



Universidade Federal da Paraíba
Centro de Ciências Sociais Aplicadas
Graduação em Ciência de Dados para Negócios

Modelo Preditivo para Despesas Discricionárias sob Restrição Fiscal no Ensino Superior: Evidências a partir da UFPB

Tainá Alcantara Alves Diniz

João Pessoa - PB
2026

Tainá Alcantara Alves Diniz

**Modelo Preditivo para Despesas Discricionárias
sob Restrição Fiscal no Ensino Superior:
Evidências a partir da UFPB**

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Ciência de Dados para Negócios do Centro de Ciências Sociais Aplicadas da Universidade Federal da Paraíba (UFPB), como requisito para obtenção do grau de Bacharel em Ciência de Dados para Negócios.

Orientador: Prof. Dr. Aléssio Tony Cavalcanti de Almeida

Catálogo na publicação
Seção de Catalogação e Classificação

D585m Diniz, Taina Alcantara Alves.

Modelo preditivo para despesas discricionárias sob restrição fiscal no ensino Superior: evidências a partir da UFPB / Taina Alcantara Alves Diniz. - João Pessoa, 2026.

52 f. : il.

Orientação: Aléssio Tony Cavalcanti de Almeida.
TCC (Graduação) - UFPB/CCSA.

1. Planejamento orçamentário. 2. Despesas discricionárias. 3. Restrição fiscal. 4. Aprendizado de máquina. 5. Ensino superior. I. Almeida, Aléssio Tony Cavalcanti de. II. Título.

UFPB/CCSA

CDU 004.6:005(043)

Tainá Alcantara Alves Diniz

Modelo Preditivo para Despesas Discricionárias sob Restrição Fiscal no Ensino Superior: Evidências a partir da UFPB

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Ciência de Dados para Negócios do Centro de Ciências Sociais Aplicadas da Universidade Federal da Paraíba (UFPB), como requisito para obtenção do grau de Bacharel em Ciência de Dados para Negócios.

Trabalho Aprovado. João Pessoa - PB, 07 de abril de 2026:

Prof. Dr. Aléssio Tony Cavalcanti de Almeida
Orientador

Prof. Dr. Hilton Martins de Brito Ramalho
Examinador

Prof. Dr. Jorge Henrique Norões Viana
Examinador

João Pessoa - PB
2026

Agradecimentos

Primeiramente, agradeço a Deus, por ter me dado a oportunidade de estar concluindo esta graduação, por ter guiado meus passos, me dado força em todos os momentos e ter colocado as pessoas certas durante a minha trajetória.

Aos meus pais, por serem minha base e meu maior exemplo de dedicação e persistência. Agradeço por todo amor, apoio e incentivo ao longo dessa jornada, e por me ensinarem que o maior legado que podemos ter é a educação.

Às minhas avós, mesmo à distância, o amor e as orações diárias são o meu grande alicerce. Ao meu avô Celso (in memoriam), cuja lembrança e carinho carrego sempre comigo, eternas saudades.

À minha irmã, Heleninha, por estar presente em todas as etapas e por ter sido um apoio em cada ciclo vivenciado ao longo dessa trajetória, você me enche de orgulho.

Ao meu namorado, João Paulo, por ser minha melhor dupla na vida e nos estudos. Agradeço pelas incontáveis chamadas em que dividimos o peso da graduação, as diversas noites estudando, e por sempre ser apoio e estar ao meu lado.

Aos meus colegas da graduação da nossa turma pioneira do curso Ciência de Dados para Negócios, em especial a Gabi, Elaine, Matheus e a Lívia, pelos momentos compartilhados, apoio e companheirismo ao longo dessa jornada.

À minha amiga Júlia, por compartilhar comigo essa etapa final da graduação com leveza e com todo o apoio.

Ao meu orientador, Prof. Aléssio, minha maior referência na área de Ciência de Dados. Agradeço pela orientação, paciência e dedicação, sempre apresentando diferentes caminhos e soluções para cada desafio. Em cada encontro, saí com novos aprendizados e uma nova perspectiva.

Por fim, agradeço a todo o corpo docente do curso de Ciência de Dados para Negócios da UFPB, pela dedicação, pelos ensinamentos e por contribuírem de forma significativa para minha formação acadêmica e profissional.

Resumo

O orçamento público é instrumento central de planejamento estatal, mas, nas universidades federais brasileiras, sua efetividade é tensionada por restrições fiscais recorrentes, pela natureza autorizativa da execução e pela permanência de práticas incrementais de projeção. Nesse contexto, o uso de médias históricas corrigidas pela inflação mostra-se insuficiente para captar a dinâmica das despesas discricionárias. Diante disso, este trabalho analisa a dinâmica entre demanda institucional e restrição orçamentária na Universidade Federal da Paraíba (UFPB) e desenvolve um modelo preditivo para estimar a execução mensal dessas despesas. A pesquisa caracteriza-se como aplicada, de abordagem quantitativa, com procedimentos técnicos de pesquisa documental, desenvolvidos no âmbito de um estudo de caso sobre a UFPB. Foi construída uma base mensal integrada para o período de janeiro de 2015 a dezembro de 2025, reunindo informações orçamentárias, macroeconômicas, institucionais e acadêmicas, tendo o valor empenhado como variável dependente. Foram comparados modelos de séries temporais e de aprendizado de máquina, com destaque para ARIMAX, SARIMAX, Random Forest e XGBoost. A comparação entre os algoritmos revelou que o modelo SARIMAX apresentou o melhor desempenho preditivo, registrando um Erro Absoluto Médio (MAE) de R\$ 7,91 milhões, uma Raiz do Erro Quadrático Médio (RMSE) de R\$ 9,97 milhões e um Erro Absoluto Escalonado Médio (MASE) de 0,55, superando os modelos ARIMAX, Random Forest e XGBoost. Adicionalmente, a análise de importância revelou que a dinâmica das despesas é fortemente influenciada por variáveis macroeconômicas e acadêmicas, destacando-se a taxa de câmbio (14,4%), o INCC (10,3%) e o número de beneficiários de assistência estudantil (9,7%). Conclui-se que o planejamento orçamentário não deve se apoiar apenas em parâmetros incrementais, mas requer ferramentas preditivas que modelem a tensão estrutural entre necessidades institucionais crescentes e severas restrições fiscais externas.

Palavras-chave: Planejamento Orçamentário. Despesas Discricionárias. Restrição Fiscal. Aprendizado de Máquina. Ensino Superior.

Abstract

Public budgeting is a central instrument of government planning; however, in Brazilian federal universities, its effectiveness is constrained by recurring fiscal restrictions, the authorizing nature of budget execution, and the persistence of incremental forecasting practices. In this context, the use of historical averages adjusted for inflation proves insufficient to capture the dynamics of discretionary expenditures. In light of this, this study analyzes the dynamics between institutional demand and budgetary constraints at the Federal University of Paraíba (UFPB) and develops a predictive model to estimate the monthly execution of these expenditures. The research is characterized as applied, with a quantitative approach and documentary research procedures, developed within the scope of a case study on UFPB. An integrated monthly database was built for the period from January 2015 to December 2025, bringing together budgetary, macroeconomic, institutional, and academic information, with committed expenditure as the dependent variable. Time series and machine learning models were compared, with emphasis on ARIMAX, SARIMAX, Random Forest, and XGBoost. The comparison among the algorithms revealed that the SARIMAX model presented the best predictive performance, recording a Mean Absolute Error (MAE) of R\$ 7.91 million, a Root Mean Squared Error (RMSE) of R\$ 9.97 million, and a Mean Absolute Scaled Error (MASE) of 0.55, outperforming the ARIMAX, Random Forest, and XGBoost models. Additionally, the feature importance analysis revealed that expenditure dynamics are strongly influenced by macroeconomic and academic variables, especially the exchange rate (14.4%), the National Construction Cost Index (INCC) (10.3%), and the number of student assistance beneficiaries (9.7%). It is concluded that budget planning should not rely solely on incremental parameters, but rather requires predictive tools capable of modeling the structural tension between growing institutional needs and severe external fiscal constraints.

Keywords: Budget Planning. Discretionary Expenditures. Fiscal Constraint. Machine Learning. Higher Education.

Lista de tabelas

Tabela 1 – Fontes de Dados Utilizadas	29
Tabela 2 – Análise descritiva dos dados	40
Tabela 3 – Resultados Modelos Testados	41

Lista de ilustrações

Figura 1 – Modelo Analítico da Pesquisa	24
Figura 2 – Série Mensal das Despesas Empenhadas (2015-2025)	39
Figura 3 – Distribuição da variável alvo	40
Figura 4 – Previsão Modelo SARIMAX (2025)	43
Figura 5 – Visão Elemento de Despesa	43
Figura 6 – Visão - Elemento de Despesas II	44
Figura 7 – Visão - Previsão do Modelo Dashboard	45

Sumário

1	INTRODUÇÃO	11
1.1	Contextualização	11
1.2	Problematização	12
1.3	Objetivos da Pesquisa	13
1.3.1	Objetivo Geral	13
1.3.2	Objetivos Específicos	13
1.4	Justificativa	13
2	REVISÃO DA LITERATURA	15
2.1	Teoria do Orçamento Público	15
2.1.1	Perspectiva Normativa do Orçamento	15
2.1.2	Perspectiva Descritiva do Orçamento	16
2.1.3	Revisões Críticas do Incrementalismo e Fundamentação Teórica da Pesquisa	16
2.2	Dinâmica do Orçamento nas Universidades Federais	17
2.3	Estrutura do Orçamento Público e os Estágios da Execução da Despesa	19
2.4	Despesas Obrigatórias versus Discricionárias	21
2.5	Ciência de Dados como Suporte à Decisão Pública	21
2.6	Modelo Analítico da Pesquisa: Aceleradores, Freios e Ajustadores	22
2.6.1	Aceleradores: variáveis endógenas	22
2.6.2	Freios: restrições exógenas	22
2.6.3	Ajustadores	23
2.6.4	Interação entre os três vetores	23
2.7	Fundamentação dos Modelos de Previsão	24
3	METODOLOGIA	27
3.1	Dados	27
3.1.1	Tratamento dos Dados e Preparação dos Dados	29
3.2	Modelo de Previsão	32
3.2.1	Modelos Clássicos	32
3.2.1.1	Modelo ARIMA com Variáveis Exógenas (ARIMAX)	32
3.2.1.2	Modelo SARIMAX	33
3.2.2	Modelos Baseados em Árvores	34
3.2.2.1	Random Forest	34
3.2.2.2	XGBoost	35
3.3	Métricas de Avaliação e Validação dos Modelos	36

3.4	Visualização dos Dados	38
4	RESULTADOS	39
4.1	Análise Descritiva	39
4.2	Resultados Modelos de Previsão	41
4.3	Visualização Analítica	43
4.4	Implicações para a Gestão Orçamentária	45
5	CONSIDERAÇÕES FINAIS	47
	REFERÊNCIAS	50

1 Introdução

1.1 Contextualização

O orçamento público constitui um dos principais instrumentos de planejamento, coordenação e execução da ação estatal, uma vez que materializa, em termos financeiros, as prioridades governamentais e os limites de atuação do Estado. No caso brasileiro, sua dinâmica historicamente revela forte influência de uma lógica incremental, em que a definição das dotações tende a se apoiar no comportamento passado das despesas, com ajustes marginais entre exercícios. Além disso, trata-se de um orçamento autorizativo, o que significa que a previsão das despesas não garante sua execução integral.

No contexto público universitário, essa característica assume contornos ainda mais sensíveis, pois a autonomia institucional convive com forte dependência das definições orçamentárias e financeiras do Governo Federal. A execução orçamentária, nesse sentido, torna-se elemento central da governança universitária, especialmente em um cenário de restrições fiscais enfrentado pelo Estado brasileiro. Nas últimas décadas, observa-se o crescimento dos contingenciamentos sobre os recursos discricionários das Instituições Federais de Ensino Superior (IFES), comprometendo sua capacidade de planejamento e de manutenção das atividades essenciais (ANDIFES, 2025).

Essa conjuntura tornou-se ainda mais expressiva em 2025, quando a Associação Nacional dos Dirigentes das Instituições Federais de Ensino Superior (Andifes) destacou que o orçamento discricionário aprovado para o exercício sofreu reduções adicionais na Lei Orçamentária Anual (LOA) e passou a estar submetido a limitações mensais de execução, estabelecidas pelo Decreto nº 12.448/2025, afetando despesas fundamentais, como a assistência estudantil, com corte de R\$ 100 milhões (ANDIFES, 2025). Na Universidade Federal da Paraíba (UFPB), isso repercutiu diretamente sobre sua capacidade de assegurar não apenas a continuidade das operações básicas, mas também o cumprimento de sua missão institucional de formar profissionais (CAETANO, 2019).

Outrossim, a elaboração da proposta orçamentária nas universidades federais também segue essa lógica de limitação. Embora o processo seja formalmente construído de baixo para cima (bottom-up), ele é fortemente condicionado pelos limites definidos pelo Ministério da Educação por meio da Matriz OCC, que define critérios objetivos de distribuição dos recursos discricionários entre as universidades federais. Na prática, as demandas das unidades são coletadas e consolidadas, mas o resultado costuma ser apenas uma redistribuição de recursos escassos. Assim, o orçamento tende a se basear nos valores de anos anteriores, com ajustes pela inflação, reforçando seu caráter incremental e reativo.

1.2 Problematização

Diante do exposto, emerge a problemática central dessa pesquisa, pois a utilização de médias históricas acrescidas de inflação como base para a elaboração da proposta orçamentária pode ser operacionalmente simples, mas revela-se insuficiente para captar a complexidade da dinâmica financeira das universidades federais. Esse ajuste desconsidera variações reais de custos, pressões macroeconômicas e regulatórias, bem como mudanças na própria demanda institucional, reduzindo a capacidade preditiva do orçamento e fragilizando seu papel como instrumento de governança.

Essa insuficiência pode ser melhor compreendida a partir da abordagem analítica adotada neste estudo, estruturada em três forças interdependentes: **aceleradores**, **freios** e **ajustadores**. Os aceleradores correspondem às demandas endógenas da missão universitária, como expansão do acesso, aumento do número de estudantes, fortalecimento da pesquisa, da extensão, ampliação da infraestrutura física e tecnológica.

Os freios dizem respeito às restrições exógenas impostas pela política fiscal do Governo Federal, materializadas em cortes, contingenciamentos e limites rígidos de execução via Matriz OCC e a LOA. Já os ajustadores referem-se aos fatores que alteram o poder de compra e a capacidade de execução dos recursos, como inflação, reajustes contratuais e oscilações nos preços de insumos. Desse modo, a interação entre as três dimensões produz uma tensão permanente entre a expansão das necessidades institucionais e a limitação dos recursos disponíveis.

