

Modelagem preditiva para otimização da execução orçamentária: um estudo de caso sobre restos a pagar

Luís Cláudio Fernandes, Flávio Roldão de Carvalho Lelis, Rafael Lima de Carvalho

Universidade Federal do Tocantins, Instituto Federal de Goiás

Palavras-chave: execução orçamentária, restos a pagar, modelagem preditiva, machine learning, estudo de caso

Introdução

A execução orçamentária no setor público brasileiro é regida, entre outros, pelo princípio da anualidade, que estabelece que as autorizações de despesa possuem vigência limitada ao exercício financeiro. Contudo, despesas legalmente empenhadas, mas não pagas até o encerramento do exercício, são transpostas para o ano subsequente sob a rubrica de Restos a Pagar (RP), conforme disciplina a Lei nº 4.320/1964.

Giacomoni (2012, p. 336) descreve a prática comum na gestão pública brasileira de se concentrar liberações financeiras no final do ano de exercício orçamentário na tentativa de se aproveitar os créditos ainda existentes, o que aumenta a quantidade de RP. Abraham (2025, p. 192) aponta que o acúmulo excessivo de RP acaba por tornar ineficaz o princípio da anualidade, que preconiza a execução de despesas dentro de um mesmo ano de exercício. Esta plurianualidade orçamentária atípica representa um risco à credibilidade e à transparência do orçamento em todos os níveis de governo (Aquino & Azevedo, 2017).

Dados recentes do Tesouro Transparente (Tesouro Nacional, 2025) revelam que, nos últimos cinco anos, observa-se uma manutenção de altos volumes de RP, especialmente no que tange às despesas discricionárias (investimentos e outras despesas correntes), em contraste com as despesas obrigatórias, como pessoal e encargos sociais, que possuem um fluxo de execução mais estável e previsível. Na Figura 1 tem-se a evolução do percentual de RP em relação ao volume total empenhado, separado em despesas de pessoal e discricionárias, nos três Poderes da União, de 2020 a 2024.

Neste contexto, o presente trabalho investiga a aplicação de técnicas de Machine Learning para mitigar este problema. O objetivo é desenvolver e avaliar um modelo preditivo capaz de calcular a probabilidade de uma nota de empenho (NE) ser inscrita em RP ao final do exercício. Tal ferramenta visa prover os gestores públicos com um instrumento de apoio à decisão, permitindo ações proativas para otimizar a execução orçamentária.

Metodologia

Este estudo caracteriza-se como uma pesquisa quantitativa aplicada, desenvolvida por meio de um estudo de caso no Tribunal Regional Eleitoral de Goiás. A metodologia foi estruturada em quatro etapas principais: (1) coleta e tratamento de dados; (2) definição das variáveis e estruturação do dataset; (3) treinamento e seleção do modelo preditivo; e (4) análise e interpretação dos resultados.

Coleta e Estruturação dos Dados

A base de dados para o estudo foi construída a partir da extração de informações de dois sistemas corporativos do órgão: o Sistema Integrado de Administração Financeira (SIAFI), via Tesouro Gerencial, para dados orçamentários e financeiros do período de 2022 a 2024, e o Sistema Eletrônico de Informações (SEI), para dados relativos à tramitação dos processos administrativos de despesa. Na Figura 2, apresenta-se o processo de coleta e tratamento dos dados utilizado no presente trabalho.

A unidade de análise é a nota de empenho (NE). A variável alvo (TARGET_RP) foi criada como uma classificação binária, recebendo o valor "1" para NEs que foram inscritas em RP ao final de seu respectivo exercício e "0" para aquelas que foram pagas ou canceladas dentro do exercício.

Variáveis do Estudo

O dataset foi estruturado com as seguintes variáveis preditoras, que representam as características da despesa e de seu processamento:

- ID_NE: Identificador único da nota de empenho (utilizado como chave).
- VALOR: Montante financeiro da nota de empenho.
- TIPO_DA_DESPESA: Categoria do empenho (ordinário, global, estimativo).
- MODALIDADE_LICITACAO: Modalidade do processo de contratação (e.g., Pregão, Inexigibilidade).
- ANO_ELEITORAL: Variável binária (1 para ano par, 0 para ímpar), para capturar o efeito de sobrecarga de trabalho em anos eleitorais.
- QTDE_TRAMITACOES: Número de movimentações do processo administrativo associado.

- QTDE_MAX_PROCESSO_PARADO: Tempo máximo, em dias, de permanência do processo em uma única unidade.
- QTDE_REENVIOS: Frequência com que um processo retorna a uma unidade pela qual já tramitou.
- MES_EMISSAO_NE: Mês de emissão da nota de empenho, variável proxy para a proximidade do encerramento do exercício.
- NATUREZA_DESPESA: Código da classificação da despesa, que detalha o objeto do gasto.

