruídos do conjunto de treinamento se não forem adequadamente podadas (*pruning*).

O funcionamento de uma Árvore de Decisão pode ser compreendido por meio de divisões sucessivas do conjunto de dados baseadas em regras lógicas. Considere um pequeno conjunto de 6 notas de empenho fictícias, avaliadas segundo duas variáveis: **Mês de Emissão** e **Valor da NE**.

Tabela 5 – Conjunto de dados fictício para o exemplo da Árvore de Decisão

ID	Mês de Emissão	Valor (R$)	Classe
1	Novembro	5.000	Não-RP
2	Dezembro	15.000	RP
3	Outubro	2.000	Não-RP
4	Dezembro	3.000	Não-RP
5	Novembro	25.000	RP
6	Dezembro	50.000	RP

Fonte: Elaborado pelo autor (2026).

O algoritmo busca a variável e o ponto de corte que melhor separam as classes (RP e Não-RP), maximizando o ganho de informação (ou minimizando a impureza de Gini).

No primeiro passo (nó raiz), o algoritmo pode identificar que o **Mês de Emissão** é um bom discriminador. Uma regra inicial poderia ser: *"O mês de emissão é Dezembro?"*.

- **Sim (IDs 2, 4, 6):** Neste subgrupo, temos 2 casos RP e 1 caso Não-RP. A impureza ainda existe, então o algoritmo faz uma nova divisão.
- **Não (IDs 1, 3, 5):** Neste subgrupo (Outubro e Novembro), temos 1 caso RP e 2 casos Não-RP.

Para refinar a classificação no ramo do "Sim" (Dezembro), o algoritmo testa a variável **Valor**. Ele pode encontrar o ponto de corte de R$ 10.000:

- *Valor > 10.000?* **Sim (IDs 2, 6):** Ambos são RP. O nó torna-se puro (folha).
- *Valor > 10.000?* **Não (ID 4):** É Não-RP. O nó torna-se puro (folha).

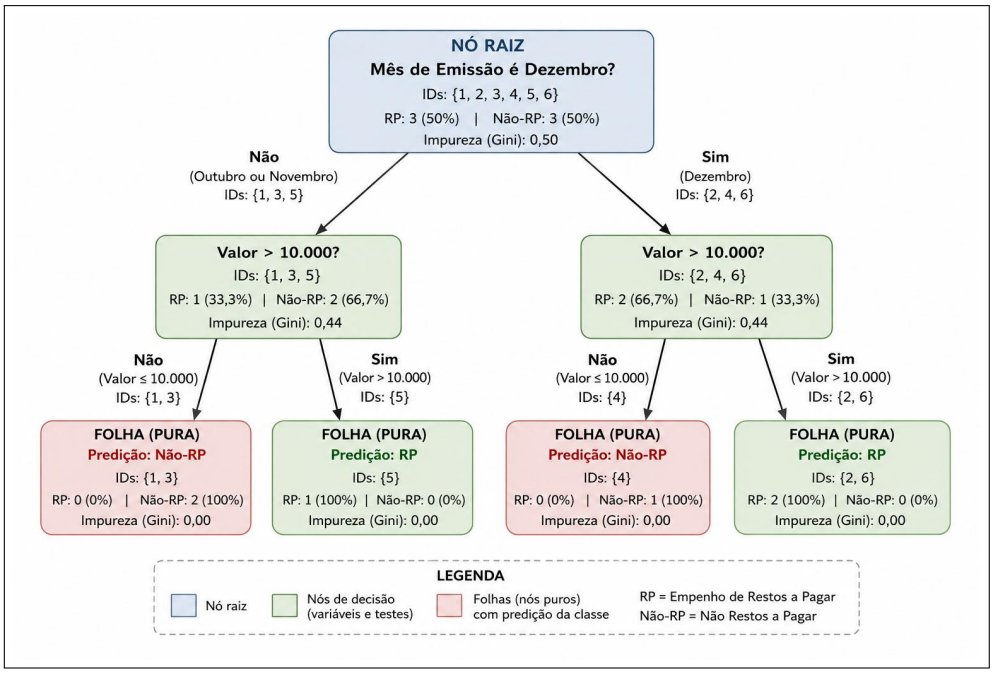
[1] *Overfitting*, ou ajuste excessivo, é um comportamento indesejável de aprendizado de máquina que ocorre quando o modelo preditivo fornece previsões precisas para dados de treinamento, mas não para novos dados. Em outras palavras, o modelo está excessivamente ajustado aos dados de treinamento.

No ramo do "Não" (Outubro/Novembro), a variável **Valor** também pode ser testada com o mesmo corte de R$ 10.000:

- *Valor > 10.000?* **Sim (ID 5):** É RP. Nó puro.
- *Valor > 10.000?* **Não (IDs 1, 3):** Ambos são Não-RP. Nó puro.

Na Figura 5, ilustra-se a árvore de decisão construída a partir do exemplo fornecido.

Figura 5 – Estrutura hierárquica de uma Árvore de Decisão



Fonte: Elaborado pelo autor (2026), com o apoio da ferramenta ChatGPT

Assim, a árvore construída estabelece regras claras e interpretáveis: se a nota for emitida em dezembro e tiver valor superior a R$ 10.000, será classificada como RP. Caso contrário, se for de meses anteriores mas com valor alto, também será RP. Valores baixos, independentemente do mês, tendem a ser Não-RP. Este mecanismo de particionamento recursivo é a base do algoritmo.

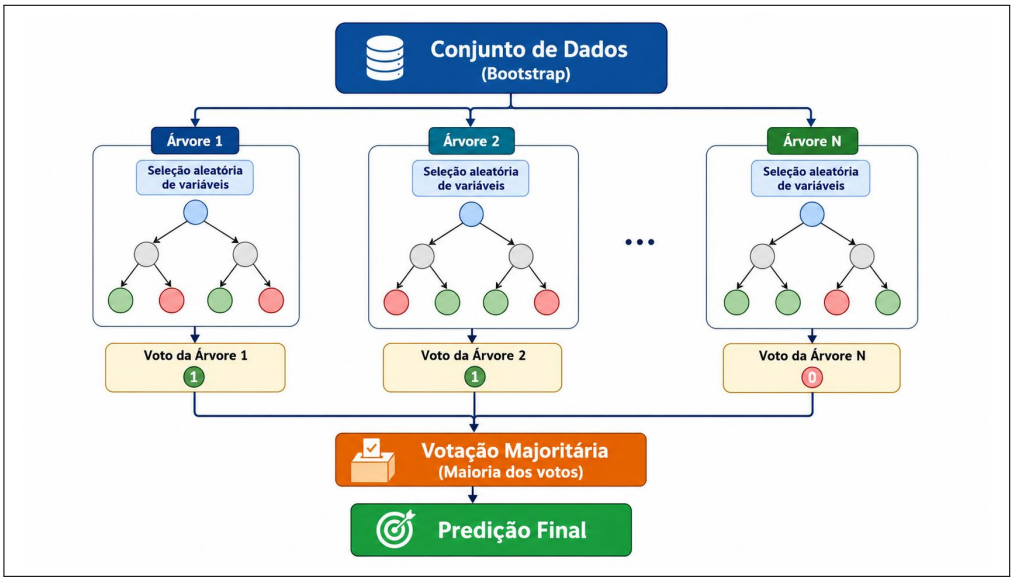
Random Forest

O *Random Forest* (Floresta Aleatória) é um método de aprendizado de conjunto (*ensemble*) proposto por Breiman (2001). O algoritmo constrói múltiplas árvores de decisão independentes durante o treinamento e combina suas previsões por meio de votação majoritária (para classificação) ou média (para regressão).

Para garantir a diversidade entre as árvores e reduzir a variância do modelo final, o *Random Forest* utiliza duas técnicas de aleatoriedade: o *bagging* (*Bootstrap Aggregating*), que consiste na amostragem com substituição dos dados de treinamento para cada árvore; e a seleção aleatória de um subconjunto de variáveis preditoras em cada divisão de nó.

Essa abordagem mitiga significativamente o problema de *overfitting* das árvores individuais, resultando em um modelo muito preciso, frequentemente adotado como padrão-ouro em dados tabulares, como os dados orçamentários do TRE-GO. Na Figura 6, ilustra-se um exemplo de Random Forest.

Figura 6 – Arquitetura do *Random Forest*



Fonte: Elaborado pelo autor (2026), com o apoio da ferramenta ChatGPT

A representação visual mostra como o conjunto de dados original é reamostrado para treinar múltiplas árvores de decisão independentes, cujas predições individuais (0 ou 1) são posteriormente combinadas por meio de votação majoritária. Essa arquitetura explica a capacidade do modelo de reduzir a variância e evitar o *overfitting* comum em árvores isoladas.

O *Random Forest* representa algoritmo *ensemble*, ou seja, reúne ou combina múltiplas árvores de decisão para produzir predições mais precisas. Raschka e Mirjalili (2019, p. 10) dizem o seguinte sobre este algoritmo:

> A ideia por trás de uma *random forest* é calcular a média de múltiplas árvores de decisão (profundas) que individualmente sofrem de alta variância para construir um modelo mais robusto que tenha um melhor desempenho de generalização e seja menos suscetível a *overfitting*.

Imagine um *ensemble* (floresta) composto por apenas 3 árvores de decisão, avaliando uma nota de empenho emitida em **Novembro**, com **Valor de R$ 30.000** e **Natureza de Despesa = Equipamentos**.

O Random Forest agrega os resultados por votação majoritária: 2 votos para RP e 1 voto para Não-RP. A decisão final da floresta para esta nota de empenho será **RP**, como mostra a Tabela 6. A combinação de múltiplas árvores reduz o risco de *overfitting*

Tabela 6 – Votação das árvores no exemplo do Random Forest

Árvore	Variáveis consideradas / Regra aplicada	Voto
Árvore 1	Mês e Valor: notas de Novembro acima de R$ 20.000 são RP	RP
Árvore 2	Natureza e Mês: Equipamentos antes de Dezembro têm tempo hábil	Não-RP
Árvore 3	Valor e Natureza: Equipamentos acima de R$ 25.000 costumam atrasar	RP
Resultado (votação majoritária: 2 × RP, 1 × Não-RP)		**RP**

Fonte: Elaborado pelo autor (2026).

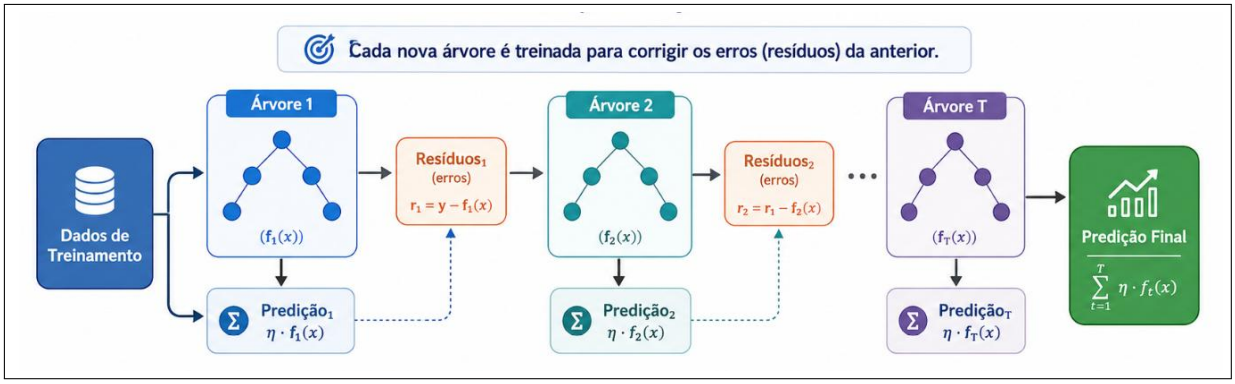
(sobreajuste) que uma única árvore poderia sofrer, tornando o modelo mais robusto e generalizável.

XGBoost (*eXtreme Gradient Boosting*)

O XGBoost é uma implementação otimizada e escalável do algoritmo *Gradient Boosting*, introduzida por Chen e Guestrin (2016). Diferentemente do *Random Forest*, que constrói árvores de forma independente e paralela, o *boosting* constrói árvores sequencialmente. Cada nova árvore é treinada para prever e corrigir os erros residuais (gradientes da função de perda) cometidos pelo conjunto das árvores anteriores.

O XGBoost destaca-se por incorporar regularização matemática (penalidades L_1[2] e L_2[3] nos pesos das folhas) na sua função objetivo para evitar *overfitting*, além de otimizações de *hardware* e processamento paralelo que o tornam extremamente rápido e eficiente. Possui boa capacidade de lidar com dados esparsos e valores ausentes. Na Figura 7, tem-se um exemplo do funcionamento do algoritmo XGBoost.

Figura 7 – Arquitetura do XGBoost



Fonte: Elaborado pelo autor (2026), com o apoio da ferramenta ChatGPT

Diferente do Random Forest, o diagrama mostra que cada nova árvore é treinada especificamente para corrigir os erros (resíduos) da árvore anterior, somando-se as predições com uma taxa de aprendizado (η). A inclusão de termos de regularização (L_1 e

[2] A penalidade L1 adiciona à função de custo a soma dos valores absolutos dos coeficientes.
[3] A penalidade L2 adiciona à função de custo a soma dos quadrados dos coeficientes.

L_2) na função objetivo ilustra como o XGBoost controla a complexidade do modelo para otimizar a generalização.

A Tabela 7 ilustra as três primeiras iterações do processo para uma nota que de fato virou RP (valor real = 1), partindo de uma previsão inicial de 0,50 e adotando taxa de aprendizado de 0,1.

Tabela 7 – Iterações do processo de *boosting* no exemplo do XGBoost

Iteração	Prev. atual	Regra aprendida pela árvore	Ajuste	Nova previsão
Inicial	0,50	—	—	0,50
Árvore 1	0,50	Notas de Dezembro têm resíduo positivo (+0,3)	+0,03	0,53
Árvore 2	0,53	Notas de alto valor precisam de ajuste (+0,4)	+0,04	0,57
Árvore 3	0,57	Natureza Equipamentos reforça RP (+0,5)	+0,05	0,62

Fonte: Elaborado pelo autor (2026).

Após centenas de iterações, a previsão acumulada é passada por uma função sigmoide. Se o resultado final for, por exemplo, 0,82, a nota será classificada como **RP**. A taxa de aprendizado (0,1 no exemplo) controla o quanto cada árvore contribui, evitando ajustes bruscos que levariam ao sobreajuste.

2.6.3 Modelos Baseados em Distância e Geometria

Os algoritmos KNN e SVC são apresentados a seguir.

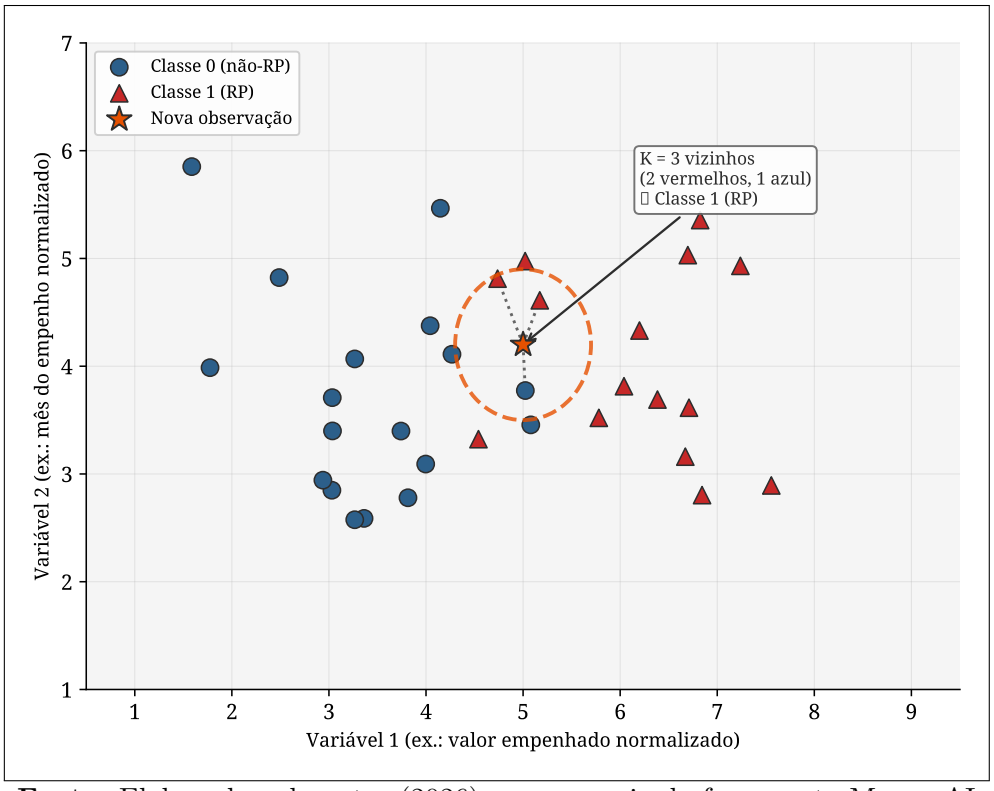
K-Vizinhos Mais Próximos (KNN - *K-Nearest Neighbors*)

O KNN é um dos algoritmos de classificação mais antigos e intuitivos, com raízes nos trabalhos de Fix e Hodges (1951) e formalizado por Cover e Hart (1967). Trata-se de um método de aprendizado baseado em instâncias (ou *lazy learning*), que não constrói um modelo interno explícito durante a fase de treinamento, apenas armazena os dados.

Para classificar uma nova observação (ex.: uma nova nota de empenho), o algoritmo calcula a distância (geralmente Euclidiana ou Manhattan) entre essa observação e todos os pontos do conjunto de treinamento, identificando os K vizinhos mais próximos no espaço de características. A classe final é determinada por votação majoritária entre esses vizinhos. O desempenho do KNN é altamente sensível à dimensionalidade dos dados e exige a normalização rigorosa das variáveis contínuas. Na Figura 8, ilustra-se um exemplo de funcionamento do algoritmo KNN.

Ao posicionar uma nova observação (estrela) no espaço de características, o modelo traça um raio de proximidade para identificar os K vizinhos mais próximos (neste

Figura 8 – Ilustração do algoritmo KNN



Fonte: Elaborado pelo autor (2026), com o apoio da ferramenta Manus AI

caso, $K = 3$). A classificação final é definida pela classe majoritária dentro desse raio, evidenciando a dependência do algoritmo em relação à métrica de distância euclidiana.

Para ilustrar o funcionamento do algoritmo KNN no contexto desta pesquisa, considere um cenário simplificado com apenas duas variáveis preditoras: o **Valor da Nota de Empenho** (em milhares de reais) e o **Tempo de Tramitação** (em dias). O objetivo é prever se uma nova nota de empenho será inscrita em Restos a Pagar (RP) ou não (Não-RP).

A Tabela 8 apresenta um conjunto de dados fictício com 5 observações históricas (treinamento).

Tabela 8 – Conjunto de dados fictício para o exemplo do KNN

ID	Valor (R$ mil)	Tempo (dias)	Classe (RP)
1	10	5	Não-RP
2	15	7	Não-RP
3	50	20	RP
4	45	18	RP
5	12	25	RP

Fonte: Elaborado pelo autor (2026)

Suponha que uma nova nota de empenho (ID 6) possua **Valor = 14** e **Tempo = 8**. Para classificá-la utilizando o KNN com $K = 3$, o algoritmo calcula a distância

(por exemplo, Euclidiana) entre a nova observação e todas as observações do conjunto de treinamento.

Calculando as distâncias (de forma simplificada, sem normalização prévia para fins didáticos):

- Distância para ID 1: $\sqrt{(14-10)^2+(8-5)^2} = \sqrt{16+9} = 5$
- Distância para ID 2: $\sqrt{(14-15)^2+(8-7)^2} = \sqrt{1+1} \approx 1,41$
- Distância para ID 3: $\sqrt{(14-50)^2+(8-20)^2} = \sqrt{1296+144} \approx 37,94$
- Distância para ID 4: $\sqrt{(14-45)^2+(8-18)^2} = \sqrt{961+100} \approx 32,57$
- Distância para ID 5: $\sqrt{(14-12)^2+(8-25)^2} = \sqrt{4+289} \approx 17,11$

Os 3 vizinhos mais próximos ($K = 3$) são os IDs 2, 1 e 5. Observando as classes desses vizinhos, temos: Não-RP (ID 2), Não-RP (ID 1) e RP (ID 5). Por votação majoritária (2 votos contra 1), o algoritmo classifica a nova nota de empenho (ID 6) como **Não-RP**. Este exemplo evidencia a importância da proximidade no espaço de características para a tomada de decisão do modelo.

Máquina de Vetores de Suporte (SVC - *Support Vector Classifier*)

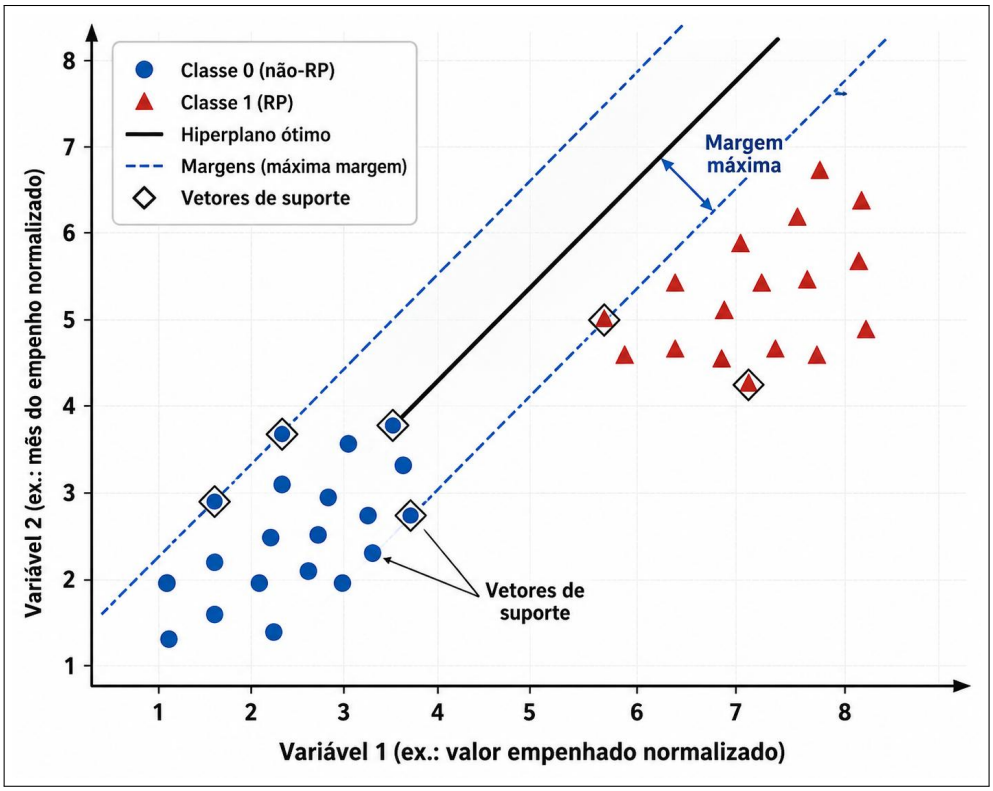
As Máquinas de Vetores de Suporte, propostas em sua formulação moderna por Cortes e Vapnik (1995), buscam encontrar o hiperplano[4] ótimo que separa as classes no espaço de características com a margem máxima possível. Os pontos de dados mais próximos ao hiperplano, que definem essa margem e são críticos para a construção do modelo, são chamados de vetores de suporte.