Sob o ponto de vista gerencial, essa fragilidade compromete a tomada de decisão orçamentária, uma vez que amplia a incerteza, reduz a capacidade de antecipação de desequilíbrios e dificulta a adoção de medidas corretivas pela gestão universitária. Com isso, a abordagem de técnicas de Aprendizagem de Máquina aliada a modelos de séries temporais amplia a capacidade preditiva ao capturar relações dinâmicas e padrões nos dados, oferecendo um suporte decisório mais robusto do que os métodos tradicionais.

Apesar dos avanços na literatura sobre execução orçamentária (AMARAL, 2018; CAETANO, 2019), observa-se uma lacuna na aplicação de modelos preditivos que integrem variáveis macroeconômicas e institucionais em nível mensal, limitando a capacidade das IFES de antecipar cenários críticos de execução.

Diante disso, formula-se a seguinte questão de pesquisa: como estruturar um modelo preditivo que integre variáveis endógenas da demanda acadêmica e variáveis exógenas do cenário fiscal, de modo a aprimorar a previsão da execução das despesas discricionárias da Universidade Federal da Paraíba e contribuir para a melhoria do planejamento orçamentário institucional?

1.3 Objetivos da Pesquisa

1.3.1 Objetivo Geral

Para responder à problemática levantada, essa pesquisa tem como objetivo geral analisar a dinâmica entre a demanda por recursos e a restrição orçamentária na Universidade Federal da Paraíba (UFPB). A partir dessa compreensão, busca-se desenvolver um modelo preditivo a partir de comparativos envolvendo abordagens estatísticas e de aprendizado de máquina capaz de traduzir essa relação em estimativas fundamentadas em dados, fornecendo subsídios informacionais para o planejamento estratégico da instituição.

1.3.2 Objetivos Específicos

- a) Analisar a dinâmica da execução das despesas discricionárias da UFPB a partir da evolução da série mensal dos empenhos;
- b) Contextualizar no período de 2015 a 2025, o comportamento da série mensal dos empenhos da UFPB;
- c) Estruturar uma base de dados que integre informações orçamentárias, acadêmicas e macroeconômicas relevantes para a modelagem preditiva;
- d) Desenvolver modelos de previsão para estimar o valor mensal empenhado das despesas discricionárias da UFPB, avaliando seu desempenho com base em métricas estatísticas apropriadas;
- e) Prototipar um dashboard que apresente as previsões da execução orçamentária e permita fazer análises do comportamento dos principais elementos de despesa da UFPB, evidenciando a composição das despesas e os itens de maior representatividade no orçamento.

1.4 Justificativa

A presente pesquisa justifica-se, em primeiro lugar, por sua relevância empírica, em um contexto de restrição fiscal, contingenciamentos e crescente pressão sobre os recursos discricionários das IFES, torna-se fundamental dispor de instrumentos analíticos mais robustos para apoiar o planejamento orçamentário. No caso da UFPB, a proposição de um modelo preditivo que considere simultaneamente a dinâmica da demanda acadêmica e as restrições fiscais externas pode contribuir para a antecipação de cenários de insuficiência financeira.

Além disso, contribui para o aprimoramento da análise orçamentária, reduzindo a dependência de estimativas baseadas no comportamento passado. Nesse sentido, esse

estudo apresenta potencial contributivo ao oferecer subsídios analíticos para uma gestão orçamentária mais informada e orientada por evidências. Em continuidade, é justificado por sua contribuição teórica, pois embora a literatura já discuta a execução orçamentária das universidades federais e o financiamento do ensino superior, ainda permanece limitada a incorporação de abordagens preditivas que integrem variáveis macroeconômicas e institucionais em nível mensal.

Ao propor a articulação entre o modelo analítico dos aceleradores, freios e ajustadores e técnicas de modelos preditivos, o estudo busca avançar na compreensão dos determinantes da execução das despesas discricionárias em IFES, ampliando o diálogo entre gestão pública, finanças públicas e ciência de dados aplicada. Desse modo, o trabalho não apenas enfrenta um problema prático de gestão, mas também contribui para um avanço de uma abordagem analítica ainda pouco explorada no campo acadêmico.

Em síntese, a relevância desse estudo decorre de dois aspectos de contribuição: a empírica, ao oferecer uma ferramenta aplicável ao planejamento orçamentário universitário, e, no plano teórico, ao contribuir para avanço de estudos na literatura sobre previsão orçamentária em instituições federais de ensino superior. Essa combinação reforça a pertinência da pesquisa e sustenta sua originalidade no campo da administração pública e da ciência de dados aplicada ao setor público universitário.

2 Revisão da Literatura

2.1 Teoria do Orçamento Público

A literatura sobre orçamento público não se desenvolveu a partir de uma teoria única, mas sim por diferentes matrizes explicativas, principalmente da economia, da ciência política e da administração pública. Ao longo desse processo, destacam-se duas perspectivas principais: a normativa, que busca definir como os recursos deveriam ser alocados de forma racional, e a descritiva, que procura entender como o orçamento realmente é elaborado na prática. Essa distinção é importante, pois revela que o debate orçamentário oscila entre prescrições sobre como o orçamento deveria ser elaborado e interpretações sobre como ele de fato é construído.

2.1.1 Perspectiva Normativa do Orçamento

A perspectiva normativa tem como referência clássica o trabalho de (KEY, 1940), que levantou uma questão central: com base em que critérios os recursos devem ser distribuídos entre diferentes áreas públicas? A preocupação de Key Jr. estava voltada à despesa pública e à busca por critérios mais objetivos para orientar decisões em um cenário de escassez. Esse trabalho foi importante por tentar trazer mais racionalidade ao debate orçamentário, embora tenha dado menos atenção a fatores como o papel da política e da receita na definição do orçamento.

Na mesma linha, (LEWIS, 1952) buscou aprofundar a ideia de uma alocação mais racional dos gastos públicos, inspirado no conceito de utilidade marginal, o autor propõe que as decisões orçamentárias poderiam ser feitas comparando os custos e benefícios de cada alternativa. Assim, o desafio seria identificar em que medida a ampliação de recursos em determinada categoria de gasto produziria maior retorno social do que sua destinação a outra finalidade. Essa abordagem parte de uma suposição exigente ao supor que os agentes públicos seriam capazes de avaliar, de maneira ampla, as múltiplas possibilidades de alocação orçamentária.

Essa visão normativa do orçamento influenciou várias reformas posteriores, entre elas, destacam-se o **Planning-Programming-Budgeting System (PPBS)**, associado (SCHICK, 1966), e o Orçamento Base Zero, proposto por (PYHRR, 1973). No primeiro caso, buscava integrar planejamento, execução e orçamento, dando mais importância à avaliação de resultados e à organização das decisões. Nesse contexto, já o Orçamento Base Zero propõe que todos os gastos sejam revistos do zero, exigindo que cada despesa seja justificada novamente. Assim, em ambos os casos, permanece a crença de que o processo

orçamentário pode ser aperfeiçoado por meio de mecanismos mais objetivos de análise, hierarquização de prioridades e mensuração de resultados.

Apesar de sua importância, a perspectiva normativa apresenta algumas limitações, seu principal problema está na ideia de que os gestores públicos teriam informação suficiente, plena capacidade de análise e atuariam de forma totalmente neutra ao tomar decisões. Na prática, porém, o processo orçamentário ocorre em um ambiente marcado por conflitos distributivos, limitações de informação, negociações políticas e assimetria de poder. Por isso, embora esses modelos tenham contribuído para tornar o orçamento mais técnico, eles não são suficientes para explicar como o processo realmente acontece.

2.1.2 Perspectiva Descritiva do Orçamento

Diferentemente da tradição normativa, a perspectiva descritiva possui o foco da pergunta sobre como o orçamento deveria ser formulado para a análise de como ele efetivamente é construído. Nesse contexto, ganha destaque a contribuição de (WILDAVSKY, 1984), que entende o orçamento não como um processo racional, mas como um processo político. Para o autor, as decisões orçamentárias envolvem estratégias, negociações e limitações dos próprios gestores envolvidos. Assim, o orçamento deixa de ser visto apenas como um cálculo técnico ideal e passa a ser compreendido como um espaço de negociação em torno da preservação e da expansão de bases previamente estabelecidas.

A base dessa interpretação está no conceito de incrementalismo, desenvolvido por (LINDBLUM, 1959) e depois aplicado ao orçamento por Wildavsky. Na visão do autor, os tomadores de decisão não conseguem analisar todas as alternativas possíveis, por isso, tendem a decidir nas margens, mediante pequenos ajustes sucessivos, pois enfrentam racionalidade limitada. No orçamento público, isso significa que as decisões geralmente partem do orçamento do ano anterior, e as disputas políticas se concentram nas mudanças marginais, como aumentos, cortes ou redistribuições.

Desse modo, a contribuição de Wildavsky é importante porque rompe com a ideia de que o orçamento é totalmente revisto a cada ano. Nesse sentido, o autor retrata que a estabilidade da base orçamentária exerce função institucional relevante, inclusive na mitigação de conflitos distributivos. Dessa forma, os atores buscam principalmente em dois fatores, preservar os recursos já garantidos e, quando possível, ampliar os valores adicionais.

2.1.3 Revisões Críticas do Incrementalismo e Fundamentação Teórica da Pesquisa

Apesar de sua influência, o incrementalismo clássico passou a receber críticas importantes. LÉLOUP (1978) aponta que, quando usado como explicação geral do

orçamento, o modelo pode perder força, já que qualquer mudança em relação ao passado pode ser interpretada como “incremento”, reduzindo seu poder explicativo. Além disso, estudos posteriores mostram que o orçamento não pode ser entendido apenas pelas decisões internas das organizações, sendo necessário considerar fatores como centralização, relações entre instituições, choques fiscais e mudanças no ambiente político e econômico.

Nesse contexto, SCHICK (1966) destaca a chamada era decremental, marcada pela necessidade de lidar com cortes e restrições orçamentárias, e não mais com expansão dos gastos. Em virtude disso, RUBIN (1990) reforça que a análise do orçamento deve estar mais próxima da realidade, incorporando sua complexidade institucional e política. Desse modo, essas contribuições ampliam a compreensão do orçamento, mostrando que ele também é influenciado por fatores externos, relações de poder e condições econômicas.

Diante desse debate, este estudo adota uma perspectiva descritiva do orçamento público, com base em WILDAVSKY (1984), mas com uma abordagem atualizada. O objetivo não é propor um modelo ideal de alocação ótima de recursos, mas compreender e prever o comportamento das despesas discricionárias em um contexto marcado por restrições fiscais, inflação e outros fatores externos.

Assim, a pesquisa reconhece a teoria que possui influência do incrementalismo, mas vai além dela ao incorporar elementos econômicos, institucionais e fiscais que afetam o orçamento atualmente. Dessa forma, essa abordagem permite compreender a execução orçamentária universitária não apenas como repetição do passado, mas como resultado da interação entre a trajetória histórica, os limites institucionais e as pressões econômicas atuais.

2.2 Dinâmica do Orçamento nas Universidades Federais

A dinâmica orçamentária das universidades federais pode ser entendida a partir de três pontos principais, o uso do orçamento passado como base (incrementalismo), o caráter autorizativo e o cenário de restrições fiscais. Por conseguinte, esses elementos operam juntos e ajudam a explicar por que o orçamento dessas instituições costuma ser mais reativo, limitado por fatores externos e nem sempre acompanha o crescimento de suas necessidades institucionais.

Embora a teoria incrementalista clássica não seja suficiente, por si só, para explicar toda a complexidade do orçamento contemporâneo, ela continua relevante para descrever práticas em que o passado funciona como referência para a projeção do exercício seguinte. Nessa lógica, o orçamento costuma ser construído a partir do que já foi executado no ano anterior, com ajuste na correção inflacionária. Logo, as discussões se concentram mais em aumentos ou cortes pontuais do que em mudanças mais profundas na estrutura dos gastos.

Ainda que o processo de elaboração do orçamento envolva a participação das uni-

dades acadêmicas e administrativas no *bottom-up*, com posterior consolidação e aprovação interna. Todavia, esse processo ocorre sob limitação prévia, imposta pelos limites globais estabelecidos externamente, especialmente pela Matriz OCC. Dessa forma, o orçamento deixa de ser um processo totalmente estratégico e passa, muitas vezes, a funcionar como uma redistribuição de recursos escassos dentro da própria instituição (CAETANO, 2019; JUNIOR; FLACH; MATTOS, 2020).

Essa característica se intensifica porque, na prática, o orçamento costuma apoiar-se em parâmetros históricos, notadamente na média dos exercícios anteriores acrescida de uma previsão inflacionária. Esse procedimento reforça o caráter incremental, pois mantém o histórico como principal referência. No entanto, apesar de ser mais simples de aplicar, esse modelo tem limitações, já que pode ignorar mudanças reais nos custos, alterações nas regras e variações econômicas que impactam diretamente as despesas discricionárias.

Em primeiro lugar, a análise do orçamento não pode se limitar apenas à lógica incremental. No cenário atual, torna-se fundamental considerar, também, o seu caráter autorizativo. Sob essa ótica, a LOA permite a realização das despesas, contudo, não garante que elas serão executadas em sua totalidade. Na prática, a execução efetiva depende da disponibilidade de recursos e das decisões de gestão fiscal tomadas ao longo do exercício. Conseqüentemente, mesmo despesas devidamente autorizadas podem ser limitadas, adiadas ou contingenciadas.

A literatura sobre financiamento do ensino superior e execução orçamentária universitária aponta que essa condição não pode ser entendida como problema meramente conjuntural. Ao contrário, trata-se de uma tensão estrutural entre a ampliação da missão universitária e a limitação dos recursos disponíveis. A expansão do acesso, o fortalecimento da pesquisa e da extensão, bem como a necessidade de ampliação e manutenção da infraestrutura, elevam continuamente a pressão sobre o orçamento das universidades federais. Em contrapartida, a política fiscal e os mecanismos de contenção de gasto restringem a capacidade institucional de resposta, produzindo um quadro de vulnerabilidade orçamentária (AMARAL, 2018; CAETANO, 2019).

Ao transpor a lógica de toda teoria para a realidade das universidades federais, constata-se que os métodos tradicionais, mostram-se insuficientes. Embora apresentem baixa complexidade, tais modelos falham ao não antecipar cenários de retração financeira, tampouco oferecer subsídios para decisões estratégicas em ambientes de alta volatilidade. Em decorrência dessa limitação, o orçamento tende a ser conduzido muito mais como uma medida de adaptação à escassez do que propriamente como um instrumento de gestão estratégica.