Modelagem e Avaliação

Para a tarefa de classificação binária, optou-se pela Regressão Logística como modelo inicial, devido à sua alta interpretabilidade. Algoritmos de ensemble, como Random Forest e LightGBM, também foram considerados para testes futuros, visando ganhos de performance (Hastie et al., 2009, p. 203). A avaliação do desempenho do modelo não se restringiu à acurácia, mas empregou métricas como Precisão, Recall, F1-Score e a curva ROC (AUC), mais adequadas para contextos de classificação potencialmente desbalanceada.

Resultados parciais e discussão

A implementação do modelo de Regressão Logística sobre o dataset permitiu extrair insights valiosos sobre os fatores determinantes para a inscrição de uma despesa em Restos a Pagar.

Análise dos Coeficientes do Modelo

O modelo final apresentou um F1-Score de 0.81 para a classe negativa (não inscrição em RP) e 0.44 para a classe positiva (inscrição em RP), provavelmente devido ao desbalanceamento entre essas duas classes (maior quantidade de não inscrição em RP). O AUC atingiu o valor de 0.82. Este resultado significa que, ao selecionar aleatoriamente uma NE que de fato foi inscrita em Restos a Pagar e outra que não foi, o modelo tem 82% de chance de atribuir uma probabilidade de risco maior à nota que efetivamente se tornou um Resto a Pagar. Apresenta-se, na Figura 3, a matriz de confusão e o gráfico da curva ROC obtida no presente estudo.

A análise dos coeficientes gerados, que representam o impacto de cada variável na chance de uma NE virar RP, revelou os seguintes fatores como os mais influentes:

- modalidade_licitacao_5 (Coeficiente: -2.39): Correspondente a "Suprimento de Fundos", esta variável teve o maior impacto negativo, indicando que despesas desta modalidade, por seu ciclo de execução curto, raramente se tornam RP.
- natureza_despesa_459061 (Coeficiente: +2.26): Representando "Aquisição de Imóveis", esta variável teve o maior impacto positivo. A complexidade inerente a esses processos torna a inscrição em RP altamente provável.
- mes_emissao_ne (Coeficiente: +1.08): Confirmando a percepção empírica, quanto mais tarde no ano a despesa é empenhada, maior a sua chance de virar RP devido à menor janela de tempo para pagamento.
- qtde_tramitacoes (Coeficiente: -1.13): De forma contraintuitiva, um maior número de tramitações diminuiu a chance de RP, sugerindo que processos que se movimentam estão, de fato, progredindo no fluxo de trabalho.
- qtde_reenvios (Coeficiente: +1.01): Ao contrário das tramitações, os reenvios (retrabalho) mostraram-se como um forte indicador de risco, pois sinalizam problemas e atrasos no processo.

Implicações para a Gestão Orçamentária

A análise dos resultados transcende a simples predição e oferece um diagnóstico gerencial. As conclusões indicam que a inscrição em RP é influenciada por três dimensões principais:

1. A natureza do objeto do gasto: Despesas de alta complexidade exigem planejamento e execução mais diligentes, alinhando-se à visão de que a má qualidade do planejamento é causa fundamental do acúmulo de RP.

2. A dinâmica do processo administrativo: A "saúde" do fluxo processual (medida por tramitações, paradas e reenvios) é um preditor poderoso, deslocando o foco da gestão de uma análise puramente orçamentária para uma análise de processos, corroborando a necessidade de modernização da gestão pública para além dos controles tradicionais (Giacomoni, 2012, p. 185).

3. O fator temporal: O modelo quantifica a criticidade do tempo, permitindo ao gestor priorizar com mais eficácia as despesas empenhadas no final do exercício. Esses resultados validam a premissa de que é possível, por meio de um modelo de Machine Learning, identificar com bom grau de precisão as despesas com alto risco de inscrição em RP e, mais importante, fornecer uma base de evidências para a ação gerencial proativa.

Considerações finais e trabalhos futuros

Este trabalho demonstra o potencial da aplicação de técnicas de Machine Learning para a solução de problemas complexos da administração pública. O modelo desenvolvido oferece uma ferramenta prática para aprimorar a capacidade gerencial existente, permitindo a identificação e mitigação de riscos na execução orçamentária.

O AUC de 0.82 encontrado, juntamente com a análise dos coeficientes gerados, indicam que o artefato desenvolvido é uma ferramenta eficaz para a triagem e o ranqueamento de despesas, permitindo que os gestores foquem seus esforços de controle e aceleração de liquidação nas notas de empenho com maior propensão a impactar o orçamento do exercício seguinte, otimizando assim a execução orçamentária e financeira.

Como próximos passos, elencamos os seguintes:

- pretende-se ampliar o dataset para abranger também o ano de 2021, para se ter dois ciclos completos de alternância entre anos eleitorais e não eleitorais (estes dados encontram-se em um sistema de tramitação processual em modo consulta);
- outras técnicas de predição, como Random Forest e LightGBM, devem também ser aplicadas, com o potencial de aumentar ainda mais o AUC. Estas técnicas também poderão permitir a confirmação do impacto das variáveis apontadas pela técnica Logistic Regression, ou até a descoberta da relevância de outras variáveis;
- Pretende-se, também utilizar técnicas como o SMOTE para tratar especificamente do desbalanceamento existente entre NE em RP e não RP.