Para lidar com dados que não são linearmente separáveis, o SVC utiliza o "truque do *kernel*" (*kernel trick*), mapeando os dados originais para um espaço de dimensão superior onde a separação linear se torna possível (utilizando funções como RBF - *Radial Basis Function* ou polinomial). É um modelo matematicamente rigoroso e eficaz em espaços de alta dimensionalidade, embora computacionalmente custoso para grandes volumes de dados. Na Figura 9, tem-se a representação gráfica de um SVC.

Esta representação visual do SVC destaca o conceito de hiperplano de margem máxima em um espaço bidimensional. O gráfico mostra a linha de separação ótima entre as duas classes e as margens paralelas que tangenciam os pontos mais críticos, denominados vetores de suporte. A figura ilustra o objetivo matemático do algoritmo:

[4]Em geometria, um hiperplano é a generalização do conceito de reta (em duas dimensões) e de plano (em três dimensões) para espaços de dimensão arbitrária. Em um espaço com p variáveis, o hiperplano é uma superfície de dimensão $p-1$ definida por uma equação linear, que divide o espaço em duas regiões. No contexto de algoritmos de classificação, essa superfície constitui a fronteira de decisão que separa as classes.

Figura 9 – Máquina de Vetores de Suporte (SVC)



Fonte: Elaborado pelo autor (2026), com o apoio da ferramenta Manus AI

encontrar a fronteira de decisão que maximize a distância (margem) entre as observações mais próximas de classes distintas.

Por exemplo, imagine um gráfico 2D onde o eixo X representa o **Mês de Emissão** (1 a 12) e o eixo Y representa o **Valor da NE**. As notas Não-RP estão concentradas no centro do gráfico (meses intermediários, valores baixos), enquanto as notas RP formam um anel ao redor delas (valores muito altos em qualquer mês, ou qualquer valor no final do ano). Nenhuma linha reta consegue separar esses dois grupos.

Ao aplicar um *Kernel* RBF (*Radial Basis Function*), o SVC calcula, para cada ponto, a sua distância em relação ao centro do grupo Não-RP. Essa distância funciona como uma nova dimensão (eixo Z), conforme ilustrado na Tabela 9.

Tabela 9 – Projeção dos dados via *Kernel* RBF no exemplo do SVC

ID	Mês	Valor (R$ mil)	Dist. ao centro	Classe
1	6	10	Baixa (0,1)	Não-RP
2	7	8	Baixa (0,2)	Não-RP
3	12	60	Alta (0,9)	RP
4	1	55	Alta (0,8)	RP
5	11	45	Alta (0,7)	RP

Fonte: Elaborado pelo autor (2026).

No espaço 3D resultante, o algoritmo consegue traçar um plano horizontal (hiper-

plano) cortando o eixo Z. Tudo que estiver abaixo do plano é classificado como Não-RP; tudo que estiver acima é classificado como RP. Os pontos mais próximos a esse plano de corte são os *vetores de suporte*, que definem a posição exata da fronteira de decisão e maximizam a margem entre as classes.

2.6.4 Redes Neurais Artificiais

O algoritmo MLP (*Multilayer Perceptron*), apresentado a seguir, também foi utilizado no desenvolvimento desta dissertação.

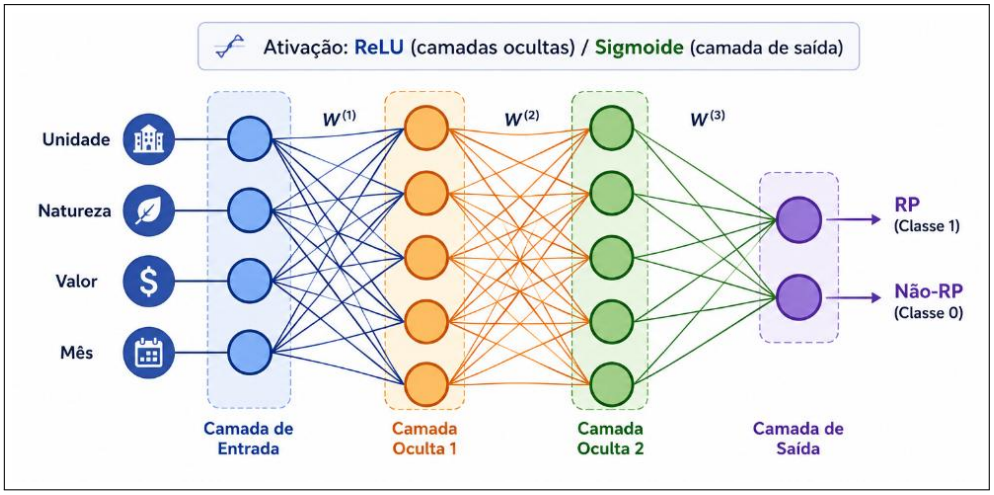
Perceptron Multicamadas (MLP - *Multilayer Perceptron*)

O Perceptron Multicamadas é uma arquitetura clássica de Redes Neurais Artificiais *Feedforward*. O modelo ganhou viabilidade prática para problemas complexos com a popularização do algoritmo de retropropagação (*backpropagation*) por Rumelhart et al. (1986).

Um MLP é composto por uma camada de entrada (recebendo as características da despesa), uma ou mais camadas ocultas e uma camada de saída (fornecendo a probabilidade de inscrição em Restos a Pagar). Cada conexão entre os neurônios possui um peso ajustável, e funções de ativação não lineares (como ReLU nas camadas ocultas e Sigmoide na saída) permitem que a rede aprenda representações complexas e altamente não lineares dos dados.

Embora possuam grande capacidade de aproximação universal, os MLPs exigem ajuste cuidadoso de hiperparâmetros e normalização de dados, e são frequentemente considerados "caixas-pretas" devido à dificuldade de interpretação de seus pesos internos. Na Figura 10, ilustra-se a arquitetura de um MLP.

Figura 10 – Arquitetura de um Perceptron Multicamadas (MLP)



Fonte: Elaborado pelo autor (2026), com o apoio da ferramenta Manus AI

Neste diagrama, tem-se a topologia de uma rede neural artificial *feedforward* clássica. A estrutura é composta por uma camada de entrada recebendo as variáveis orçamentárias, duas camadas ocultas com funções de ativação não-lineares (ReLU) para extração de padrões complexos, e uma camada de saída com ativação sigmoide para a classificação binária. As conexões densas entre todos os neurônios ilustram a alta capacidade de aproximação de funções do modelo.

Para exemplificar, considere uma rede simplificada para prever RP, composta por: (a) **camada de entrada** com 2 neurônios (X_1: Valor normalizado = 0,8; X_2: Mês normalizado = 0,9, representando Dezembro); (b) **camada oculta** com 2 neurônios (H_1 e H_2) e função de ativação ReLU; e (c) **camada de saída** com 1 neurônio e função sigmoide.

A Tabela 10 resume os pesos e vieses hipotéticos da rede após o treinamento.

Tabela 10 – Pesos e vieses hipotéticos da rede MLP para o exemplo didático

Conexão	Peso (W)	Viés (b)	Ativação
$X_1 \to H_1$	0,50	$-0{,}20$	ReLU
$X_2 \to H_1$	0,60		
$X_1 \to H_2$	$-0{,}30$	$+0{,}10$	ReLU
$X_2 \to H_2$	0,80		
$H_1 \to$ Saída	0,70	$-0{,}50$	Sigmoide
$H_2 \to$ Saída	0,40		

Fonte: Elaborado pelo autor (2026).

No processamento (*forward pass*), os cálculos são realizados camada a camada:

Camada oculta — neurônio H_1:

$$\text{Soma}_{H_1} = (0{,}8 \times 0{,}50) + (0{,}9 \times 0{,}60) - 0{,}20 = 0{,}40 + 0{,}54 - 0{,}20 = 0{,}74$$

$$\text{Saída}_{H_1} = \text{ReLU}(0{,}74) = 0{,}74$$

Camada oculta — neurônio H_2:

$$\text{Soma}_{H_2} = (0{,}8 \times -0{,}30) + (0{,}9 \times 0{,}80) + 0{,}10 = -0{,}24 + 0{,}72 + 0{,}10 = 0{,}58$$

$$\text{Saída}_{H_2} = \text{ReLU}(0{,}58) = 0{,}58$$

Camada de saída:

$$\text{Soma}_{\text{saída}} = (0{,}74 \times 0{,}70) + (0{,}58 \times 0{,}40) - 0{,}50 = 0{,}518 + 0{,}232 - 0{,}50 = 0{,}25$$

$$P(\text{RP}) = \frac{1}{1 + e^{-0{,}25}} \approx 0{,}562 \;\; (56{,}2\%)$$

Como $56{,}2\% > 50\%$, a rede classifica a nota como **RP**. Se a nota fosse, na realidade,

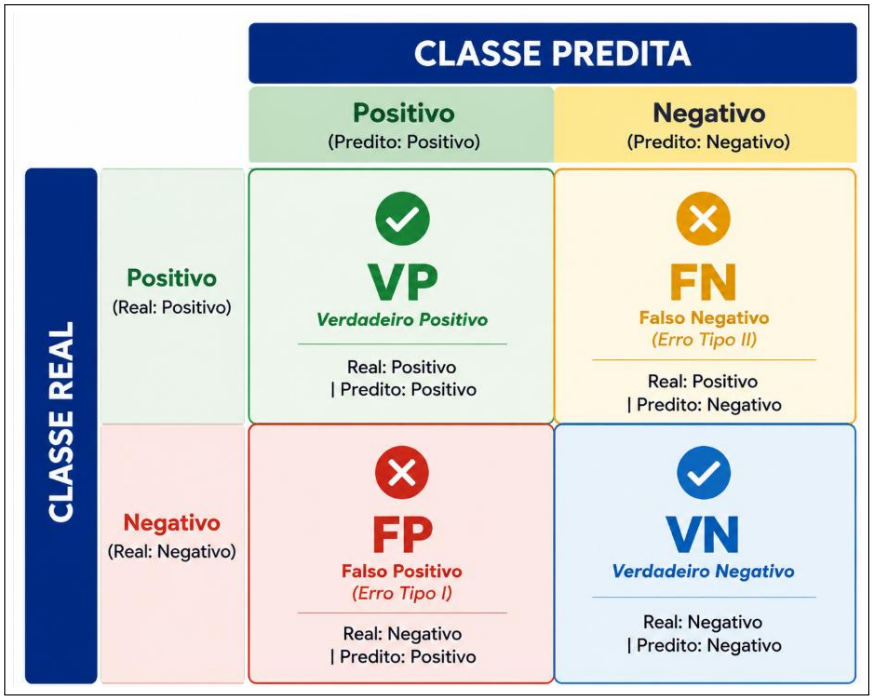
Não-RP, o erro seria propagado de volta (*backpropagation*) para ajustar ligeiramente todos os pesos e vieses da rede, melhorando a previsão nas iterações seguintes.

2.6.5 Métricas de Avaliação para Classificação

A avaliação adequada de modelos de classificação requer métricas específicas que capturem diferentes aspectos do desempenho preditivo. Para problemas de classificação binária, especialmente em contextos de classes desbalanceadas, destacam-se métricas que vão além da simples acurácia, que mede a taxa de acertos do modelo preditivo - a soma de verdadeiros positivos e verdadeiros negativos dividido pelo total de predições (Raschka; Mirjalili, 2019, p. 213).

Em um problema de classificação binária, temos, de um lado, o total de classes positivas e negativas reais e, do outro lado, as classes positivas e negativas preditas. Assim, pode-se construir uma matriz, chamada matriz de confusão (Figura 11), para facilitar a análise dos resultados.

Figura 11 – Matriz de confusão para um problema de duas classes



Fonte: Elaborado pelo autor (2026), com o apoio da ferramenta ChatGPT

A figura apresenta a estrutura padrão de uma matriz de confusão 2x2, ferramenta essencial para avaliação de classificadores. O diagrama cruza a classe real (positiva/negativa) com a classe predita pelo modelo, definindo visualmente os quatro quadrantes possíveis: Verdadeiros Positivos (VP), Verdadeiros Negativos (VN), Falsos Positivos (FP - Erro Tipo I) e Falsos Negativos (FN - Erro Tipo II), servindo de base para o cálculo de métricas como precisão e *recall*.

As quatro intersecções entre os valores da classe real e os da classe predita podem gerar os seguintes resultados:

- Verdadeiros Positivos (VP): o número de exemplos da classe positiva classificados corretamente;
- Verdadeiros Negativos (VN): o número de exemplos da classe negativa classificados corretamente;
- Falsos Positivos (FP): o número de exemplos cuja classe real é a negativa, mas que foram classificados incorretamente como pertencentes à classe positiva. Também chamados de erro tipo I;
- Falsos Negativos (FN): o número de exemplos pertencentes originalmente à classe positiva que foram classificados incorretamente como da classe negativa. Também conhecidos como erro tipo II.

A **precisão** mede a proporção de verdadeiros positivos entre todas as predições positivas, respondendo à pergunta: "Das instâncias que o modelo classificou como positivas, quantas realmente são positivas?". É calculada pela seguinte fórmula:

$$\text{Precisao} = \frac{VP}{VP + FP} \tag{3}$$

Em termos práticos, uma alta precisão indica que, quando o modelo sinaliza que uma NE será inscrita em RP, essa predição tende a estar correta. Erros de precisão correspondem a falsos alarmes: empenhos classificados incorretamente como futuros RP.

O ***recall***, ou sensibilidade, mede a proporção de verdadeiros positivos identificados corretamente, respondendo: "Das instâncias realmente positivas, quantas o modelo identificou corretamente?". O *recall* é dado pela seguinte fórmula

$$\text{Recall} = \frac{VP}{VP + FN} \tag{4}$$

Um alto *recall* indica que o modelo consegue detectar a maior parte dos empenhos que efetivamente serão inscritos em Restos a Pagar. Erros de recall correspondem a casos não detectados: empenhos que se tornarão RP, mas que o modelo classificou incorretamente como FN. Estas NE deixariam de ser monitoradas pela administração por não serem consideradas problemáticas e aumentariam o volume de RP, acarretando em prejuízo para a gestão orçamentária.

O *F1-Score* constitui média harmônica entre precisão e *recall*, fornecendo métrica única que equilibra ambos os aspectos. Raschka e Mirjalili (2019, p. 156) destacam que "o F1-Score é particularmente útil quando se busca equilibrar a capacidade de identificar casos positivos com a precisão dessas identificações". O *F1-Score* é calculado desta forma:

$$\text{F1-Score} = 2 * \frac{Precisão * Recall}{Precisão + Recall} \tag{5}$$

A escolha pela média harmônica em detrimento da média aritmética é deliberada: ela penaliza de forma mais severa as situações em que há grande assimetria entre as duas métricas, atribuindo ao F1-Score um valor elevado apenas quando precisão e *recall* são simultaneamente satisfatórios. Assim, o F1-Score constitui a métrica mais adequada para comparar modelos em cenários de classes desbalanceadas, como o deste estudo, em que a classe de interesse (empenhos inscritos em Restos a Pagar) representa uma parcela minoritária do conjunto de dados.

Para consolidar a avaliação do compromisso entre precisão e *recall*, adotou-se também a métrica AUC-PR (*Area Under the Precision-Recall Curve*). Diferentemente de métricas pontuais que dependem de um limiar de decisão fixo (tradicionalmente 0,5), a curva Precisão-Recall ilustra o comportamento do classificador ao longo de todos os limiares de probabilidade possíveis, plotando a precisão no eixo das ordenadas e o recall no eixo das abscissas. O cálculo da área sob essa curva (AUC-PR) resulta em um valor único que resume a capacidade global do modelo em identificar corretamente a classe de interesse sem gerar um volume excessivo de falsos alarmes.

Em contextos de dados desbalanceados, o AUC-PR é preferível ao AUC-ROC como métrica de avaliação, pois não é distorcido pelo grande volume de verdadeiros negativos da classe majoritária: enquanto o AUC-ROC pode apresentar valores artificialmente elevados mesmo quando o modelo tem desempenho ruim sobre a classe minoritária, o AUC-PR foca exclusivamente no comportamento do modelo sobre os casos de interesse (Davis e Goadrich (2006); Hastie et al. (2009)).

O AP é uma estimativa alternativa da mesma área, calculada por meio de uma soma ponderada dos incrementos de *recall* em cada limiar de decisão, em vez da regra do trapézio utilizada pelo AUC-PR. Por essa razão, o AP tende a ser ligeiramente mais preciso como estimativa da área real sob a curva e é o padrão adotado pela documentação do scikit-learn (Pedregosa et al., 2011).

2.6.6 Tratamento de Classes Desbalanceadas

Problemas de classificação com classes desbalanceadas — nos quais uma classe é significativamente mais frequente do que outra — requerem técnicas específicas para garantir eficácia preditiva adequada (He; Garcia, 2009). Nesse cenário, algoritmos treinados sem qualquer tratamento tendem a classificar praticamente todas as observações como pertencentes à classe majoritária, atingindo alta acurácia global às custas de um *recall* baixo para a classe de interesse.

A técnica SMOTE (*Synthetic Minority Oversampling Technique*), proposta por Chawla et al. (2002), constitui a abordagem de sobreamostragem mais amplamente ado-

tada na literatura para o tratamento desse problema. Em vez de simplesmente duplicar exemplos existentes da classe minoritária, o que levaria ao *overfitting*, o SMOTE gera **exemplos sintéticos** por meio de interpolação entre instâncias reais da classe minoritária e seus k vizinhos mais próximos no espaço de características.

O mecanismo de geração funciona da seguinte forma: dado um exemplo x_i da classe minoritária e um de seus k vizinhos mais próximos x_{nn}, um novo exemplo sintético x_{novo} é criado segundo a equação:

$$x_{novo} = x_i + \lambda \cdot (x_{nn} - x_i), \quad \lambda \in [0, 1] \tag{6}$$

onde λ é um número aleatório uniforme entre 0 e 1. O novo ponto é, portanto, posicionado aleatoriamente **ao longo do segmento de reta** que une x_i a x_{nn}, gerando variações realistas que enriquecem a representação da classe minoritária sem simplesmente replicar observações existentes.

Exemplo numérico ilustrativo.

Considere duas notas de empenho inscritas em RP descritas por duas variáveis: valor logarítmico do empenho (V) e mês de emissão (M). Suponha $x_i = (3{,}2;\ 10)$ e seu vizinho mais próximo $x_{nn} = (4{,}0;\ 11)$. Sorteando $\lambda = 0{,}6$, o SMOTE gera:

$$x_{novo} = (3{,}2;\ 10) + 0{,}6 \cdot \Big[(4{,}0;\ 11) - (3{,}2;\ 10)\Big] = (3{,}2 + 0{,}48;\ 10 + 0{,}6) = (3{,}68;\ 10{,}6)$$

O ponto sintético (3,68; 10,6) representa uma nota de empenho "hipotética" com valor e mês intermediários aos dos dois exemplos reais, enriquecendo a fronteira de decisão aprendida pelo classificador sem introduzir ruído artificial.

É importante ressaltar que, neste trabalho, o SMOTE foi aplicado **exclusivamente ao conjunto de treinamento**, após a divisão treino/teste. Essa precaução é fundamental para evitar o *data leakage*: se amostras sintéticas derivadas de observações do conjunto de teste fossem incluídas no treinamento, a avaliação de desempenho seria artificialmente inflada, comprometendo a validade dos resultados (Chawla et al., 2002).

Variantes do SMOTE.

A literatura registra extensões do algoritmo original para contextos específicos. O **Borderline-SMOTE** (Han et al., 2005) concentra a geração de exemplos sintéticos nas regiões de fronteira entre as classes, onde a discriminação é mais difícil.

O **SMOTE-NC** (*Nominal and Continuous*) (Chawla et al., 2002) adapta o mecanismo de interpolação para conjuntos de dados mistos, com variáveis numéricas e categóricas simultâneas — situação relevante para o presente trabalho, dado que variáveis como

modalidade de licitação e natureza da despesa são categóricas.

A escolha pelo SMOTE original, após a etapa de codificação (*one-hot encoding*), justifica-se pela ampla adoção na literatura e pela compatibilidade com o *pipeline* de pré-processamento adotado.

Este capítulo apresentou a fundamentação teórica necessária ao desenvolvimento da presente pesquisa. O próximo capítulo apresenta um resumo dos trabalhos relacionados a este por meio de mapeamento de literatura e apresenta o posicionamento desta pesquisa em relação aos trabalhos encontrados.

3 TRABALHOS RELACIONADOS

Este capítulo apresenta o mapeamento de literatura sobre Restos a Pagar, organizando os trabalhos conforme suas abordagens metodológicas e focos temáticos. A análise dos estudos relacionados permite posicionar a presente investigação no contexto acadêmico atual, demonstrando a evolução do conhecimento na área e as oportunidades de contribuição científica.

Procurou-se concentrar a pesquisa no período de 2020 a 2025 com o objetivo de evidenciar a relevância e atualidade do tema. No entanto, Aquino e Azevedo (2017) e Nonaka (2019) também foram incluídos, dada a importância destes para o presente trabalho.

3.1 Estudos sobre Fatores Determinantes dos Restos a Pagar

Influência de Fatores Político-Eleitorais

A literatura revelou interesse na investigação de como o ciclo político-eleitoral influencia a gestão orçamentária, particularmente no que se refere à inscrição de despesas em Restos a Pagar. Estes estudos fundamentam-se na Teoria dos Ciclos Político-Econômicos (Preussler, 2001), que postula que gestores públicos do Poder Executivo podem adotar comportamentos oportunistas durante períodos eleitorais.

Silva e Junior (2024) conduziu estudo sobre a influência do período eleitoral na inscrição de Restos a Pagar nas 26 capitais brasileiras, utilizando dados do período 2012-2023. Através de análise de regressão com dados em painel, o autor identificou que todo o ciclo eleitoral exerce influência estatisticamente significativa na inscrição de Restos a Pagar. Os resultados revelaram que nos anos eleitorais e pós-eleitorais há tendência de diminuição, enquanto no ano pré-eleitoral observa-se aumento na inscrição. Adicionalmente, o estudo demonstrou que capitais com gestores de ideologias de direita e das regiões Sul e Sudeste, bem como a população, influenciam no maior volume inscrito.

Araújo (2022) investigou os incentivos eleitorais e o gerenciamento de resultados orçamentários através de Restos a Pagar em municípios brasileiros. O trabalho, fundamentado na teoria do oportunismo político, analisou de que forma gestores municipais utilizam mecanismos orçamentários para postergar despesas como estratégia eleitoral. Os achados corroboram a hipótese de que períodos eleitorais intensificam o uso de Restos a Pagar como instrumento de manipulação fiscal.

Em linha semelhante de investigação, Araújo et al. (2023) expandiram a análise dos incentivos eleitorais e gerenciamento de resultados, fornecendo evidências adicionais sobre como o ciclo político influencia as decisões orçamentárias. O estudo reforça a importância de considerar variáveis político-eleitorais na compreensão dos determinantes dos Restos a

Pagar.

Almeida et al. (2023) contribuíram para esta vertente ao examinar especificamente os incentivos eleitorais em contextos municipais, oferecendo perspectiva complementar sobre como fatores políticos moldam as práticas orçamentárias locais.

Características Organizacionais e Setoriais

Estudos recentes têm explorado de que forma características específicas de diferentes tipos de organizações públicas influenciam a gestão de Restos a Pagar, revelando padrões distintos conforme o setor e a natureza organizacional.

Maurício (2025) desenvolveu investigação pioneira sobre cancelamentos de Restos a Pagar Não Processados no Exército Brasileiro, utilizando métodos mistos sequenciais para analisar 87.190 empenhos dos exercícios 2023 e 2024. O estudo identificou que o fenômeno resulta de um ecossistema de fragilidades interdependentes, incluindo insuficiências nos controles internos, falhas de planejamento, desafios na responsabilização e fiscalização contratual, deficiências na gestão de riscos, aspectos culturais e limitações de capital humano. A pesquisa evidenciou que tais determinantes atuam de maneira combinada, refletindo debilidades nos mecanismos de governança organizacional.