Diante disso, a compreensão plena do orçamento universitário exige a análise conjunta de um desses determinantes: a herança do histórico orçamentário, o caráter autorizativo da execução e as recorrentes restrições fiscais. Esses fatores não apenas

explicam as limitações das metodologias vigentes, como também reforçam a necessidade de abordagens mais aprofundadas, capazes de conferir sustentabilidade ao planejamento em contextos de incerteza (AMARAL, 2018).

2.3 Estrutura do Orçamento Público e os Estágios da Execução da Despesa

O orçamento público é o principal instrumento por meio do qual o Estado transforma suas prioridades em ações concretas. Sob essa ótica, para além de seu caráter financeiro, ele representa um processo que envolve escolhas políticas, definição de prioridades e organização dos recursos disponíveis. Nesse sentido, o orçamento se estrutura como um sistema integrado de planejamento e execução, no qual decisões governamentais são convertidas em programas, ações e dotações orçamentárias.

A Constituição Federal de 1988 organizou esse sistema a partir de três instrumentos principais, o Plano Plurianual (PPA), a Lei de Diretrizes Orçamentárias (LDO) e a LOA, com o objetivo de articular planejamento, definição de metas e execução do gasto público. O PPA estabelece as diretrizes e objetivos de médio prazo da ação governamental, a LDO orienta a elaboração do orçamento anual, definindo prioridades e metas fiscais, e a LOA detalha as receitas e despesas autorizadas para cada exercício. Esse arranjo reforça a ideia de que o orçamento não é apenas uma autorização de despesas, mas parte de um processo amplo de gestão pública.

No campo do orçamento público, geralmente se aponta que ele possui três funções principais: alocar recursos, distribuir renda e contribuir para a estabilidade econômica. Essa classificação é amplamente utilizada nos estudos de finanças públicas (GIAMBIAGI; ALÉM, 2016). A LOA define os valores autorizados para os gastos, sem implicar obrigatoriedade de execução, uma vez que o orçamento brasileiro possui caráter autorizativo e está sujeito a restrições fiscais e financeiras ao longo do exercício.

Na prática, a execução do orçamento depende de outros fatores, como a programação financeira, o cumprimento de metas fiscais, os limites de empenho e a disponibilidade de caixa. Por isso, diz-se que o orçamento tem caráter autorizativo, pois ele permite a realização da despesa, mas não garante que ela será executada integralmente. Essa diferença é especialmente importante no caso das universidades federais, que, embora tenham autonomia formal, enfrentam limitações reais impostas pelo processo de execução orçamentária do Governo Federal.

Esse ponto se torna evidente quando a alocação de recursos é analisada sob a lógica top-down. Nessa abordagem, o processo orçamentário é estruturado de forma hierárquica, em que as decisões sobre o volume total de gastos são definidas previamente à distribuição entre setores e instituições (LJUNGMAN, 2009). Assim, é estabelecido inicialmente um

limite global de despesas, que posteriormente é desdobrado em cotas setoriais e, por fim, distribuído internamente entre as unidades executoras.

A execução da despesa pública consiste na aplicação dos recursos previstos no orçamento, seguindo rigorosamente as etapas e formalidades exigidas pela lei. No Brasil, a Lei nº 4.320/1964 regulamenta esse processo, enquanto o Manual de Contabilidade Aplicada ao Setor Público (MCASP) orienta sua estruturação para separar os impactos orçamentário, patrimonial e financeiro de cada ação. Dessa forma, a observância rigorosa dessas normas garante a transparência governamental e a geração de dados sobre a destinação dos recursos públicos (BRASIL, 1964).

Sob essa perspectiva, o processo se consolida por meio de três estágios fundamentais: o empenho, a liquidação e o pagamento. O empenho constitui o primeiro ato da execução da despesa, formalizando a reserva de recursos no orçamento e a decisão administrativa de comprometer uma parcela do crédito disponível. Nesse sentido, essa etapa representa o compromisso de gasto por parte do gestor público, mesmo que o bem ou serviço ainda não tenha sido entregue. No âmbito do orçamento, destaca-se que essa autorização prévia é o mecanismo que impede o ente público de assumir obrigações acima de sua capacidade orçamentária.

Posteriormente, o processo avança para a liquidação, momento em que se comprova que o serviço contratado foi executado conforme as condições estabelecidas, gerando o direito adquirido do credor. Somente após essa conferência técnica é que ocorre o pagamento, que traduz o desembolso financeiro propriamente dito. Portanto, essa sequência é indispensável para o controle do gasto público, pois estabelece uma separação clara entre a autorização, a comprovação e a quitação final da despesa (GIACOMONI, 2021; JR.; REIS, 2016).

Na análise da execução orçamentária, a escolha da métrica de despesa não é apenas uma decisão técnica, mas uma definição do fenômeno que se pretende observar. Para estudos voltados ao comportamento do gestor e à priorização de políticas públicas, o valor empenhado consolida-se como a variável de previsão mais fidedigna, funcionando como uma proxy da decisão de gasto efetiva. Nesse contexto, enquanto o pagamento representa apenas a extinção da obrigação, frequentemente sujeita a disponibilidades de caixa, contingenciamentos e restos a pagar, o empenho revela a intenção real da administração (NACIONAL, 2023).

Dessa forma, a decisão discricionária do gestor materializa-se no exato momento em que ele vincula a dotação autorizada a um credor específico, convertendo a autorização legislativa em compromisso institucional. Dessa maneira, ao adotar o valor empenhado como variável de análise, a pesquisa ganha maior consistência analítica ao identificar o ponto de transição em que a política pública deixa a esfera da intenção orçamentária para ingressar no campo do compromisso.

2.4 Despesas Obrigatórias versus Discricionárias

A diferenciação entre despesas obrigatórias e discricionárias torna-se fundamental para a análise da gestão fiscal, uma vez que evidencia que a capacidade de decisão do gestor público não se distribui de forma homogênea sobre o orçamento. As despesas obrigatórias são aquelas em que o ente público não possui autonomia de decisão sobre o seu pagamento, pois o seu fundamento reside em normas de hierarquia superior. Desse modo, constituem, primordialmente, os gastos pessoais e encargos sociais, benefícios previdenciários e assistenciais, além do serviço da dívida e transferências constitucionais a entes subnacionais.

Em contrapartida, as despesas discricionárias são gastos públicos que não são obrigatórios por lei, permitindo ao gestor decidir sobre sua execução conforme a disponibilidade de recursos. Essas despesas incluem, sobretudo, investimentos e outros gastos necessários ao funcionamento da administração pública. Todavia, a literatura adverte que a "discricionariedade" é na prática limitada, uma vez que a paralisação desses gastos pode levar à descontinuidade de serviços públicos essenciais, configurando o que se denomina "despesa discricionária de manutenção obrigatória" (GIACOMONI, 2021).

Nessa linha, o funcionamento cotidiano da universidade depende justamente da parcela mais vulnerável do orçamento que são os contratos de energia, água, vigilância, limpeza, manutenção predial, insumos laboratoriais, assistência estudantil. Quando a compressão recai sobre a despesa discricionária, atinge a base material que sustenta ensino, pesquisa e extensão.

As notas publicadas da Andifes reiteram esse diagnóstico ao apontarem redução nominal de recursos, insuficiência orçamentária e dificuldades recorrentes de execução e fechamento do exercício, mesmo diante da continuidade das atividades acadêmicas (ANDIFES, 2025).

2.5 Ciência de Dados como Suporte à Decisão Pública

A ciência de dados configura-se como um campo interdisciplinar voltado à extração de conhecimento a partir de dados, combinando estatística, computação e entendimento do contexto analisado. Nesse sentido, o núcleo reside na capacidade de estruturar problemas reais, permitindo não apenas a identificação de padrões, mas também a geração de previsões e o suporte à tomada de decisão (PROVOST; FAWCETT, 2013).

Nesse contexto, o uso estratégico de dados transcende ganhos de eficiência operacional, ao reforçar a coordenação institucional, a transparência e a fundamentação técnica das decisões governamentais. No setor público, a relevância do uso de modelos de previsão se intensifica sobretudo em ambientes caracterizados por escassez de recursos, condicionamento normativo e necessidade de definição de prioridades. Não se propõe a

substituição do juízo administrativo por algoritmos, mas sua qualificação por meio de uma base mais consistente (CAMPBELL; DEBLOIS; OBLINGER, 2007).

Por isso, a aplicação de ciência de dados à gestão pública exige interpretação, governança dos dados e clareza sobre os limites do modelo. Em vez de prometer neutralidade, a abordagem mais consistente é tratá-la como mecanismo de apoio à decisão sob incerteza. Desse modo, o ponto central não reside na possibilidade de produzir estimativas futuras, mas na capacidade de compreender padrões de comportamento da despesa e de analisar como fatores acadêmicos, administrativos e fiscais se articulam na conformação do gasto discricionário (GUPTA; CHOUDHARY, 2015).

2.6 Modelo Analítico da Pesquisa: Aceleradores, Freios e Ajustadores

A presente pesquisa propõe um modelo analítico para compreender e prever a execução das despesas discricionárias nas universidades federais, considerando a interação entre três elementos principais: aceleradores, freios e ajustadores. Esse modelo parte do entendimento de que a forma tradicional de elaboração do orçamento, baseada na média dos anos anteriores com um ajuste inflacionário, é limitada para explicar a dinâmica atual.

2.6.1 Aceleradores: variáveis endógenas

Os aceleradores correspondem às forças endógenas intrínsecas à missão institucional da universidade. Dessa forma, tratam-se de fatores que não apenas ampliam, como também pressionam e tornam mais complexa a necessidade de gastos discricionários com o passar do tempo. Dentre esses aspectos, destacam-se a expansão do acesso ao ensino superior, assistência estudantil, o fortalecimento das atividades de pesquisa e programas de monitoria.

Sob uma perspectiva analítica, os aceleradores assumem papel central ao expressarem a magnitude das demandas institucionais por recursos. Esses fatores refletem a intensificação das atividades associadas à missão universitária e, por conseguinte, contribuem para a ampliação das necessidades de gasto discricionário.

2.6.2 Freios: restrições exógenas

Os freios, por outro lado, correspondem às restrições exógenas impostas ao orçamento universitário, especialmente no que tange à política fiscal do Governo Federal e ao arranjo institucional da execução orçamentária brasileira. Tais entraves materializam-se por meio de cortes, contingenciamentos e bloqueios, além de restrições mensais de execução e condicionamentos impostos pela LOA e pela Matriz OCC.

Substancialmente, os freios representam a dimensão externa da escassez, ou seja, são os mecanismos que restringem a disponibilidade de recursos em função do ambiente fiscal e do controle centralizado do gasto. No modelo analítico desta pesquisa, esses freios atuam como vetores de contenção, visto que, mesmo diante do aumento da demanda, a execução da despesa não depende apenas de necessidades internas.

Ao contrário, a execução permanece condicionada ao espaço fiscal e à liberação efetiva de recursos. Nesse contexto, os freios explicam a divergência entre a trajetória da execução e a necessidade institucional, evidenciando a assimetria entre a demanda universitária e o empenho efetivo. No modelo proposto, esses fatores apresentam associação negativa, atuando como elementos de compressão sobre a despesa.

2.6.3 Ajustadores

Os ajustadores não atuam como fatores de expansão da demanda, como os aceleradores, nem como mecanismos diretos de contenção fiscal, como os freios. Seu papel é ajustar a relação entre a dotação disponível e a capacidade efetiva de execução. Nessa categoria, incluem-se a inflação, os reajustes contratuais e as variações nos preços de insumos essenciais. Esses fatores afetam diretamente o poder de compra do orçamento, gerando impactos mesmo quando os valores nominais permanecem estáveis.

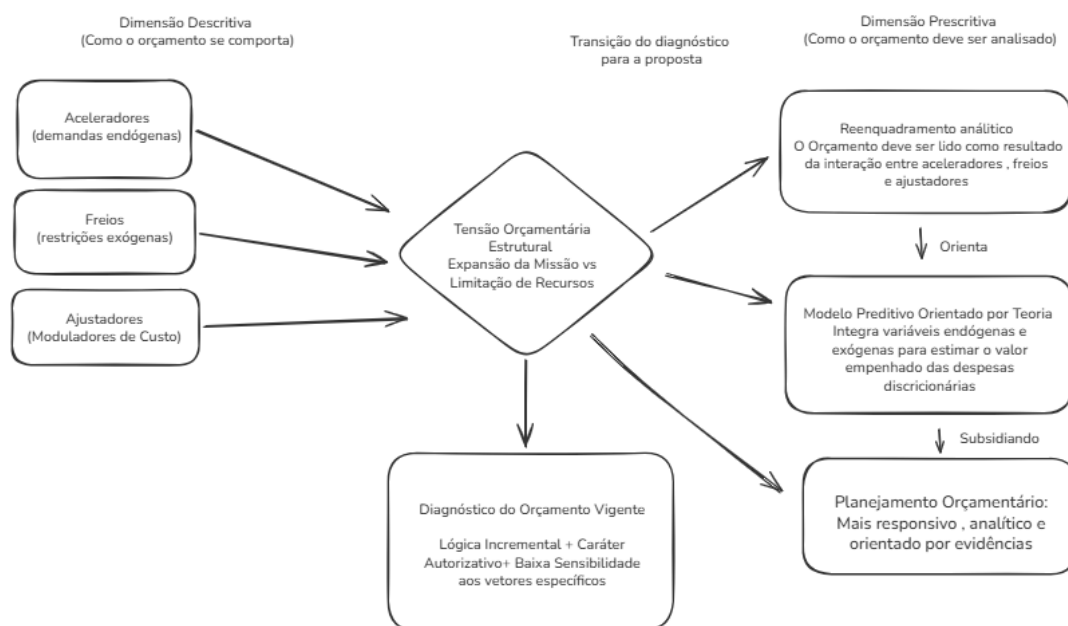
A relevância desses elementos reside na sua capacidade de explicar por que exercícios com valores nominais semelhantes podem apresentar diferentes níveis de execução. Em contextos inflacionários, por exemplo, um mesmo volume de recursos perde poder de compra, comprometendo a sustentação das atividades institucionais. Assim, os ajustadores introduzem a dimensão do custo real da operação, ajustando a relação entre necessidade e execução. Em última análise, evidenciam que o orçamento não deve ser analisado apenas em termos nominais ou com base em médias históricas.

2.6.4 Interação entre os três vetores

A principal contribuição do modelo proposto reside na compreensão das interações entre seus elementos, e não na análise isolada de cada um deles. Neste estudo, a execução das despesas discricionárias é interpretada como resultado da interação entre três forças: aceleradores, freios e ajustadores. Enquanto os aceleradores ampliam a demanda por recursos, os freios restringem sua disponibilidade ou execução, e os ajustadores afetam o poder de compra do orçamento.

