

UNIVERSIDADE FEDERAL DE ITAJUBÁ

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

**Meta-otimização dos Hiperparâmetros do Algoritmo
XGBoost para classificação binária: uma integração entre
Planejamento de Experimentos e o Método da Interseção Normal
à Fronteira**

Caio Tertuliano Ribeiro

ITAJUBÁ
FEVEREIRO DE 2026

Caio Tertuliano Ribeiro

**Meta-otimização dos Hiperparâmetros do Algoritmo
XGBoost para classificação binária: uma integração entre
Planejamento de Experimentos e o Método da Interseção Normal
à Fronteira**

Dissertação submetida para **defesa** do tema ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Itajubá, como parte dos requisitos para a obtenção do título de **Mestre em Ciências em Engenharia de Produção**.

Área de concentração: Engenharia de Produção.

Orientador: Prof. Dr. Anderson Paulo de Paiva

ITAJUBÁ
FEVEREIRO DE 2026

UNIVERSIDADE FEDERAL DE ITAJUBÁ

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

Caio Tertuliano Ribeiro

**Meta-otimização dos Hiperparâmetros do Algoritmo
XGBoost para classificação binária: uma integração entre
Planejamento de Experimentos e o Método da Interseção Normal
à Fronteira**

Dissertação submetida para **defesa** do tema ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Itajubá, como parte dos requisitos para a obtenção do título de **Mestre em Ciências em Engenharia de Produção**.

Banca examinadora:

Prof. Dr. Lilian Berton (UNIFESP / ITA)

Prof. Dr. Juliana Helena Daroz Gaudêncio (UNIFEI)

Prof. Dr. Matheus Brendon Francisco (UNIFEI)

Prof. Dr. Anderson Paulo de Paiva (Orientador)

ITAJUBÁ

FEVEREIRO DE 2026

DEDICATÓRIA

Aos meus pais, que são o alicerce de toda a minha vida. Por todo o amor, apoio incondicional e ensinamentos que me sustentam e inspiram diariamente.

AGRADECIMENTOS

A Deus, pela força concedida em todos os momentos, pelo caminho digno e pelo bom combate.

Aos meus pais, pelo suporte incondicional desde muito cedo, especialmente através do incentivo aos estudos, que tornaram possível cada passo dessa jornada.

Ao meu orientador e amigo, Prof. Dr. Anderson Paulo de Paiva, pela generosidade, paciência e dedicação desde o primeiro dia, assim como pelos ensinamentos que vão muito além da vida acadêmica.

Ao colega Me. Matheus Costa, pela parceria constante, apoio e amizade que foram fundamentais ao longo desta trajetória.

Ao Prof. Dr. Matheus Brendon Francisco, cuja postura acadêmica e dedicação foram fonte de inspiração durante o programa.

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), pelo apoio financeiro por meio do Projeto BPD-01045-22, que contribuiu para o desenvolvimento das atividades relacionadas a este trabalho.

À Universidade Federal de Itajubá e à Coordenação de Aperfeiçoamento Pessoal de Nível Superior pelo apoio institucional e financeiro, tornando possível o desenvolvimento desta pesquisa, e pelas oportunidades proporcionadas ao longo dessa formação acadêmica.

Que eu consiga, cada vez mais, me espelhar em todos aqueles aqui mencionados, honrando o exemplo que a mim representam.

EPÍGRAFE

A moderação na defesa da verdade é serviço prestado à mentira.

RESUMO

A calibração de hiperparâmetros é decisiva para o desempenho preditivo e o custo computacional de modelos de aprendizagem de máquina. Em classificação binária com *Extreme Gradient Boosting*, esse ajuste é inerentemente multiobjetivo: busca-se maximizar a qualidade preditiva e, simultaneamente, minimizar o tempo de execução. Este trabalho propõe um arcabouço metodológico que integra Planejamento de Experimentos, Metodologia de Superfície de Resposta, Análise Fatorial e o método Interseção Normal à Fronteira para orientar a seleção de hiperparâmetros do *Extreme Gradient Boosting* sob um orçamento fixo de avaliações. A exploração inicial do espaço é conduzida por um delineamento composto central de face centrada fracionário, totalizando 88 configurações. As respostas (acurácia, precisão, revocação, especificidade e tempo) são obtidas sob protocolo reproduzível com replicações e sintetizadas por escores fatoriais derivados de análise de componentes principais com rotação *Varimax*. Esses escores permitem definir funções objetivo de qualidade e de custo, reduzindo a redundância entre métricas e favorecendo a interpretação de *trade-offs*. Em seguida, modelos quadráticos de Metodologia de Superfície de Resposta são ajustados às funções objetivo e empregados pela Interseção Normal à Fronteira para amostrar uma fronteira de Pareto aproximadamente uniforme; as soluções candidatas são reavaliadas no modelo real e uma configuração final de compromisso é selecionada. A proposta é comparada, sob orçamento equivalente, com busca em grade, busca aleatória, otimização bayesiana e *Hyperopt*. Os resultados indicam redução substancial do custo computacional: o método proposto alcança tempo médio por *fold* de 0,078 s e tempo médio total de 158 s, com reduções de 9% a 71% no tempo total em relação aos *benchmarks*, mantendo desempenho preditivo elevado e estável entre réplicas. Como validação externa, o *pipeline* foi replicado em duas bases adicionais com perfis contrastantes (balanceada e desbalanceada), reproduzindo o padrão custo-qualidade observado no benchmark principal e reforçando a estabilidade multivariada do método. Conclui-se que a integração entre Planejamento de Experimentos e Interseção Normal à Fronteira constitui uma alternativa parcimoniosa, interpretável e replicável para calibração multiobjetivo de hiperparâmetros, com potencial de generalização para outros algoritmos da família *Gradient Boosting Decision Trees* e cenários com restrições de custo.

Palavras-chave: Otimização de hiperparâmetros; Planejamento de experimentos; Superfície de resposta; Otimização multiobjetivo; Interseção Normal à Fronteira; *Extreme Gradient Boosting*.

ABSTRACT

Hyperparameter tuning is decisive for both predictive performance and computational cost in machine learning models. In binary classification with Extreme Gradient Boosting, tuning is inherently multiobjective: predictive quality must be maximized while execution time is minimized. This work proposes a methodological framework that integrates Design of Experiments, Response Surface Methodology, Factor Analysis, and the Normal Boundary Intersection method to guide the selection of Extreme Gradient Boosting hyperparameters under a fixed evaluation budget. The initial exploration is conducted through a fractional face-centered central composite design, totaling 88 configurations. The observed responses (accuracy, precision, recall, specificity, and runtime) are collected under a reproducible protocol and summarized through factor scores obtained via principal component analysis with Varimax rotation. These scores define quality and cost objective functions, reduce redundancy among metrics, and support an interpretable assessment of trade-offs. Quadratic response-surface models are then fitted to the objective functions and used by Normal Boundary Intersection to sample an approximately uniform Pareto frontier; candidate solutions are re-evaluated on the real model, and a final compromise configuration is selected. The proposed approach is benchmarked, under equivalent evaluation budgets, against grid search, random search, Bayesian optimization, and Hyperopt. Results show substantial computational savings: the method achieves an average runtime of 0.078 s per fold and an average total runtime of 158 s, reducing total time by 9% to 71% relative to the benchmarks while maintaining high and stable predictive performance across replications. As external validation, the proposed pipeline was replicated on two additional datasets with contrasting profiles (balanced and highly imbalanced), reproducing the cost–quality trade-off observed in the canonical benchmark and reinforcing the multivariate stability of the method. Overall, integrating Design of Experiments and Normal Boundary Intersection provides a parsimonious, interpretable, and replicable alternative for multiobjective hyperparameter tuning, with potential applicability to other Gradient Boosting Decision Tree families and cost-constrained settings.