Queiroz (2020) conduziu avaliação de desempenho de Restos a Pagar Não Processados em diversos *campi* do Instituto Federal de Educação, Ciência e Tecnologia de Rondônia (IFRO). O estudo revelou que o indicador de eficácia orçamentária anual no período estudado (2016 a 2019), nos nove *campi* analisados, apresenta desempenho insatisfatório. As motivações internas e externas para tal desempenho também foram coletadas por meio de entrevistas e apresentadas no trabalho.

Barbosa e Rodrigues (2023) levantaram dados orçamentários, financeiros e contábeis do governo federal entre 2008 e 2020 e aplicaram a técnica de regressão linear múltipla com o objetivo de identificar de que forma a liberação de limites orçamentários próxima ao final do exercício financeiro, bem como a inscrição de despesas em restos a pagar, afetam os cancelamentos destas despesas. A principal motivação para o desenvolvimento do trabalho foi a preocupação com os prejuízos que esses cancelamentos trazem à Administração Pública.

3.2 Estudos sobre Eficiência na Gestão de Restos a Pagar

Análise Comparativa de Eficiência

A avaliação da eficiência na gestão de Restos a Pagar também tem sido objeto de desenvolvimento de pesquisas, especialmente através da aplicação de técnicas quantitativas que permitem comparações entre organizações similares.

Mota (2021) realizou estudo sobre eficiência relativa da gestão de Restos a Pagar nas universidades federais brasileiras no contexto do Decreto nº 9.428/2018, que limitou

a três anos o prazo para execução de Restos a Pagar Não Processados. Utilizando Análise Envoltória de Dados (DEA) e o Índice de Produtividade de Malmquist, o autor analisou 63 universidades federais no período 2018-2020. Os resultados indicaram que, de maneira geral, as universidades federais não apresentaram aumento de eficiência na gestão de Restos a Pagar devido à implementação do decreto. O estudo revelou que a média de eficiência das despesas de capital foi acentuadamente menor que a do grupo de outras despesas correntes, e observou-se pequena perda de produtividade de aproximadamente 5% nas outras despesas correntes.

Andrade (2022) contribuiu para esta vertente ao examinar a gestão de Restos a Pagar no Conselho Administrativo de Defesa Econômica (Cade). Sob o ponto de vista quantitativo, o autor realizou uma análise descritiva de dados extraídos do SIAFI.

Estudos de Caso e Análises Descritivas

Alguns trabalhos adotaram abordagem de estudo de caso para investigar a evolução e características dos Restos a Pagar em contextos específicos, fornecendo percepções específicas sobre realidades locais.

Pinheiro e Hollanda (2020) analisaram a evolução dos Restos a Pagar no município de Ladário-MS, oferecendo perspectiva longitudinal sobre como fatores locais influenciam a gestão orçamentária municipal. O estudo revelou padrões específicos de inscrição e execução de Restos a Pagar em contexto municipal de pequeno porte.

Santos et al. (2021) investigaram a situação dos Restos a Pagar no estado de Alagoas, contribuindo para a compreensão dos desafios enfrentados por governos estaduais na gestão destes recursos. Os resultados do estudo indicaram uma boa evolução na gestão orçamentária do estado de Alagoas a partir de 2014, fruto dos esforços de gestão na administração dos Restos a Pagar.

Nonaka (2019) desenvolveu estudo específico sobre Restos a Pagar Não Processados, explorando suas características distintivas e impactos na gestão orçamentária. A pesquisa contribui para a compreensão das diferenças entre as modalidades de Restos a Pagar e suas implicações gerenciais.

3.3 Estudos sobre Impactos e Consequências dos Restos a Pagar

Impactos na Credibilidade Orçamentária

Um estudo sobre as consequências da quantidade excessiva de despesas em Restos a Pagar na credibilidade e na transparência do orçamento público foi encontrado.

Aquino e Azevedo (2017) conduziram investigação sobre Restos a Pagar e perda de credibilidade orçamentária, analisando dados do governo federal, 26 estados e Distrito Federal, e cerca de 4.100 municípios. Os autores identificaram o surgimento de um “or-

çamento paralelo" nos três níveis de governo, evidenciando que o crescente uso de Restos a Pagar está reduzindo seriamente a credibilidade e transparência do orçamento.

O estudo revelou que o total de Restos a Pagar do governo federal aumentou 277% em valores atualizados de 2003 a 2014, saltando de R$ 33 para R$ 227 bilhões. Os autores alertam que a fraca regulação, sobretudo dos Restos a Pagar Não Processados, compromete a sustentabilidade fiscal e a confiabilidade das informações orçamentárias.

Ferreira e Souza (2022), abordado em mais detalhes a seguir, também traz como resultado a constatação de que a falta de transparência é um problema do orçamento público brasileiro.

Gestão de Cancelamentos

Foram encontrados estudos específicos sobre cancelamento de Restos a Pagar, que exploram tanto as causas quanto as implicações destes cancelamentos para a gestão orçamentária.

Ferreira e Souza (2022) investigaram os determinantes e processos de cancelamento de Restos a Pagar, contribuindo para a compreensão de quando e por que despesas empenhadas são posteriormente canceladas. Os resultados da pesquisa indicaram uma relação significativa entre o volume de saldo inscrito em restos a pagar e o cancelamento de despesas inscritas em RP. O estudo conclui que o "uso" de Restos a Pagar no Brasil, como forma de flexibilizar a anualidade orçamentária, está gerando endividamento sem transparência, tornando o orçamento uma "peça de ficção" e perdendo sua credibilidade como ferramenta de planejamento e controle

Araújo et al. (2022) exploraram o fenômeno *use it or lose it* em relação aos Restos a Pagar. Em uma tradução idiomática livre, podemos entender como "usa senão vai perder". Este estudo analisa como a pressão temporal para execução orçamentária influencia as decisões de gestores públicos. O trabalho contribui para a compreensão dos incentivos perversos criados pela rigidez da anualidade orçamentária, podendo levar a uma execução orçamentária pouco criteriosa e a gastos de baixa prioridade e qualidade.

3.4 Estudos sobre a aplicação de IA ao orçamento público

Dois trabalhos sobre a aplicação de técnicas de IA ao contexto de orçamento público foram encontrados.

Inteligência artificial aplicada ao orçamento público mexicano

Valle-Cruz et al. (2020) focou especificamente na etapa de planejamento do ciclo orçamentário, buscando entender como a IA pode apoiar a tomada de decisão governamental. A abordagem metodológica fundamentou-se em uma perspetiva algorítmica baseada

em dados públicos do orçamento federal do governo do México referentes ao período de 2012 a 2018, incluindo variáveis de Produto Interno Bruto (PIB) e inflação anual.

O estudo aplicou Redes Neurais Artificiais (RNA) para identificar os efeitos de 12 variáveis de entrada (como saúde, educação, energia e transportes) sobre as variáveis de saída (desenvolvimento social, desenvolvimento econômico, despesas governamentais e orçamento não programável). As equações resultantes foram otimizadas por meio de Algoritmos Genéticos (AG) para encontrar a proporção ideal de alocação orçamentária que maximizasse o PIB e minimizasse a inflação.

A IA identificou que os setores com maior influência no orçamento público e na economia mexicana são os assuntos financeiros e fiscais, serviços gerais, comunicações, despesas governamentais e o setor de energia. Para uma gestão orçamentária mais "inteligente", o modelo sugeriu um aumento nos investimentos em desenvolvimento social e econômico, além de um incremento no orçamento para despesas não programáveis (para ativar a economia).

Concluiu-se que a IA é viável e promissora para o setor público, não para substituir o gestor, mas como uma ferramenta de análise de dados capaz de gerar ideias, oportunidades e simulações de cenários que os métodos estatísticos tradicionais (como regressões lineares) podem não capturar.

Aprendizado de máquina para minimização de perda orçamentária por cancelamento

Santos (2023) investiga o uso de inteligência artificial para aprimorar a execução orçamentária no setor público brasileiro. O objetivo primordial do estudo foi simular o processo de inscrição de Restos a Pagar (RP), focando na classificação de obras contratadas pela Diretoria de Obras Militares (DOM) do Exército Brasileiro entre 2016 e 2018.

A pesquisa visou identificar empenhos com alta probabilidade de serem cancelados, permitindo que o gestor atue preventivamente para reduzir perdas de créditos orçamentários autorizados na LOA por meio de realocação de recursos, reduzindo em 30% o valor destes empenhos identificados e destinando-os a empenhos com baixa probabilidade de cancelamento.

Para isso, a metodologia empregou o aprendizado de máquina supervisionado através do algoritmo *Random Forest*. A base de dados integrou informações financeiras e de engenharia do sistema OPUS, informações orçamentárias do Tesouro Gerencial (SIAFI) e dados qualitativos sobre fornecedores da empresa Trovale Informática, que trabalha com soluções *Big Data* e *Analytics*.

O modelo foi treinado com dados dos anos de 2016 e 2017 e testado por meio de uma simulação da inscrição de RP referente ao exercício de 2018. Técnicas como o *random undersampling* foram aplicadas para balancear a amostra de treino, e diferentes combinações de variáveis preditivas (como características da obra e do fornecedor) foram

avaliadas para otimizar o desempenho.

Os resultados alcançados demonstraram a viabilidade de uso do modelo, com a melhor configuração de variáveis atingindo uma sensibilidade de 92% e precisão de 93%. A simulação de aplicação prática indicou que o uso da ferramenta geraria uma redução de aproximadamente 21% nos cancelamentos de Restos a Pagar Não Processados (RPNP).

O autor concluiu que a aplicação de modelos econométricos baseados em aprendizado de máquina pode traduzir-se em melhorias significativas na eficiência da gestão de recursos públicos.

3.5 Posicionamento da Pesquisa

Síntese dos Achados da Literatura

A análise da literatura revela convergência em torno de alguns achados fundamentais. Primeiramente, há consenso sobre a influência significativa do ciclo político-eleitoral na gestão de Restos a Pagar, com evidências consistentes de comportamento oportunista por parte de mandatários de cargos eletivos do Poder Executivo.

Segundo, estudos demonstram que características organizacionais e setoriais exercem influência determinante nos padrões de inscrição e gestão de Restos a Pagar.

Terceiro, há reconhecimento de que o uso excessivo de Restos a Pagar compromete tanto a credibilidade e a transparência orçamentária como a possibilidade de bom uso dos recursos públicos.

Na Tabela 11, apresenta-se a síntese dos trabalhos relacionados e a da presente pesquisa, para fins comparativos.

Contribuição da Presente Pesquisa

A presente pesquisa propõe-se a oferecer, como contribuição metodológica, de que forma a integração de dados orçamentários (SIAFI) e de tramitação processual (SEI) permite uma análise mais abrangente dos fatores determinantes dos Restos a Pagar.

Do ponto de vista temático, a análise do impacto do ciclo eleitoral como variável preditiva em contexto organizacional específico pode oferecer novas percepções sobre como características institucionais influenciam a gestão orçamentária.

Este trabalho também se propõe a contribuir para o avanço do conhecimento na intersecção entre gestão orçamentária e inteligência artificial aplicada ao setor público, área emergente com potencial significativo de impacto na modernização da administração pública brasileira.

Tabela 11 – Síntese dos trabalhos relacionados e posicionamento da pesquisa

Autor(es) / Ano	Fator Eleitoral	Dados Processuais	Aprendizado de Máquina	Predição de RP	Foco Temático
Aquino e Azevedo (2017)	Não	Não	Não	Não	Credibilidade orçamentária
Nonaka (2019)	Não	Não	Não	Não	Características dos RPNP
Queiroz (2020)	Não	Não	Não	Não	Eficiência e desempenho
Pinheiro e Hollanda (2020)	Não	Não	Não	Não	Evolução dos RP
Santos et al. (2021)	Não	Não	Não	Não	Gestão orçamentária
Mota (2021)	Não	Não	Não	Não	Eficiência (DEA)
Ferreira e Souza (2022)	Não	Não	Não	Não	Cancelamento de RP
Araújo (2022)	Sim	Não	Não	Não	Incentivos eleitorais
Araújo et al. (2022)	Não	Não	Não	Não	Fenômeno *use it or lose it*
Andrade (2022)	Não	Não	Não	Não	Eficiência e desempenho
Araújo et al. (2023)	Sim	Não	Não	Não	Incentivos eleitorais
Almeida et al. (2023)	Sim	Não	Não	Não	Incentivos eleitorais
Barbosa e Rodrigues (2023)	Não	Não	Não	Não	Cancelamento de RP
Silva e Junior (2024)	Sim	Não	Não	Não	Inscrição de RP
Maurício (2025)	Não	Não	Não	Não	Cancelamento de RP
Valle-Cruz et al. (2020)	Não	Não	Sim	Não	Planejamento orçamentário
Santos (2023)	Não	Não	Sim	Sim	Cancelamento de RP
Presente pesquisa	Sim	Sim	Sim	Sim	**Inscrição de RP**

Fonte: Elaborado pelo autor (2026).

4 PROCEDIMENTOS METODOLÓGICOS

O presente estudo caracteriza-se como uma pesquisa aplicada de natureza quantitativa, desenvolvida através de um estudo de caso no Tribunal Regional Eleitoral de Goiás. A abordagem quantitativa justifica-se pela necessidade de analisar dados estruturados e desenvolver modelos preditivos baseados em técnicas estatísticas e de aprendizado de máquina.

Adotou-se como estratégia de pesquisa o emprego de dados longitudinais, abrangendo o período de 2022 a 2025. A unidade de análise é a nota de empenho, considerando como variável dependente a inscrição ou não em Restos a Pagar ao final do exercício financeiro. O método principal empregado é o aprendizado de máquina supervisionado, especificamente algoritmos de classificação binária.

4.1 Contexto e Local da Pesquisa

A execução orçamentária do Tribunal Regional Eleitoral de Goiás, órgão de justiça especializada em matéria eleitoral, foi escolhida como objeto de estudo. Órgãos dessa classe apresentam um ciclo orçamentário diferenciado, alternando entre anos eleitorais (pares) e não eleitorais (ímpares), o que resulta em variações significativas na carga de trabalho e na complexidade da execução orçamentária. Estes elementos enriquecem e particularizam a análise da problemática dos Restos a Pagar.

Durante anos eleitorais, o tribunal opera com dois orçamentos distintos: o ordinário, destinado às atividades rotineiras, e o eleitoral, específico para a realização das eleições. Esta dualidade orçamentária, combinada com a intensificação das atividades durante o período eleitoral, cria um ambiente propício para a análise dos fatores que influenciam a inscrição de despesas em Restos a Pagar.

A escolha do TRE-GO também se fundamenta na disponibilidade e qualidade dos dados nos sistemas corporativos, bem como no acesso facilitado às informações necessárias para a pesquisa. O período de análise (2022-2025) foi definido considerando-se a integridade dos dados nos sistemas atuais e a necessidade de abranger um ciclo eleitoral completo.

4.2 Modelagem do Problema

A identificação de notas de empenho que podem ser inscritas em restos a pagar se encaixa em um problema de classificação binária supervisionada. O objetivo é desenvolver um modelo, f, que mapeie um vetor de características $\mathbf{x}_i \in \mathbb{R}^p$ de uma nota de empenho para uma variável alvo binária $y_i \in \{0, 1\}$. A variável y_i assume o valor 1 se a nota de

empenho foi inscrita em restos a pagar e 0, caso contrário. Formalmente, o modelo busca estimar a probabilidade condicional $P(Y = 1|\mathbf{X} = \mathbf{x})$.

Seja $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$ o conjunto de dados de treinamento, onde:

- $\mathbf{x}_i \in \mathbb{R}^p$ é o vetor de características do i-ésimo empenho;
- $y_i \in \{0, 1\}$ é o rótulo binário indicando a inscrição em Restos a Pagar;
- n é o número de observações; e
- p é o número de variáveis preditoras.

O objetivo é encontrar uma função $f : \mathbb{R}^p \to [0, 1]$ tal que:

$$f(\mathbf{x}) = P(Y = 1 \mid \mathbf{X} = \mathbf{x}), \tag{7}$$

onde $P(Y = 1 \mid \mathbf{X} = \mathbf{x})$ representa a probabilidade condicional de uma nota de empenho ser inscrita em Restos a Pagar dado o vetor de características $\mathbf{x}$.

A modelagem das características da nota de empenho é feita a seguir.

Modelagem das Características para Predição de Restos a Pagar

As características das notas de empenho para o desenvolvimento da presente pesquisa foram escolhidas com base na sua relevância em contribuir para a predição de inscrição em Restos a Pagar. As características escolhidas podem ser organizadas em três categorias principais.

A primeira categoria compreende as características da despesa:

- **VALOR**: valor monetário da nota de empenho;
- **NATUREZA DA DESPESA**: código da classificação orçamentária. A Natureza da Despesa é detalhada na Seção 4.7;
- **TIPO**: categoria do empenho - ordinário, global, estimativo;
- **MODALIDADE DA LICITAÇÃO**: modalidade do processo de contratação, como pregão, tomada de preços, inexigibilidade, suprimento de fundos, entre outros.

A segunda categoria abrange os aspectos temporais:

- **MÊS DE EMISSÃO DA NOTA DE EMPENHO**: o mês em que a nota de empenho foi criada. Os valores possíveis compreendem a faixa entre janeiro e dezembro; e
- **ANO**: o ano de criação da nota de empenho. A importância de sua individualização reside no fato de que pode-se inferir a partir de seu valor se se trata de ano eleitoral ou não.

A terceira categoria engloba as métricas processuais, derivadas da análise da tramitação administrativa, todas elas consideradas dentro do ano de exercício orçamentário em análise:

- **QUANTIDADE DE TRAMITAÇÕES**: número total de movimentações do processo, ou seja, a quantidade total de vezes em que o processo saiu de uma unidade do TRE-GO e ingressou em outra;
- **QUANTIDADE DE REENVIOS**: frequência com que o processo retornou a uma unidade pela qual já havia tramitado; e
- **QUANTIDADE MÁXIMA DE DIAS "PARADO"**: maior período, em dias, que o processo permaneceu em uma única unidade organizacional.

4.3 Coleta e Fonte de Dados

A coleta de dados foi realizada através de dois sistemas corporativos do Tribunal Regional Eleitoral de Goiás. O primeiro sistema utilizado foi o SIAFI (Sistema Integrado de Administração Financeira), acessado através do módulo Tesouro Gerencial, do qual foram extraídas informações orçamentárias e financeiras relacionadas às notas de empenho, incluindo dados sobre valores, natureza da despesa, modalidade de licitação e fase de pagamento (empenho, liquidação, pagamento, restos a pagar).

O segundo sistema utilizado foi o SEI (Sistema Eletrônico de Informações), responsável pela tramitação dos processos administrativos do órgão. Deste sistema foram extraídas informações sobre o histórico de tramitação dos processos relacionados às notas de empenho, incluindo dados sobre movimentações, tempos de permanência em cada unidade organizacional e ocorrências de reenvios.

O processo de extração de dados foi conduzido através de consultas estruturadas aos bancos de dados corporativos, seguindo critérios rigorosos de inclusão e exclusão. Foram incluídas todas as notas de empenho emitidas no período de 2022 a 2025, excluindo-se apenas aquelas canceladas ou com dados incompletos que pudessem comprometer a qualidade da análise.

Registra-se que, até agosto de 2022, o sistema administrativo de tramitação processual utilizado pelo TRE-GO era o Processo Administrativo Digital (PAD), e como a vinculação entre o número da nota de empenho e o número do processo era feita manualmente, ocorreram muitos erros e inconsistências por erros de digitação. Por esse motivo, não foi possível a utilização de dados de notas de empenho e suas respectivas tramitações processuais anteriores a 2022.

4.4 Estruturação do *Dataset*

O *dataset* foi estruturado tendo como unidade de observação a nota de empenho, resultando em um conjunto de dados com as seguintes características: a variável dependente TARGET_RP foi definida como binária, assumindo valor 1 para notas de empenho inscritas em Restos a Pagar e valor 0 para as não inscritas.

As variáveis independentes foram organizadas em três categorias principais, de acordo com a modelagem descrita na Seção 4.2. Apresenta-se, na Tabela 12, a descrição detalhada das variáveis preditoras.

Tabela 12 – Descrição das variáveis do conjunto de dados

Variável	Descrição	Tipo	Exemplo
NE	Número identificador da Nota de Empenho	Identificador	20230439
TARGET_RP	Variável alvo: indica se a NE foi inscrita em RP (1) ou não (0)	Binário	1
ND	Natureza da Despesa (código completo)	Categórico	33903948
GND	Grupo de Natureza da Despesa (derivado da ND)	Categórico	3
ELEMENTO	Elemento de Despesa (derivado da ND)	Categórico	39
VALOR_NE	Valor total da Nota de Empenho em reais (R$)	Numérico	9700,00
ANO_ELEICAO	Indica se o ano de emissão da NE foi eleitoral (1) ou não (0)	Binário	0
MES_EMISSAO_NE	Mês de emissão da Nota de Empenho (1 a 12)	Numérico	10
QTDE_-TRAMITACOES	Quantidade total de tramitações do processo administrativo	Numérico	41
QTDE_REENVIOS	Quantidade de vezes que o processo foi reenviado entre unidades	Numérico	37
QTDE_MAX_-DIAS_-PROCESSO_-PARADO_UNIDADE	Número máximo de dias que o processo ficou parado em uma unidade	Numérico	42
TIPO	Tipo do processo (ex.: Aquisição, Serviço)	Categórico	0
MODALIDADE	Modalidade de licitação (ex.: Pregão, Inexigibilidade)	Categórico	3

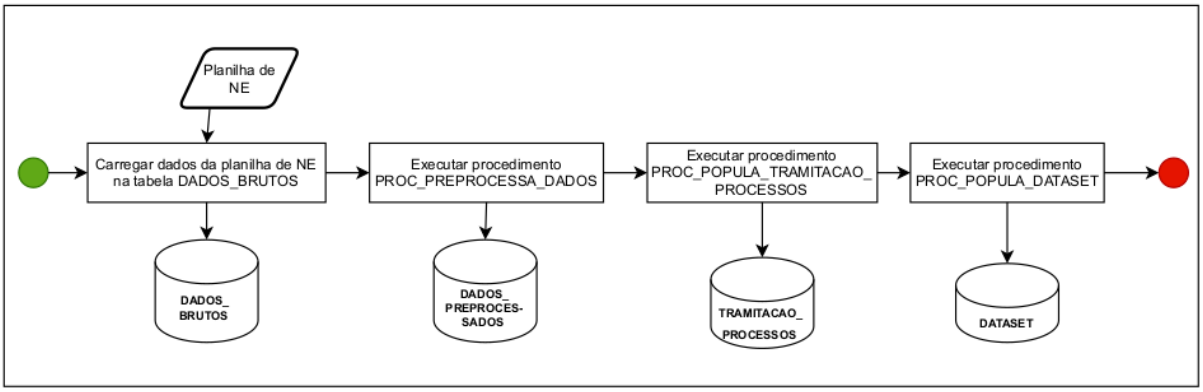
Fonte: Elaborado pelo autor (2026).

4.5 Pré-processamento e Tratamento de Dados

Para a obtenção do *dataset*, foram necessárias quatro etapas distintas, ilustradas na Figura 12:

1. Carga dos dados brutos da planilha de notas de empenho gerada pelo SIAFI;
2. Pré-filtragem de notas de empenho;
3. Busca de informações de tramitação processual das notas de empenho; e
4. População da tabela do dataset.