A Figura 1 apresenta o modelo analítico da pesquisa em duas dimensões complementares, na dimensão descritiva, evidencia-se como aceleradores, freios e ajustadores produzem uma tensão orçamentária estrutural que, no modelo vigente, é tratada de forma insuficiente por uma lógica incremental e autorizativa. Na dimensão prescritiva, propõe-se um novo enquadramento analítico, no qual o orçamento é interpretado a partir dessas

Figura 1 – Modelo Analítico da Pesquisa



Fonte: Elaboração própria (2026).

interações, servindo de base para um modelo preditivo teoricamente orientado e para um planejamento baseado em evidências.

2.7 Fundamentação dos Modelos de Previsão

A previsão de séries temporais pode ser desenvolvida por diferentes estratégias de modelagem, que incluem desde modelos estatísticos tradicionais até métodos baseados em aprendizado de máquina. Nesse campo, a escolha do método não deve ser tratada como decisão técnica ou automática, pois depende das propriedades da série, das informações disponíveis, da presença de variáveis explicativas e do propósito da análise. Assim, além de optar entre metodologias distintas, o desafio consiste em selecionar a abordagem mais aderente com a problemática e com as condições empíricas em que a previsão será produzida (JANUSCHOWSKI et al., 2020; SHMUELI; KOPPIUS, 2011).

No caso das despesas discricionárias, a literatura indica que sua modelagem pode exigir abordagens capazes de representar simultaneamente regularidades temporais e a influência de fatores externos ao comportamento passado da série. Pois, séries mensais de empenho podem incorporar tendência, sazonalidade e autocorrelação, mas também responder a condicionantes institucionais, fiscais e macroeconômicos (ZHENG; CASARI, 2018).

O modelo ARIMAX (AutoRegressive Integrated Moving Average with Exogenous

Variables) admite que o comportamento futuro da variável de estudo seja explicado tanto por sua própria trajetória passada quanto por variáveis externas (BOX; JENKINS; REINSEL, 2016). Trata-se de uma extensão do modelo ARIMA, na qual os componentes autorregressivos, de diferenciação e de médias móveis são combinados à inclusão de regressores exógenos, permitindo captar simultaneamente a persistência temporal da série e os efeitos de fatores observáveis que afetam sua dinâmica (GUJARATI, 2011; WOOLDRIDGE, 2016).

$$y_t = \phi(B)(1 - B)^d y_t + \theta(B)\varepsilon_t + \beta X_t \quad (2.1)$$

em que B é o operador de defasagem, ou seja, os valores passados da série, d representa o grau de diferenciação necessário para estacionarizar a série, $\phi(B)$ e $\theta(B)$ correspondem aos polinômios autorregressivo e de médias móveis, e X_t corresponde ao conjunto de variáveis exógenas incluídas no modelo.

O modelo SARIMAX (Seasonal AutoRegressive Integrated Moving Average with Exogenous Variables) constitui uma extensão do ARIMAX que incorpora explicitamente componentes sazonais. Sua formulação geral é dada por:

$$\phi(B)\Phi(B^s)(1 - B)^d(1 - B^s)^D y_t = \theta(B)\Theta(B^s)\varepsilon_t + \beta X_t \quad (2.2)$$

em que B os termos associados indicam como valores passados da série influenciam o presente, s corresponde à periodicidade da série e $\Phi(B^s)$ e $\Theta(B^s)$ são utilizados para tornar a série mais estável, removendo tendências e padrões que se repetem ao longo do tempo. Os componentes sazonais capturam comportamentos periódicos, como variações mensais ou anuais.

Por fim, o termo βX_t representa o efeito de variáveis externas sobre a série, permitindo que o modelo incorpore informações adicionais que ajudam a explicar e prever seu comportamento. Dessa forma, o SARIMAX integra, em uma única estrutura, efeitos do passado, padrões sazonais e influências externas.

No âmbito dos métodos de aprendizado de máquina, o modelo Random Forest, proposto por (BREIMAN, 2001), consiste em um método de agregação de múltiplas árvores de decisão construídas a partir de subconjuntos aleatórios dos dados e das variáveis. O funcionamento consiste em combinar previsões originadas de diferentes árvores, estimadas a partir de subconjuntos aleatórios dos dados e das variáveis explicativas, de modo a reduzir a variância do modelo e aumentar sua robustez. A previsão é obtida pela média das árvores individuais:

$$\hat{y}_t = \frac{1}{M} \sum_{m=1}^M T_m(\mathbf{X}_t)$$

onde T_m representa a m -ésima árvore e M o número total de árvores. Esse procedimento reduz a variância do modelo e aumenta sua robustez, sendo capaz de capturar relações não lineares e interações complexas entre variáveis.

O XGBoost (Extreme Gradient Boosting) adota uma abordagem baseada em boosting, na qual as árvores são construídas de forma sequencial, sendo cada nova árvore ajustada para corrigir os erros das anteriores. Sua representação geral pode ser expressa como:

$$\hat{y}_t = \sum_{m=1}^M \gamma_m h_m(\mathbf{X}_t)$$

em que h_m são funções base (árvores de decisão) e γ_m são parâmetros ajustados iterativamente por meio da minimização de uma função de perda (FRIEDMAN, 2001). O modelo incorpora mecanismos de regularização que penalizam estruturas excessivamente complexas, contribuindo para melhor capacidade de generalização.

No contexto deste estudo, o XGBoost mostra-se uma alternativa promissora porque a execução das despesas discricionárias pode ser influenciada por choques administrativos, eventos institucionais, fatores fiscais e condições macroeconômicas cujos efeitos não são necessariamente lineares nem constantes ao longo do tempo.

A comparação entre modelos estatísticos com estrutura temporal explícita, como ARIMAX e SARIMAX, e métodos de aprendizado de máquina, como Random Forest e XGBoost, mostra-se coerente com a natureza da problemática de despesas discricionárias. Pois, a dinâmica das despesas discricionárias pode refletir, de um lado, persistência, autocorrelação e sazonalidade e, de outro, relações complexas entre múltiplos fatores explicativos. Enquanto os modelos da família ARIMA com variáveis exógenas oferecem maior transparência analítica e aderência à tradição econométrica, os métodos de aprendizado de máquina ampliam a capacidade de capturar não linearidades, interações e efeitos de alta complexidade.

3 Metodologia

Para alcançar os objetivos propostos, foi adotado um percurso metodológico alinhado à natureza do problema e à hipótese formulada. Trata-se de uma investigação de caráter aplicado, focada na resolução do desafio prático da gestão orçamentária da UFPB (GIL, 2010). No tocante aos objetivos, o estudo assume um perfil descritivo, ao contextualizar e examinar o comportamento histórico da série de empenhos, analítico-comparativo, ao confrontar diferentes abordagens de modelagem, e preditivo, ao desenvolver e avaliar modelos capazes de estimar valores futuros da execução orçamentária (VERGARA, 2016; JR. et al., 2009).

Quanto aos procedimentos técnicos, a pesquisa é classificada como documental e estudo de caso. Essa abordagem fundamenta-se na utilização de fontes secundárias, como relatórios oficiais e bases públicas de dados, permitindo uma análise aprofundada voltada especificamente para a unidade institucional da UFPB (LAKATOS; MARCONI, 2003; YIN, 2015).

Adicionalmente, adotou-se uma abordagem quantitativa, escolhida por sua capacidade de oferecer análises objetivas, mensuráveis e replicáveis (RICHARDSON, 2017). Nesse sentido, serão empregadas técnicas estatísticas e ferramentas de modelagem computacional para a coleta, o tratamento e a análise dos dados, o que permitirá identificar padrões relevantes e realizar previsões baseadas em evidências (CRESWELL, 2014).

Dessa maneira, buscou-se garantir rigor analítico à investigação e fornecer subsídios objetivos para a tomada de decisão no âmbito da gestão pública. Em última análise, a articulação entre dados quantitativos e modelagem preditiva constitui a base analítica deste estudo, permitindo descrever o comportamento histórico da execução orçamentária e estimar tendências futuras no contexto da UFPB

3.1 Dados

Os dados utilizados nesta pesquisa foram obtidos em bases públicas e documentos institucionais, com a finalidade de compor uma base mensal integrada contendo informações orçamentárias, acadêmicas, institucionais e macroeconômicas relacionadas às despesas discricionárias da UFPB. O período de análise compreende janeiro de 2015 a dezembro de 2025, recorte definido em função da disponibilidade e da compatibilidade temporal das diferentes fontes consultadas.

A variável dependente do estudo corresponde ao valor mensal empenhado pela Universidade Federal da Paraíba em despesas discricionárias. As informações orçamentárias foram extraídas do Portal da Transparência do Governo Federal, mantido pela

Controladoria-Geral da União (CGU), a partir da base de Execução de Despesas. Posteriormente, os registros foram filtrados de modo a isolar os dados referentes à UFPB e aos elementos de despesa classificados como discricionários.

As variáveis macroeconômicas empregadas nesta pesquisa foram obtidas por meio de consultas à API do Banco Central do Brasil, a partir do Sistema Gerenciador de Séries Temporais (SGS). Por essa fonte, foram coletadas as séries mensais referentes à taxa Selic, ao salário mínimo e ao Índice Nacional de Custo da Construção (INCC). Entre essas variáveis, a taxa Selic corresponde à taxa básica de juros da economia brasileira, divulgada pelo Banco Central como a taxa média ajustada das operações compromissadas de um dia, lastreadas em títulos públicos federais. No presente estudo, essa variável é utilizada como indicador das condições monetárias da economia e do custo macroeconômico do crédito (Banco Central do Brasil, 2023).

Adicionalmente, o salário mínimo foi incorporado ao conjunto de variáveis explicativas como referência para a evolução do piso remuneratório nacional, podendo refletir, ainda que indiretamente, alterações em estruturas de custo associadas à dinâmica institucional. Por sua vez, o INCC foi utilizado como indicador da variação dos custos relacionados à construção civil e à manutenção da infraestrutura.

Além disso, para a obtenção da variável câmbio mensal, foi utilizada a API de cotações do Banco Central, representando potencial efeito sobre os custos de bens insumos sensíveis ao mercado internacional. Já a variável de inflação, representada pelo Índice Nacional de Preços ao Consumidor Amplo (IPCA), foi obtida por meio do Sistema IBGE de Recuperação Automática (SIDRA) e foi incorporada por permitir captar a variação do nível geral de preços e, conseqüentemente, a alteração do poder de compra real dos recursos orçamentários ao longo do tempo (ESTATÍSTICA, 2026).

No que se refere às variáveis que representam características internas da universidade e que podem afetar a demanda por gastos discricionários, por captar as pressões internas, os dados foram obtidos junto à PROPLAN (Pró-Reitoria de Planejamento e Desenvolvimento) (Universidade Federal da Paraíba, 2025). Entre essas informações estão a quantidade de beneficiários da assistência estudantil, o número de estudantes bolsistas inseridos em programas de monitoria, pesquisa e extensão, bem como os gastos destinados à manutenção desses bolsistas nos programas.

Adicionalmente, para a construção de variáveis institucionais relacionadas ao período da pandemia, à suspensão das atividades presenciais, ao retorno das aulas e à identificação dos meses correspondentes aos períodos letivos, foram utilizados os calendários acadêmicos da UFPB. Essas variáveis foram inseridas por capturar choques operacionais e padrões sazonais específicos do funcionamento universitário.

Os calendários foram obtidos a partir do portal da Pró-Reitoria de Graduação (PRG), mediante extração via HTTP dos arquivos em formato PDF referentes ao período de

janeiro de 2015 a dezembro de 2025. Após a coleta, foi realizado um processo de raspagem e extração de informações textuais diretamente dos arquivos, permitindo identificar, de forma estruturada, datas de início e término dos períodos letivos, interrupções acadêmicas e eventos pontuais, como a suspensão e retomada das atividades presenciais durante a pandemia.

As variáveis dummies institucionais foram construídas a partir da identificação de eventos relevantes com potencial de impactar a gestão e o ambiente orçamentário da universidade. A definição desses eventos foi realizada com base em fontes documentais, incluindo notícias, comunicados oficiais e registros institucionais, o que permitiu o mapeamento temporal de ocorrências como períodos de instabilidade administrativa, contingenciamentos fiscais e mudanças na gestão da reitoria. A Tabela 1 sintetiza as bases de dados utilizadas e suas respectivas origens.

Tabela 1 – Fontes de Dados Utilizadas

Variável	Fonte de Dados
valor_empenhado	Portal da Transparência (CGU)
ipca	IBGE (SIDRA)
incc	BCB
salário mínimo	BCB
câmbio mensal	BCB
selic mensal	BCB
dias úteis	Calendário
dummy suspensão presencial	Calendários acadêmicos (PRG/UFPB)
dummy aula	Calendários acadêmicos (PRG/UFPB)
dummy instabilidade gestão	Notícias e comunicados oficiais
dummy mudança de gestão	Site UFPB
beneficiários extensão	PROPLAN(UFPB)
beneficiários pesquisa	PROPLAN(UFPB)
beneficiários assistência	PROPLAN(UFPB)
bolsa extensão	PROPLAN(UFPB)
bolsa pesquisa	PROPLAN(UFPB)
bolsa assistência	PROPLAN(UFPB)

Fonte: Elaboração própria (2026).

3.1.1 Tratamento dos Dados e Preparação dos Dados

Na definição da variável dependente, foi adotado o valor empenhado mensal como proxy da execução das despesas discricionárias. Essa escolha justificou-se pelo fato do empenho representar o primeiro estágio formal da execução da despesa pública, momento em que a administração assume a obrigação orçamentária, sendo, portanto, uma medida mais aderente à decisão efetiva de gasto do que o pagamento, que pode ser afetado por restrições financeiras e postergações de caixa.

Por fim, após a realização da consolidação das bases provenientes das diferentes fontes documentais e administrativas, os dados foram organizados em frequência mensal. Em seguida, a série foi particionada em dois subconjuntos: treinamento, compreendendo as observações anteriores a janeiro de 2025, e teste, composto pelos meses de janeiro a dezembro de 2025. Esse ponto de corte foi definido, por corresponder ao período mais recente da série.

Para garantir comparabilidade intertemporal, os valores monetários foram trabalhados em termos reais, com correção inflacionária baseada no IPCA. Além disso, na etapa de modelagem, o valor negativo do mês de julho de 2022, interpretado como reflexo de estorno ou regularização contábil, foi substituído por zero. Tal procedimento foi adotado por duas razões principais: primeiro, porque observações negativas não correspondem à execução típica da despesa discricionária, segundo, porque sua permanência na série inviabilizaria a transformação $\log(1 + y)$. Não obstante, reconhece-se que essa escolha pode reduzir a capacidade do modelo de capturar eventos abruptos de ajuste orçamentário, configurando uma limitação metodológica do estudo.