A validação final do modelo será realizada em conjunto com os gestores da área orçamentária, avaliando sua utilidade e usabilidade em um ambiente simulado. A disseminação dos resultados busca contribuir para a comunidade de pesquisa e para os profissionais da área de orçamento e finanças.

Referências

Abraham, M. (2025). AFO e orçamento público (1a. ed.). Forense.

Afonso, J. R., & Mello, G. M. (2019). A crise fiscal dos estados: Causas, consequências e soluções. FGV IBRE.

Aquino, A. C. B. D., & Azevedo, R. R. D. (2017). Restos a pagar e a perda da credibilidade orçamentária. Revista de Administração Pública, 51(4), 580-595.

Giacomoni, J. (2012). Orçamento público (16a. ed.). Atlas.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.

Tesouro Nacional (2025). Relatório de avaliação dos restos a pagar.
<https://tesourotransparente.gov.br/temas/execucao-orcamentaria-e-financeira/restos-a-pagar/>

Apêndice - Figuras

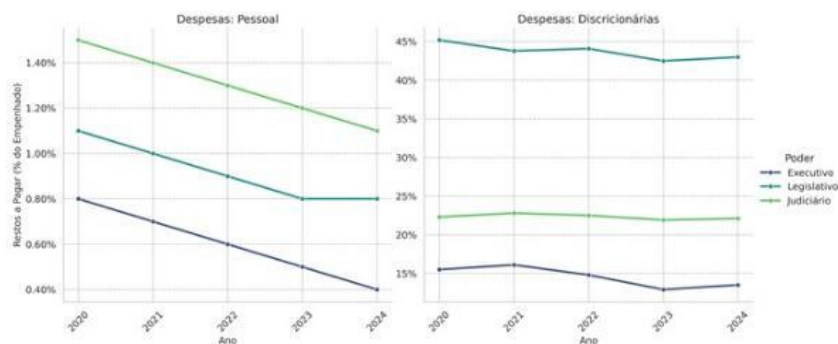


Figura 1. Percentual de RP sobre empenhado por Tipo de Despesa e Poder (2020-2024)

Fonte: Tesouro Nacional (2025)

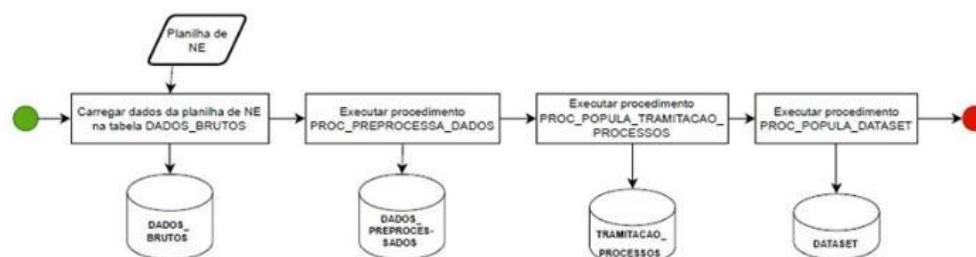


Figura 2. Processo de coleta e tratamento de dados.

Fonte: Elaborado pelos autores.

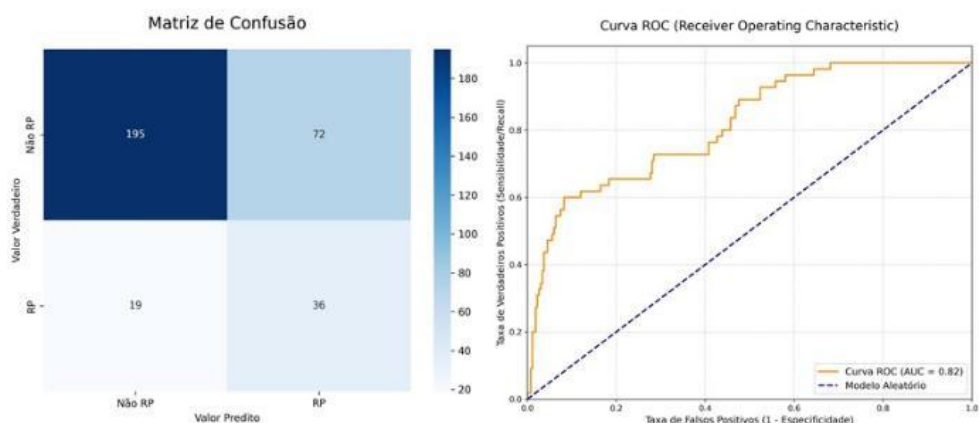


Figura 3. Matriz de Confusão e Curva ROC do modelo preditivo de NE inscritas em RP.

Fonte: Elaborado pelos autores.