Figura 12 – Fluxo de Preparação do Dataset



Fonte: Elaborado pelo autor (2025)

O diagrama detalha a sequência de procedimentos armazenados (*stored procedures*) que transformam os dados brutos extraídos do SIAFI e do SEI, passando pelas etapas de pré-processamento e consolidação da tramitação de processos, até a geração da tabela final (DATASET) que alimenta os algoritmos de aprendizado de máquina. Cada uma destas fases é detalhada a seguir[1].

Carga de Dados das Notas de Empenho

Na primeira fase, fez-se a carga dos dados da planilha de notas de empenho de 2022 a 2025, extraída do SIAFI - Tesouro Gerencial, para a tabela DADOS_BRUTOS. Um total de 157.755 linhas foram importadas.

[1]Utilizou-se, para esta fase, um Sistema Gerenciador de Banco de Dados (SGBD) Oracle versão 19 para armazenamento dos dados e rotinas (procedures) em linguagem PL/SQL, a linguagem procedimental nativa deste SGBD.

Pré-processamento

Em seguida, o procedimento PROC_PREPROCESSA_DADOS populou a tabela DADOS_PREPROCESSADOS a partir dos dados brutos. As linhas da tabela DADOS_-BRUTOS que se enquadram nas seguintes condições foram excluídas:

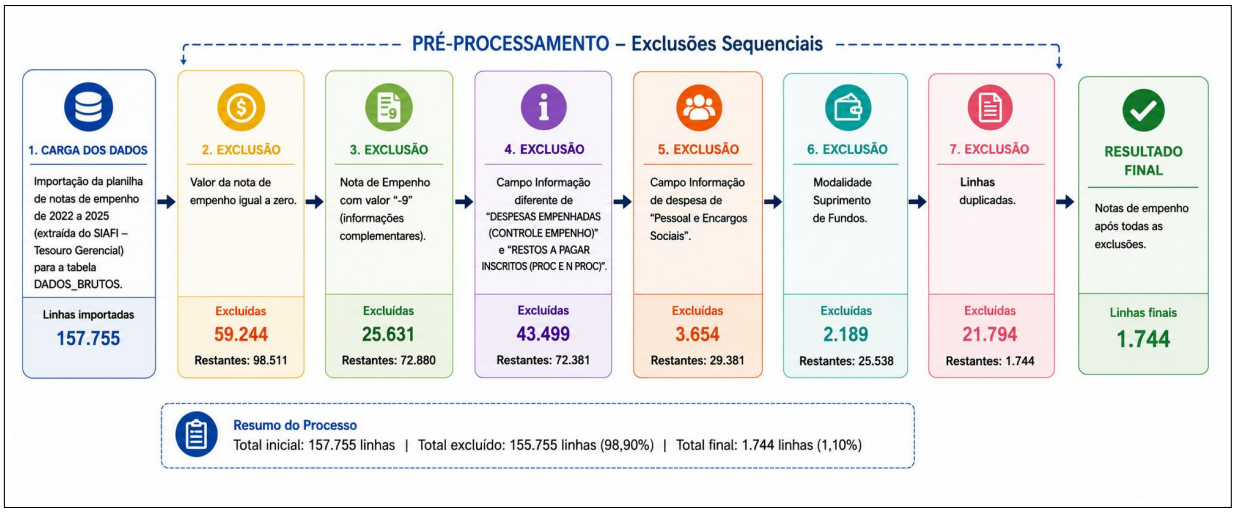
- **Valor da nota de empenho igual a zero**: notas de empenho com valor zero não interessam ao estudo, pois não são passíveis de serem inscritas em RP. 59.244 linhas foram excluídas neste passo;
- **Nota de Empenho com valor "-9"**: Estas linhas correspondem a informações complementares, como valor de crédito disponível ou valor pré-empenhado, que não fazem parte do escopo deste trabalho. 25.631 linhas foram excluídas neste passo;
- **Campo Informação diferente de *DESPESAS EMPENHADAS (CONTROLE EMPENHO)* e *RESTOS A PAGAR INSCRITOS (PROC E N PROC)*:** estas informações significam, respectivamente, notas de empenho não inscritas em RP e inscritas em RP, que são as duas situações que interessam ao presente trabalho. Um total de 43.499 linhas foram excluídas neste passo;
- **Grupo de natureza de despesa de *Pessoal* e *Encargos Sociais*:** por serem despesas com baixíssimo índice histórico de inscrição em RP, por se tratarem de categorias de despesas obrigatórias, optou-se por desconsiderá-las; assim, para este trabalho apenas as notas de empenho de despesas **discricionárias**, ou seja, aquelas sobre as quais a Administração possui poder de decisão, representam interesse de estudo. 3.654 linhas foram excluídas;
- **Modalidade *Suprimento de Fundos***: notas de empenho relacionadas a despesas pagas com suprimentos de fundos devem, segundo o regramento orçamentário, ser liquidadas dentro do mesmo exercício (Brasil, 1986), e portanto não podem ser inscritas em RP. Nesta etapa, foram excluídas 2.189 linhas;
- **Linhas duplicadas**: Diversas linhas estavam duplicadas na planilha de dados brutos, devido a andamentos de notas de empenho não capturados na exportação da planilha. Nesta etapa final, foram excluídas 21.794 linhas;

Após a aplicação destas exclusões, a população da tabela DADOS_PREPROCESSADOS resultou em 1.744 linhas. Na Figura 13, representa-se a sequência de exclusões da fase de pré-processamento.

Carga dos Dados de Tramitação Processual

A próxima fase corresponde à busca de informações de tramitação dos processos referentes às notas de empenho selecionadas. O procedimento PROC_POPULA_TRA-

Figura 13 – Exclusões do pré-processamento



Fonte: Elaborado pelo autor (2026), com o apoio da ferramenta ChatGPT

MITACAO_PROCESSOS foi construído para filtrar as tramitações destes processos e popular a tabela TRAMITACAO_PROCESSOS. Esta tabela atingiu um total de 65.971 linhas.

Na Tabela 13, apresentam-se as descrições de como variáveis preditoras relacionadas à tramitação processual (QTDE_TRAMITACOES, QTDE_REENVIOS e QTDE_MAX_DIAS_PROCESSO_PARADO_UNIDADE) foram calculadas.

Tabela 13 – Descrição do cálculo das variáveis de tramitação processual

Variável	Descrição
QTDE_TRAMITACOES	Contagem do total de tramitações processuais para cada processo associado a uma nota de empenho
QTDE_REENVIOS	Identificação e soma de envios do processo a unidades pelas quais o processo já havia passado
QTDE_MAX_DIAS_PROCESSO_PARADO_UNIDADE	Identificação de todas as diferenças entre data de saída e data de entrada de cada processo nas unidades por onde tramitou e seleção do maior valor

Fonte: Elaborado pelo autor (2026).

População da tabela DATASET

A última fase do pré-processamento e carga de dados refere-se à população da tabela DATASET, cujas colunas correspondem exatamente às variáveis preditoras descritas na Tabela 12. O procedimento utilizado nesta fase, PROC_POPULA_DATASET, relaciona as tabelas DADOS_PRE_PROCESSADOS e TRAMITACAO_PROCESSOS por meio da coluna número do processo administrativo (PROCESSO).

Caracterização do *Dataset*

O *dataset* final compreendeu um total de 1.744 notas de empenho emitidas no período de 2022 a 2025. A distribuição temporal destas notas de empenho foi a seguinte: 2022 (ano eleitoral) apresentou 501 notas de empenho, 2023 (ano não eleitoral) registrou 397 notas, 2024 (ano eleitoral) contabilizou 352 notas e 2025 (ano não eleitoral), totalizou 494 notas de empenho.

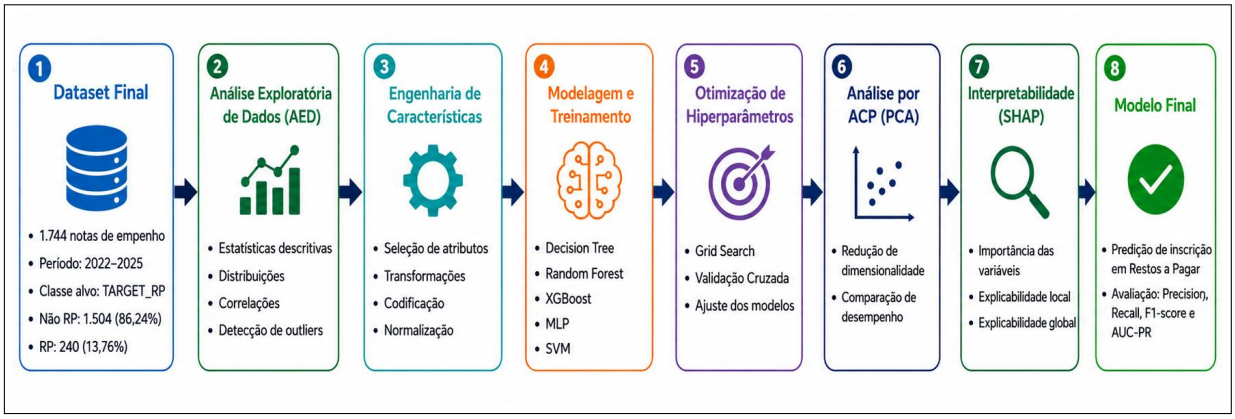
A variável dependente TARGET_RP apresentou distribuição desbalanceada, característica esperada para este contexto. Do total de notas de empenho analisadas, 1504 (86,24%) não foram inscritas em Restos a Pagar, enquanto 240 (13,76%) foram inscritas.

Esta proporção de aproximadamente 5:1 justifica a necessidade de técnicas específicas para tratamento do desbalanceamento, pois a maioria dos algoritmos de aprendizado padrão pressupõe distribuições de classes balanceadas e, na presença de dados desbalanceados, tende a produzir regras de indução mais fracas para a classe minoritária, resultando em acurácias desfavoráveis (He; Garcia, 2009).

4.6 O *Pipeline* Preditivo

Após a obtenção do conjunto de dados, adotou-se a construção de um *pipeline* de aprendizado de máquina para o desenvolvimento da solução tecnológica. O *pipeline* possui uma estrutura sequencial e iterativa, o que proporciona rastreabilidade e repetibilidade. O *pipeline* preditivo deste trabalho é composto por seis fases distintas: análise exploratória de dados (AED), engenharia de características (*feature engineering*), modelagem e treinamento,otimização de hiperparâmetros, análise de componentes principais (ACP) e interpretabilidade com SHAP. Tem-se, na Figura 14, a representação do *pipeline* preditivo adotado no presente trabalho, que é descrito em detalhes nas seções seguintes.

Figura 14 – *Pipeline preditivo*



Fonte: Elaborado pelo autor (2026), com o apoio da ferramenta ChatGPT.

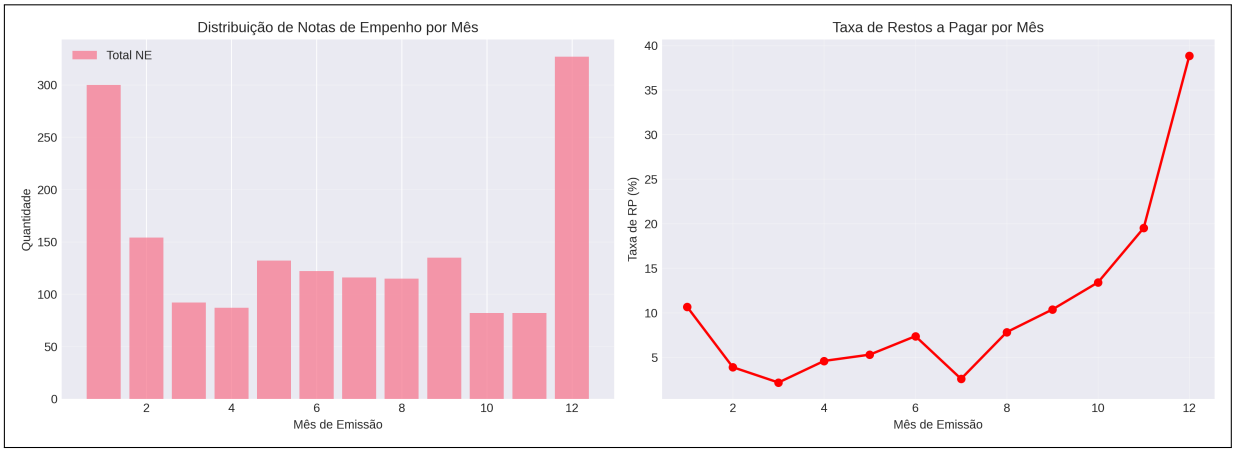
4.6.1 Análise Exploratória de Dados (AED)

A primeira fase consistiu em uma análise inicial dos dados para extrair insights iniciais e guiar as etapas subsequentes de pré-processamento. A AED revelou alguns padrões, que são descritos em seguida. A maioria destes padrões afeta a fase subsequente de modelagem.

Padrão Temporal

A análise da distribuição de RP ao longo do ano (Figura 15) mostrou um aumento drástico e inequívoco na taxa de inscrição em RP nos últimos meses do ano, especialmente em dezembro, como mostra a Figura 15.

Figura 15 – Distribuição mensal das notas de empenho (2022 a 2025)



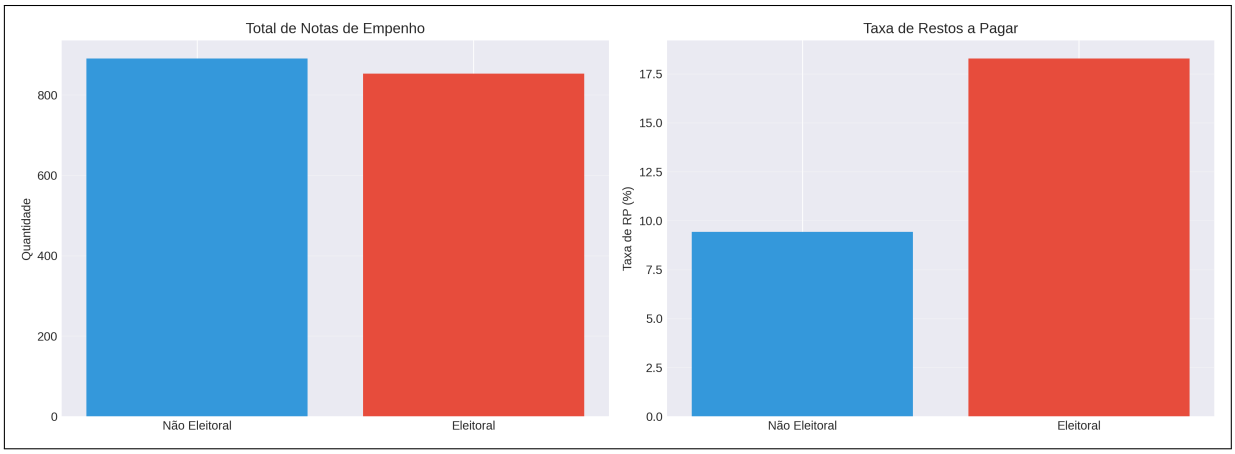
Fonte: Elaborado pelo autor (2026)

A figura contrapõe o volume de emissão de empenhos à taxa de conversão em Restos a Pagar ao longo dos meses do ano. Enquanto o gráfico de barras mostra uma distribuição de emissões relativamente uniforme (com picos no início e no fim do ano), o gráfico de linhas revela um crescimento exponencial da taxa de RP no último trimestre. Essa divergência visual comprova que o risco de inscrição em RP não é proporcional ao volume empenhado, mas sim fortemente dependente da sazonalidade de fim de exercício.

Além disso, observou-se um aumento percentual de notas de empenho inscritas em RP em anos eleitorais, como mostra a Figura 16.

Os gráficos de barras evidenciam o impacto do ciclo eleitoral na execução orçamentária do TRE-GO. Embora o volume total de notas de empenho emitidas seja estatisticamente similar entre anos eleitorais e não eleitorais, a taxa de conversão em Restos a Pagar quase dobra em anos de pleito (saltando de aproximadamente 8% para 17,5%). Essa constatação empírica justifica a inclusão de variáveis temporais relacionadas às eleições no modelo preditivo.

Figura 16 – Taxa de notas de empenho inscritas em RP

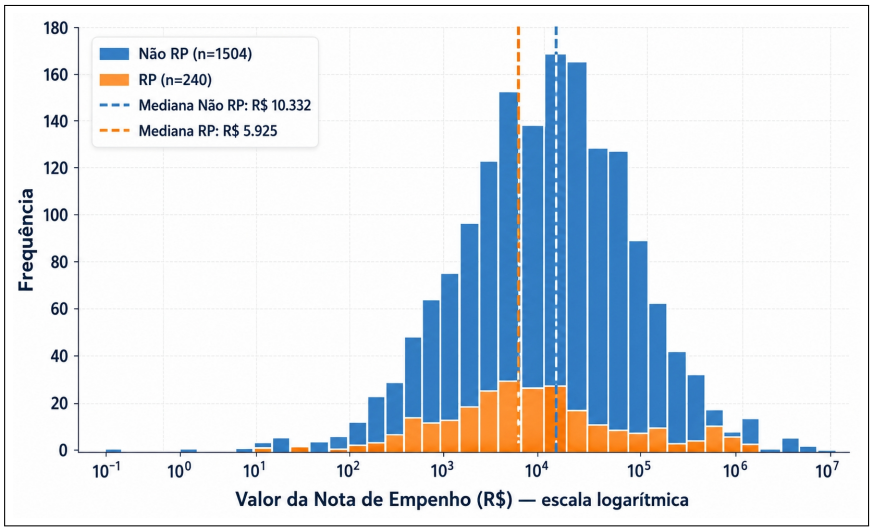


Fonte: Elaborado pelo autor (2026)

Distribuição de Valores

A variável VALOR_NE apresentou uma forte assimetria à direita, com a maioria dos empenhos concentrada em valores mais baixos e alguns poucos empenhos de valor muito elevado, o que é ilustrado na Figura 17. Tal distribuição sugere a necessidade de uma transformação (como a logarítmica) para normalizar a escala e reduzir a influência dos *outliers*[2].

Figura 17 – Distribuição de valores de notas de empenho



Fonte: Elaborado pelo autor (2026)

A visualização mostra que a grande maioria das despesas concentra-se em valores entre mil e cem mil real, enquanto algumas despesas ultrapassam o valor de um milhão de reais, atingindo até o valor superior a dez milhões de reais.

[2]Valores atípicos em um conjunto de dados, muito acima ou muito abaixo da maioria dos valores.

Correlações

Os resultados da análise de correlação entre as variáveis preditoras numéricas e a variável-alvo `TARGET_RP` são apresentados na Tabela 14, revelam que todas as variáveis numéricas do *dataset* exibem correlações de baixa magnitude com a variável-alvo.

Tabela 14 – Correlação entre variáveis numéricas e a variável-alvo `TARGET_RP`

Variável	Correlação (r)
`QTDE_REENVIOS`	+0,044
`QTDE_TRAMITACOES`	+0,031
`QTDE_MAX_DIAS_PROCESSO_PARADO_UNIDADE`	+0,021
`VALOR_NE`	−0,010

Fonte: Elaborado pelo autor (2026).

A variável `QTDE_REENVIOS` apresentou a maior correlação positiva ($r = 0{,}044$), indicando que notas de empenho cujos processos administrativos sofreram maior número de reenvios entre unidades tendem, ainda que de forma modesta, a ser inscritas em Restos a Pagar com maior frequência.

As variáveis `QTDE_TRAMITACOES` ($r = 0{,}031$) e `QTDE_MAX_DIAS_PROCESSO_PARADO_UNIDADE` ($r = 0{,}021$) também apresentaram correlação positiva, embora ainda mais tênue, sugerindo que processos mais longos e com maiores períodos de paralisação estão levemente associados à inscrição em RP.

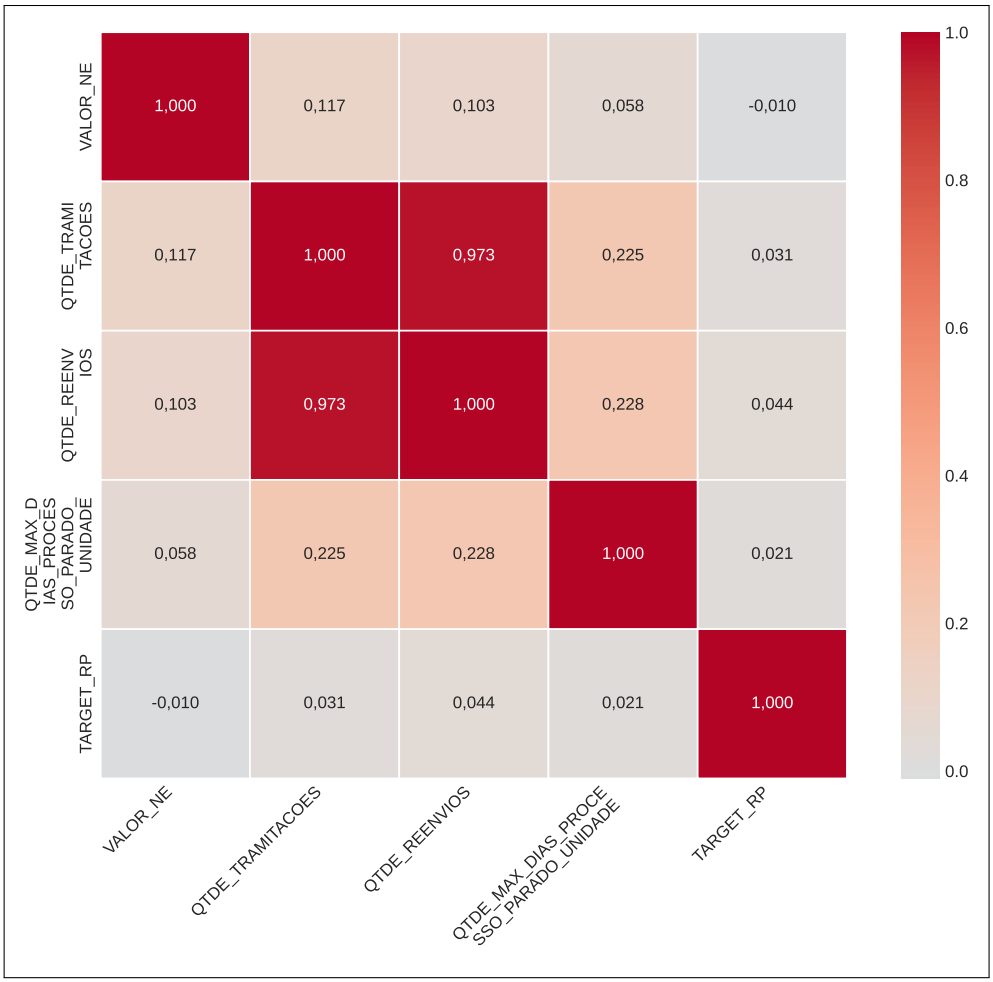
O `VALOR_NE` exibiu correlação negativa e próxima de zero ($r = -0{,}010$), indicando ausência de associação linear direta entre o valor da nota de empenho e a probabilidade de inscrição em Restos a Pagar.

A baixa magnitude geral dos coeficientes não implica, contudo, que as variáveis sejam irrelevantes para a predição: relações não lineares e interações entre variáveis, capturadas pelos modelos de aprendizado de máquina, especialmente pelo Random Forest e pelo XGBoost, podem conferir poder preditivo mesmo na ausência de correlação linear expressiva (Breiman, 2001).