A variável dependente (valor empenhado) do conjunto de treino foi transformada por meio de $\log(1 + y)$, com o objetivo de reduzir assimetria, estabilizar a variância e tornar o processo de estimação mais robusto diante da presença de picos de execução. Após a geração das previsões, os valores estimados foram reconvertidos à escala original por meio da transformação inversa **expm1**.

Com o fito de evitar vazamento temporal, as variáveis derivadas alvo foram construídas com base em informações passadas. Para isso, os valores da variável `valor_empenhado` a partir da data de corte foram temporariamente mascarados durante a etapa de engenharia de atributos, impedindo que observações futuras contaminassem o conjunto de treinamento. A partir desse ajuste, foram geradas defasagens temporais, médias móveis, somas móveis, desvios-padrão móveis, diferenças sazonais e acumulados anuais deslocados, sempre com referência ao período anterior ao mês previsto.

A engenharia de atributos contemplou diferentes grupos de variáveis, o primeiro grupo reuniu atributos de sazonalidade e tendência, incluindo mês, trimestre, semestre, indicadores de início e fim de ano, tendência linear, tendência quadrática e termos de Fourier. O segundo grupo foi composto por atributos autorregressivos da variável-alvo, tais como defasagens de 1, 2, 3, 6, 12 e 24 meses, médias móveis, somas móveis, desvios-padrão móveis e diferenças em relação ao período imediatamente anterior e ao mesmo mês do ano anterior. Adicionalmente, foram construídas variáveis acumuladas no ano (year-to-date), permitindo capturar a progressão intra-anual da execução orçamentária.

A sazonalidade da série foi representada, adicionalmente, por componentes de Fourier, construídos a partir de funções trigonométricas seno e cosseno aplicadas à periodicidade mensal da base. Esse procedimento foi adotado para captar padrões sazonais

recorrentes de forma contínua e flexível.

Em termos práticos, para cada observação mensal t , foram gerados pares de variáveis do tipo seno e cosseno considerando a periodicidade de 12 meses, correspondente ao ciclo anual da série. Foram utilizadas quatro frequências harmônicas ($k = 1, 2, 3, 4$), de modo que cada componente foi calculado pelas seguintes expressões.

$$\sin\left(\frac{2\pi k \cdot m}{12}\right) \quad (3.1)$$

$$\cos\left(\frac{2\pi k \cdot m}{12}\right) \quad (3.2)$$

em que m representa o mês da observação e k indica a frequência harmônica considerada.

A inclusão desses termos permitiu representar oscilações sazonais com diferentes intensidades e formatos ao longo do ano. A primeira harmônica capta movimentos sazonais mais amplos e suaves, enquanto as harmônicas seguintes permitem modelar variações mais complexas e menos lineares no interior do ciclo anual.

Assim, os termos de Fourier funcionam como uma forma parsimoniosa de incorporar sazonalidade à modelagem, permitindo que o modelo aprenda padrões cíclicos sem a necessidade de criar excessivo número de parâmetros independentes para cada mês.

No caso das variáveis exógenas contínuas, como IPCA, INCC, dias úteis, salário mínimo, taxa Selic e câmbio mensal, foram geradas defasagens, médias móveis e diferenças temporais, com o objetivo de incorporar efeitos contemporâneos retardados e padrões de curto e médio prazo. Também foram construídas interações entre variáveis institucionais e atributos temporais, a exemplo de interações com o quarto trimestre e com eventos específicos, como pandemia e calendário letivo.

Após a etapa de pré-tratamento e consolidação dos dados, procedeu-se à realização de uma exploração dos dados com o objetivo de compreender o comportamento das despesas discricionárias ao longo do tempo. Considerando que os dados estavam estruturados por elemento de despesa, realizou-se inicialmente o somatório mensal de todos os elementos, agregando-os por mês e ano, de modo a construir uma série temporal representativa do total empenhado mensalmente. Além disso, foram calculadas estatísticas descritivas fundamentais da série, permitindo sintetizar quantitativamente as características da distribuição de dados.

As métricas calculadas foram a média, o desvio padrão, a amplitude e o coeficiente de variação. A média foi utilizada para representar o valor médio mensal das despesas discricionárias ao longo do período analisado, funcionando como uma medida de tendência central da série. O desvio padrão foi calculado para mensurar o grau de dispersão dos valores em relação à média, permitindo avaliar o quanto os valores mensais se afastam

do comportamento médio da série. Ademais, foi calculada a amplitude, definida como a diferença entre o maior e o menor valor observado na série temporal.

3.2 Modelo de Previsão

Nesta seção, detalham-se os quatro algoritmos principais utilizados para a geração das previsões: Random Forest, XGBoost, ARIMAX e SARIMAX. A abordagem combina modelos de aprendizado de máquina baseados em árvores com modelos econométricos clássicos de séries temporais.

Nos modelos clássicos, utilizou-se uma abordagem de Seleção Automatizada de Ordens, em uma classe que integra pré-processamento e seleção de variáveis exógenas.

3.2.1 Modelos Clássicos

3.2.1.1 Modelo ARIMA com Variáveis Exógenas (ARIMAX)

O modelo ARIMA foi implementado em sua forma estendida com variáveis exógenas (ARIMAX), com o objetivo de capturar tanto a dinâmica temporal interna da série quanto os efeitos de variáveis externas potencialmente relevantes para a explicação das despesas discricionárias.

Diferentemente de uma especificação puramente univariada, o modelo adotado permite incorporar variáveis explicativas adicionais, ampliando sua capacidade de representação do processo gerador dos dados. No entanto, para evitar problemas de sobreparametrização e inclusão de variáveis irrelevantes, foi implementado um procedimento automatizado de seleção de variáveis exógenas.

Inicialmente, as variáveis candidatas foram avaliadas com base na correlação de Spearman em relação à variável dependente, sendo selecionadas apenas aquelas que apresentaram correlação mínima de 0,15. Além disso, foi imposto um limite máximo de oito variáveis exógenas, de modo a controlar a complexidade do modelo e evitar instabilidade na estimação. Após a seleção, as variáveis foram padronizadas por meio de transformação *z-score*, garantindo comparabilidade de escala e estabilidade numérica no processo de estimação.

No que se refere à variável dependente, foi aplicada uma transformação de winsorização no conjunto de treino, com parâmetro $k = 2,5$, com o objetivo de reduzir a influência de valores extremos. Esse procedimento foi fundamental para evitar que picos atípicos de empenho distorcessem a identificação da estrutura autorregressiva e de médias móveis do modelo.

A estimação dos parâmetros do ARIMA foi realizada por meio de um procedimento automatizado de busca, utilizando a função *auto_arima*. Nesse processo, foram testadas

diferentes combinações das ordens (p, d, q) dentro dos limites estabelecidos ($p \leq 5$, $q \leq 5$, $d \leq 2$), sendo selecionada a especificação que minimizou o Critério de Informação de Akaike (AIC). A busca foi conduzida pelo algoritmo *stepwise*, que explora o espaço de parâmetros de forma eficiente, reduzindo o custo computacional sem comprometer a qualidade da seleção.

Destaca-se que o modelo foi estimado considerando apenas componentes não sazonais (ARIMA), uma vez que a sazonalidade da série foi tratada separadamente por meio da inclusão de componentes de Fourier no conjunto de variáveis explicativas. Essa estratégia permitiu maior flexibilidade na modelagem sazonal, evitando a rigidez da especificação sazonal paramétrica tradicional.

Durante o treinamento, o modelo foi ajustado com dados do conjunto de treino, respeitando a ordenação temporal da série e evitando vazamento de informação. Na etapa de previsão, as variáveis exógenas foram transformadas utilizando os parâmetros previamente estimados no treino, garantindo consistência entre as etapas.

Adicionalmente, foi aplicado um mecanismo de controle de valores extremos nas previsões, limitando os resultados a um intervalo definido com base na média e no desvio padrão da série de treino. Esse procedimento teve como objetivo evitar projeções implausíveis, especialmente em contextos de elevada volatilidade.

3.2.1.2 Modelo SARIMAX

O modelo *SARIMAX* considerou parâmetros não sazonais (p, d, q) e sazonais (P, D, Q) , definidos com periodicidade $m = 12$, compatível com a frequência mensal dos dados e com a hipótese de sazonalidade anual. Os limites máximos estabelecidos para a busca automática foram $p, q \leq 3$, $d \leq 2$, $P, Q \leq 2$ e $D \leq 1$. A escolha do modelo final foi realizada com base no Critério de Informação de Akaike (AIC), por meio do algoritmo *stepwise*, utilizando a função *auto_arima*.

Além da estrutura temporal e sazonal, o modelo permitiu a inclusão de variáveis exógenas. Entretanto, para evitar sobreparametrização, foi realizado um procedimento prévio de seleção de variáveis com base na correlação de Spearman com a variável dependente. Foram mantidas apenas as covariáveis com correlação mínima de 0,15, respeitando um limite máximo de seis variáveis exógenas. Após essa seleção, as variáveis escolhidas foram padronizadas por meio de transformação *z-score*, de modo a melhorar a estabilidade numérica da estimação.

Antes do ajuste do modelo, a série dependente foi submetida à winsorização, com parâmetro $k = 2,5$, aplicada apenas ao conjunto de treino. Esse procedimento teve como objetivo reduzir a influência de valores extremos, preservando a estrutura geral da série sem eliminar observações. Em séries de execução orçamentária, é relevante esse processo, pois meses com picos atípicos de empenho podem distorcer a estimação dos parâmetros

autorregressivos, das médias móveis e da componente sazonal.

3.2.2 Modelos Baseados em Árvores

Além dos modelos estatísticos da família ARIMA, foram avaliadas abordagens baseadas em aprendizado de máquina, em especial os algoritmos Random Forest e XGBoost, com o objetivo de capturar relações não lineares e interações complexas entre as variáveis explicativas e o comportamento das despesas discricionárias da UFPB.

A execução de despesas públicas não depende apenas de padrões temporais passados, mas também de fatores institucionais, operacionais e macroeconômicos que interagem entre si de forma não linear. Por exemplo, variações no número de beneficiários, mudanças no calendário acadêmico e condições macroeconômicas podem afetar simultaneamente o nível de empenho, gerando dinâmicas que dificilmente são capturadas por modelos lineares tradicionais.

Nesse contexto, os modelos baseados em árvores foram treinados com o objetivo de aprender diretamente essas relações a partir dos dados, sem impor previamente uma forma funcional específica.

3.2.2.1 Random Forest

O modelo Random Forest foi utilizado como uma abordagem baseada em agregação (*bagging*), na qual múltiplas árvores de decisão são treinadas de forma independente a partir de subconjuntos dos dados. Cada árvore aprende diferentes padrões presentes na série, e a previsão final é obtida pela média das previsões individuais.

No contexto deste estudo, essa característica é particularmente relevante, pois a execução orçamentária pode apresentar elevada variabilidade mensal, com comportamentos distintos ao longo do ano. A agregação de múltiplas árvores contribui para reduzir a sensibilidade do modelo a oscilações específicas de determinados períodos, tornando as previsões mais estáveis.

O treinamento do Random Forest foi conduzido por meio de uma busca em grade (*grid search*) sobre um conjunto pré-definido de hiperparâmetros. Foram avaliadas combinações dos parâmetros $n_estimators$, max_depth e $min_samples_leaf$, considerando os valores $n_estimators \in \{100, 200\}$, $max_depth \in \{3, 5, 8\}$ e $min_samples_leaf \in \{2, 4, 8\}$.

O parâmetro $n_estimators$, que define o número de árvores do conjunto, foi calibrado com o objetivo de aumentar a estabilidade das previsões por meio da agregação de múltiplos modelos. Já o parâmetro max_depth foi utilizado para controlar a complexidade das árvores, limitando sua profundidade e evitando estruturas excessivamente ajustadas aos dados de treino. De forma complementar, o parâmetro $min_samples_leaf$ impõe um número mínimo de observações em cada nó terminal, restringindo a criação de partições muito específicas.

No contexto da previsão de despesas discricionárias, essa calibração assume papel importante, uma vez que a série apresenta episódios de elevada volatilidade, incluindo picos de empenho associados a eventos administrativos e operacionais. A combinação de profundidade controlada e restrição no tamanho dos nós terminais contribui para evitar que o modelo memorize esses comportamentos atípicos, favorecendo a captura de padrões estruturais mais robustos da execução orçamentária.

3.2.2.2 XGBoost

O modelo XGBoost foi utilizado como uma abordagem baseada em *boosting*, na qual as árvores são construídas sequencialmente com o objetivo de corrigir os erros das previsões anteriores. Diferentemente do Random Forest, que busca reduzir a variância, o XGBoost atua principalmente na redução do viés, ajustando progressivamente os resíduos do modelo.

Essa característica torna o XGBoost especialmente adequado para o problema analisado, uma vez que a execução de despesas discricionárias pode ser influenciada por múltiplos fatores simultâneos, cujos efeitos não são homogêneos ao longo do tempo. Ao aprender iterativamente com os erros, o modelo é capaz de capturar padrões mais sutis, como mudanças no comportamento da série em períodos específicos, transições entre regimes de execução orçamentária e efeitos combinados de variáveis institucionais e macroeconômicas.

O treinamento do modelo XGBoost foi realizado por meio de uma busca em grade (*grid search*), com o objetivo de identificar a combinação de hiperparâmetros que melhor se ajustasse à dinâmica das despesas discricionárias. Foram avaliadas diferentes combinações dos parâmetros $n_estimators$, $learning_rate$, max_depth e min_child_weight , considerando os valores $n_estimators \in \{50, 100, 200\}$, $learning_rate \in \{0,01, 0,05, 0,1\}$, $max_depth \in \{2, 3, 5\}$ e $min_child_weight \in \{3, 5, 10\}$.

O parâmetro $n_estimators$ define o número de árvores construídas sequencialmente, sendo responsável pela capacidade do modelo de refinar progressivamente suas previsões. Já o $learning_rate$ controla a contribuição de cada nova árvore ao modelo final, permitindo um aprendizado mais gradual e reduzindo o risco de sobreajuste. Valores menores desse parâmetro tornam o modelo mais conservador, exigindo maior número de árvores para alcançar bom desempenho.

O parâmetro max_depth foi utilizado para limitar a complexidade das árvores individuais, evitando estruturas excessivamente profundas que poderiam capturar ruídos específicos da amostra. Por sua vez, o min_child_weight atua como mecanismo de regularização, exigindo um número mínimo de observações ponderadas para a criação de novos nós, o que torna o modelo mais robusto a variações pontuais dos dados.