Keywords: *Hyperparameter tuning. Design of Experiments. Response Surface Methodology. Multi-objective optimization. Normal Boundary Intersection. XGBoost.*

LISTA DE FIGURAS

Figura 1 - Representação do CCDFC no hipercubo codificado e decomposição em pontos fatoriais, axiais e centrais.....	43
Figura 2 - Fronteira de Pareto obtida via <i>Normal Boundary Intersection</i> (NBI). Adaptado de Brito <i>et al.</i> (2014), com base em Das e Dennis (1998).	55
Figura 3 - Encadeamento conceitual: <i>Simplex Lattice</i> {p,q}, Otimização Multiobjetivo, Eficiência MODM, Superfície de Pareto.....	56
Figura 4 - Fluxograma do <i>pipeline</i> proposto (DoE + RSM + PCA/Varimax + NBI).	69
Figura 5 - Pseudocódigo do <i>pipeline</i> proposto.	70
Figura 6 - Extrato da matriz experimental do CCDFC, com fatores (hiperparâmetros) e respostas originais.	87
Figura 7 - Gráfico de Pareto de efeitos padronizados, contorno e efeitos principais para Acurácia.....	88
Figura 8 - Gráfico de Pareto de efeitos padronizados, contorno e efeitos principais para Precisão.	89
Figura 9 - Gráfico de Pareto de efeitos padronizados, contorno e efeitos principais para Revocação.....	91
Figura 10 - Gráfico de Pareto de efeitos padronizados, contorno e efeitos principais para Especificidade.....	92
Figura 11 - Gráfico de Pareto de efeitos padronizados, contorno e efeitos principais para Tempo de execução.....	93
Figura 12 - Análise Fatorial com rotação Varimax e Cluster de métricas de performance e escores rotacionados de fator.....	96
Figura 13 - Histogramas das métricas de performance do XGBoost e dos Escores rotacionados de fator.....	97
Figura 14 - Gráfico de Pareto dos efeitos padronizados e contorno para FS ₁ (VRF ₁).....	98
Figura 15 - Resumo do modelo e coeficientes codificados para FS ₁ (VRF ₁).....	99
Figura 16 - Gráfico de Pareto dos efeitos padronizados e contorno para FS ₂ (VRF ₂).	99
Figura 17 - Resumo do modelo e coeficientes codificados para FS ₂ (VRF ₂).....	100
Figura 18 - Gráfico de Pareto dos efeitos padronizados e contorno para FS ₃ (VRF ₃).	101
Figura 19 - Resumo do modelo e coeficientes codificados para FS ₃ (VRF ₃).....	101
Figura 20 - Arranjo <i>Simplex-Lattice</i> {3, 10} usado no caso XGBoost.	107

Figura 21 - Superfície de Pareto para os Índices Multivariados (VRF) relativo às métricas de desempenho do XGBoost para os pesos definidos pelo arranjo <i>Simplex-Lattice</i> {3, 10}.....	107
Figura 22 - Superfícies de Resposta de Misturas para as métricas de desempenho do XGBoost baseado no arranjo <i>Simplex-Lattice</i> {3, 10}.....	108
Figura 23 - Superfícies de Resposta de Misturas para o tempo de processamento do XGBoost e os indicadores multivariados das métricas de desempenho avaliadas nos pontos do arranjo <i>Simplex-Lattice</i> {3, 10}.....	109
Figura 24 - <i>Overlaid Contour Plot</i> (Sobreposição das Superfícies de Resposta para o Arranjo <i>Simplex-Lattice</i> {3, 10} das métricas de desempenho do XGBoost e os respectivos indicadores multivariados (VRF).....	109
Figura 25 - Arranjo <i>Simplex-Lattice</i> {3, 10} do Índice Escalar de Qualidade do XGBoost..	110
Figura 26 - <i>Boxplot</i> do tempo total da otimização por método (n = 30).	112
Figura 27 - <i>Boxplot</i> do tempo total de execução por método (n = 30).....	112
Figura 28 - <i>Boxplot</i> do tempo médio por fold por método (escala logarítmica; n = 30).....	113
Figura 29 - ANOVA ($\log_{10}T$) e comparações de Tukey para o tempo total da otimização...	113
Figura 30 - ANOVA ($\log_{10}T$) e comparações de Tukey para o tempo total de execução.....	114
Figura 31 - <i>Boxplot</i> da acurácia média por método (n = 30).....	115
Figura 32 - Dispersão entre índice de qualidade e tempo médio por <i>fold</i>	116
Figura 33 - Saída da ANOVA e comparações de Tukey para o índice de qualidade.....	117
Figura 34 - Perfil dos hiperparâmetros ao longo de β_2 (ênfase em custo/tempo).....	122
Figura 35 - Perfil dos hiperparâmetros ao longo de β_1 (ênfase em qualidade).....	123
Figura 36 - Superfície RSM (custo/tempo): “subsample” × “colsample_bytree”.....	124
Figura 37 - Superfície RSM (custo/tempo): “n_estimators” × “max_depth”.....	124
Figura 38 - Superfície RSM (custo/tempo): “n_estimators” × “learning_rate”	125
Figura 39 - Superfície RSM (qualidade): “learning_rate” × “max_depth”.....	126
Figura 40 - Superfície RSM (qualidade): “subsample” × “colsample_bytree”.....	126

LISTA DE TABELAS

Tabela 1 - Caracterização descritiva do conjunto de dados <i>MAGIC Gamma Telescope</i> utilizado nos experimentos	71
Tabela 2 - Especificação das variáveis preditoras do conjunto de dados <i>MAGIC Gamma Telescope</i> utilizado nos experimentos, com resumo conceitual e respectivas unidades físicas.	72
Tabela 3 - Resumo dos modelos e coeficientes codificados para as métricas de desempenho do XGBoost e suas variáveis latentes (Escores Rotacionados de Fator).....	102
Tabela 4 - Configuração do método VRF-NBI.	104
Tabela 5 - Estatísticas descritivas do custo computacional por método (média ± desvio padrão; B = 138; n = 30).....	111
Tabela 6 - Estatísticas descritivas de qualidade por método (média ± desvio padrão; n = 30).	114
Tabela 7 - Distâncias de Mahalanobis ao perfil do método proposto e contagem de <i>outliers</i> ($d^2 > \chi_{0,9752}^2$).....	118
Tabela 8 - Teste de Mardia (assimetria e curtose) por método.	119
Tabela 9 - Correlações (Spearman) em DoE + NBI (proposto) e p-valores com ajuste de Holm.	120
Tabela 10 - Pós-testes (Hotelling T^2): DoE + NBI versus <i>benchmarks</i> (com Holm).	120
Tabela 11 - Métricas multiobjetivo (média, desvio padrão e quantis por réplica; quando aplicável).	122
Tabela 12 - Estatísticas descritivas de qualidade por método (média ± desvio padrão; n = 30; B = 138) em “ <i>pepper_species</i> ”.....	128
Tabela 13 - Estatísticas descritivas do custo computacional por método (média ± desvio padrão; n = 30; B = 138) em “ <i>pepper_species</i> ”.....	128
Tabela 14 - Robustez em “ <i>pepper_species</i> ”: distâncias de Mahalanobis ao perfil do método proposto e contagem de outliers (limiar $\chi_{(5; 0,975)}^2 = 12,833$).	129
Tabela 15 - Correlações (Spearman) no método proposto em “ <i>pepper_species</i> ” e p-valores com ajuste de Holm.	130
Tabela 16 - Estatísticas descritivas de qualidade por método (média ± desvio padrão; n = 30; B = 138) em “ <i>credit_card_default</i> ”.....	131
Tabela 17 - Estatísticas descritivas do custo computacional por método (média ± desvio padrão; n = 30; B = 138) em “ <i>credit_card_default</i> ”.....	131