Identificou-se uma alta correlação positiva entre as variáveis QTDE_TRAMITACOES e QTDE_REENVIOS (Figura 18), indicando que processos com mais trâmites também tendem a ser mais reenviados. Essa observação inspirou a criação de uma nova variável, composta por estas duas, na fase seguinte: a taxa de reenvios.

O mapa de calor (*heatmap*) apresenta as correlações lineares entre as variáveis numéricas do modelo. O destaque visual é a fortíssima correlação positiva (0,970) entre a quantidade de tramitações e a quantidade de reenvios do processo, indicando redundância informacional. Por outro lado, a variável de valor do empenho apresenta correlações muito baixas com as métricas processuais, sugerindo que ela captura uma dimensão independente do risco de inscrição em RP.

Figura 18 – Matriz de correlação das variáveis preditoras



Fonte: Elaborado pelo autor (2026)

4.6.2 Engenharia de Características (*Feature Engineering*)

Com base nos achados da AED, a segunda fase focou na preparação e transformação dos dados para otimizar o aprendizado dos modelos. As principais etapas são apresentadas a seguir.

Transformação Logarítmica

Para mitigar a assimetria da variável VALOR_NE, foi aplicada uma transformação logarítmica na forma $\log(1 + x)$, resultando na nova variável VALOR_NE_LOG. O uso de $\log(1 + x)$ em vez de $\log(x)$ garante que valores nulos sejam tratados sem indefinição matemática (Hastie et al., 2009).

Criação de Novas Variáveis

As seguintes variáveis derivadas foram criadas a partir das preditoras originais:

- **TAXA_REENVIO**: Calculada como a razão entre QTDE_REENVIOS e QTDE_-

TRAMITACOES, com o objetivo de capturar a eficiência de tramitação do processo administrativo;

- **MES_ANO_ELEICAO_INTER**: Uma variável de interação entre o mês de emissão e a ocorrência de ano eleitoral, para testar a hipótese de que pode haver diferenças de comportamento entre despesas em anos eleitorais e em anos não eleitorais.

Pré-processamento

Foi construído um ***ColumnTransformer*** da biblioteca de aprendizado de máquina **Scikit-learn** (Pedregosa et al., 2011) da linguagem de programação **Python** (Rossum; Drake, 2009) para aplicar transformações distintas a diferentes tipos de colunas.

Variáveis numéricas foram padronizadas com ***StandardScaler*** (colocadas na mesma escala), enquanto variáveis categóricas foram convertidas em um formato numérico através do ***OneHotEncoder***. Este último passo transforma cada categoria em uma nova coluna binária, resultando em uma matriz de dados esparsa e de alta dimensionalidade.

Balanceamento de Classes (SMOTE)

Para lidar com o desbalanceamento da variável TARGET_RP, foi aplicada a técnica SMOTE (Synthetic Minority Over-sampling Technique). O SMOTE (Chawla et al., 2002) cria exemplos sintéticos da classe minoritária (NE inscritas em RP), balanceando o conjunto de dados. É importante ressaltar que esta técnica foi aplicada apenas no conjunto de treinamento, após a divisão dos dados, para evitar o vazamento de informações (*data leakage*) para o conjunto de teste, o que invalidaria a avaliação do modelo.

4.6.3 Modelagem e Treinamento (*Baseline*)

Nesta fase, foram treinados sete algoritmos de aprendizado de máquina com suas configurações padrão para estabelecer uma linha de base (*baseline*) de desempenho. Os modelos escolhidos representam diferentes abordagens de aprendizado e foram implementados por meio das bibliotecas *scikit-learn* (Pedregosa et al., 2011) e *XGBoost* (Chen; Guestrin, 2016):

- **K-Vizinhos Mais Próximos** (KNN - *K-Nearest Neighbors*);
- **Regressão Logística**
- **Árvore de Decisão**;
- **Random Forest**;

- **XGBoost** (*Extreme Gradient Boosting*);
- **Máquina de Vetores de Suporte** (SVC - *Support Vector Classifier*);
- **Perceptron Multicamadas** (MLP - *Multilayer Perceptron*);

4.6.4 Otimização de Hiperparâmetros

Para extrair o máximo potencial de cada algoritmo, foi conduzida uma busca exaustiva pelos melhores hiperparâmetros (as configurações internas de cada modelo) utilizando a ferramenta ***GridSearchCV*** da biblioteca *scikit-learn* (Pedregosa et al., 2011).

Para cada modelo, uma grade de possíveis combinações de hiperparâmetros foi definida, e a ferramenta treinou e avaliou cada combinação usando validação cruzada de 5 partições (*5-fold cross-validation*). A métrica utilizada para selecionar a melhor combinação foi o **F1-Score**. Na Tabela 15, apresentam-se os melhores hiperparâmeros encontrados pelo *GridSearchCV*.

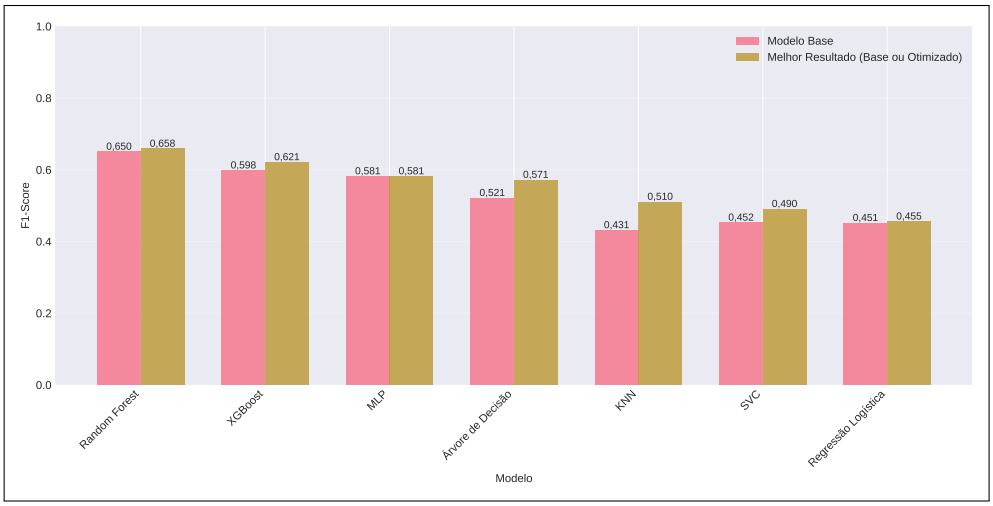
Tabela 15 – Melhores hiperparâmetros para cada modelo

Modelo	Melhores Hiperparâmetros
Regressão Logística	`C = 100`, `penalty = l2`, `solver = liblinear`
Árvore de Decisão	`criterion = entropy`, `max_depth = 20`, `min_samples_leaf = 1`, `min_samples_split = 2`
Random Forest	`max_depth = None`, `max_features = log2`, `min_samples_leaf = 1`, `min_samples_split = 2`, `n_estimators = 100`
XGBoost	`colsample_bytree = 0,7`, `learning_rate = 0,3`, `max_depth = 10`, `n_estimators = 100`, `subsample = 0,9`
MLP	`activation = relu`, `alpha = 0,0001`, `hidden_layer_sizes = (100, 100)`, `learning_rate = constant`, `learning_rate_init = 0,01`, `max_iter = 500`
SVC	`C = 10`, `degree = 2`, `gamma = auto`, `kernel = rbf`
KNN	`metric = manhattan`, `n_neighbors = 3`, `p = 1`, `weights = distance`

Fonte: Elaborado pelo autor (2026).

Na Figura 19, apresenta-se a comparação entre o desempenho dos modelos base e o dos otimizados. A visualização demonstra que a maioria dos modelos obteve ganhos marginais de desempenho com a otimização, consolidando o Random Forest como o algoritmo superior em ambos os cenários. A exceção notável é o MLP, que não apresentou melhora de desempenho, possivelmente devido à dificuldade de otimização em um espaço de características esparso.

Figura 19 – Comparação entre o desempenho de modelos base e otimizados



Fonte: Elaborado pelo autor (2026)

4.6.5 Análise de Componentes Principais (ACP)

Conduziu-se em seguida um experimento com Análise de Componentes Principais (ACP), uma técnica de redução de dimensionalidade, para avaliar se o desempenho dos modelos poderia ser mantido ou melhorado com um número menor de variáveis. O ACP transforma as variáveis originais em um novo conjunto de variáveis não correlacionadas (os “ componentes principais”).

Os resultados mostraram que, mesmo utilizando componentes que explicavam 99% da variância dos dados, houve uma queda de desempenho em comparação com o uso de todas as características originais, como mostra a Figura 20.

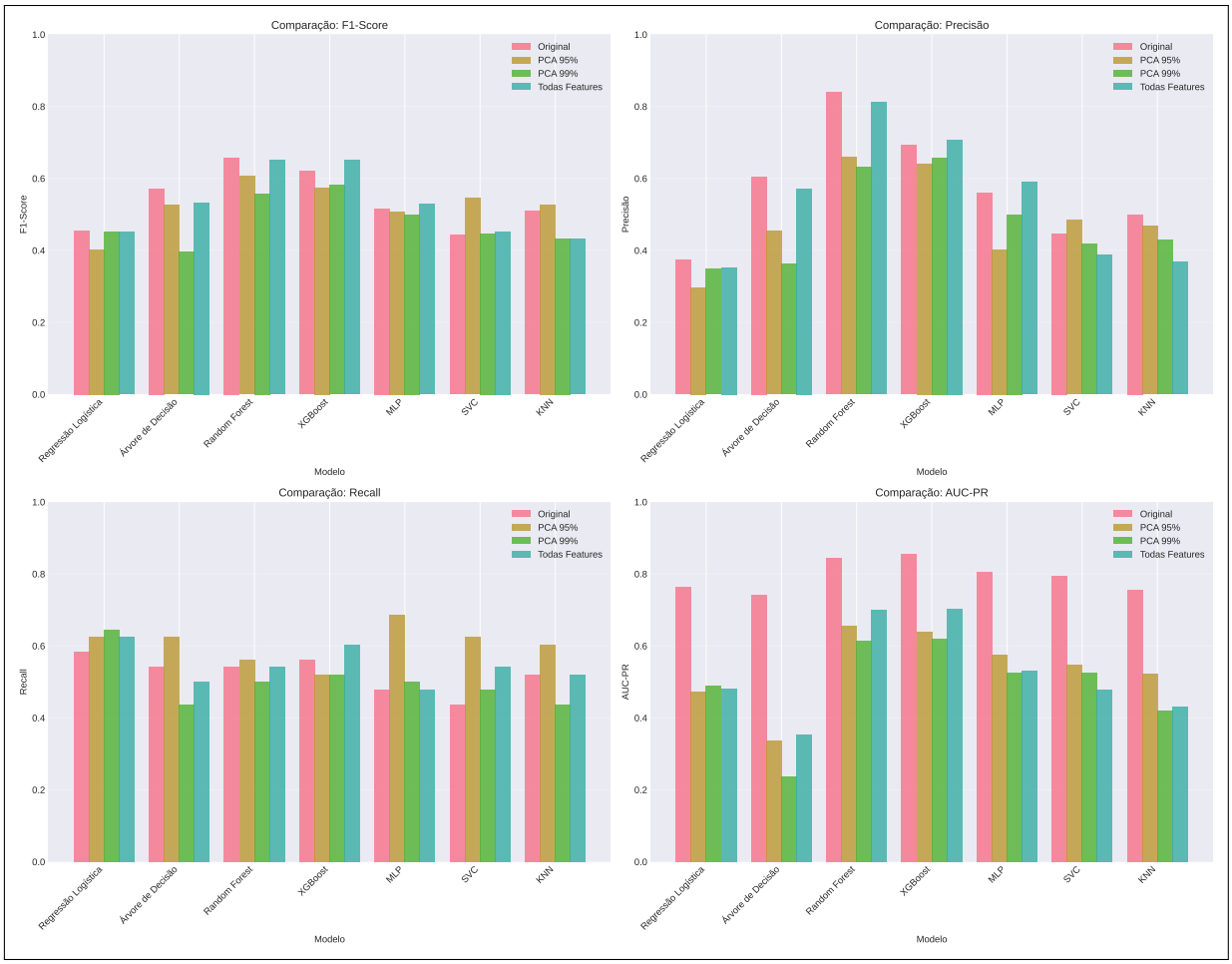
O painel com quatro subgráficos avalia o impacto da Análise de Componentes Principais (ACP) nas métricas de desempenho dos modelos. As barras evidenciam que a redução de dimensionalidade, mesmo retendo 99% da variância dos dados, resultou em degradação generalizada do F1-Score, especialmente para os modelos baseados em árvores (Random Forest e XGBoost). Isso comprova visualmente que a transformação linear do ACP dilui informações categóricas cruciais para a discriminação das classes.

O *Random Forest* treinado com os componentes ACP alcançou um F1-Score de 0,6067, inferior aos 0,65 do modelo original. Este resultado indicou que as variáveis originais, mesmo com sua alta dimensionalidade após o *one-hot encoding*, continham informações valiosas que eram parcialmente perdidas na transformação ACP.

4.6.6 Interpretabilidade com SHAP e Análise de Erros

A sexta e última fase do *pipeline* aplicou a metodologia SHAP (*SHapley Additive exPlanations*) ao modelo *Random Forest* otimizado, com o objetivo de identificar quais variáveis mais influenciam as predições e em que direção (Lundberg; Lee, 2017). A análise foi conduzida sobre o conjunto de teste, produzindo valores SHAP individuais para cada

Figura 20 – Comparação de desempenho: modelos originais vs. ACP



Fonte: Elaborado pelo autor (2026)

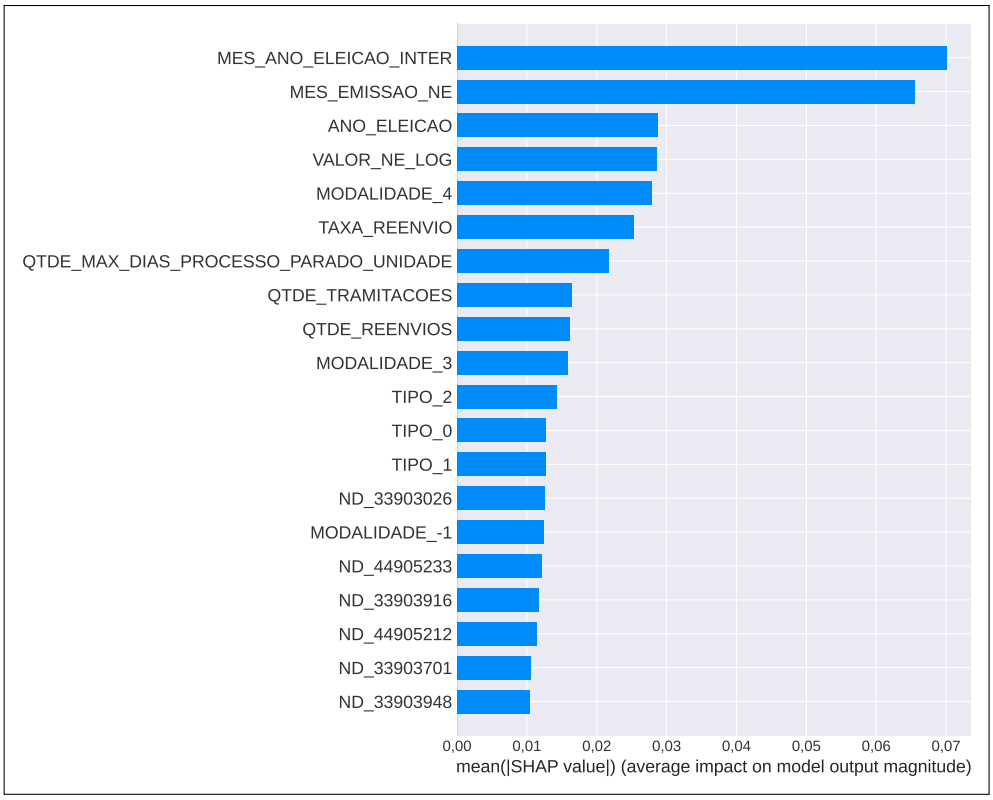
observação e cada variável.

O gráfico de importância global (*bar plot*) revelou que as duas variáveis de maior impacto médio absoluto são MES_ANO_ELEICAO_INTER (SHAP médio = 7,0%) e MES_EMISSAO_NE (SHAP médio = 6,6%), seguidas por ANO_ELEICAO, VALOR_-NE_LOG e MODALIDADE_4, com importâncias consideravelmente menores.

O eixo X do gráfico de barras SHAP representa a importância global de cada variável, medida pela média do valor absoluto de suas contribuições individuais (SHAP values) sobre todas as observações. Quanto maior a barra, maior é o impacto médio daquela variável nas predições do modelo, independentemente de a influência ser no sentido de aumentar ou reduzir a probabilidade de inscrição em Restos a Pagar. Apresenta-se na Figura 21 o gráfico de importância global obtido.

O gráfico de barras horizontais ranqueia as variáveis preditoras com base no impacto médio absoluto (valores SHAP) nas decisões do modelo. A visualização deixa claro o protagonismo absoluto das variáveis temporais, com a interação entre mês e ano eleitoral liderando o ranking, seguida pelo mês de emissão. O valor do empenho e métricas

Figura 21 – Gráfico de importância global (SHAP) - *Random Forest*



Fonte: Elaborado pelo autor (2026)

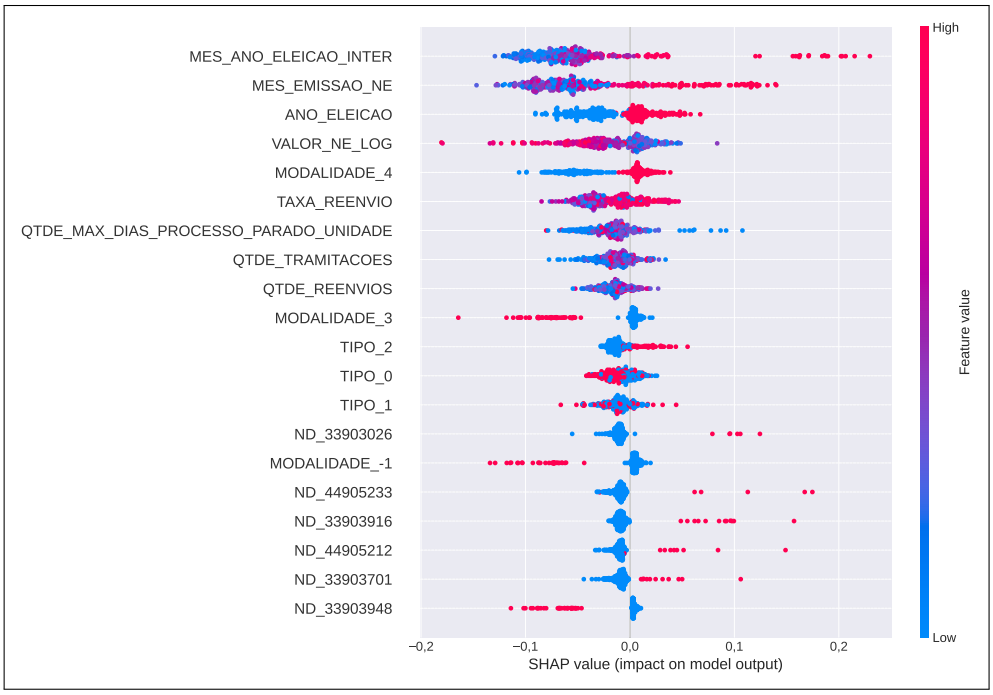
de ineficiência processual (taxa de reenvio) também se destacam, enquanto a maioria das naturezas de despesa específicas apresenta contribuição marginal.

A Figura 22 mostra o gráfico de enxame, ou *summary plot* para o modelo Random Forest otimizado. Ele auxilia no aprofundamento dessa análise ao revelar a direção do efeito de cada variável.

O summary plot (gráfico de enxame) oferece uma visão tridimensional da interpretabilidade do modelo, combinando a importância da variável com a direção do seu efeito. A dispersão de cores revela padrões causais claros: pontos rosas (valores altos) nas variáveis de mês empurram a predição fortemente para a direita (maior risco de RP), enquanto a modalidade 4, o Pregão (ponto rosa) empurra a predição para a esquerda (menor risco). Esta é a representação visual mais rica do comportamento do algoritmo.

Para MES_EMISSAO_NE, pontos rosas (valores altos, correspondentes aos meses finais do ano) concentram-se à direita do eixo, com valores SHAP positivos de até +0,14, enquanto pontos azuis (meses iniciais) acumulam-se à esquerda com SHAP negativo, confirmando o efeito de sazonalidade de fim de ano. O *dependence plot* desta variável reforça esse padrão: há uma transição clara de SHAP negativo (−0,15 a −0,05) nos meses iniciais para SHAP fortemente positivo (+0,05 a +0,14) nos meses finais, sugerindo que empenhos emitidos no último trimestre do ano apresentam risco substancialmente maior de inscrição em Restos a Pagar.

Figura 22 – Gráfico resumo para o modelo *Random Forest* otimizado



Fonte: Elaborado pelo autor (2026)

Em uma fase intermediária do desenvolvimento deste trabalho, observou-se, junto à equipe de especialistas em execução orçamentária, que havia um componente da Natureza de Despesa (ND) com bom potencial preditivo: o elemento de despesa, por meio da constatação de que o elemento de despesa 92 (despesas de exercícios anteriores) possui baixa probabilidade de inscrição em RP. Esta constatação levou à seguinte pergunta: o desmembramento da Natureza de Despesa pode melhorar a capacidade discriminatória do modelo preditivo? Para respondê-la, desenvolveu-se um experimento comparativo, descrito na Seção 4.7.

4.7 Experimento: Desmembramento da Natureza de Despesa

A ND é formada por quatro componentes (Brasil, 2026):

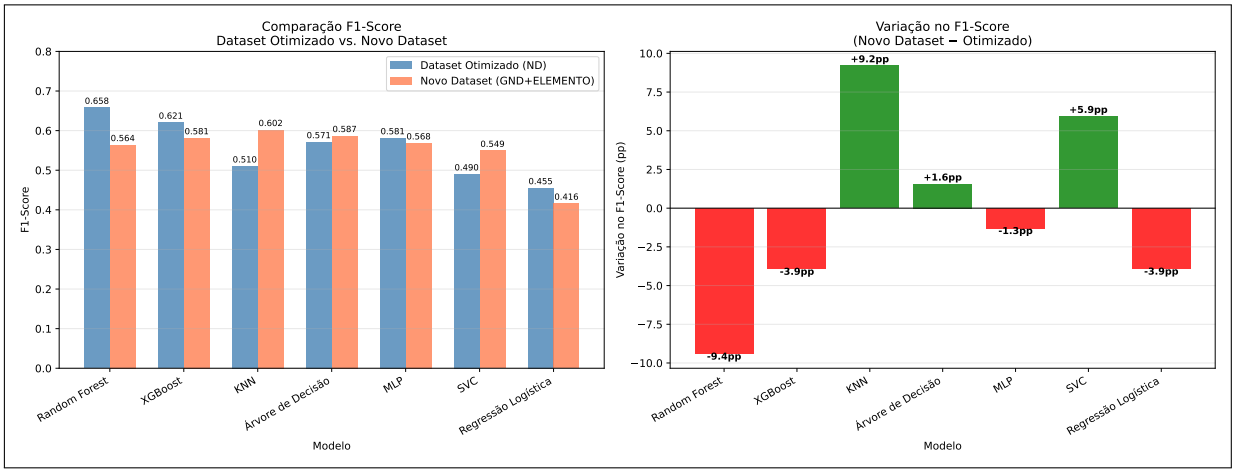
- Grupo de Natureza de Despesa (dois dígitos): identifica despesas de pessoal (11), correntes (33) ou investimentos (44);
- Modalidade de Aplicação dois dígitos: indica se a aplicação dos recursos será direta, isto é, dentro do próprio órgão (90) ou se será uma transferência de recursos para outros órgãos (91);
- Elemento de Despesa (dois dígitos) e Subelemento de Despesa (dois dígitos): Por exemplo, o código orçamentário 33.90.40.01 representa despesas correntes (33) de aplicação no próprio órgão (90) de locação de equipamentos de TIC (40) de ativos de rede (01).