A inclusão dos modelos Random Forest e XGBoost teve como objetivo avaliar a

capacidade de métodos orientados por dados em capturar a complexidade da dinâmica das despesas discricionárias da UFPB. Enquanto os modelos da família ARIMA se baseiam em estrutura temporal explícita e pressupostos de linearidade, os modelos baseados em árvores permitem explorar relações não lineares e interações entre variáveis, oferecendo uma perspectiva complementar de modelagem.

3.3 Métricas de Avaliação e Validação dos Modelos

No interior do conjunto de treinamento, a seleção e a comparação entre especificações foram realizadas por meio de validação cruzada temporal do tipo walk-forward. Nessa estratégia, cada modelo foi sucessivamente reestimado com base apenas em informações passadas e validado em um bloco temporal subsequente. Como o conjunto de treinamento é progressivamente ampliado a cada iteração, o procedimento adotado caracteriza-se como validação com janela expansiva (*expanding window*) (MOLNAR, 2020).

A adoção desse procedimento é importante porque, em problemas preditivos com dados temporais, a utilização de validação aleatória convencional pode gerar estimativas de desempenho artificialmente otimistas, ao violar a estrutura cronológica da série. Ao empregar validação temporal com origem móvel, assegura-se que cada previsão seja produzida apenas com base em informações efetivamente disponíveis até o período imediatamente anterior ao bloco validado.

Em cada partição da validação, o desempenho dos modelos foi mensurado por meio do Erro Absoluto Médio (MAE), e a média dos erros observados nas diferentes iterações foi utilizada como critério de comparação interna entre especificações. Após a escolha dos modelos mais adequados com base nessa etapa de validação, procedeu-se à estimação final no conjunto completo de treinamento e à avaliação no conjunto de teste correspondente ao ano de 2025.

A avaliação do desempenho dos modelos preditivos foi realizada por meio de um conjunto de métricas complementares, com o objetivo de capturar diferentes dimensões do erro de previsão. Especificamente, foram utilizadas as métricas Erro Absoluto Médio (MAE), Raiz do Erro Quadrático Médio (RMSE), Erro Percentual Absoluto Médio Simétrico (sMAPE) e Erro Absoluto Escalonado Médio (MASE).

O MAE foi utilizado como medida direta da magnitude média dos erros, sendo particularmente relevante no contexto deste estudo por expressar o erro em termos monetários, permitindo interpretar o desvio médio das previsões em relação aos valores efetivamente empenhados. Sua formulação é dada por:

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (3.3)$$

em que y_t representa o valor observado da despesa empenhada no período t , \hat{y}_t

corresponde ao valor previsto pelo modelo, e n indica o número total de observações avaliadas.

O RMSE, por sua vez, foi empregado como uma métrica sensível a erros de maior magnitude, penalizando de forma mais intensa discrepâncias elevadas, o que é importante em séries que apresentam picos de execução orçamentária.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (3.4)$$

Diferentemente do MAE, que atribui peso linear aos erros, o RMSE eleva os desvios ao quadrado antes de calcular a média, tornando-se mais sensível a erros extremos.

No contexto da execução orçamentária, essa característica é relevante porque grandes desvios de previsão podem comprometer a capacidade de antecipação da gestão diante de meses com maior concentração de empenhos.

Adicionalmente, foi utilizado o sMAPE, que permite avaliar o erro em termos relativos, sendo menos sensível à escala da variável e adequado para comparação entre períodos com diferentes níveis de despesa. Essa métrica é especialmente útil no contexto de despesas públicas, nas quais a magnitude dos valores pode variar significativamente ao longo do tempo. A métrica é definida por:

$$sMAPE = \frac{100}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{(|y_t| + |\hat{y}_t|) / 2} \quad (3.5)$$

Essa métrica permite interpretar o erro médio em termos percentuais, favorecendo a comparação do desempenho preditivo em diferentes períodos da série, mesmo quando há variações relevantes no nível de execução das despesas.

Por fim, foi incorporado o MASE, que avalia o desempenho do modelo em relação a uma previsão ingênua baseada na sazonalidade da série.

$$MASE = \frac{\frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|}{\frac{1}{T-1} \sum_{t=2}^T |y_t - y_{t-1}|} \quad (3.6)$$

em que o numerador representa o erro absoluto médio do modelo avaliado, enquanto o denominador corresponde ao erro médio de uma previsão ingênua calculada no conjunto de treinamento. Além disso, os valores de MASE inferiores a 1 indicam que o modelo apresenta desempenho superior ao método ingênuo, ao passo que valores superiores a 1 sugerem desempenho inferior à regra de referência.

Nesse caso, foi adotada uma referência sazonal com defasagem de 12 períodos, compatível com a frequência mensal dos dados. O MASE permite interpretar se o modelo apresenta ganho efetivo de desempenho em relação a uma abordagem simples baseada na repetição do padrão anual.

3.4 Visualização dos Dados

Após o processo de estimação dos modelos, procedeu-se à etapa de visualização interativa dos resultados por meio do desenvolvimento de uma aplicação analítica em Streamlit. Essa aplicação foi concebida como parte integrante do desenho metodológico, com o objetivo de organizar, explorar e comunicar os resultados da modelagem de forma estruturada, dinâmica e acessível.

A aplicação foi estruturada em quatro abas principais, cada uma com uma função analítica específica. A primeira aba e segunda aba, denominadas visão geral das despesas, foi desenvolvida com o propósito de apresentar uma caracterização descritiva da execução orçamentária da universidade, permitindo compreender o comportamento histórico das despesas discricionárias. Nessa seção, são exibidos indicadores agregados, como média mensal, valor total no período, desvio padrão e valores mínimos e máximos, além da possibilidade de exploração de diferentes elementos de despesa. Essa visão inicial permite contextualizar o fenômeno analisado, evidenciando padrões de volatilidade, concentração e amplitude da execução orçamentária ao longo do tempo.

A terceira e quarta aba foi dedicada aos resultados dos modelos preditivos, reunindo os principais produtos da modelagem. Nessa seção, são apresentados a série histórica observada, os valores previstos pelos modelos estimados e a comparação entre valores realizados e projetados. Adicionalmente, são disponibilizados os indicadores de desempenho utilizados na avaliação dos algoritmos, tais como MAE, RMSE, sMAPE e MASE, permitindo uma análise detalhada da capacidade preditiva das diferentes abordagens testadas.

Do ponto de vista analítico, a aplicação em Streamlit foi concebida como um instrumento de apoio à interpretação dos resultados, possibilitando a organização de gráficos, tabelas e painéis comparativos de forma interativa. A ferramenta também permite a aplicação de filtros e a segmentação das análises, facilitando a comparação entre diferentes especificações dos modelos e a avaliação do impacto de distintos conjuntos de variáveis explicativas.

A inserção dessa etapa no desenho metodológico justifica-se pela necessidade de ampliar a interpretabilidade dos resultados, permitindo uma leitura mais clara tanto do comportamento temporal da despesa quanto do desempenho dos modelos preditivos. Dessa forma, o uso do Streamlit não se limita a um recurso de apresentação, mas constitui uma ferramenta analítica que contribui para a compreensão dos achados e para a potencial aplicação prática do modelo desenvolvido no apoio à tomada de decisão.

4 Resultados

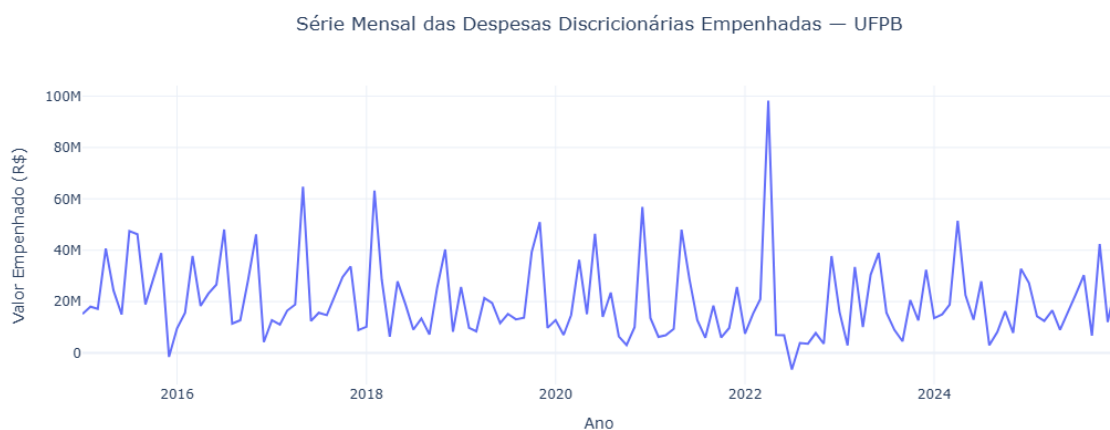
A apresentação dos resultados foi dividida em duas fases, a primeira, consiste na análise descritiva da série histórica de despesas discricionárias da UFPB, com ênfase na evolução do valor empenhado, identificação de sazonalidade e detecção de oscilações. Na sequência, o estudo consiste na avaliação dos resultados dos modelos preditivos estimados.

Diferentemente de uma abordagem técnica, a comparação entre modelos foi conduzida de modo a extrair implicações substantivas sobre o processo orçamentário investigado. Assim, mais do que identificar o menor erro do modelo, buscou-se examinar quais conjuntos de variáveis efetivamente agregam capacidade preditiva e o que isso revela sobre os determinantes da execução das despesas discricionárias no âmbito da UFPB.

4.1 Análise Descritiva

A Figura 2 representa a análise inicial descritiva realizada indicando que o comportamento mensal dos gastos públicos discricionários empenhados pela UFPB (2015-2025) é caracterizado por elevada volatilidade. Ao longo do horizonte analisado, nota-se a ausência de uma tendência linear definida, predominando fortes oscilações entre os meses. Os registros oscilam de valores abaixo de R\$ - 5 milhões a picos superiores a R\$ 90 milhões.

Figura 2 – Série Mensal das Despesas Empenhadas (2015-2025)



Fonte: Elaboração própria (2026).

Complementando a análise gráfica da série, a Tabela 2 apresenta as estatísticas descritivas do valor empenhado mensal das despesas discricionárias da UFPB, a média mensal do empenho é de aproximadamente R\$ 20,36 milhões, porém a medida deve ser

analisada com critérios, uma vez que a elevada dispersão dos dados indica que a média não reflete o comportamento adequado da série em diversos momentos.

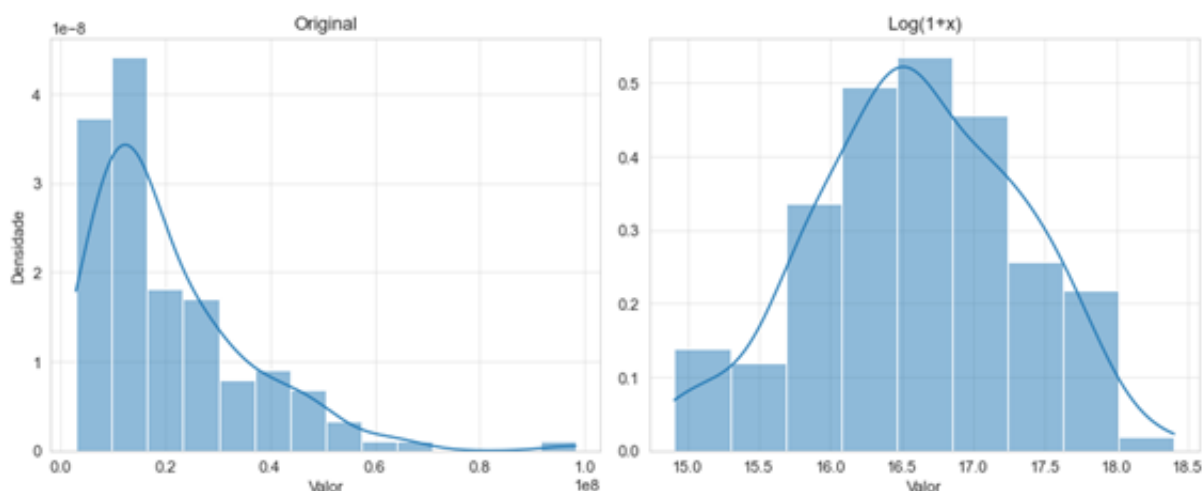
Tabela 2 – Análise descritiva dos dados

Métrica	Valor
Média	R\$ 20,36 mi
Desvio Padrão	R\$ 15,51 mi
Valor Mínimo	R\$ -6,43 mi
Valor Máximo	R\$ 98,25 mi
Amplitude	R\$ 104,67 mi
Coef de Variação (%)	76,17 %

Fonte: Elaboração própria (2026).

O desvio padrão de R\$ 15,51 milhões indica uma elevada dispersão dos valores em relação à média, enquanto o coeficiente de variação de 76,17% reflete a instabilidade observada, evidenciando uma dispersão equivalente a aproximadamente três quartos do valor médio da série. Sob a perspectiva estatística descritiva, esse resultado mostra um comportamento heterogêneo, caracterizado por alta variabilidade. Ademais, destaca-se a amplitude expressiva de R\$ 104,67 milhões, resultante da diferença entre o valor máximo (R\$ 98,25 milhões) e o mínimo (R\$ -6,43 milhões), o que reforça altas oscilações. A presença de valores negativos nos empenhos indica a ocorrência de ajustes contábeis, como cancelamentos ou estornos de registros previamente realizados.

Figura 3 – Distribuição da variável alvo



Fonte: Elaboração própria (2026).

Conforme apresentando na Figura 3 a distribuição dos dados apresenta uma forte assimetria à direita. Isso significa que a maior parte dos registros se concentra em valores mais baixos de empenho, enquanto uma cauda longa se estende em direção aos

gastos extremos. Estatisticamente, essa assimetria indica que os dados não seguem uma distribuição normal.

A aplicação da transformação logarítmica, contudo, muda esse cenário uma vez que, como mostra o painel à direita, Figura 3, a distribuição torna-se simétrica e se aproxima da normalidade. Na prática, o logaritmo reduz o peso dos valores extremos, encurta a cauda direita e estabiliza a variância dos dados. Portanto, a Figura 3 confirma que a dinâmica das despesas discricionárias da UFPB é caracterizada por assimetria e concentração em valores mais baixos, intercalados por picos expressivos.