Tabela 18 - Robustez em “ <i>credit_card_default</i> ”: distâncias de Mahalanobis ao perfil do método proposto e contagem de <i>outliers</i> (limiar $\chi^2_{(5; 0,975)} = 12,833$).....	131
Tabela 19 - Correlações (Spearman) no método proposto em “ <i>credit_card_default</i> ” e p-valores com ajuste de Holm.	132
Tabela 20 - Pós-testes (Hotelling T^2) entre o método proposto e benchmarks no vetor y , com ajuste de Holm.	133

LISTA DE ABREVIATURAS E SIGLAS

ANOVA	Análise de Variância (<i>Analysis of Variance</i>)
ARM64	Arquitetura ARM de 64 bits
AUC	Área sob a curva (<i>Area Under the Curve</i>)
AUC-PR	Área sob a curva precisão-revocação (<i>Area Under the Precision-Recall Curve</i>)
AUC-ROC	Área sob a curva característica de operação do receptor (<i>Area Under the Receiver Operating Characteristic Curve</i>)
CCD	Delineamento Composto Central (<i>Central Composite Design</i>)
CCDFC	Delineamento Composto Central de Face Centrada (<i>Central Composite Design Face-Centered</i>)
CHIM	Fecho convexo dos mínimos individuais (<i>Convex Hull of Individual Minima</i>)
CPU	Unidade Central de Processamento (<i>Central Processing Unit</i>)
CSV	Valores separados por vírgula (<i>Comma-Separated Values</i>)
CUDA	Arquitetura Unificada de Dispositivos de Computação (<i>Compute Unified Device Architecture</i>)
DoE	Planejamento de Experimentos (<i>Design of Experiments</i>)
DP	Desvio padrão
EI	Melhoria esperada (<i>Expected Improvement</i>)
F1	Medida F1 (<i>F1-score</i>)
GL	Graus de liberdade
GPU	Unidade de Processamento Gráfico (<i>Graphical Processing Unit</i>)
HPO	Otimização de Hiperparâmetros (<i>Hyperparameter Optimization</i>)
IGD	<i>Distância Geracional Invertida (Inverted Generational Distance).</i>
JSON	<i>JavaScript Object Notation</i>
KDD	<i>Knowledge Discovery in Databases</i>
LCB	Limite inferior de confiança (<i>Lower Confidence Bound</i>)
LLM	Modelo de Linguagem de Grande Porte (<i>Large Language Model</i>)
MAGIC	Conjunto de dados “MAGIC Gamma Telescope” (<i>Major Atmospheric Gamma Imaging Cherenkov</i>)
MANOVA	Análise de Variância Multivariada (<i>Multivariate Analysis of Variance</i>).
MCC	Coefficiente de Correlação de Matthews (<i>Matthews Correlation Coefficient</i>)
ML	Aprendizado de Máquina (<i>Machine Learning</i>)

MMSE	Erro Quadrático Médio Mínimo (<i>Minimum Mean Squared Error</i>)
MODM	Tomada de decisão multiobjetivo (<i>Multi-Objective Decision Making</i>)
MOO	Otimização multiobjetivo (<i>Multi-Objective Optimization</i>)
MQO	Mínimos Quadrados Ordinários (<i>Ordinary Least Squares</i>)
NBI	Interseção Normal à Fronteira (<i>Normal Boundary Intersection</i>)
NBI-MMSE	Interseção Normal à Fronteira com MMSE
NVMe	Expressão de memória não volátil (<i>Non-Volatile Memory Express</i>)
PCA	Análise de Componentes Principais (<i>Principal Component Analysis</i>)
PI	Probabilidade de melhoria (<i>Probability of Improvement</i>)
PIP	Gerenciador de pacotes da linguagem Python (<i>Python Installs Packages</i>)
PR	Precisão-Revocação (<i>Precision-Recall</i>)
QM	Quadrado médio
RAM	Memória de Acesso Aleatório (<i>Random Access Memory</i>)
ROC	Curva ROC (<i>Receiver Operating Characteristic Curve</i>)
RSM	Metodologia de Superfície de Resposta (<i>Response Surface Methodology</i>)
SLSQP	Programação de Mínimos Quadrados Sequenciais (<i>Sequential Least Squares Programming</i>)
SMAC	Configuração de algoritmos sequenciais baseados em modelos (<i>Sequential Model-based Algorithm Configuration</i>)
SQ	Soma de quadrados
SSD	Unidade de Estado Sólido (<i>Solid State Drive</i>)
TPE	Estimador de Parzen em árvore (<i>Tree-structured Parzen Estimator</i>)
UCB	Limite superior de confiança (<i>Upper Confidence Bound</i>)
XGBoost	<i>Extreme Gradient Boosting</i>

LISTA DE SÍMBOLOS

Nota: por convenção, vetores e matrizes (arranjos) são representados em negrito ao longo do texto (inclusive nesta lista).

$\mathbf{D} = (\mathbf{X}, \mathbf{y})$	Conjunto de dados, em que \mathbf{X} é a matriz de atributos e \mathbf{y} é o vetor de rótulos.
H	Espaço de busca de hiperparâmetros (conjunto de configurações admissíveis).
θ	Vetor de parâmetros do modelo (parâmetros ajustados durante o treinamento).
k	Número de partições na validação cruzada estratificada (<i>k-fold</i>).
B	Orçamento de avaliações do modelo, número de configurações avaliadas (<i>budget</i>).
β	Parâmetro (ou vetor) de varredura utilizado na geração de direções no método NBI.
z^U	Ponto de utopia (melhores valores dos objetivos, por componente).
\mathbf{Y}	Matriz de métricas/respostas (dimensões $n \times p$) associada às execuções experimentais.
$\hat{y}(x)$	Resposta predita por um modelo de superfície de resposta no ponto experimental x .
x_i	Variável codificada do i -ésimo fator no planejamento de experimentos.
γ	Penalidade mínima (ganho mínimo) requerida para realizar uma divisão (<i>split</i>) no XGBoost.
$m(t)$	Métrica monitorada na iteração t (parada antecipada; <i>early stopping</i>).

$VRF_{i,k}$	Escore fatorial rotacionado do i -ésimo experimento associado ao k -ésimo fator ($k = 1, 2, 3$).
(G_j, H_j)	Estatísticas agregadas (gradiente e hessiana) na folha j .
(P^t)	Subconjunto de atributos amostrados na iteração t .
(y_i)	Rótulo da i -ésima instância.
A	Matriz de cargas rotacionadas.
$C(h)$	Função objetivo de custo associada à configuração h .
$\hat{C}(h)$	Aproximação por superfície de resposta para $C(h)$.
$C(\theta)$	Índice de custo computacional (tempo) associado a θ .
\hat{C}_i	Custo normalizado via <i>min-max</i> (i -ésima solução).
d_{\max}	Profundidade máxima de cada árvore.
f	Vetor de fatores comuns (análise fatorial).
$f(\mathbf{x}; \theta)$	Preditor paramétrico ajustado.
FN	Falsos negativos (<i>false negatives</i>).
FP	Falsos positivos (<i>false positives</i>).
h	Vetor de hiperparâmetros.
h_i^2	Comunalidade da variável i .
$\ell(\cdot)$	Função de perda.
m	Número de componentes/fatores retidos.
n	Número de observações.
P	Matriz de cargas (<i>loadings</i>) das componentes principais.
p	Número de atributos.
\mathbb{P}	Conjunto de índices de atributos.
$Q(h)$	Função objetivo de qualidade associada à configuração h .
$\hat{Q}(h)$	Aproximação por superfície de resposta para $Q(h)$.