Desta forma, conduziu-se um experimento em que a coluna ND (Natureza de Despesa), que possui alta cardinalidade (muitos valores únicos), foi desmembrada em duas colunas mais agregadas e com maior significado semântico: GND (Grupo de Natureza da Despesa) e ELEMENTO.

As colunas Modalidade de Aplicação e Subelemento de Despesas foram suprimidas neste experimento: a primeira pois observou-se que todas as NEs do conjunto de dados possuem apenas o código 90 para esse componente, e a segunda para se verificar o efeito da redução de valores únicos, considerando-se apenas o agregador Elemento de Despesa.

Todos os sete modelos foram retreinados e otimizados com este novo conjunto de dados. Os resultados apresentados na Figura 23 revelaram uma relação de compromisso:

Figura 23 – Comparação: modelo otimizado vs modelo com ND desmembrada



Fonte: Elaborado pelo Autor (2026)

- **Modelos baseados em árvores (XGBoost, Random Forest)**: Tiveram uma ligeira queda de desempenho (entre 1 e 4 pontos percentuais no F1-Score). Esses modelos são excelentes em encontrar interações complexas e criar regras específicas para categorias individuais. Ao desmembrar a ND em GND e ELEMENTO, parte dessa granularidade fina foi perdida, prejudicando sua capacidade de criar partições ótimas nos dados.

- **Modelos baseados em distância (KNN, SVC) e Redes Neurais (MLP)**: Apresentaram uma melhora de desempenho significativa (entre 6 e 13 pontos percentuais no F1-Score). Esses modelos possivelmente se beneficiaram com a redução de colunas na etapa de aplicação do algoritmo one-hot encoding: a ND original, com uma grande quantidade de valores únicos, gera 149 colunas. Esta redução no número de colunas pode tornar as distâncias entre os pontos de dados menos significativas. A representação mais compacta com GND e ELEMENTO criou um espaço de características mais denso e informativo para esses algoritmos.

Os gráficos ilustram o experimento de redução de cardinalidade da Natureza de Despesa. O gráfico de variação (barras verdes e vermelhas) revela um fenômeno teórico interessante: enquanto os modelos de árvore (Random Forest, XGBoost) sofreram leve perda de desempenho ao perderem a granularidade fina das categorias, os modelos baseados em distância (KNN, SVC) e o MLP apresentaram ganhos expressivos (barras verdes longas) ao se beneficiarem de um espaço de características menos esparso.

Este experimento demonstra que a escolha da representação das características não é neutra e interage diretamente com o tipo de algoritmo de aprendizado de máquina utilizado. Para a gestão pública, isso significa que a forma como os dados orçamentários são estruturados e disponibilizados pode ter um impacto direto nos resultados obtidos e na tomada de decisões gerenciais.

O Capítulo 5 aborda os resultados do presente trabalho e traz discussões relacionadas a eles.

5 RESULTADOS E DISCUSSÃO

Este capítulo apresenta os resultados detalhados do *pipeline* de modelagem, focando no desempenho do modelo que apresentou melhores resultados, o Random Forest, na interpretação de suas decisões e na análise aprofundada dos padrões de acertos e erros.

5.1 Desempenho dos Modelos e Seleção do Algoritmo Final

Na Tabela 16, apresentam-se os resultados iniciais dos sete modelos testados, com o objetivo de servir como linha de base para as fases posteriores de desenvolvimento do trabalho.

Tabela 16 – Desempenho dos modelos *baseline*

Modelo	F1-Score	Precisão	Recall
Random Forest	0,6500	0,8125	0,5417
XGBoost	0,5977	0,6667	0,8585
Árvore de Decisão	0,5208	0,5208	0,5208
MLP	0,5000	0,5227	0,4792
KNN	0,5102	0,5000	0,5208
Regressão Logística	0,4511	0,3529	0,6250
SVC	0,4421	0,4468	0,4375

Fonte: Elaborado pelo autor (2026)

O modelo Random Forest obteve o melhor desempenho inicial, com um F1-Score de 0,65; seguido de perto pelo XGBoost com 0,5977. Estes resultados iniciais indicam a possibilidade de predição maior do que em relação a um classificador ingênuo ou a um classificador aleatório. O ajuste dos hiperparâmetros[1] dos modelos pode auxiliar na melhoria das predições.

Após a fase de otimização de hiperparâmetros, foram consolidados os resultados de sete algoritmos distintos. Na Tabela 17, compara-se o desempenho final de todos os modelos no conjunto de teste, utilizando as métricas de Precisão, *Recall*, F1-Score, AUC-PR (*Area Under the Precision-Recall Curve*) e AP (*Average Precision*).

O AUC-PR é a área sob a curva que relaciona a precisão (eixo Y) ao recall (eixo X) ao longo de todos os limiares de decisão possíveis. O valor de referência do AUC-PR para um classificador aleatório é igual à proporção da classe positiva no conjunto de dados (neste estudo, aproximadamente 14%), tornando qualquer valor acima desse patamar um ganho real de discriminação.

[1] Hiperparâmetros são configurações externas ao modelo que não são aprendidas a partir dos dados, mas definidas antes do treinamento e que controlam como os algoritmos aprendem.

Tabela 17 – Comparação de resultados dos modelos preditivos

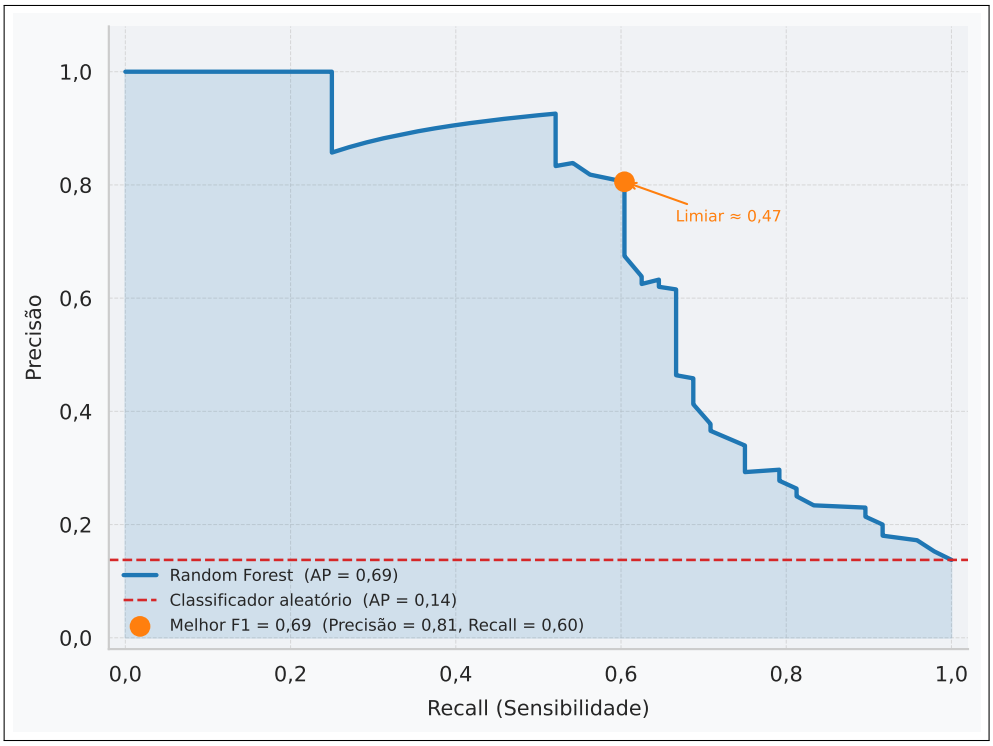
Modelo	F1-Score	Precisão	Recall	AUC-PR	AP
Random Forest	66%	84%	54%	69%	69%
Árvore de Decisão	57%	61%	54%	60%	39%
MLP	52%	56%	48%	57%	57%
KNN	51%	50%	52%	53%	43%
Regressão Logística	46%	37%	58%	44%	44%
SVC	44%	45%	44%	48%	49%
XGBoost	24%	14%	100%	24%	24%

Fonte: Elaborado pelo autor (2026).

O AP (*Average Precision*), como já explicado na fundamentação teórica, também calcula a área sob a curva por meio de uma soma ponderada dos incrementos de *recall* em cada limiar de decisão. Ambas as métricas são reportadas na Tabela 17 para fins de completude e comparabilidade com a literatura.

A Figura 24 apresenta a curva Precisão-*Recall* (PR) do modelo Random Forest otimizado, obtida pela variação contínua do limiar de classificação entre 0 e 1.

Figura 24 – Gráfico AUC-PR para o modelo *Random Forest* otimizado



Fonte: Elaborado pelo autor (2026)

A *Average Precision* (AP) alcançada foi de 0,69, valor que representa um ganho expressivo de 55 pontos percentuais em relação ao classificador aleatório de referência (AP = 0,14, indicado pela linha tracejada vermelha), cujo desempenho equivale à prevalência da classe positiva no conjunto de dados. Esse resultado confirma que o modelo possui

capacidade discriminativa substancial para identificar notas de empenho com risco de inscrição em Restos a Pagar, mesmo diante do acentuado desbalanceamento de classes (86% Não-RP e 14% RP).

A análise do formato da curva revela um comportamento característico de modelos com boa calibração para a classe minoritária. Para limiares elevados (região à esquerda do gráfico, *recall* $< 0{,}2$), o modelo opera com precisão próxima a 1,0, identificando apenas os empenhos com maior certeza de inscrição em RP, ainda que a cobertura seja reduzida. À medida que o limiar decresce, o *recall* cresce progressivamente, incorporando casos de maior ambiguidade, com consequente queda na precisão. Esse comportamento é esperado e reflete a relação de compromisso inerente a qualquer classificador binário: ampliar a cobertura implica aceitar mais alertas incorretos.

O ponto de operação escolhido, correspondente ao limiar $\approx 0{,}47$ (marcado em laranja no gráfico), representa o melhor equilíbrio entre precisão e *recall*, com F1-Score de 0,69, precisão de 0,81 e *recall* de 0,60. A precisão de 81% indica que, a cada 10 alertas emitidos pelo modelo, aproximadamente 8 correspondem a notas que de fato serão inscritas em Restos a Pagar, o que confere à ferramenta credibilidade operacional suficiente para apoiar a tomada de decisão do gestor orçamentário. O *recall* de 60%, por sua vez, significa que o modelo identifica cerca de 6 em cada 10 notas que efetivamente virarão RP, deixando 4 sem alerta — um resultado que, embora não ideal, representa ganho significativo frente à ausência de qualquer sistema preditivo. A calibração do limiar pode ser ajustada conforme a estratégia institucional: reduzi-lo amplia a cobertura ao custo de mais falsos positivos; elevá-lo aumenta a confiabilidade dos alertas ao custo de menor cobertura.

5.1.1 Análise Comparativa do Desempenho dos Algoritmos

O Random Forest foi confirmado como o modelo de melhor desempenho geral, com o F1-Score mais alto (66%) e a maior área sob a curva Precisão-Recall (AUC-PR = 69%), indicando uma capacidade superior de discriminar a classe positiva em um cenário desbalanceado. Sua precisão de 84% é particularmente expressiva: de cada 100 empenhos classificados pelo modelo como prováveis Restos a Pagar, 84 de fato se tornaram RP, um grau de confiabilidade considerável para uso como ferramenta de alerta. O *recall* de 54% indica que o modelo identificou corretamente 54% de todos os empenhos que efetivamente se tornaram RP.

A superioridade do Random Forest sobre os demais modelos encontra respaldo na literatura sobre aprendizado de máquina em dados tabulares. Conforme demonstrado por estudos empíricos em larga escala (Couronné et al., 2018), algoritmos baseados em árvores de decisão frequentemente superam modelos lineares em conjuntos de dados estruturados. O mecanismo subjacente a esse sucesso é a combinação de *bagging* (agregação de *bootstrap*) com a amostragem aleatória de características (*feature subsampling*).

Enquanto uma única Árvore de Decisão (que neste estudo obteve F1-Score de 57%) sofre com alta variância e tendência ao *overfitting*, o Random Forest constrói centenas de árvores independentes e agrega suas predições. Essa agregação reduz a variância do modelo sem aumentar seu viés (Breiman, 2001), permitindo que o algoritmo capture interações não-lineares complexas de forma eficaz.

Em contraste, a Regressão Logística apresentou um comportamento assimétrico: obteve um *recall* superior ao do Random Forest (58% contra 54%), mas ao custo de uma precisão substancialmente menor (37% contra 84%). Esse fenômeno pode ser explicado pela natureza linear de sua fronteira de decisão combinada com o efeito da técnica SMOTE. Ao gerar amostras sintéticas da classe minoritária, o SMOTE desloca o hiperplano de separação da Regressão Logística em direção à classe majoritária.

Como resultado, o modelo passa a classificar um volume muito maior de observações como positivas, o que eleva o *recall*. No entanto, por não conseguir modelar as interações não-lineares presentes nos dados orçamentários, grande parte dessas classificações positivas são incorretas (falsos positivos), o que derruba a precisão para níveis incompatíveis com a aplicação gerencial.

Um resultado que merece atenção especial é o colapso do XGBoost, que apresentou *recall* perfeito (1,000) ao custo de uma precisão extremamente baixa (14%) e o pior F1-Score do conjunto (24%). Na prática, isso significa que o modelo classificou corretamente todos os empenhos como Restos a Pagar, o que, embora garanta a detecção de todos os casos positivos, gera um volume inviável de falsos alarmes. Esse comportamento anômalo decorre da interação entre o algoritmo de *gradient boosting* e a técnica SMOTE.

O XGBoost constrói árvores sequencialmente, com cada nova árvore focada em corrigir os erros das anteriores. Quando o SMOTE introduz amostras sintéticas da classe minoritária (que muitas vezes se sobrepõem à região da classe majoritária no espaço de características), o algoritmo de *boosting* pode sofrer *overfitting* severo em relação a esses exemplos artificiais (Chawla et al., 2003).

A otimização de hiperparâmetros acabou convergindo para uma configuração que reduziu o limiar de decisão a ponto de classificar quase todo o conjunto de teste como positivo. Esse achado reforça a importância de avaliar o F1-Score e o AUC-PR em conjunto, em vez de considerar o *recall* isoladamente.

Os modelos baseados em distância (KNN e SVC) e a rede neural (MLP) apresentaram desempenhos intermediários, com F1-Scores variando entre 44% e 52%. A dificuldade desses algoritmos está diretamente relacionada à representação dos dados. A aplicação da técnica de *one-hot encoding* na variável Natureza da Despesa (ND) gerou um espaço de características de alta dimensionalidade (mais de 140 colunas) e extrema esparsidade.

Em espaços esparsos, a distância euclidiana utilizada pelo KNN perde sua capacidade discriminatória. Este fenômeno é conhecido como a "maldição da dimensionalidade" (Bellman, 1961). Segundo Beyer et al. (1999), a diferença relativa entre a distância ao

vizinho mais próximo e ao mais distante tende a zero à medida que a dimensionalidade cresce.

Já o hiperplano de margem máxima do SVC tem dificuldade em encontrar uma separação eficiente quando o número de características supera substancialmente o número de amostras (Vapnik, 1995; Hastie et al., 2009). O MLP, por sua vez, requer volumes de dados substancialmente maiores para aprender representações robustas em espaços de alta dimensão (Bishop, 2006), uma vez que o número de amostras necessário para cobrir adequadamente o espaço de características cresce exponencialmente com a dimensionalidade.

5.1.2 O Impacto da Redução de Dimensionalidade (ACP)

Conduziu-se um experimento com Análise de Componentes Principais (ACP) para avaliar se a performance dos modelos poderia ser mantida com um número menor de variáveis. Os resultados mostraram que, mesmo utilizando componentes que explicavam 99% da variância dos dados, houve uma queda de performance. O Random Forest treinado com os componentes ACP alcançou um F1-Score de 61%, inferior aos 65% do modelo original.

Esse declínio de desempenho possui uma explicação teórica. O ACP é uma técnica de transformação não-supervisionada que busca maximizar a variância explicada, o que não necessariamente corresponde às direções de maior poder discriminativo entre as classes. Mais criticamente, ao transformar variáveis categóricas esparsas (geradas pelo *one-hot encoding*) em componentes principais contínuos, o ACP dilui a informação semântica específica de cada categoria. O Random Forest se beneficia enormemente da capacidade de realizar divisões binárias precisas em categorias específicas (por exemplo, "se a Natureza da Despesa for Locação de Mão-de-Obra, aumente o risco"). Quando essa informação é misturada em combinações lineares contínuas pelo ACP, o algoritmo perde a capacidade de isolar esses sinais categóricos fortes, resultando em degradação preditiva.

5.2 Interpretação do modelo: Fatores para a formação de RP

Utilizando a metodologia SHAP no modelo Random Forest, foi possível identificar e quantificar a importância de cada variável para as predições, como mostrado na Figura 4. A análise do gráfico SHAP revela os principais fatores que o modelo aprendeu para prever a inscrição em RP:

1. **MES_ANO_ELEICAO_INTER e MES_EMISSAO_NE**: As variáveis temporais são, de longe, as mais influentes. Valores altos (meses próximos ao fim do ano, como novembro e dezembro) empurram fortemente a predição para a classe 1 (RP), confirmando o padrão sazonal identificado na AED. A interação com o ano eleitoral também se mostrou relevante.

2. **VALOR_NE_LOG**: O valor do empenho é a segunda variável mais importante. O gráfico mostra que valores mais altos (pontos vermelhos) tendem a ter um impacto SHAP positivo, aumentando a chance de inscrição em RP. Isso pode ocorrer porque despesas de maior vulto frequentemente estão associadas a contratos mais complexos e de execução mais longa.

3. **TAXA_REENVIO**: A taxa de reenvios do processo administrativo também possui alto poder preditivo. Valores altos (processos com muitos reenvios) estão associados a um maior impacto SHAP positivo, sugerindo que a ineficiência ou complexidade processual é um forte indicador de que a despesa não será paga dentro do exercício.

4. **MODALIDADE_4 (Pregão)**: A modalidade de licitação "Pregão" aparece como uma variável importante. O gráfico indica que ser um pregão (valor 1, em vermelho) tem um impacto SHAP negativo, ou seja, reduz a probabilidade de a despesa virar RP. Isso pode ser explicado pela natureza do pregão, que geralmente se aplica a bens e serviços comuns, de execução mais rápida e padronizada.

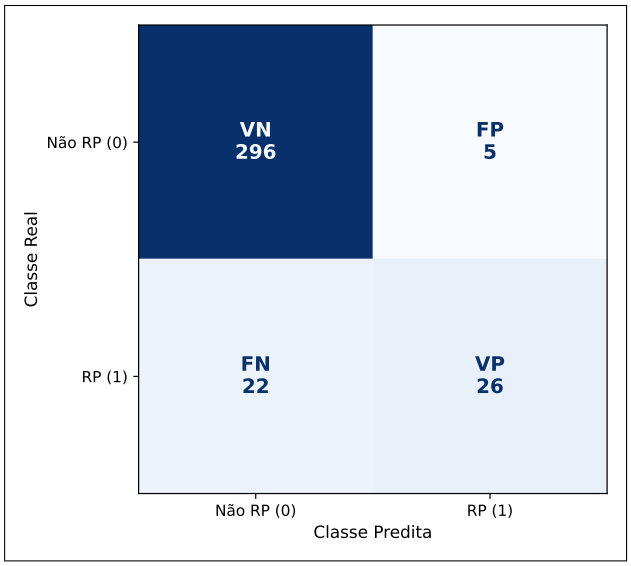
5. **Natureza da Despesa (ND)**: Diversas naturezas de despesa específicas, como ND_33903701 (Locação de Mão-de-Obra) e ND_33903923 (Serviços de Telecomunicações), apareceram no ranking de importância, indicando que o tipo de gasto, por si só, carrega informação preditiva relevante.

5.3 Análise de Domínio: Investigação dos erros do modelo

A análise dos 22 (vinte e dois) Falsos Negativos (FN) e 05 (cinco) Falsos Positivos (FP) gerados pelo modelo Random Forest (Figura 25) foi realizada em conjunto com os especialistas da área e revelou padrões que não foram capturados pelas variáveis disponíveis, indicando algumas limitações do modelo e a complexidade do processo real. Na Tabela 18, tem-se a relação das observações dos especialistas em orçamento do TRE-GO para os FN e FP gerados pelo modelo Random Forest.

A matriz de confusão final apresenta os resultados absolutos do Random Forest no conjunto de teste (349 observações). A visualização confirma a alta precisão do modelo, evidenciada pelo baixo número de Falsos Positivos (apenas 5 casos no quadrante superior direito). Contudo, o quadrante inferior esquerdo revela 22 Falsos Negativos, indicando que o modelo, embora altamente confiável quando emite um alerta, ainda deixa de identificar cerca de metade dos empenhos que efetivamente se tornarão Restos a Pagar. As causas de FN e FP são analisadas a seguir.

Figura 25 – Matriz de confusão do modelo *random forest* otimizado



Fonte: Elaborado pelo Autor (2026)

Causas de FN (o modelo errou ao prever NEs como *não RP*)

A análise dos 22 FNs revelou que a principal causa de erro não foi a sazonalidade atípica de início de ano, mas sim decisões discricionárias e características contraintuitivas no final do exercício. Metade dos casos (11 empenhos) ocorreu em novembro e dezembro (meses de alta probabilidade histórica de RP). Nesses casos, o modelo previu que a despesa seria liquidada no ano, possivelmente devido a características como baixo valor ou poucas tramitações, mas a inscrição em RP acabou ocorrendo.

Outro padrão relevante (8 casos) foram empenhos inscritos em RP por recomendação expressa da unidade responsável para garantir pagamentos ou ressarcimentos de exercícios anteriores, uma variável administrativa não capturada pelos dados estruturados do modelo. Houve também casos pontuais de ações de capacitação adiadas para o exercício seguinte.

Tabela 18 – Casos para investigação identificados pelo modelo Random Forest

Tipo de Erro	NE	Observação
Falso Negativo	20230519	Empenho emitido em dezembro — alta probabilidade de inscrição em Restos a Pagar.
	20230488	
	20230500	
	20230501	
	20230485	
	20230487	
	20230503	

(continua)

Tipo de Erro	NE	Observação
	20230520	
Falso Negativo	20240572	Empenho emitido em novembro — alta probabilidade de inscrição em Restos a Pagar.
	20220552	
	20220559	
	20230245	
Falso Negativo	20220039	Valor inscrito por recomendação da unidade responsável, para pagamentos de exercício anterior.
	20220131	
	20220514	
	20230045	
	20220021	
	20230067	
Falso Negativo	20230439	Ação de capacitação com data inicial adiada, executada no exercício seguinte.
	20240195	
Falso Negativo	20230271	Valor inscrito relativo aos serviços prestados em dezembro de 2023, liquidados no exercício seguinte.
Falso Negativo	20240418	Aparente demora na liquidação do empenho — material permanente.
Falso Positivo	20230493	Aparente rapidez na liquidação do empenho — material permanente.
	20240544	
Falso Positivo	20220534	Despesa de natureza contínua iniciada em 2022, com pagamentos bimestrais; o pagamento de 2022 foi realizado no próprio mês de emissão da nota de empenho.
Falso Positivo	20220353	Despesa de exercício anterior liquidada no exercício da emissão da nota.
Falso Positivo	20220191	Material de consumo liquidado no exercício.