4.2 Resultados Modelos de Previsão

Foram testados diferentes modelos, como os clássicos ARIMAX e SARIMAX e algoritmos de machine learning, como o Random Forest e o XGBoost. É possível observar que o desempenho preditivo variou de forma consistente entre os modelos analisados. Abaixo, verificam-se de forma consolidada, os resultados dos modelos testados na Tabela 3:

Tabela 3 – Resultados Modelos Testados

Modelo	MAE (R\$)	RMSE (R\$)	sMAPE (%)	MASE
SARIMAX	R\$ 7.91 M	R\$ 9.97 M	43.27%	0.55
ARIMAX	R\$ 8.06 M	R\$ 10,08 M	44.50 %	0.56
Random Forest	R\$ 7.99 M	R\$ 10.67 M	41.63%	0.56
XGboost	R\$ 8.28 M	R\$ 11.22 M	43.08 %	0.57

Fonte: Elaboração própria (2026).

Os resultados apresentados na Tabela 3 mostram que o modelo SARIMAX foi o que melhor desempenhou no geral entre os modelos analisados, com um MAE de R\$ 7.91 mi, o que significa que, em média, as previsões ficam cerca de R\$ 7,91 milhões acima ou abaixo do valor real das despesas. Em termos práticos, quanto menor esse valor, melhor o modelo. Portanto, entre os modelos testados, o SARIMAX foi o que mais se aproximou dos valores observados.

Em relação ao RMSE de R\$ 9.97 milhões, é uma métrica que penaliza mais fortemente desvios elevados. Assim, além de apresentar erro médio menor, o SARIMAX também demonstrou maior capacidade de lidar com meses mais atípicos, nos quais a despesa foge do comportamento habitual.

Já o MASE de 0,55 mostra que o modelo teve desempenho melhor do que uma previsão ingênua. Em geral, o MASE é considerado aceitável quando é menor que 1. Valores abaixo de 1 indicam que o modelo supera a referência básica, valores próximos de 0,5, como nesse trabalho, podem ser interpretados como bons resultados, pois mostram que o erro do modelo foi consideravelmente menor do que o erro de uma previsão simples.

No que se refere ao MASE em geral, todos os modelos permaneceram abaixo de 1, com valores entre 0.55 e 0.57, o que mostra um desempenho superior ao de uma previsão ingênua como referência de comparação. Esses resultados indicam, portanto, um quadro de equilíbrio competitivo, no qual os modelos tradicionais de séries temporais foram ligeiramente mais eficazes em erros absolutos, ao passo que o Random Forest demonstrou pequena vantagem na métrica percentual.

Do ponto de vista analítico, esses resultados são coerentes com a natureza da variável estudada. O objetivo da pesquisa não é prever uma despesa estável, regida exclusivamente por tendências regulares e previsíveis, mas sim uma série fortemente condicionada por fatores institucionais, orçamentários e administrativos. A execução das despesas discricionárias em universidades federais depende da interação entre necessidades internas da instituição e condicionantes externos associados ao regime fiscal, à liberação orçamentária e ao processo decisório do setor público.

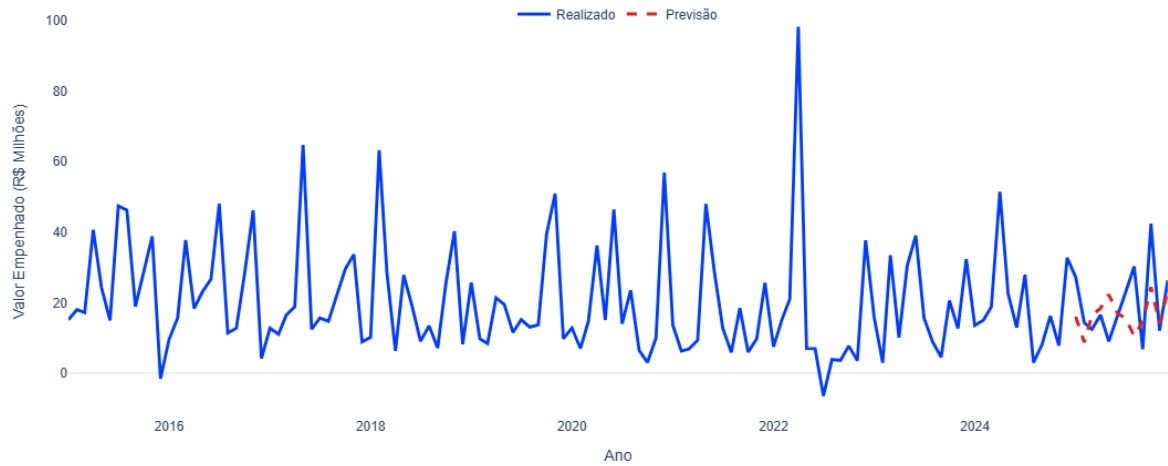
Além disso, é importante ressaltar que parte da variação observada decorre de um componente decisório que não é capturado pelos modelos. A alocação dos recursos envolve escolhas da gestão quanto à priorização de despesas, ao momento da execução e à distribuição interna dos limites disponíveis. Como essas decisões dependem do fator de escolha do gestor, que nem sempre se convertem em variáveis mensuráveis, sua influência permanece apenas refletida nos dados históricos utilizados para estimação. Desse modo, parte do erro de previsão também pode refletir decisões da própria gestão sobre como e quando os recursos são alocados ao longo do exercício.

Ademais, complementar à comparação entre modelos, foi observado e a importância relativa das variáveis. Os resultados indicaram maior relevância para `cambio_mensal` (14,4%), `incc` (10,3%), `beneficiarios_assistencia` (9,7%), `valor_assistencia` (8,7%) e `valor_ensino` (8,2%). Em conjunto, essas evidências sugerem que a previsão das despesas discricionárias depende não apenas da trajetória histórica da série, mas também de fatores macroeconômicos, do custo real do orçamento e de variáveis associadas à atividade acadêmica e à política de assistência estudantil.

Desse modo, os achados da importância das variáveis reforçam, em caráter exploratório, a hipótese de que a execução dessas despesas é influenciada por múltiplas dimensões, nas quais se articulam condicionantes econômicos externos e pressões institucionais internas. Ainda que tais importâncias não devam ser interpretadas como relações causais, elas oferecem indícios relevantes de que a dinâmica orçamentária da UFPB é sensível tanto aos ajustadores macroeconômicos quanto aos aceleradores vinculados à missão universitária.

Como ilustrado na Figura 4, apresenta a comparação entre os valores realizados e os previstos pelo modelo SARIMAX para as despesas discricionárias da UFPB. Em geral, o modelo capturou bem o comportamento da série, mas apresentou dificuldades em acompanhar os picos. Devido à forte sazonalidade das despesas, o modelo tende a suavizar

Figura 4 – Previsão Modelo SARIMAX (2025)



Fonte: Elaboração própria (2026).

os valores extremos de execução.

4.3 Visualização Analítica

Com o objetivo de ampliar a interpretabilidade dos resultados e proporcionar uma visão dos dados analisados, foi desenvolvida uma aplicação analítica interativa utilizando a biblioteca Streamlit.

Figura 5 – Visão Elemento de Despesa



Fonte: Elaboração própria (2026).

A ferramenta foi feita como um painel de apoio à análise exploratória e à validação dos resultados do modelo preditivo, permitindo a visualização dinâmica da execução das

despesas discricionárias da UFPB.

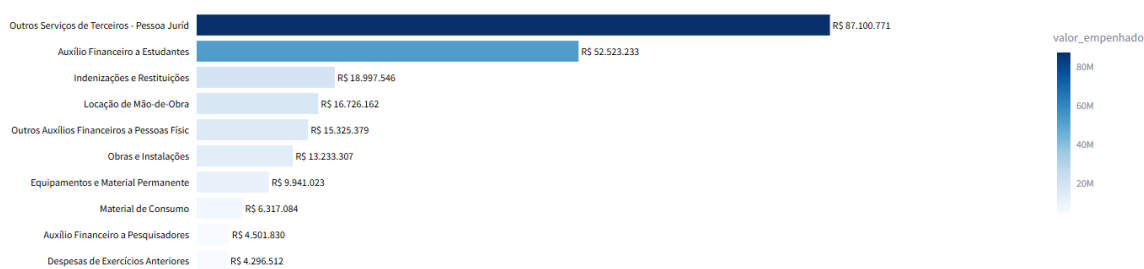
Por meio do dashboard, é possível examinar a distribuição temporal dos gastos, bem como realizar filtros por elementos específicos de despesa, possibilitando identificar padrões mensais, sazonalidades e comportamentos atípicos ao longo do período analisado.

Na Figura 6, é possível analisar também a evolução mensal do empenho "Auxílio Financeiro a Estudantes", evidenciando a variabilidade dos valores empenhados ao longo dos meses de 2025. Além disso, o painel apresenta estatísticas descritivas das despesas empenhadas, como média mensal, total executado no período, desvio padrão e valores mínimos e máximos.

Figura 6 – Visão - Elemento de Despesas II



Top 10 Maiores Despesas no Período Selecionado



Fonte: Elaboração própria (2026).

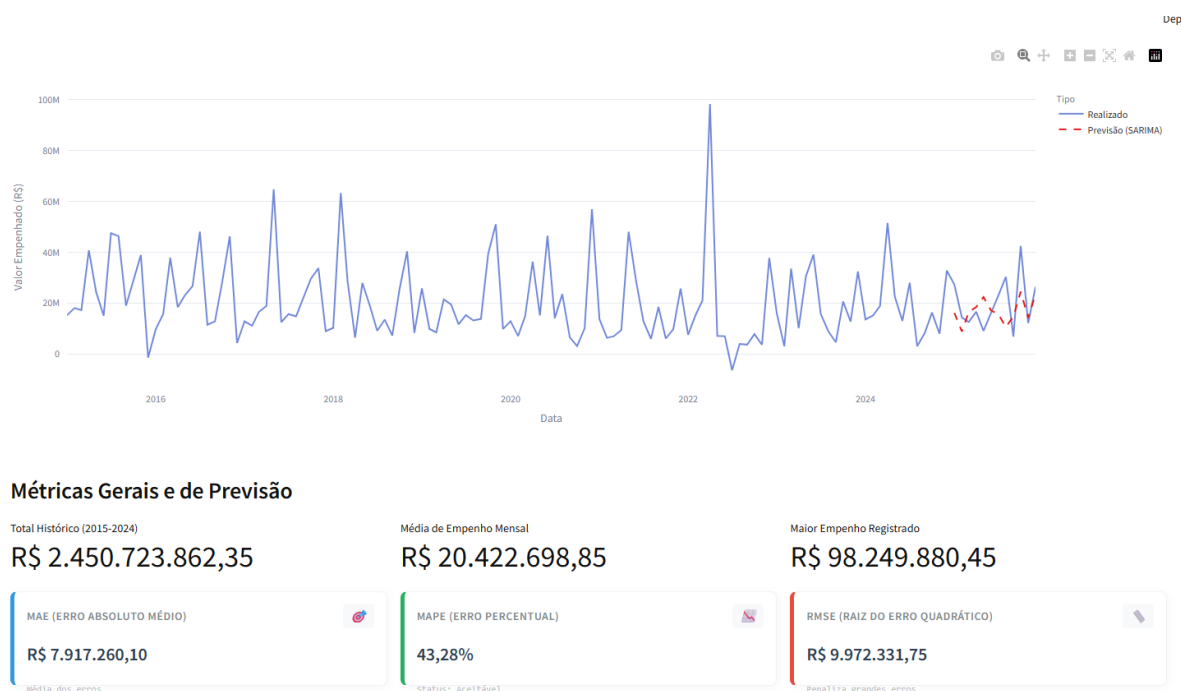
Complementarmente, foi incluída uma visualização dos principais elementos de despesa, por meio de um ranking dos maiores gastos no período selecionado, permitindo identificar quais categorias concentram maior volume de recursos.

Adicionalmente com o intuito de facilitar a análise dos resultados do modelo preditivo, foi desenvolvida uma visualização que integra a série histórica dos valores realizados com as previsões geradas pelo modelo SARIMAX para o período de teste.

A interpretação dessa ilustração permite avaliar, de forma intuitiva, o desempenho do modelo ,visão descritiva e de métricas de previsão, uma vez que a proximidade entre as duas linhas indica maior capacidade preditiva. Observa-se que o modelo consegue acompanhar o comportamento médio da série, evidenciando sua capacidade de capturar a estrutura temporal dos dados.

De forma geral, os resultados indicam que o modelo SARIMAX é eficaz na captura

Figura 7 – Visão - Previsão do Modelo Dashboard



Fonte: Elaboração própria (2026).

da tendência central e da estrutura temporal da série, embora apresente limitações na previsão de eventos extremos. Ainda assim, o modelo se mostra útil como instrumento de apoio à análise e ao planejamento das despesas públicas.

4.4 Implicações para a Gestão Orçamentária

Os resultados deste estudo mostram que a gestão orçamentária da UFPB não deve continuar apoiada exclusivamente em projeções baseadas na média dos três últimos anos acrescidas da correção da inflação. Em um ambiente marcado por volatilidade, contingenciamentos e forte assimetria entre necessidade institucional e capacidade efetiva de execução, o uso de um modelo preditivo e de um dashboard analítico desloca o orçamento de uma lógica reativa para uma lógica gerencial mais preventiva, comparativa e orientada por evidências.

O fato de o modelo SARIMAX ter apresentado o melhor desempenho geral, ainda que com limitações para captar picos extremos, mostra que a previsão não deve ser lida como promessa de exatidão, mas como instrumento de apoio à decisão em contexto de incerteza. Em termos administrativos, isso significa que o valor previsto passa a funcionar como referência técnica para organizar prioridades, testar a coerência do ritmo de execução e apoiar decisões de curto prazo ao longo do exercício.

O gestor deve utilizar o dashboard e as previsões para decidir antes que o problema

apareça no encerramento do exercício. Na prática, isso implica comparar, no início de cada mês, quatro dimensões: o valor previsto de empenho, o valor efetivamente executado, o limite orçamentário-financeiro disponível e a composição da despesa por elemento. Se a previsão indicar pressão de execução acima da capacidade real de empenho, a resposta não deve ser aguardar a insuficiência se materializar, mas reprogramar o orçamento, revisar cronogramas de contratação, escalonar despesas adiáveis e proteger, com prioridade, os gastos discricionários de manutenção obrigatória.

Se, ao contrário, a previsão apontar espaço de execução inferior ao limite disponível, a gestão deve aproveitar essa margem para acelerar despesas estratégicas represadas, reduzir concentração de empenhos no final do ano e melhorar a qualidade do gasto. O uso do modelo, portanto, não é meramente informativo, ele deve induzir ação administrativa concreta. No plano do monitoramento da execução, o dashboard deve ser incorporado à rotina de acompanhamento da PROPLAN e das áreas responsáveis como ferramenta de gestão mensal.