$Q(\theta)$	Índice de qualidade (escore principal) associado a θ .
\hat{Q}_i	Qualidade normalizada via <i>min-max</i> (i-ésima solução).
R	Matriz de correlações.
S	Matriz de escores rotacionados.
T	Matriz de escores (scores) das componentes principais.
T	Número de iterações/árvores no processo de <i>boosting</i> .
TN	Verdadeiros negativos (<i>true negatives</i>).
TP	Verdadeiros positivos (<i>true positives</i>).
u	Termo específico/erro (análise fatorial).
v_i	i-ésimo autovetor associado a λ_i .
v_i	i-ésimo autovetor associado a λ_i .
W	Matriz de rotação ortogonal (<i>Varimax</i>).
w^+	Peso atribuído à classe positiva (<i>scale positive weight</i>).
w_i	Peso associado à instância i na função de perda.
w_j^*	Peso ótimo associado à folha j .
X	Matriz de atributos (instâncias por variáveis; covariáveis).
x_i	Vetor de atributos da i-ésima instância.
y	Vetor resposta (rótulos da classe).
Z	Matriz de dados padronizados.
$z(\mathbf{h})$	Vetor de objetivos no espaço multiobjetivo.
$z(\theta)$	Vetor de objetivos no espaço multiobjetivo.
z^N	Ponto de nadir (piores valores observados, por componente).
α	Termo de regularização L1 aplicado aos pesos das folhas.
$\beta_0, \beta_i, \beta_{ii}, \beta_{ij}$	Coefficientes do modelo quadrático de superfície de resposta.
Δ	Tolerância mínima de melhora para parada antecipada.

ε	Erro aleatório (resíduo) no modelo de superfície de resposta.
η	Taxa de aprendizagem (<i>learning rate</i>) no XGBoost.
κ_{level}	Frações de atributos nível.
κ_{node}	Frações de atributos por nó.
κ_{tree}	Frações de atributos por estimador (árvore).
Λ	Matriz de cargas fatoriais (análise fatorial).
λ	Termo de regularização L2 aplicado aos pesos das folhas.
λ_i	i-ésimo autovalor (variância explicada).
λ_i	i-ésimo autovalor.
μ	Média amostral.
ρ	Fração de subamostragem de instâncias (<i>subsample</i>).
Σ	Matriz de covariâncias.
σ	Desvio padrão amostral.
Φ	Matriz <i>payoff</i> (ótimos mono-objetivo) no NBI.
Ψ	Matriz diagonal de unicidades (análise fatorial).
$\Omega(\theta)$	Termo de regularização associado a θ .
ω_{min}	Peso mínimo do nó-filho (<i>minimal child weight</i>).
d^2	Distância de Mahalanobis (usada no critério de <i>outlier</i> multivariado).
χ^2	Distribuição qui-quadrado (quantil usado como limiar para outliers multivariados).
$b_{1,p}$	Estatística de assimetria de Mardia (normalidade multivariada).
$b_{2,p}$	Estatística de curtose de Mardia (normalidade multivariada).
T^2	Estatística T^2 de Hotelling (pós-teste multivariado).
ρ_s	Coefficiente de correlação de Spearman

SUMÁRIO

1. Introdução	1
1.1 Contexto de Pesquisa	4
1.2 Objetivos	6
1.2.1 Objetivo Geral	6
1.2.2 Objetivos Específicos	7
1.3 Contribuições Esperadas	7
1.4 Delimitações da Pesquisa	8
1.5 Estrutura do Trabalho.....	9
2. Referencial Teórico	10
2.1 <i>Machine learning</i> , hiperparâmetros e motivação metodológica	10
2.2 Abordagens de busca de hiperparâmetros: <i>benchmarks</i> e panorama.....	11
2.2.1 Busca em grade	12
2.2.2 Busca aleatória	13
2.2.3 Otimização bayesiana.....	15
2.2.4 Hyperopt e o estimador de Parzen em árvore.....	17
2.2.5 Síntese e lacunas.....	18
2.3 <i>Extreme Gradient Boosting</i>	19
2.3.1 Notação e modelo aditivo por árvores.....	19
2.3.2 Função objetivo regularizada	20
2.3.3 Otimização incremental via aproximação de segunda ordem.....	20
2.3.4 Pesos ótimos de folhas e ganho de divisão.....	21
2.3.5 Considerações práticas e relação com hiperparâmetros	22
2.4 Hiperparâmetros do XGBoost: papel, efeitos e compromissos (<i>trade-offs</i>)	23
2.4.1 Número de estimadores (árvores).....	24
2.4.2 Taxa de aprendizado.....	24
2.4.3 Profundidade máxima da árvore.....	25
2.4.4 Peso mínimo do nó-filho	25
2.4.5 Perda mínima para divisão	26
2.4.6 Fração de amostragem de instâncias	27
2.4.7 Fração de colunas por árvore; e variantes por nível e por nó.....	27
2.4.8 Regularizações Alpha e Lambda	27

2.4.9	Peso da classe positiva	28
2.4.10	Cr�terios auxiliares: parada antecipada, m�trica de avalia�o e fun�o objetivo	28
2.5	M�tricas de avalia�o e custo computacional	29
2.6	Planejamento de Experimentos: princ�pios, arranjos e an�lise	31
2.6.1	Princ�pios gerais	32
2.6.2	Arranjos cl�ssicos e superf�cies de resposta.....	34
2.6.3	Metodologia de Superf�cie de Resposta	35
2.7	Metodologia de Superf�cie de Resposta e Delineamento Central Composto Face- Centrado	37
2.7.1	Formula�o do modelo quadr�tico completo e interpreta�o dos termos	38
2.7.2	Codifica�o de fatores, escalas e regi�o experimental	39
2.7.3	Delineamento Central Composto: estrutura, estima�o e replica�o no centro	39
2.7.4	Delineamento Central Composto Face-Centrado: defini�o, posi�o dos pontos axiais e implica�es	40
2.7.5	Compara�o conceitual entre Delineamento Central Composto de Face Centrada e Box-Behnken: crit�rio de escolha	41
2.7.6	Metodologia de Superf�cie de Resposta como base para otimiza�o mono- objetivo e multiobjetivo e conex�o com o M�todo de Interse�o Normal � Fronteira ...	42
2.8	An�lise fatorial, componentes principais e rota�o <i>Varimax</i> para s�ntese de m�tricas	43
2.8.1	Nota�o e padroniza�o das vari�veis	44
2.8.2	Modelo de an�lise fatorial comum	44
2.8.3	An�lise de Componentes Principais: formula�o e propriedades	45
2.8.4	Rota�o <i>Varimax</i> : motiva�o e formula�o	46
2.8.5	Escores fatoriais rotacionados e indetermina�o de sinal	47
2.8.6	Constru�o de �ndices compostos: “Qualidade” e “Tempo”	48
2.8.7	Discuss�o cr�tica: explicabilidade, vantagens e limita�es	49
2.9	M�todo da Interse�o Normal � Fronteira.....	50
2.9.1	Formula�o do problema multiobjetivo	51
2.9.2	Elementos geom�tricos do NBI: pontos �ncora, linha utopia e CHIM	51
2.9.3	Normaliza�o e formula�o b�sica do subproblema NBI	52
2.9.4	Caso bidimensional: rela�es geom�tricas e redu�o alg�brica.....	53