Fonte: Elaborado pelo autor (2026).

Causas de FP (o modelo errou ao prever NEs como *vai ser inscrita em RP)*

Nos 5 FPs, o modelo superestimou o risco de inscrição em RP. O padrão predominante não se concentrou no final do ano, mas sim na natureza da despesa e na celeridade atípica da liquidação. Destacam-se casos de aparente rapidez na liquidação de materiais permanentes e de consumo, que contrariaram o tempo médio de tramitação aprendido pelo algoritmo.

Além disso, o modelo classificou como RP despesas de natureza contínua com pagamentos bimestrais e Despesas de Exercício Anterior (DEA) liquidadas no próprio exercício da emissão da nota. Por possuírem um rito de pagamento mais célere ou dinâmicas espe-

cíficas, essas despesas “enganaram” o modelo, que não possuía variáveis suficientes para inferir essa celeridade excepcional.

Implicações da análise dos erros do modelo

A análise de erros demonstra que o modelo preditivo é uma ferramenta que não substitui a análise do gestor, principalmente nos casos em que a predição do modelo contradiz a expectativa (por exemplo, um empenho de dezembro previsto como “Não RP”).

Adicionalmente, o experimento que comparou o uso da Natureza de Despesa (ND) com sua forma decomposta e reduzida (GND + ELEMENTO) resultou da degradação leve de desempenho dos dois modelos preditivos mais bem posicionados no ranking: o Random Forest e o XGBoost. Este achado desfavorável pode ser interpretado como um motivo para a continuação da investigação de métodos para a obtenção de melhores resultados.

Por fim, o resultado de *recall* (100%) do XGBoost otimizado, indicando baixíssima ou nula produção de falsos negativos, pode sugerir a adoção de estratégias híbridas, ou seja, a utilização de mais de um modelo preditivo a depender da estratégia de monitoração de NE que a organização desejar adotar.

5.4 Implicações para a gestão pública

A implementação de um modelo preditivo como o desenvolvido neste estudo oferece a possibilidade de melhoria da gestão orçamentária, permitindo o monitoramento proativo de notas de empenho que estão em vias de se tornarem Restos a Pagar. A confirmação das predições por especialistas em orçamento é indispensável, após o que a administração pode elaborar um plano de ação para tratar estas notas de empenho e, assim, minimizar os potenciais efeitos associados.

O principal valor da solução tecnológica reside em sua aplicação como um sistema de alerta e apoio à decisão para os ordenadores de despesa e equipes de planejamento. Ao invés de uma análise manual e subjetiva no final do exercício, o gestor pode receber informações sobre a classificação dos empenhos como “vai ser inscrito ou não em RP”. Isso permite focar a atenção e os recursos limitados nos casos mais críticos.

Um empenho de alto valor, emitido em novembro para um serviço complexo e que o modelo aponta como provável inscrição em RP, torna-se um ponto de atenção imediato. O gestor pode, então, contatar a unidade responsável, verificar o andamento do processo licitatório ou da execução contratual e tomar ações para acelerar o processo. Por outro lado, um empenho de baixo valor para material de consumo, também de novembro, mas que o modelo prediz a não inscrição em RP, pode ser considerado de baixo risco, liberando o gestor para focar em outros problemas.

5.5 Recomendações para mitigação de riscos

A partir dos conhecimentos obtidos com a análise SHAP e pela investigação dos erros, é possível derivar recomendações concretas para a administração:

- **Monitoramento Intensificado no Fim do Exercício**: A confirmação da criticidade dos últimos meses do ano sugere a implementação de um "comitê de crise" ou um rito processual acelerado para todas as despesas empenhadas a partir de outubro, visando garantir a liquidação e o pagamento em tempo hábil.
- **Revisão de Processos para Despesas de Alto Valor**: Dado que o valor do empenho é um forte preditor de RP, contratos de maior vulto deveriam ter um plano de execução e acompanhamento mais rigoroso desde o início, com marcos e prazos claros para evitar atrasos que levem à inscrição em RP.
- **Análise da Eficiência Processual**: A variável TAXA_REENVIO mostrou-se um relevante indicador de problemas. Processos com alta taxa de reenvio deveriam ser analisados para identificar gargalos, sejam eles burocráticos, de falha na instrução processual ou de complexidade excessiva, permitindo a otimização dos fluxos de trabalho.
- **Atenção às Modalidades de Licitação**: O fato de a modalidade Pregão reduzir a chance de RP sugere que, sempre que a legislação permitir, sua utilização deve ser incentivada para bens e serviços comuns, dada a sua celeridade.

6 CONSIDERAÇÕES FINAIS

Este capítulo apresenta as principais conclusões da pesquisa, sintetizando os achados obtidos e suas implicações para a teoria e prática da gestão orçamentária pública. São discutidas as contribuições científicas e gerenciais do trabalho, suas limitações, e as perspectivas para investigações futuras na intersecção entre inteligência artificial e administração pública.

6.1 Síntese dos Principais Achados

A presente investigação demonstrou a viabilidade e a eficácia da aplicação de técnicas de aprendizado de máquina para a predição de Restos a Pagar (RP) no contexto da execução orçamentária de um tribunal eleitoral. A avaliação comparativa de sete algoritmos distintos revelou que o modelo *Random Forest*, combinado com a técnica SMOTE para o tratamento do desbalanceamento de classes, alcançou o melhor desempenho global.

O modelo com o melhor desempenho obteve um F1-Score de 66% e uma Área sob a Curva Precisão-Recall (AUC-PR) de 69%, com uma precisão de 84% na identificação da classe minoritária. Este resultado valida a hipótese central da pesquisa, confirmando que é possível prever a inscrição de notas de empenho em RP com alta confiabilidade, viabilizando seu uso como ferramenta de alerta gerencial.

A interpretabilidade do modelo, conduzida por meio da metodologia SHAP (*SHapley Additive exPlanations*), revelou padrões consistentes com o conhecimento empírico sobre a dinâmica orçamentária pública. A análise identificou que as variáveis de maior impacto preditivo possuem natureza temporal e contextual. Especificamente, a interação entre o mês de emissão da nota de empenho e a ocorrência de ano eleitoral (`MES_ANO_ELEICAO_INTER`), bem como o mês de emissão isolado (`MES_EMISSAO_NE`), despontaram como os fatores mais determinantes.

O valor logarítmico do empenho (`VALOR_NE_LOG`) e a modalidade de licitação também demonstraram relevância considerável, evidenciando que despesas de maior vulto e processos licitatórios mais complexos tendem a apresentar maior risco de não liquidação no exercício corrente.

Adicionalmente, a pesquisa revelou importantes nuances metodológicas quanto à representação dos dados orçamentários. O experimento de decomposição da Natureza da Despesa (ND) em Grupo de Natureza da Despesa (GND) e Elemento de Despesa demonstrou que a granularidade da informação afeta os algoritmos de maneiras distintas.

Enquanto modelos baseados em árvores de decisão, como o *Random Forest* e o *XGBoost*, sofreram leve degradação de desempenho ao perderem a granularidade fina da ND completa, algoritmos baseados em distância (KNN, SVC) e redes neurais (MLP)

apresentaram ganhos significativos de performance (entre 6 e 13 pontos percentuais no F1-Score) devido à redução da dimensionalidade e da esparsidade gerada pelo *one-hot encoding*.

A análise de erros (falsos positivos e falsos negativos) evidenciou que, embora o modelo capture com precisão o risco contextual e temporal, ele possui limitações diante de exceções operacionais, como empenhos emitidos no final do ano que são liquidados em tempo recorde, ou despesas de exercícios anteriores (DEA) que possuem rito de pagamento acelerado.

A adoção do modelo em ambiente operacional requer atenção ao equilíbrio entre os tipos de erro de classificação. Um **falso positivo** ocorre quando o modelo sinaliza uma nota de empenho como provável Resto a Pagar, mas ela é liquidada dentro do exercício; um **falso negativo** ocorre quando o modelo não emite alerta para uma nota que, de fato, não será liquidada a tempo.

Do ponto de vista da adoção prática, alertas falsos em excesso tendem a desgastar a credibilidade da ferramenta junto ao gestor: se a maioria dos avisos não se concretiza, o usuário passa a ignorá-los.

Por outro lado, a tolerância excessiva a falsos negativos implica subestimar o volume de Restos a Pagar, comprometendo o planejamento orçamentário. O limiar de classificação adotado neste trabalho ($\approx 0{,}47$) foi calibrado para maximizar o F1-Score, que pondera igualmente precisão e *recall*.

A depender da estratégia institucional do TRE-GO, pode ser preferível deslocar esse limiar, reduzindo-o para ampliar a cobertura (maior *recall*, mais alertas) ou elevando-o para aumentar a confiabilidade dos alertas emitidos (maior precisão, menos falsos positivos).

Tais achados reforçam o fato de que a ferramenta preditiva atua como um sistema de apoio à decisão que direciona a atenção do gestor para os casos de maior risco, mas não substitui a análise qualitativa e o conhecimento tácito das equipes de planejamento e execução orçamentária do Tribunal Regional Eleitoral de Goiás.

6.2 Contribuições Teóricas

A pesquisa contribui para o avanço do conhecimento na intersecção entre gestão orçamentária e inteligência artificial aplicada ao setor público, área emergente com potencial significativo de impacto na modernização da administração pública. A abordagem metodológica desenvolvida acompanha a tendência de aplicação de técnicas preditivas para problemas específicos da gestão fiscal, o que vai além de análises descritivas tradicionais para oferecer capacidade prospectiva.

A integração de dados orçamentários (SIAFI) com informações processuais (SEI) constitui contribuição metodológica relevante, demonstrando compreensão mais ampla

dos fatores que influenciam a execução orçamentária e oferece elementos adicionais para a tomada de decisões gerenciais.

6.3 Contribuições Práticas

A análise dos fatores determinantes fundamenta estratégias específicas de mitigação baseadas em evidências empíricas. A forte correlação entre mês de emissão e inscrição em RP indica necessidade de distribuição mais equilibrada das emissões ao longo do exercício, sugerindo estabelecimento de metas mensais de empenho e monitoramento sistemático para evitar concentração no final do ano.

O impacto das variáveis de tramitação processual sugere necessidade de monitoramento ativo dos fluxos processuais, com implementação de alertas automáticos para processos parados por períodos superiores a limites estabelecidos e análise sistemática de causas de reenvios para redução do retrabalho administrativo.

6.4 Limitações da Pesquisa

É importante mencionar as limitações às quais o presente trabalho está submetido. Esta seção aborda as suas limitações contextuais e metodológicas, bem como tece considerações sobre causalidade que podem impactar na produção e análise dos resultados obtidos.

6.4.1 Limitações Contextuais

A presente análise baseia-se em dados de um único órgão público, o que pode limitar a generalização dos achados para outros contextos organizacionais. As particularidades do Tribunal Regional Eleitoral de Goiás, especialmente o ciclo eleitoral e as características específicas de sua execução orçamentária, podem não ser representativas de outros órgãos públicos com características distintas.

O período de análise, embora abranja quatro anos e inclua um ciclo eleitoral completo, é relativamente curto para capturar variações de longo prazo ou eventos excepcionais que possam influenciar a execução orçamentária. A expansão temporal do *dataset* poderá fortalecer a robustez dos achados e permitir análise de tendências temporais mais amplas.

6.4.2 Limitações Metodológicas

O desbalanceamento natural do *dataset*, mesmo após aplicação da técnica SMOTE, permanece como desafio metodológico. Embora as métricas utilizadas sejam apropriadas para este contexto, a baixa prevalência de Restos a Pagar pode limitar a capacidade do modelo de identificar padrões sutis na classe minoritária, especialmente para naturezas de despesa com poucas ocorrências.

6.4.3 Considerações sobre Causalidade

É importante ressaltar que os resultados apresentados indicam correlações estatísticas, não necessariamente relações causais. A interpretação dos coeficientes deve considerar possíveis variáveis não observadas e a complexidade inerente aos processos administrativos públicos, que podem envolver fatores não capturados pelos sistemas corporativos utilizados.

6.5 Sugestões para Trabalhos Futuros

A presente pesquisa abre um conjunto de frentes para investigação futura, tanto no aprimoramento técnico dos modelos quanto na ampliação do escopo institucional e na incorporação de novas fontes de dados. As sugestões a seguir decorrem, em parte, das observações formuladas pelos membros da banca de defesa e das limitações identificadas ao longo do desenvolvimento do trabalho.

Refinamento das variáveis de processo. A variável de tempo de paralisação foi construída com base no valor máximo de dias parados em uma única unidade administrativa ao longo da tramitação do processo. Embora essa escolha capture situações extremas de bloqueio, ela pode ser sensível a eventos pontuais e não representativos do padrão geral de lentidão. Trabalhos futuros poderiam explorar funções de agregação alternativas, como a média, a mediana ou o desvio padrão dos dias parados por unidade, comparando sistematicamente seu poder preditivo e avaliando qual delas melhor captura a dinâmica processual associada ao risco de inscrição em Restos a Pagar.

Aplicação do SMOTE antes do *One-Hot Encoding*. Uma limitação identificada no *pipeline* deste trabalho diz respeito à ordem e ao escopo de aplicação do SMOTE. Os dados utilizados na Fase de otimização foram carregados já com o balanceamento aplicado (arquivos `X_train_smote.pkl` e `y_train_smote.pkl`), o que implica que o SMOTE foi executado previamente ao *GridSearchCV* e, portanto, fora dos *folds* de validação cruzada.

Essa abordagem pode introduzir vazamento de dados (*data leakage*), uma vez que amostras sintéticas geradas a partir de todo o conjunto de treino podem influenciar indiretamente a avaliação nos *folds* de validação (Kapoor; Narayanan, 2023).

Adicionalmente, como o SMOTE foi aplicado após a codificação *one-hot* da variável Natureza da Despesa, o que gerou mais de 140 colunas binárias, o algoritmo tratou essas colunas como valores numéricos contínuos, podendo gerar amostras sintéticas com valores fracionários que violam a natureza categórica original da variável (Chawla et al., 2002).

Trabalhos futuros poderiam corrigir ambas as limitações simultaneamente por meio da construção de um *pipeline* integrado com a classe `ImbPipeline` da biblioteca *imbalanced-learn*, inserindo a etapa de reamostramento com SMOTE-NC — variante projetada para dados mistos numéricos e categóricos — dentro dos *folds* de treino da valida-

ção cruzada, garantindo que nenhuma informação do conjunto de validação contamine o processo de geração de amostras sintéticas.

Incorporação de variáveis não estruturadas via LLMs. Os dados utilizados neste trabalho são inteiramente estruturados (campos numéricos e categóricos extraídos do SIAFI e do SEI). Contudo, os processos administrativos registrados no SEI contêm informações textuais ricas(despachos, termos de referência, editais de licitação) que podem carregar sinais preditivos não capturados pelas variáveis estruturadas. Trabalhos futuros poderiam explorar o uso de modelos de linguagem de grande escala (*Large Language Models* — LLMs) para extração de *features* a partir desses documentos, integrando-as ao *pipeline* de classificação.

Inclusão do usuário no ciclo de aprendizado (*user-in-the-loop*). O modelo preditivo desenvolvido opera de forma estática: é treinado com dados históricos e aplicado a novos empenhos sem mecanismo de atualização contínua. Uma extensão natural seria incorporar o *feedback* dos gestores orçamentários ao longo do tempo, registrando se os alertas gerados pelo modelo foram confirmados ou refutados pela execução real, o que permitiria o aprendizado incremental e a calibração do modelo com base na experiência acumulada da instituição.

Validação em outros órgãos e transferência de aprendizado. Os modelos foram treinados e avaliados exclusivamente com dados do TRE-GO, o que limita a generalização dos resultados. Pesquisas futuras poderiam investigar a replicabilidade da metodologia em outros tribunais regionais eleitorais ou em órgãos de esferas distintas de governo, avaliando em que medida os padrões identificados são específicos ao contexto do TRE-GO ou refletem dinâmicas mais amplas da execução orçamentária federal. Técnicas de *transfer learning* poderiam ser exploradas para adaptar o modelo a instituições com menor volume histórico de dados.

Exploração de algoritmos adicionais. O conjunto de algoritmos avaliados neste trabalho abrangeu sete abordagens representativas das principais famílias de classificação. Estudos futuros poderiam incluir o LightGBM — alternativa ao XGBoost com menor custo computacional e desempenho competitivo em dados tabulares — bem como modelos de aprendizado por conjunto (*stacking* e *blending*) que combinem as predições de múltiplos classificadores. A exploração de estratégias de aprendizado sensível ao custo (*cost-sensitive learning*), que penalizam assimetricamente os erros de falso positivo e falso negativo conforme as preferências do gestor, também constitui uma frente promissora.

Integração com iniciativas de melhoria de processos. Os fatores identificados pela análise SHAP — em particular a taxa de reenvio e o tempo máximo de paralisação em unidades administrativas — revelam possíveis gargalos de tramitação processual que transcendem o escopo preditivo e podem orientar iniciativas de redesenho e simplificação de processos (*Business Process Management* — BPM) no TRE-GO. Trabalhos futuros poderiam investigar a relação causal entre essas variáveis e a inscrição em RP, subsidiando

intervenções organizacionais voltadas à redução dos gargalos identificados.

6.6 Conclusões

A presente pesquisa conclui pela viabilidade da aplicação de técnicas de aprendizado de máquina para a predição de Restos a Pagar no contexto de um tribunal eleitoral. A avaliação comparativa de sete algoritmos distintos (Regressão Logística, Árvore de Decisão, Random Forest, XGBoost, KNN, SVC e MLP) demonstrou que o Random Forest, combinado com a técnica SMOTE para tratamento do desbalanceamento de classes, alcançou o melhor desempenho global, com F1-Score de 66%, AUC-PR de 69% e precisão de 84% na identificação de empenhos inscritos em Restos a Pagar. Esse resultado responde afirmativamente à questão de pesquisa e valida a hipótese central do trabalho.

A integração de dados orçamentários do SIAFI com dados de tramitação processual do SEI representou a adoção de uma abordagem metodológica que tem potencial de ampliar a capacidade preditiva dos modelos. Variáveis derivadas do SEI, como a quantidade de tramitações, o número de reenvios e o tempo máximo de paralisação em uma unidade, mostraram-se preditores complementares às informações financeiras, evidenciando que a qualidade da execução processual impacta diretamente a execução orçamentária.

A investigação dos erros do modelo, realizada em conjunto com especialistas em execução orçamentária do TRE-GO, revelou que os falsos negativos e falsos positivos concentram-se em situações que dependem de contexto administrativo não capturado pelos dados estruturados, como inscrições por recomendação expressa da unidade responsável ou celeridade excepcional na liquidação de determinadas despesas. Esse achado reforça que o modelo preditivo deve atuar como instrumento de apoio à decisão, e não como substituto do julgamento qualificado do gestor público.

O experimento de decomposição da Natureza da Despesa demonstrou que a forma como os dados orçamentários são estruturados e disponibilizados não é neutra: ela interage diretamente com a arquitetura do algoritmo utilizado, favorecendo modelos baseados em árvores quando a granularidade é preservada e beneficiando modelos baseados em distância quando a dimensionalidade é reduzida. Essa constatação possui implicação prática para a governança de dados do TRE-GO, sinalizando que o *design* dos sistemas de informação orçamentária deve considerar os requisitos analíticos dos instrumentos de apoio à decisão que deles se alimentam.

6.6.1 Atingimento dos objetivos

Para se fazer uma análise do atingimento dos objetivos específicos e do objetivo geral do trabalho, eles serão aqui transcritos para evitar a remissão ao início do trabalho. Os objetivos específicos são:

1. Verificar se dados orçamentários e processuais disponíveis no TRE-GO permitem antecipar a inscrição de notas de empenho em Restos a Pagar;

2. Identificar o algoritmo de aprendizado de máquina que oferece o melhor equilíbrio entre precisão e sensibilidade na predição de Restos a Pagar;

3. Identificar os fatores que mais influenciam a probabilidade de inscrição de uma nota de empenho em Restos a Pagar no TRE-GO;

4. Avaliar em que medida as predições do modelo são úteis como ferramenta de apoio à tomada de decisão orçamentária.

A integração de dados do SIAFI (valor, natureza da despesa, modalidade de licitação, mês e ano de emissão) com dados de tramitação processual do SEI (quantidade de tramitações, taxa de reenvio e tempo máximo de paralisação em uma unidade) produziu um conjunto de variáveis com poder preditivo mensurável e estatisticamente superior ao de um classificador aleatório (cuja AUC-PR de referência seria de 14%, equivalente à prevalência da classe positiva). O ganho de 55 pontos percentuais sobre o acaso aponta para a suficiência dos dados disponíveis nos sistemas corporativos do TRE-GO para a antecipação do fenômeno, o que leva ao atendimento do primeiro e segundo objetivos específicos do trabalho.

O terceiro objetivo específico do trabalho decorre da análise de interpretabilidade conduzida via metodologia SHAP, que revelou que os fatores determinantes possuem natureza predominantemente temporal e contextual. A interação entre o mês de emissão da nota de empenho e a ocorrência de ano eleitoral emergiu como o preditor de maior impacto global, seguida pelo mês de emissão isolado, pelo valor logarítmico do empenho, pela taxa de reenvio do processo administrativo e pelo tempo máximo de paralisação em uma unidade. A modalidade Pregão apresentou efeito protetor: empenhos realizados por essa modalidade têm menor probabilidade de inscrição em RP, o que é consistente com a natureza de bens e serviços comuns que caracteriza essa forma de contratação e com a maior celeridade processual que ela propicia.

O quarto e último objetivo específico nos leva à reflexão da utilidade da aplicação do modelo como ferramenta de apoio à tomada de decisão orçamentária. Do ponto de vista quantitativo, a precisão de 84% significa que, a cada cem empenhos sinalizados como prováveis Restos a Pagar, oitenta e quatro efetivamente se confirmam, conferindo ao instrumento confiabilidade suficiente para orientar ações preventivas sem gerar volume excessivo de alertas falsos.

Do ponto de vista qualitativo, a investigação dos erros do modelo, realizada em conjunto com especialistas em execução orçamentária do TRE-GO, revelou que os falsos negativos e falsos positivos concentram-se em situações que dependem de contexto administrativo não capturado pelos dados estruturados, como inscrições por recomendação

expressa da unidade responsável ou celeridade excepcional na liquidação de determinadas despesas. Esse achado reforça que o modelo deve atuar como instrumento de apoio à decisão, e não como substituto do julgamento qualificado do gestor público.

O objetivo geral, "desenvolver e avaliar comparativamente modelos preditivos de aprendizado de máquina para a detecção antecipada de notas de empenho a serem inscritas em Restos a Pagar no TRE-GO", foi, portanto, atingido em decorrência do atingimento articulado dos quatro objetivos específicos, que percorreram desde a verificação da viabilidade preditiva dos dados disponíveis até a análise de interpretabilidade dos fatores determinantes do fenômeno.

Do ponto de vista da eficiência do gasto público, a identificação antecipada das notas de empenho com maior risco de inscrição em RP permite ao gestor público agir preventivamente e, com isso, ampliar a proporção do orçamento efetivamente executada no próprio exercício financeiro. Nesse sentido, o modelo preditivo desenvolvido, em conjunto com a análise de gestores e especialistas em orçamento público, representa uma ferramenta de apoio à eficiência alocativa do gasto público, alinhada ao princípio constitucional da eficiência administrativa e às diretrizes de modernização da gestão pública.