Como o painel permite filtrar elementos de despesa, observar a distribuição temporal dos gastos, acompanhar estatísticas descritivas e comparar valores realizados com valores previstos, ele oferece uma base concreta para identificar desvios relevantes, mudanças de padrão e concentrações anormais de execução. Por isso, a contribuição mais importante do trabalho para a gestão não está em trazer previsão exata, mas em oferecer uma tecnologia de gestão capaz de organizar informação, disciplinar o acompanhamento da execução e trazer racionalidade ao processo de replanejamento orçamentário.

5 Considerações finais

Esse trabalho de conclusão de curso alcançou o objetivo geral ao analisar a dinâmica entre a demanda por recursos e a restrição orçamentária na UFPB e desenvolver um modelo preditivo capaz de traduzir essa relação em estimativas ao planejamento institucional.

Nesse contexto, a pesquisa procurou responder à questão de como estruturar um modelo preditivo capaz de integrar variáveis endógenas da demanda acadêmica e variáveis exógenas do cenário fiscal, de modo a aprimorar a previsão da execução das despesas discricionárias da Universidade Federal da Paraíba e contribuir para o planejamento orçamentário institucional.

Inicialmente, foi analisada a dinâmica da execução das despesas discricionárias a partir da evolução da série mensal de empenhos. Em seguida, contextualizou-se o comportamento dessa série no período de 2015 a 2025, evidenciando oscilações associadas. Posteriormente, foi estruturada uma base de dados com informações orçamentárias, acadêmicas, institucionais e macroeconômicas.

Além disso, foram desenvolvidos e comparados diferentes modelos de previsão, o que possibilitou identificar a abordagem com melhor desempenho para estimar o valor mensal empenhado. Por fim, foi prototipado um dashboard analítico, concebido para apoiar a visualização das previsões e da composição das despesas, ampliando o potencial prático dos resultados obtidos.

A execução discricionária revelou-se como um processo condicionado por múltiplas determinações, no qual a demanda interna por recursos convive com limites exógenos impostos pelo ambiente fiscal, pela temporalidade da execução orçamentária e pelas regras institucionais que estruturam a gestão pública universitária.

Em termos analíticos, isso significa que a despesa efetivamente executada não expressa apenas aquilo de que a universidade necessita, mas aquilo que se torna viável dentro de um arranjo de restrições que extrapola a esfera decisória da própria instituição. Tal conclusão está em consonância com o modelo teórico adotado na pesquisa, fundado na interação entre aceleradores, freios e ajustadores, e com a própria formulação do objetivo geral do estudo.

Na pipeline desenvolvida, os modelos foram treinados com um conjunto de atributos, tendo o modelo SARIMAX apresentado o melhor resultado relativo entre as abordagens comparadas. Este resultado demonstra que a série das despesas discricionárias da UFPB é representada de forma mais eficiente por uma estrutura parcimoniosa, apoiada em sazonalidade e variáveis exógenas, do que por modelos complexos.

Ao mesmo tempo, a análise complementar da importância relativa das variáveis

indicou maior relevância preditiva para fatores como câmbio, inflação setorial e variáveis associadas à assistência e ao ensino, o que reforça, em caráter exploratório, a hipótese de que a execução da despesa resulta da interação entre condicionantes macroeconômicos e pressões institucionais vinculadas à missão universitária.

Assim, a principal contribuição do trabalho está em evidenciar empiricamente a assimetria entre necessidade e execução. A universidade amplia suas demandas em função de sua missão acadêmica, social e administrativa, mas sua capacidade de materializar essas demandas em despesa depende de um espaço orçamentário rigidamente condicionado por fatores externos ao seu controle direto.

Sob a perspectiva da gestão, essa conclusão possui implicações relevantes, o estudo mostra que o planejamento orçamentário universitário não pode permanecer restrito a mecanismos retrospectivos e simplificados de projeção, como o ajuste com a correção pela inflação, porque essa prática desconsidera a heterogeneidade temporal da execução, os efeitos de sazonalidade, os condicionantes institucionais e a própria instabilidade do cenário fiscal.

A relevância aplicada da pesquisa reside em oferecer fundamentos para a consolidação de uma cultura de gestão orientada por dados no âmbito universitário. Adicionalmente, mais do que produzir um modelo preditivo sobre a execução das despesas discricionárias, o estudo aponta para a necessidade de institucionalizar práticas de planejamento baseadas em evidências, capazes de qualificar a alocação de recursos, aprimorar o acompanhamento financeiro e ampliar a capacidade decisória da administração superior.

Nesse sentido, o dashboard prototipado e as visualizações analíticas construídas ao longo do trabalho representam a tradução dos resultados técnicos em informação gerencial mais acessível e útil à governança universitária, ao trazer, além de informações dos resultados dos modelos, visões sobre como estão distribuídas as despesas discricionárias na UFPB.

Isso, contudo, não afasta as limitações inerentes ao estudo, os resultados evidenciam que a previsão das despesas discricionárias é sensível, especialmente em períodos marcados por oscilações, choques institucionais e mudanças no ambiente macroeconômico. A principal limitação da pesquisa, portanto, não reside no escopo dos objetivos, mas na precisão restrita dos modelos diante da complexidade do fenômeno, e também no fato de que a execução das despesas discricionárias incorpora um componente decisório que não é observável na base de dados, o poder de escolha do gestor para alocação dos recursos, o que abre espaço para pesquisas futuras.

Para trabalhos futuros, sugere-se o aprofundamento do modelo mediante a incorporação de variáveis que representem com maior precisão tanto o espaço fiscal disponível quanto a margem de escolha do gestor na alocação das despesas discricionárias. Nessa direção, sugere-se incluir variáveis relacionadas ao orçamento anual autorizado da uni-

versidade, ao saldo orçamentário remanescente ao longo do exercício, ao percentual do orçamento anual já executado em cada mês e à participação de cada mês no total anual empenhado, de modo a captar a forma como o orçamento é distribuído temporalmente.

Adicionalmente, recomenda-se também a construção de uma base em painel com informações mensais de diferentes universidades federais. A incorporação de múltiplas instituições tende a ampliar substancialmente o volume de observações disponíveis, permitindo testar a validade externa do modelo e verificar em que medida os padrões identificados na UFPB se reproduzem em organizações com perfis institucionais, estruturas de gasto e dinâmicas de execução distintas.

Além disso, a integração do modelo aos sistemas institucionais de planejamento e acompanhamento orçamentário, permitindo a atualização contínua das previsões à medida que novos dados forem incorporados e oferecendo suporte mais dinâmico e aderente às necessidades da tomada de decisão gerencial.

Por fim, a aplicação entre ciência de dados e gestão orçamentária, conforme proposta neste estudo, representa um avanço relevante tanto do ponto de vista aplicado quanto teórico, contribuindo para o fortalecimento de práticas de planejamento mais eficientes, transparentes e orientadas por evidências no âmbito das universidades federais brasileiras.

Referências

- AMARAL, N. C. d. Financiamento da educação superior: impasses e perspectivas. *Educação & Sociedade*, Campinas, v. 39, n. 144, p. 1115–1136, 2018. Out./dez. 12, 18, 19
- ANDIFES. *Nota da ANDIFES sobre os cortes no orçamento aprovado pelo Congresso Nacional para as universidades federais*. 2025. Disponível em: <https://www.andifes.org.br/2025/12/23/nota-da-andifes-sobre-os-cortes-no-orcamento-aprovado-pelo-congresso-nacional-para-as-universidades-federais/>. Acesso em: 29 mar. 2026. 11, 21
- Banco Central do Brasil. *Taxa Selic*. 2023. Disponível em: <https://www.bcb.gov.br/controleinflacao/taxaselic>. Acesso em: 22 mar. 2026. 28
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. *Time series analysis: forecasting and control*. 5. ed. New Jersey: Wiley, 2016. 25
- BRASIL. *Lei nº 4.320, de 17 de março de 1964*. Brasília, DF: [s.n.], 1964. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/14320.htm. Estatui normas gerais de direito financeiro. Acesso em: 25 mar. 2026. 20
- BREIMAN, L. Random forests. *Machine Learning*, v. 45, n. 1, p. 5–32, 2001. 25
- CAETANO, E. F. d. S. A autonomia das universidades federais no exercício de sua autonomia financeira assegurada constitucionalmente. *Revista Brasileira de Educação*, Rio de Janeiro, v. 24, p. e240017, 2019. 11, 12, 18
- CAMPBELL, J. P.; DEBLOIS, P. B.; OBLINGER, D. G. Academic analytics: a new tool for a new era. *EDUCAUSE Review*, v. 42, n. 4, p. 40–57, 2007. 22
- CRESWELL, J. W. *Research design: qualitative, quantitative, and mixed methods approaches*. 4. ed. Thousand Oaks: SAGE Publications, 2014. 27
- ESTATÍSTICA, I. B. de Geografia e. *SIDRA: Sistema IBGE de Recuperação Automática*. 2026. Disponível em: <https://sidra.ibge.gov.br/home/ipca/brasil>. Acesso em: 15 mar. 2026. 28
- FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, v. 29, n. 5, p. 1189–1232, 2001. 26
- GIACOMONI, J. *Orçamento público*. 18. ed. São Paulo: Atlas, 2021. 20, 21
- GIAMBIAGI, F.; ALÉM, A. C. *Finanças públicas: teoria e prática no Brasil*. 5. ed. Rio de Janeiro: Elsevier, 2016. 19
- GIL, A. C. *Métodos e técnicas de pesquisa social*. 6. ed. São Paulo: Atlas, 2010. 27
- GUJARATI, D. N. *Econometria básica*. 5. ed. Porto Alegre: AMGH, 2011. 25
- GUPTA, S.; CHOUDHARY, J. Academic analytics: actionable intelligence in teaching and learning for higher education in indian institutions. In: *Proceedings of the International Conference on Skill Development & Technological Innovations for Economic Growth (ICST-2015)*. Ghaziabad, India: [s.n.], 2015. 22

- JANUSCHOWSKI, T. et al. Criteria for classifying forecasting methods. *International Journal of Forecasting*, v. 36, n. 1, p. 167–177, 2020. ISSN 0169-2070. M4 Competition. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0169207019301529>. 24
- JR., J. F. H. et al. *Análise multivariada de dados*. 6. ed. Porto Alegre: Bookman, 2009. 27
- JR., J. T. M.; REIS, H. d. C. *A Lei 4.320 comentada*. Rio de Janeiro: IBAM, 2016. 20
- JUNIOR, D. D. H.; FLACH, L.; MATTOS, L. K. d. The efficiency of public expenditure on higher education: a study with brazilian federal universities. *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, v. 28, n. 106, p. 123–145, 2020. Jan./mar. 18
- KEY, V. O. J. The lack of a budgetary theory. *American Political Science Review*, v. 34, n. 6, p. 1137–1144, dez. 1940. 15
- LAKATOS, E. M.; MARCONI, M. d. A. *Fundamentos de metodologia científica*. 7. ed. São Paulo: Atlas, 2003. 27
- LELOUP, L. T. The myth of incrementalism: Analytical choices in budgetary theory. *Polity*, v. 10, n. 4, p. 488–509, 1978. 16
- LEWIS, V. B. Toward a theory of budgeting. *Public Administration Review*, v. 12, n. 1, p. 42–54, 1952. 15
- LINDBLOM, C. E. The science of “muddling through”. *Public Administration Review*, v. 19, n. 2, p. 79–88, 1959. 16
- LJUNGMAN, G. *Top-Down Budgeting—An Instrument to Strengthen Budget Management*. Washington, DC, 2009. Disponível em: <https://www.imf.org/external/pubs/ft/wp/2009/wp09243.pdf>. 19
- MOLNAR, C. *Interpretable machine learning*. 2. ed. [S.l.]: Independently published, 2020. 36
- NACIONAL, B. S. do T. *Manual de Contabilidade Aplicada ao Setor Público (MCASP)*. Brasília, 2023. Disponível em: <https://www.tesourotransparente.gov.br/publicacoes/manual-de-contabilidade-aplicada-ao-setor-publico-mcasp>. Acesso em: 25 mar. 2026. 20
- PROVOST, F.; FAWCETT, T. *Data science for business: what you need to know about data mining and data-analytic thinking*. Sebastopol, CA: O’Reilly Media, 2013. 21
- PYHRR, P. A. *Zero-Base Budgeting: A Practical Management Tool for Evaluating Expenses*. New York: John Wiley & Sons, 1973. ISBN 9780471702344. 15
- RICHARDSON, R. J. *Pesquisa social: métodos e técnicas*. 3. ed. São Paulo: Atlas, 2017. 27
- RUBIN, I. S. Budget theory and budget practice: How good the fit? *Public Administration Review*, v. 50, n. 2, p. 179–189, 1990. 17
- SCHICK, A. The road to ppb: The stages of budget reform. *Public Administration Review*, v. 26, n. 4, p. 243–258, 1966. 15, 17

- SHMUELI, G.; KOPPIUS, R. O. Predictive analytics in information systems research. *MIS Quarterly*, v. 35, n. 3, p. 553–572, 2011. 24
- Universidade Federal da Paraíba. *UFPB em Números*. João Pessoa, PB: [s.n.], 2025. Desenvolvido pela Coordenação de Informação (Codeinfo/Proplan). Disponível em: <https://sigov.ufpb.br/ufpb-em-numeros>. Acesso em: 29 mar. 2026. 28
- VERGARA, S. C. *Projetos e relatórios de pesquisa em administração*. 17. ed. São Paulo: Atlas, 2016. 27
- WILDAVSKY, A. *The politics of the budgetary process*. Boston: Little, Brown, 1984. 16, 17
- WOOLDRIDGE, J. M. *Introdução à econometria: uma abordagem moderna*. 4. ed. São Paulo: Cengage Learning, 2016. 25
- YIN, R. K. *Estudo de caso: planejamento e métodos*. 5. ed. Porto Alegre: Bookman, 2015. 27
- ZHENG, A.; CASARI, A. *Feature engineering for machine learning: principles and techniques for data scientists*. Sebastopol: O'Reilly Media, 2018. 24

Emitido em 05/05/2026

DOCUMENTO Nº 1/2026 - CCDN (11.00.52.09)
(Nº do Documento: 1)

(Nº do Protocolo: NÃO PROTOCOLADO)

(Assinado digitalmente em 07/05/2026 12:40)
ANDREA ALVES BORBA
ASSISTENTE EM ADMINISTRACAO
1517808

Para verificar a autenticidade deste documento entre em <https://sipac.ufpb.br/documentos/> informando seu número: **1**,
ano: **2026**, documento (espécie): **DOCUMENTO**, data de emissão: **07/05/2026** e o código de verificação:
c3ba407b51