2.9.5	Discretização do <i>simplex</i> de pesos e número de subproblemas	54
2.9.6	Procedimento resumido para aplicação do NBI	55
2.9.7	Conexões com a literatura e implicações	55
2.10	Otimização de Hiperparâmetros em XGBoost.....	57
2.10.1	Natureza de “caixa-preta”, custo e ruído	58
2.10.2	HPO como configuração de algoritmo e implicações para XGBoost.	58
2.10.3	Estrutura do espaço de busca em XGBoost.....	59
2.10.4	Interações e regimes de equivalência entre hiperparâmetros.	60
2.10.5	Importância relativa e estratégia de alocação do orçamento	61
2.10.6	Métrica e protocolo de validação: risco de sobreajuste na seleção de métricas-alvo	61
2.10.7	Custo computacional, multi-fidelidade e alocação adaptativa de recursos ...	62
2.10.8	Estocasticidade e perspectiva de otimização por simulação	63
2.10.9	Multiobjetividade: desempenho versus custo (e outros critérios).....	64
2.10.10	Implicações práticas: definição de espaços, escalas e orçamentos.....	64
2.10.11	Tendências contemporâneas e avanços metodológicos em otimização de hiperparâmetros	65
3.	Metodologia	68
3.1	Visão geral do <i>pipeline</i> e hipóteses operacionais.....	68
3.2	Definição do problema e notação.....	68
3.3	Pré-processamento, balanceamento e particionamento.....	70
3.3.1	Caracterização resumida da base de dados canônica	71
3.3.2	Pré-processamento e descrição de atributos	71
3.4	Espaço de busca de hiperparâmetros.....	72
3.5	Planejamento experimental CCDFC	73
3.5.1	Construção do delineamento e codificação	73
3.5.2	Definição da região experimental e restrições experimentais	74
3.5.3	Execução do DoE e registro das respostas	74
3.6	Métricas de avaliação e medida de custo	75
3.6.1	Matriz de confusão e métricas de qualidade.....	75
3.6.2	Tempo de execução como <i>proxy</i> de custo computacional	75
3.7	Etapa mono-objetivo: modelagem das respostas individuais.....	76
3.7.1	Procedimento iterativo de ajuste ponderado e diagnóstico dos resíduos	76

3.8 Síntese das respostas e construção de escores.....	77
3.8.1 Padronização das respostas.....	77
3.8.2 PCA, rotação <i>Varimax</i> e escores operacionais.....	77
3.9 Modelagem por Superfícies de Resposta	78
3.10 Otimização Multiobjetivo por NBI	79
3.10.1 Normalização e Pontos de referência (Utopia x Nadir).....	79
3.10.2 Geração de candidatos na fronteira de Pareto	79
3.10.3 Confirmação no modelo real e seleção de compromisso	79
3.11 <i>Benchmarks</i> e critério de justiça por orçamento	80
3.12 Replicações, controle de aleatoriedade e reprodutibilidade experimental	80
3.13 Ambiente Computacional.....	81
3.13.1 <i>Hardware</i>	81
3.13.2 Sistema Operacional.....	81
3.13.3 Ambiente de Software	81
3.13.4 Controle de Execução.....	82
3.14 Ameaças à validade.....	82
4. Resultados e Discussões	84
4.1 Visão geral da avaliação.....	84
4.1.1 Condições de comparação e orçamento computacional.....	84
4.1.2 Métricas avaliadas	84
4.2 Estudo mono-objetivo com CCDFC: modelagens individuais das respostas	85
4.2.1 Matriz experimental e codificação dos fatores.....	86
4.2.2 Modelo de Superfície de Resposta e significância dos efeitos.....	87
4.2.3 Acurácia.....	87
4.2.4 Precisão	88
4.2.5 Revocação	90
4.2.6 Especificidade	91
4.2.7 Tempo de Execução	92
4.2.8 Síntese e implicações para a formulação multiobjetivo	93
4.3 Estudo multiobjetivo - aplicação da metodologia proposta	94
4.3.1 Obtenção de VRF_1 , VRF_2 e VRF_3	94
4.3.2 Da VRF ao espaço contínuo: modelagem por Superfície de Resposta	97
4.4 Otimização multiobjetivo: do simplex NBI ao resultado otimizado.....	103

4.4.1	Formulação do problema no espaço VRF	103
4.4.2	Pontos de Anclagem e construção do <i>simplex</i> no espaço de objetivos	103
4.4.3	Do <i>simplex</i> aos hiperparâmetros candidatos: subproblemas NBI.....	105
4.4.4	Confirmação experimental e seleção do resultado otimizado	105
4.5	Encadeamento final: da obtenção dos fatores ao resultado otimizado.....	106
4.6	Análise de custo computacional do estudo multiobjetivo	111
4.6.1	Análise de variância para tempos de execução	113
4.7	Análise de qualidade preditiva do estudo multiobjetivo	114
4.7.1	Análise de variância para métricas de qualidade.....	116
4.8	Análises multivariadas de robustez e estabilidade	117
4.8.1	Vetor multivariado e pré-processamento	117
4.8.2	Robustez via distância de Mahalanobis (detecção de pontos atípicos)	117
4.8.3	Normalidade multivariada (teste de Mardia) e implicações.....	118
4.8.4	Estrutura de correlação e correção de múltiplas comparações (Holm-Bonferroni)	119
4.8.5	Comparação multivariada entre métodos (MANOVA) e pós-testes.....	120
4.8.6	Métricas multiobjetivo (IGD, <i>Spread</i> e entropia de espaçamento) a partir do conjunto NBI	121
4.8.7	Interpretação do compromisso qualidade-custo e sensibilidade dos hiperparâmetros (β_1/β_2 , superfícies RSM e síntese)	122
4.9	Validação externa em bases adicionais	127
4.9.1	Protocolo experimental preservado	127
4.9.2	Base de Espécies de Pimentas	127
4.9.3	Base de Inadimplência de Crédito	130
4.9.4	Síntese multivariada e implicações para validade externa	133
5.	Conclusões	134
5.1	Síntese dos resultados e implicações.....	134
5.2	Evidências de validade externa e estabilidade em bases adicionais	136
5.3	Contribuições do trabalho	137
5.4	Limitações e trabalhos futuros	138
	Referências	141
	Apêndice A - Disponibilidade do código-fonte e artefatos experimentais	149
	Apêndice B - Design Matrix (X) relativa ao CCD de Face Centrada	150

Apêndice C - Superfície de Pareto (<i>Simplex-Lattice</i> {3, 10} e Métricas de Performance)	151
Apêndice D - Pareto MDE (<i>Setups</i> Ótimos e não-dominados)	152
Apêndice E - Superfície de Pareto (Métricas Multivariadas)	153

1. INTRODUÇÃO

Em termos práticos, o aprendizado de máquina (*Machine Learning* - ML) pode ser visto como um processo de “aprender com exemplos”: a partir de registros históricos, como exames e diagnósticos, transações e fraudes confirmadas, ou sensores e falhas em equipamentos; busca-se construir um modelo capaz de reconhecer padrões e antecipar resultados em novos casos. Diferentemente de abordagens baseadas apenas em regras fixas, o ML ajusta seus parâmetros a partir dos dados disponíveis, de modo a capturar regularidades que nem sempre são triviais de explicitar manualmente. Essa perspectiva é especialmente útil quando o volume de observações é grande e as relações entre variáveis são complexas, exigindo modelos que generalizem bem para situações não vistas.