6.6.2 Sugestões para Trabalhos Futuros

A presente pesquisa abre um conjunto de frentes para investigação futura, tanto no aprimoramento técnico dos modelos quanto na ampliação do escopo institucional e na incorporação de novas fontes de dados. As sugestões a seguir decorrem, em parte, das observações formuladas pelos membros da banca de defesa e das limitações identificadas ao longo do desenvolvimento do trabalho.

Refinamento das variáveis de processo. A variável de tempo de paralisação foi construída com base no valor máximo de dias parados em uma única unidade administrativa ao longo da tramitação do processo. Embora essa escolha capture situações extremas de bloqueio, ela pode ser sensível a eventos pontuais e não representativos do padrão geral de lentidão. Trabalhos futuros poderiam explorar funções de agregação alternativas, como a média, a mediana ou o desvio padrão dos dias parados por unidade, comparando sistematicamente seu poder preditivo e avaliando qual delas melhor captura a dinâmica processual associada ao risco de inscrição em Restos a Pagar.

Tratamento de dados mistos com SMOTE-NC. O balanceamento de classes foi realizado com o SMOTE original, aplicado após a codificação *one-hot* das variáveis categóricas. Essa abordagem, embora amplamente adotada, ignora a natureza discreta das variáveis categóricas ao interpolar pontos sintéticos em um espaço contínuo. O SMOTE-NC (**??**), variante projetada para conjuntos de dados com variáveis mistas (numéricas e categóricas), constitui uma alternativa mais adequada ao contexto deste trabalho e merece investigação em estudos futuros, com comparação direta dos resultados obtidos por ambas as abordagens.

Utilização de outra técnica de normalização. Uma investigação adicional de interesse diz respeito à estratégia de normalização das variáveis numéricas contínuas. No presente estudo, adotou-se o *StandardScaler*, que transforma cada variável para média zero e desvio padrão unitário, pressupondo uma distribuição aproximadamente simétrica dos dados. Contudo, variáveis como o valor da nota de empenho apresentam distribuição fortemente assimétrica à direita, com presença de *outliers* de grande magnitude, contexto em que o *MinMaxScaler*, que comprime os valores para o intervalo $[0, 1]$ preservando a forma da distribuição original, pode oferecer representações mais adequadas, especialmente para os algoritmos sensíveis à escala, como KNN, SVC e MLP (Hastie et al., 2009). Recomenda-se, portanto, que estudos futuros avaliem o impacto da substituição do *StandardScaler* pelo *MinMaxScaler*.

Exploração da arquitetura do MLP. No que diz respeito à rede neural *Multilayer Perceptron* (MLP), o presente estudo adotou uma configuração de hiperparâmetros definida por *GridSearchCV*, porém restrita a um espaço de busca limitado pelas dimensões do conjunto de dados disponível. Estudos futuros poderiam explorar de forma mais sistemática a arquitetura do MLP, variando o número de camadas ocultas, a quantidade de neurônios por camada, as funções de ativação (ReLU, *tanh*, sigmoide) e os algoritmos de otimização (*Adam*, *SGD*, *L-BFGS*), bem como técnicas de regularização como *dropout* e *early stopping*, que podem mitigar o *overfitting* em conjuntos de dados de pequeno porte (Goodfellow et al., 2016).

Incorporação de variáveis não estruturadas via LLMs. Os dados utilizados neste trabalho são inteiramente estruturados (campos numéricos e categóricos extraídos do SIAFI e do SEI). Contudo, os processos administrativos registrados no SEI contêm informações textuais ricas(despachos, termos de referência, editais de licitação) que podem carregar sinais preditivos não capturados pelas variáveis estruturadas. Trabalhos futuros poderiam explorar o uso de modelos de linguagem de grande escala (*Large Language Models* — LLMs) para extração de *features* a partir desses documentos, integrando-as ao *pipeline* de classificação.

Inclusão do usuário no ciclo de aprendizado (*user-in-the-loop*). O modelo preditivo desenvolvido opera de forma estática: é treinado com dados históricos e aplicado a novos empenhos sem mecanismo de atualização contínua. Uma extensão natural seria incorporar o *feedback* dos gestores orçamentários ao longo do tempo — registrando se os alertas gerados pelo modelo foram confirmados ou refutados pela execução real —, permitindo o aprendizado incremental e a calibração do modelo com base na experiência acumulada da instituição.

Validação em outros órgãos e transferência de aprendizado. Os modelos foram treinados e avaliados exclusivamente com dados do TRE-GO, o que limita a generalização dos resultados. Pesquisas futuras poderiam investigar a replicabilidade da metodologia em outros tribunais regionais eleitorais ou em órgãos de esferas distintas de

governo, avaliando em que medida os padrões identificados são específicos ao contexto do TRE-GO ou refletem dinâmicas mais amplas da execução orçamentária federal. Técnicas de *transfer learning* poderiam ser exploradas para adaptar o modelo a instituições com menor volume histórico de dados.

Exploração de algoritmos adicionais. O conjunto de algoritmos avaliados neste trabalho abrangeu sete abordagens representativas das principais famílias de classificação. Estudos futuros poderiam incluir o LightGBM — alternativa ao XGBoost com menor custo computacional e desempenho competitivo em dados tabulares — bem como modelos de aprendizado por conjunto (*stacking* e *blending*) que combinem as predições de múltiplos classificadores. A exploração de estratégias de aprendizado sensível ao custo (*cost-sensitive learning*), que penalizam assimetricamente os erros de falso positivo e falso negativo conforme as preferências do gestor, também constitui uma frente promissora.

Integração com iniciativas de melhoria de processos. Os fatores identificados pela análise SHAP, em particular a taxa de reenvio e o tempo máximo de paralisação em unidades administrativas, revelam possíveis gargalos processuais que transcendem o escopo preditivo e podem orientar iniciativas de redesenho e simplificação de processos (*Business Process Management* — BPM) no TRE-GO. Trabalhos futuros poderiam investigar a relação causal entre essas variáveis e a inscrição em RP, subsidiando intervenções organizacionais voltadas à redução dos gargalos identificados.

6.7 Produtos e publicações

Esta seção detalha as produções acadêmicas e técnicas desenvolvidas no âmbito da pesquisa de mestrado, abrangendo capítulos de livros, artigos científicos e produtos tecnológicos. Essas produções refletem o engajamento com a pesquisa e o desenvolvimento de soluções aplicadas, consolidando o conhecimento adquirido e contribuindo para a área de Governança e Transformação Digital.

6.7.1 Capítulos de Livro, Artigos e Outros

Durante o período de desenvolvimento da pesquisa, foram elaboradas as seguintes produções no formato de capítulos de livro ou outras publicações relevantes:

- **Capítulo de livro:** “Efetividade na Execução Orçamentária do Setor Público: Uma Revisão Sistemática de Literatura”. Este capítulo foi desenvolvido no contexto da disciplina de Metodologia da Pesquisa Científica, abordando a eficácia na gestão orçamentária no setor público através de uma revisão sistemática da literatura (Em processo de submissão).

- **Artigo científico:** “Aplicação de Técnicas de Inteligência Computacional no Contexto da Saúde dos Servidores do TRE-GO”. Elaborado na disciplina de Técnicas de

Inteligência Computacional, este artigo explora a aplicação de métodos computacionais avançados para otimizar a saúde dos servidores do Tribunal Regional Eleitoral de Goiás. Este trabalho foi orientado pelo prof. Dr. Marcelo Lisboa e contou com os coautores Adenir José de Sousa e Roberto César Rodrigues.

- **Artigo científico:** "Uma Solução de Rolagem Inteligente de Texto e Cifra Musical Baseado em Reconhecimento de Fala". Desenvolvido na disciplina de Internet das Coisas, este trabalho apresenta uma solução inovadora para rolagem automática de texto e cifras musicais utilizando reconhecimento de fala, demonstrando a integração de tecnologias emergentes.

- **Resenha crítica:** "Transformação Digital - Uma Abordagem Metodológica Aplicada ao Tribunal Regional Eleitoral de Goiás". Esta resenha crítica analisa um capítulo de livro focado na transformação digital, com uma perspectiva aplicada ao contexto do Tribunal Regional Eleitoral de Goiás, destacando a relevância das abordagens metodológicas para a implementação de inovações digitais.

- **Artigo Científico submetido para congresso e aprovado:** "O paradoxo das plataformas de detecção e atribuição de iA: Análise comparativa de trabalhos acadêmicos entre 1980 e 2025". Estudo realizado sob orientação e coautoria com o Prof. Dr. Flávio Roldão de Carvalho Lelis, contando ainda com os seguintes coautores: Dr. Rafael Lima de Carvalho e Dr. Rodrigo de Melo Brustolin, e os colegas de PPGGTD Ilana Murici Ayres, José Carlos Lucio Maia, Márcio Antônio Duarte Oliveira e Vitor Carneiro Ramos. Submetido ao XXVIII Encontro Nacional de Modelagem Computacional, que será realizado em 21/10/2025, em Montes Claros/MG.

- **Submissão de resumo para congresso:** "Linguagem simples na convocação de mesários: relato de experiência de projeto de *design thinking* na Justiça Eleitoral". Trabalho orientado pelo professor Doutor Rafael Lima de Carvalho e submetido ao V Encontro Internacional de Gestão da Comunicação, evento a distancia a ser realizado de 22 a 24/10/2025. Coautores: José Carlos Lucio Maia, Katiusse Kelle de Melo Soares, Giovana Beatriz de Souza, Lucio Maia e professor Doutor Rafael Lima de Carvalho.

- **Artigo Científico submetido para congresso e aprovado:** "Análise comparativa do processo de transformação digital do TRE-GO sob a perspectiva da doutrina europeia". Estudo realizado sob orientação e coautoria com o Prof. Dr. George Lauro Ribeiro de Brito , contando ainda com os seguintes colegas de PPGGTD como coautores: Adenir José de Souza, Ilana Murici Ayres, José Carlos Lucio Maia, Katiusse Kelle de Melo Soares, Luiz Henrique Borges de Azevedo Silva e Vitor Carneiro

Ramos. Submetido ao XXVIII Encontro Nacional de Modelagem Computacional, que será realizado em 21/10/2025, em Montes Claros/MG.

- **Submissão de resumo expandido para congresso e publicação nos anais do evento:** "Modelagem preditiva para otimização da execução orçamentária: um estudo de caso sobre restos a pagar". Trabalho orientado pelo professor Doutor Rafael Lima de Carvalho e submetido ao Congresso de Gestão da Informação na Esfera Pública (Infosfera 2025,realizado em Curitiba de 20 a 21/10/2025).

- **Submissão de resumo expandido para congresso e publicação nos anais do evento:**"Aplicação de Técnicas de Inteligência Computacional no Contexto da Saúde dos Servidores do TRE-GO". Trabalho orientado pelo prof. Dr. Marcelo Lisboa e contou com os coautores Adenir José de Sousa, José Carlos Lucio Maia e Roberto César Rodrigues. Submetido e apresentado no V ENGEC - Encontro Internacional de Gestão e Comunicação, evento online realizado de 22 a 24/10/2026.

- **Submissão de artigo para revista:** " Predição de Restos a Pagar para Otimização da Execução Orçamentária". Trabalho orientado pelos professores Doutores Rafael Lima de Carvalho e Flávio Roldão de Carvalho Lelis, submetido à revista Caderno de Gestão Pública e Cidadania (CGPC), da Fundação Getúlio Vargas. O trabalho sumariza os resultados obtidos com o desenvolvimento de modelos preditivos a partir de dados orçamentários e dados de tramitação de processos administrativos do TRE-GO relacionados a notas de empenho, com o objetivo de predizer a inscrição destas em Restos a Pagar

6.7.2 Artigos Publicados em Congressos, Revista Internacional e Nacional

No que tange à disseminação do conhecimento em eventos científicos e periódicos, destaca-se a seguinte publicação:

- **Artigo Científico**: "Avaliação da implantação do Pacto Nacional pela Linguagem Simples no Tribunal Regional Eleitoral de Goiás". Este artigo, elaborado na disciplina de Gerenciamento de Políticas Públicas, teve seu resumo submetido e aceito para a III Conferência Internacional de Políticas Públicas e Ciência de Dados, com apresentação realizada em maio de 2025. A pesquisa avalia a implementação de uma iniciativa de simplificação da linguagem em documentos oficiais, visando aprimorar a comunicação institucional.

6.7.3 Produtos Técnicos e Tecnológicos

Além das produções acadêmicas, a pesquisa resultou no desenvolvimento de produtos técnicos e tecnológicos que demonstram a aplicação prática do conhecimento:

- **Cartilha: "Suas Demandas, Nosso Orçamento - Aprenda a acompanhar a execução orçamentária dos pedidos da sua unidade"**. Esta cartilha foi desenvolvida como um produto educacional, visando capacitar os servidores do Tribunal Regional Eleitoral de Goiás a compreender e acompanhar a execução orçamentária de suas unidades, promovendo a transparência e a gestão participativa.

- ***Software* Hermes**: Este software implementa um sistema de rolagem inteligente de texto e cifra musical baseado em reconhecimento de fala. O Hermes representa uma inovação tecnológica que facilita a interação com partituras e textos, utilizando recursos de inteligência artificial para otimizar a experiência do usuário.

REFERÊNCIAS

ABRAHAM, M. **AFO e orçamento público**. 1. ed. Rio de Janeiro: Forense, 2025. ISBN 9788530995553.

ALMEIDA, F. F.; SAKURAI, S. N.; ALMEIDA, R. B. Incentivos eleitorais e regras fiscais (não tão) rígidas: novas evidências para os municípios brasileiros a partir da rubrica restos a pagar. **Pesquisa e Planejamento Econômico**, IPEA, Brasília, v. 53, n. 1, p. 37–74, 2023.

ANDRADE, R. R. **Restos a pagar na administração pública**. Tese (Doutorado) — Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa, Brasília, DF, 2022.

AQUINO, A. C. B. d.; AZEVEDO, R. R. d. Restos a pagar e a perda da credibilidade orçamentária. **Revista de Administração Pública**, Fundação Getulio Vargas, Rio de Janeiro, v. 51, n. 4, p. 580–595, 2017.

ARAÚJO, J. G. R. d.; LINS, T. S. M.; DINIZ, J. A. "Use it or lose it" e restos a pagar: análise dos incentivos perversos na execução orçamentária. **Advances in Scientific and Applied Accounting**, v. 15, n. 1, p. 109–124, 2022.

ARAÚJO, R. J. R. **Incentivos eleitorais e o gerenciamento de resultados orçamentários por meio de restos a pagar: um estudo em municípios brasileiros**. Tese (Doutorado) — Universidade Federal da Paraíba, João Pessoa, PB, 2022.

ARAÚJO, R. J. R.; QUEIROZ, D. B.; PAULO, E. Incentivos eleitorais e gerenciamento de resultados orçamentários. **Revista de Administração Pública**, FGV EBAPE, Rio de Janeiro, v. 57, n. 6, p. 1–23, 2023.

BARBOSA, M. E. F.; RODRIGUES, E. C. C. Idiossincrasias associadas aos cancelamentos de despesas inscritas em restos a pagar. **Revista de Gestão e Secretariado**, São Paulo, SP, v. 14, n. 1, p. 1118–1137, 2023.

BELLMAN, R. E. **Adaptive Control Processes: A Guided Tour**. [S.l.]: Princeton University Press, 1961.

BEYER, K. et al. When is "nearest neighbor" meaningful? In: **Proceedings of the 7th International Conference on Database Theory (ICDT)**. [S.l.: s.n.], 1999. p. 217–235.

BISHOP, C. M. **Pattern Recognition and Machine Learning**. [S.l.]: Springer, 2006.

BRASIL. **Decreto n. 93.872, de 23 de dezembro de 1986**. 1986. Referência específica ao Art. 45, que dispõe sobre a concessão, aplicação e comprovação de suprimento de fundos. Disponível em: <http://www.planalto.gov.br/ccivil_03/decreto/d93872.htm>.

BRASIL. **Manual Técnico do Orçamento - MTO**. 2026. Disponível em: <https://www1.siop.planejamento.gov.br/mto/doku.php/mto2026>.

BREIMAN, L. Random forests. **Machine learning**, Springer, v. 45, n. 1, p. 5–32, 2001.

BREIMAN, L. et al. **Classification and regression trees**. Belmont, CA: Wadsworth International Group, 1984.

CHAWLA, N. V. et al. SMOTE: Synthetic Minority Over-sampling Technique. **Journal of Artificial Intelligence Research**, v. 16, p. 321–357, jun. 2002. ISSN 1076-9757. Disponível em: <https://www.jair.org/index.php/jair/article/view/10302>.

CHAWLA, N. V. et al. SMOTEBoost: Improving prediction of the minority class in boosting. In: **Proceedings of the 7th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD)**. [S.l.: s.n.], 2003. p. 107–119.

CHEN, T.; GUESTRIN, C. XGBoost: A Scalable Tree Boosting System. In: **Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: Association for Computing Machinery, 2016. (KDD '16), p. 785–794. ISBN 978-1-4503-4232-2. Disponível em: <https://dl.acm.org/doi/10.1145/2939672.2939785>.

CORTES, C.; VAPNIK, V. Support-vector networks. **Machine learning**, Springer, v. 20, n. 3, p. 273–297, 1995.

COURONNé, R.; PROBST, P.; BOULESTEIX, A.-L. Random forest versus logistic regression: a large-scale benchmark experiment. **BMC Bioinformatics**, v. 19, n. 270, 2018.

COVER, T.; HART, P. Nearest neighbor pattern classification. **IEEE transactions on information theory**, IEEE, v. 13, n. 1, p. 21–27, 1967.

COX, D. R. The regression analysis of binary sequences. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 20, n. 2, p. 215–232, 1958.

DAVIS, J.; GOADRICH, M. The relationship between Precision-Recall and ROC curves. In: **Proceedings of the 23rd international conference on Machine learning - ICML '06**. Pittsburgh, Pennsylvania: ACM Press, 2006. p. 233–240. ISBN 978-1-59593-383-6. Disponível em: <http://portal.acm.org/citation.cfm?doid=1143844.1143874>.

FACELI, K. et al. **Inteligência artificial: uma abordagem de aprendizado de máquina**. 2. ed. Rio de Janeiro: LTC, 2021. ISBN 9788521637349.

FERREIRA, M. C.; SOUZA, F. S. R. N. Cancelamento de restos a pagar: determinantes e implicações para a gestão orçamentária. **Acanto em Revista**, [Editora], p. 73–82, 2022.

FIX, E.; HODGES, J. L. Discriminatory analysis. Nonparametric discrimination: Consistency properties. **USAF School of Aviation Medicine, Randolph Field, Texas**, 1951.

GIACOMONI, J. **Orçamento público**. 16. ed. Barueri, SP: Atlas, 2012. ISBN 9788522469673.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016.

HAN, H.; WANG, W.-Y.; MAO, B.-H. Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. In: **Advances in Intelligent Computing (ICIC 2005)**. [S.l.]: Springer, 2005. p. 878–887.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **Elements of Statistical Learning: data mining, inference, and prediction. 2nd Edition.** 2009. Disponível em: <https://hastie.su.domains/ElemStatLearn/>.

HE, H.; GARCIA, E. A. Learning from Imbalanced Data. **IEEE Transactions on Knowledge and Data Engineering**, v. 21, n. 9, p. 1263–1284, set. 2009. ISSN 1558-2191. Disponível em: <https://ieeexplore.ieee.org/document/5128907>.

KAPOOR, S.; NARAYANAN, A. Leakage and the reproducibility crisis in machine-learning-based science. **Patterns**, Elsevier, v. 4, n. 9, 2023.

LUNDBERG, S. M.; LEE, S.-I. A Unified Approach to Interpreting Model Predictions. In: **Advances in Neural Information Processing Systems**. Curran Associates, Inc., 2017. v. 30. Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>.

MAURÍCIO, F. H. **Restos a pagar não processados no Exército Brasileiro: uma análise dos determinantes dos cancelamentos**. Dissertação (Mestrado) — Escola Nacional de Administração Pública, Brasília, DF, 2025.

MELO, M. A.; PEREIRA, C.; FIGUEIREDO, C. M. Political and institutional checks on corruption: explaining the performance of brazilian audit institutions. **Comparative Political Studies**, Sage Publications Sage CA: Los Angeles, CA, v. 42, n. 9, p. 1217–1244, 2009.

MOTA, S. C. **Eficiência relativa da gestão de restos a pagar nas universidades federais brasileiras**. Dissertação (Mestrado) — Universidade Federal do Ceará, Fortaleza, CE, 2021.

NONAKA, T. H. **Restos a pagar não processados: análise dos determinantes e impactos na gestão orçamentária**. Dissertação (Mestrado) — Universidade de Brasília, Brasília, DF, 2019.

PEDREGOSA, F. et al. Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, v. 12, n. 85, p. 2825–2830, 2011. ISSN 1533-7928. Disponível em: <http://jmlr.org/papers/v12/pedregosa11a.html>.

PINHEIRO, M. R. S. e.; HOLLANDA, J. E. S. Evolução dos restos a pagar no município de ladário-ms: uma análise temporal. **[Nome do Periódico]**, [Editora], [Volume], n. [Número], p. [Páginas], 2020.

PREUSSLER, A. P. S. **Um estudo empírico dos ciclos político-econômicos do Brasil**. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 2001.

QUEIROZ, A. G. A. **Avaliação de desempenho de restos a pagar não processados no instituto federal de Rondônia**. Dissertação (Mestrado) — Instituto Superior de Contabilidade e Administração do Porto, Porto Velho, RO, 2020.

RASCHKA, S.; MIRJALILI, V. **Python machine learning: machine learning and deep learning with python, scikit-learn, and tensorflow 2**. Birmingham, UK: Packt Publishing Ltd, 2019.

ROSSUM, G. V.; DRAKE, F. L. . **Python 3 Reference Manual**. CreateSpace, 2009. Disponível em: <https://dl.acm.org/doi/book/10.5555/1593511>.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **nature**, Nature Publishing Group, v. 323, n. 6088, p. 533–536, 1986.

SANTOS, L. F. D. Aplicação de machine learning na classificação de restos a pagar. 2023.

SANTOS, R. I. d.; COSTA, P. H. C. d. H.; SILVA, V. d. Gestão de restos a pagar: estudo de caso dos resultados alcançados pelo estado de alagoas no período de 2004 a 2020. **Refas - Revista Fatec Zona Sul**, Faculdade de Tecnologia da Zona Sul, São Paulo, v. 8, n. 2, p. 38–57, 2021.

SILVA, M. M. d.; JUNIOR, A. E. X. Trabalho de Conclusão de Curso (Graduação em Ciências Contábeis), **A Influência do período eleitoral na inscrição de restos a pagar nas capitais brasileiras**. Mossoró, RN: [s.n.], 2024.

Tesouro Nacional. **Relatório de gestão fiscal resumido da União (RGF)**. 2025. Acesso em: 13 ago. 2025. Disponível em: <https://www.tesourotransparente.gov.br/>.

VALLE-CRUZ, D.; GIL-GARCIA, J. R.; FERNANDEZ-CORTEZ, V. Towards Smarter Public Budgeting? Understanding the Potential of Artificial Intelligence Techniques to Support Decision Making in Government. In: **The 21st Annual International Conference on Digital Government Research**. New York, NY, USA: Association for Computing Machinery, 2020. (dg.o '20), p. 232–242. ISBN 978-1-4503-8791-0. Disponível em: <https://doi.org/10.1145/3396956.3396995>.

VAPNIK, V. N. **The Nature of Statistical Learning Theory**. [S.l.]: Springer, 1995.

ZARZÀ, I. de et al. Optimized financial planning: integrating individual and cooperative budgeting models with LLM recommendations. **AI**, v. 5, p. 91–114, dez. 2023.