A crescente adoção de técnicas de ML em tarefas de classificação e predição tem sido impulsionada pelo aumento da capacidade de processamento, pela maior disponibilidade de dados e pela necessidade de modelos capazes de apoiar decisões em contextos científicos e industriais (Bishop, 2006; Hastie; Tibshirani; Friedman, 2009; Shalev-Shwartz; Ben-David, 2014). Em problemas supervisionados, parte-se de um conjunto de dados rotulado $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$, em que $\mathbf{x}_i \in \mathbb{R}^p$ representa o vetor de atributos de entrada e $y_i \in \mathcal{Y}$ representa o rótulo associado. Deseja-se aprender uma função de predição $f(\mathbf{x}; \boldsymbol{\theta})$, parametrizada por $\boldsymbol{\theta}$, capaz de generalizar para amostras não observadas, o que usualmente é operacionalizado por procedimentos de minimização de risco empírico com regularização (Bishop, 2006; Shalev-Shwartz; Ben-David, 2014):

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n \ell(y_i, f(\mathbf{x}_i; \boldsymbol{\theta})) + \Omega(\boldsymbol{\theta}) \quad (1.1)$$

Em que $\ell(\cdot)$ é uma função de perda e $\Omega(\cdot)$ é um termo de regularização introduzido para controlar complexidade e mitigar sobreajuste (*overfitting*).

Entre os algoritmos de destaque para dados tabulares, os métodos de *gradient boosting* com árvores de decisão apresentam elevado desempenho preditivo devido à combinação aditiva de modelos fracos por meio de descida em gradiente no espaço funcional (Friedman, 2001). Nessa classe, o *Extreme Gradient Boosting* (XGBoost) consolidou-se como uma das abordagens mais utilizadas, combinando aprimoramentos algorítmicos e de engenharia para escalabilidade, além de mecanismos explícitos de regularização, amostragem e controle de complexidade (Chen; Guestrin, 2016). A predição é obtida por um modelo aditivo composto por T árvores:

$$\hat{y}_i = \sum_{t=1}^T f_t(\mathbf{x}_i), \quad f_t \in \mathcal{F} \quad (1.2)$$

Em que \mathcal{F} denota o espaço de funções correspondentes a árvores de decisão. A cada iteração t , aprende-se uma nova árvore f_t minimizando uma função objetivo que combina perda e regularização estrutural (Chen; Guestrin, 2016):

$$\mathcal{L}^{(t)} = \sum_{i=1}^n \ell(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)) + \Omega(f_t), \quad \Omega(f_t) = \gamma T_t + \frac{\lambda}{2} \sum_{j=1}^{T_t} w_j^2 \quad (1.3)$$

Em que T_t representa o número de folhas da árvore f_t , w_j representa o peso associado à folha j , e γ e λ controlam penalizações que buscam limitar a complexidade do modelo.

Apesar das propriedades atrativas de modelos baseados em árvores, o desempenho obtido em aplicações reais depende criticamente do ajuste de hiperparâmetros, isto é, de variáveis de controle selecionadas antes do treinamento que não são estimadas diretamente pela otimização de θ e que influenciam diretamente a capacidade do modelo, o grau de regularização, o regime de amostragem e a dinâmica de aprendizado (Probst; Wright; Boulesteix, 2019). No XGBoost, hiperparâmetros como taxa de aprendizado, profundidade máxima das árvores, peso mínimo do nó-filho, penalização por divisão, número de estimadores e frações de amostragem de instâncias e de atributos definem compromissos relevantes entre viés e variância, estabilidade, custo computacional e risco de sobreajuste (Chen; Guestrin, 2016). Assim, mesmo em bases relativamente pequenas, escolhas inadequadas podem induzir modelos excessivamente rígidos ou excessivamente complexos, degradando a capacidade de generalização.

De maneira abstrata, o ajuste de hiperparâmetros pode ser formulado como um problema de otimização em que se deseja selecionar $\lambda \in \Lambda$ para maximizar uma métrica estimada de desempenho $\hat{\mathcal{M}}(\cdot)$, calculada por um protocolo de validação, como validação cruzada estratificada por k-dobras (*k-fold*):

$$\hat{\lambda} = \operatorname{argmax}_{\lambda \in \Lambda} \hat{\mathcal{M}}(f_\lambda, \mathcal{D}) \quad (1.4)$$

Entretanto, a avaliação de $\hat{\mathcal{M}}$ exige treinamento repetido, pode ser tratada como uma função “caixa-preta” onerosa e pode apresentar variabilidade associada à amostragem e ao particionamento dos dados (Bergstra; Bengio, 2012; Shahriari *et al.*, 2016; Gramacy, 2020). Além disso, em cenários de engenharia e ciência de dados aplicada, a busca por ganhos marginais em métricas de qualidade frequentemente se associa a maior custo computacional, seja pelo aumento da complexidade do modelo, seja pelo incremento do orçamento de buscas. Por essa razão, o problema pode ser expresso como uma otimização multiobjetivo, em que se

deseja simultaneamente maximizar um índice de qualidade $Q(\lambda)$ e minimizar um índice de custo computacional $C(\lambda)$:

$$\max_{\lambda \in \Lambda} \mathbf{F}(\lambda) = (Q(\lambda), -C(\lambda)) \quad (1.5)$$

Em que soluções eficientes são definidas por dominância de Pareto (Deb, 2002; Morales-Hernández *et al.*, 2022). Essa perspectiva é particularmente relevante quando tempo computacional é interpretado como um *proxy* do custo de experimentação, permitindo comparar métodos em condições de orçamento justo.

No contexto do ajuste de hiperparâmetros do XGBoost, abordagens consagradas incluem busca em grade, busca aleatória, otimização bayesiana e métodos baseados em densidade, como o algoritmo estimador de *Parzen* em árvore (*tree-structured Parzen estimator*, TPE) (Bergstra; Bengio, 2012; Bergstra *et al.*, 2011; Shahriari *et al.*, 2016). A busca em grade é conceitualmente simples, mas tende a ser ineficiente em espaços de alta dimensão e sob restrições de orçamento. A busca aleatória, por sua vez, apresenta boa relação custo-benefício em muitos cenários e pode explorar com maior eficiência dimensões relevantes quando apenas parte dos hiperparâmetros tem forte influência no desempenho (Bergstra; Bengio, 2012). Abordagens bayesianas utilizam modelos probabilísticos como substitutos para orientar a exploração do espaço por meio de funções de aquisição, reduzindo o número de avaliações necessárias sob determinadas condições (Shahriari *et al.*, 2016; Gramacy, 2020). O TPE modela densidades condicionais de regiões promissoras e tem sido amplamente empregado em sistemas de ML automatizado (*AutoML*) devido à flexibilidade frente a espaços mistos (Bergstra *et al.*, 2011). Em paralelo, a literatura também explora meta-heurísticas para ajuste de hiperparâmetros, como algoritmos genéticos e otimização por enxame de partículas, usualmente motivadas por espaços não diferenciáveis e pela necessidade de exploração global (Balestrassi *et al.*, 2009; Pontes *et al.*, 2016).

Em contraste com estratégias puramente estocásticas ou baseadas em modelos probabilísticos, métodos baseados em planejamento de experimentos (*Design of Experiments*, DoE) e Superfícies de Resposta (*Response Surface Methodology*, RSM) oferecem um arcabouço estatístico clássico para explorar sistematicamente o espaço de fatores, estimar efeitos principais e interações, e construir modelos aproximadores interpretáveis (Box; Wilson, 1951; Myers; Montgomery; Anderson-Cook, 2016; Montgomery, 2017). Embora tradicionalmente associadas à estatística industrial, tais técnicas têm sido empregadas como alternativa para otimização de parâmetros em modelos de ML e redes neurais artificiais, sobretudo quando o orçamento de avaliações é restrito e quando se deseja interpretar a

influência de fatores sobre respostas (Balestrassi *et al.*, 2009; Pontes *et al.*, 2016). Ainda assim, a integração entre Planejamento de Experimentos (*Design of Experiments*, DoE) síntese multivariada de métricas e uma estratégia multiobjetivo explícita para o ajuste de hiperparâmetros permanece um campo em evolução, com oportunidades para métodos que conciliem eficiência amostral, orientação de busca e explicabilidade.

Este trabalho propõe um método de ajuste de hiperparâmetros do XGBoost orientado por DoE e otimização multiobjetivo com o Método Interseção Normal à Fronteira (*Normal Boundary Intersection*, NBI). A proposta integra: (i) um delineamento experimental do tipo Delineamento Composto Central de Face Centrada (*Central Composite Design Face-Centered*, CCDFC) para exploração estruturada do espaço de hiperparâmetros; (ii) modelagem por RSM para aproximação das respostas observadas; (iii) síntese de múltiplas métricas em escores latentes via Análise de Componentes Principais (*Principal Components Analysis*, PCA) com rotação *Varimax* (Kaiser, 1958), visando reduzir colinearidades e produzir interpretações de fatores (Kaiser, 1958); e (iv) geração de soluções de compromisso via NBI que é uma estratégia sistemática para amostrar a fronteira de Pareto em problemas multiobjetivo (Das; Dennis, 1998; Lopes *et al.*, 2016; Naves *et al.*, 2017). Ao final, a avaliação é conduzida sob orçamento computacional controlado e equiparado às abordagens de referência, permitindo comparar desempenho e custo em termos justos. Embora o detalhamento metodológico seja apresentado nos capítulos subsequentes, a motivação central é estabelecer uma alternativa orientada e explicável às estratégias exaustivas, aleatórias ou puramente probabilísticas, preservando competitividade em qualidade e ampliando eficiência em custo computacional.

1.1 Contexto de Pesquisa

A seleção de algoritmos e o ajuste de hiperparâmetros constituem etapas determinantes do desempenho em aplicações de ML. Em dados tabulares, a literatura recente aponta que métodos baseados em árvores permanecem competitivos e frequentemente superiores em diversas tarefas, seja por robustez a escalas e distribuições heterogêneas, seja por menor exigência de grandes volumes de dados e engenharia de arquiteturas quando comparados a abordagens profundas (Grinsztajn; Gabriel; Lee, 2022; Borisov *et al.*, 2023). Nesse cenário, o XGBoost destaca-se por apresentar bom desempenho, facilidade de integração em *pipelines* e mecanismos explícitos para controlar complexidade e regularização (Chen; Guestrin, 2016).

O ajuste de hiperparâmetros, entretanto, apresenta desafios específicos. Primeiro, o espaço A pode ser misto, envolvendo parâmetros contínuos, discretos e categóricos, além de restrições condicionais entre hiperparâmetros. Segundo, a função objetivo estimada $\hat{\mathcal{M}}$ costuma ser não convexa, ruidosa e cara, pois depende de treinamento repetido e validação; em muitos casos, inclui também pré-processamento e seleção de atributos. Terceiro, quando o objetivo envolve múltiplas métricas de qualidade, a decisão deve ser informada por critérios de compromisso, uma vez que maximizar uma única métrica pode gerar soluções indesejáveis sob outras medidas (Hastie; Tibshirani; Friedman, 2009; Morales-Hernández *et al.*, 2022). Assim, estratégias de ajuste devem equilibrar eficiência amostral, robustez a variabilidade experimental e adequação aos critérios de decisão relevantes para o problema.

A literatura consolidou famílias de métodos para esse problema. A busca em grade apresenta simplicidade e paralelização direta, mas escala mal com a dimensionalidade. A busca aleatória é simples e, em diversos cenários, explora com maior eficiência dimensões relevantes do espaço, especialmente quando a sensibilidade do desempenho se concentra em um subconjunto de hiperparâmetros (Bergstra; Bengio, 2012). Estratégias baseadas em modelos substitutos, como otimização bayesiana, reduzem o número de avaliações ao explorar um modelo probabilístico do desempenho e uma função de aquisição que equilibra exploração e aproveitamento; tais métodos têm apresentado bons resultados em problemas de ML, embora possam sofrer com custo de ajuste do substituto, sensibilidade a ruído e dificuldades em espaços de alta dimensionalidade ou com muitas variáveis condicionais (Shahriari *et al.*, 2016; Gramacy, 2020). O TPE, por sua vez, oferece flexibilidade para espaços estruturados e tem ampla adoção em bibliotecas práticas, como o *Hyperopt* (Bergstra *et al.*, 2011). Adicionalmente, meta-heurísticas e estratégias híbridas têm sido investigadas para lidar com espaços complexos e objetivos múltiplos, reforçando a natureza ativa do problema de pesquisa em torno de ajuste de hiperparâmetros (Balestrassi *et al.*, 2009; Ramamohan *et al.*, 2024).

Nesse panorama, o emprego de DoE e RSM apresenta diferenciais relevantes em cenários aplicados. O primeiro diferencial consiste na eficiência amostral, pois delineamentos estruturados permitem estimar efeitos principais, interações e curvaturas com um número controlado de execuções. O segundo diferencial é a interpretabilidade, já que modelos aproximadores podem ser analisados para identificar hiperparâmetros influentes e regiões promissoras do espaço de busca, apoiando decisões de engenharia e reduzindo dependência de amostragem puramente aleatória (Box; Wilson, 1951; Myers; Montgomery; Anderson-Cook, 2016; Montgomery, 2017). Trabalhos correlatos demonstram que delineamentos e superfícies

de resposta podem ser aplicados para otimizar parâmetros de redes neurais artificiais e métodos de predição, especialmente quando se busca reduzir custo de experimentação (Balestrassi *et al.*, 2009; Pontes *et al.*, 2016). Em paralelo, o uso do NBI em otimização multiresposta tem sido proposto para amostrar fronteiras eficientes de forma sistemática e com boa distribuição de soluções, inclusive com extensões que utilizam escores rotacionados para lidar com respostas correlacionadas (Das; Dennis, 1998; Lopes *et al.*, 2016; Naves *et al.*, 2017).

A validação empírica deste estudo é conduzida em um problema de classificação binária empregado como *benchmark*, derivado de dados do experimento MAGIC de telescópios Cherenkov, no qual se busca discriminar eventos de raios gama e *hádrons* (Bock *et al.*, 2004). A base apresenta características típicas de dados tabulares com atributos contínuos e ruído, exigindo protocolo de avaliação cuidadoso. Considerando que bases pequenas e balanceadas podem levar a desempenhos elevados mesmo sob ajustes simples, este trabalho adota replicações independentes e validação cruzada estratificada, além de análise estatística para investigar se diferenças observadas entre métodos são atribuíveis a variações sistemáticas ou a flutuações experimentais.

Como extensão, o mesmo *pipeline* foi aplicado a duas bases adicionais com foco em estabilidade e robustez fora do cenário controlado (Seção 4.9)

1.2 Objetivos

Esta pesquisa tem como objetivo geral propor e avaliar um método sistemático de ajuste de hiperparâmetros para o XGBoost, formulado como um problema multiobjetivo que considera simultaneamente qualidade preditiva e custo computacional. Para alcançar esse objetivo, estabelecem-se os objetivos específicos descritos a seguir.

1.2.1 Objetivo Geral

Propor, implementar e validar um método de ajuste de hiperparâmetros do XGBoost orientado por DoE e otimização multiobjetivo com NBI, visando identificar soluções de compromisso entre desempenho preditivo e custo computacional sob orçamento de avaliação controlado.

1.2.2 Objetivos Específicos

1. Definir o espaço de busca Λ para um conjunto representativo de hiperparâmetros do XGBoost e construir um delineamento experimental CCDFC para explorar esse espaço de forma estruturada;
2. Executar o delineamento experimental sob validação cruzada estratificada k -fold, registrando métricas de qualidade (acurácia, precisão, revocação e especificidade) e métricas de custo (tempo médio por *fold*, tempo total de otimização e tempo total de execução);
3. Aplicar padronização das respostas e síntese multivariada via PCA com rotação *Varimax* (Kaiser, 1958), obtendo escores latentes que representem, de forma compacta e interpretável, dimensões de qualidade e custo;
4. Ajustar modelos quadráticos de RSM para os escores latentes e realizar análise de efeitos principais com PCA, interações e termos quadráticos; interpretando a influência dos hiperparâmetros;
5. Empregar o método NBI para gerar um conjunto de configurações candidatas na fronteira de Pareto aproximada e selecionar uma solução de compromisso segundo critérios de decisão previamente definidos;
6. Comparar o método proposto com abordagens de referência de ajuste, incluindo busca em grade, busca aleatória, otimização bayesiana e TPE (com a disponibilização na biblioteca *Hyperopt*), utilizando orçamento de avaliações equiparado;
7. Realizar experimentos de confirmação e análise estatística baseada em múltiplas replicações independentes com 30 sementes pseudoaleatórias para verificar robustez e variabilidade, com foco em diferenças de custo computacional e em possíveis diferenças de desempenho;
8. Disponibilizar artefatos de reprodutibilidade, incluindo código, configurações experimentais e relatórios de saída, visando facilitar replicação e extensão do estudo.

1.3 Contribuições Esperadas

As contribuições esperadas deste trabalho são:

1. Proposição de um procedimento de ajuste de hiperparâmetros baseado em DoE e RSM, integrando síntese multivariada (PCA com rotação *Varimax*) e otimização multiobjetivo por NBI, com foco em eficiência amostral e interpretação de efeitos;

2. Formulação explícita do ajuste como problema multiobjetivo qualidade versus custo, com geração de soluções de compromisso e discussão da fronteira eficiente em cenários com objetivos conflitantes (Deb, 2001; Das; Dennis, 1998; Morales-Hernández *et al.*, 2022);
3. Aplicação do método NBI em um contexto de ajuste de hiperparâmetros, incorporando a ideia de uso de escores rotacionados como respostas multivariadas para robustez e interpretação, conforme abordagens correlatas em otimização multiresposta (Lopes *et al.*, 2016; Naves *et al.*, 2017);
4. Avaliação experimental sob orçamento computacional justo em comparação com métodos de referência, contribuindo para evidências sobre custo e desempenho de estratégias de ajuste em um cenário padronizado de classificação binária com XGBoost;
5. Produção de artefatos e procedimentos reprodutíveis para execução de experimentos, com exportação de resultados intermediários e finais, permitindo auditoria e reuso em cenários futuros.

1.4 Delimitações da Pesquisa

Esta pesquisa está delimitada aos seguintes aspectos:

1. Considera-se um problema de classificação binária em base de dados tabular derivada do experimento MAGIC, com pré-processamento incluindo balanceamento de classes por subamostragem (*undersampling*), de modo a mitigar efeitos de desbalanceamento na avaliação (Bock *et al.*, 2004). O protocolo experimental foi replicado em outras duas bases públicas para avaliação de estabilidade;
2. A investigação restringe-se ao algoritmo XGBoost no contexto de árvores de decisão por *gradient boosting*, não abrangendo outros modelos de aprendizado supervisionado, tais como redes neurais artificiais profundas, máquinas de vetor de suporte ou métodos lineares (Friedman, 2001; Chen; Guestrin, 2016);
3. O espaço de busca Λ é composto por um conjunto específico de hiperparâmetros selecionados por relevância prática e por representarem compromissos conhecidos entre desempenho e custo, não contemplando a totalidade de parâmetros disponíveis;
4. O orçamento de avaliações para comparação com métodos de referência é controlado e equiparado ao total de avaliações realizadas pelo método proposto, assegurando comparabilidade em termos de custo experimental;

5. As conclusões fundamentam-se em experimentos executados em ambiente computacional específico e em um número finito de replicações independentes (30 sementes); assim, a generalização para outros ambientes, bases, algoritmos, tamanhos amostrais, regimes de pré-processamento ou cenários multiobjetivo adicionais constitui tópico de investigação futura.

1.5 Estrutura do Trabalho

Além desta introdução, o Capítulo 2 apresenta o referencial teórico, abrangendo fundamentos do XGBoost, métricas de avaliação, abordagens de ajuste de hiperparâmetros e princípios de planejamento de experimentos aplicados à otimização. O Capítulo 3 descreve a metodologia proposta, detalhando o delineamento CCDFC, a síntese de métricas por PCA com rotação *Varimax* (Kaiser, 1958), a modelagem por RSM e a geração de soluções por NBI. O Capítulo 4 apresenta os resultados e discussões, incluindo comparações com métodos de referência e análise estatística sobre custo computacional e desempenho. Por fim, o Capítulo 5 sistematiza as conclusões e indica direções para trabalhos futuros.

2. REFERENCIAL TEÓRICO

2.1 *Machine learning*, hiperparâmetros e motivação metodológica

A evolução recente de ML consolidou a posição de algoritmos supervisionados em aplicações críticas (diagnóstico médico, análise de crédito, detecção de fraudes, previsão de séries temporais). Em classificação binária, parte-se de um conjunto rotulado de observações (x_i, y_i) , com $y_i \in \{0,1\}$, buscando-se um classificador $h(\cdot)$ que maximize o desempenho preditivo sobre dados não vistos. Para além do desenvolvimento de novos algoritmos, a literatura reconhece que a seleção de hiperparâmetros é um determinante primário do desempenho, pois controla a capacidade de generalização, o grau de regularização, a estrutura do modelo e o custo computacional de treino e predição (Probst; Wright; Boulesteix, 2019).

Hiperparâmetros são variáveis de decisão extrínsecas ao problema de estimação paramétrica: diferentemente de parâmetros ajustados durante o treinamento (e.g., pesos de redes neurais, coeficientes de regressão), são definidos *ex ante* e moldam o regime de aprendizado. Em *ensembles* baseados em árvores de decisão, como *XGBoost*, hiperparâmetros regulam a profundidade e o número de árvores, a taxa de aprendizado, a amostragem de instâncias e de atributos, bem como penalizações de complexidade. Esse conjunto determina o *trade-off* viés-variância e o tempo de execução por configuração (Chen; Guestrin, 2016).

A inexistência de uma configuração universalmente ótima; consequência do *no free lunch theorem* (Wolpert; Macready, 1997), justifica o esforço em *hyperparameter tuning*. O problema apresenta três características relevantes: (i) não convexidade e frequente rugosidade da superfície de resposta (métricas de avaliação em função dos hiperparâmetros); (ii) custo de avaliação não negligenciável (cada ponto testado implica treinar/validar um modelo); e (iii) potencial multiobjetividade (maximização de métricas da matriz de confusão (*confusion matrix*)) e minimização de tempo de execução e, quando pertinente, de viés algorítmico (*algorithmic bias*). Esses elementos demandam estratégias de busca que combinem eficiência estatística, exploração inteligente do espaço e controle de custo computacional.

Em dados tabulares, análises recentes indicam que modelos baseados em árvores e, em particular, *ensembles* do tipo *gradient boosting*, tendem a apresentar desempenho superior ao de redes neurais profundas em grande parte dos cenários práticos, sobretudo sob orçamentos computacionais moderados e tamanhos de amostra típicos de aplicações industriais. Entre as razões apontadas estão a adequação indutiva aos atributos tabulares, a regularização implícita

e o custo de ajuste mais previsível (Grinsztajn; Gabriel; Lee, 2022; McElfresh *et al.*, 2023). Essa evidência empírica justifica o foco no *XGBoost* como algoritmo de referência para o estudo-piloto.

Para um panorama abrangente do desempenho de redes profundas em dados tabulares e suas limitações práticas frente a modelos baseados em árvores, ver também a síntese de Borisov *et al.* (2023).

2.2 Abordagens de busca de hiperparâmetros: *benchmarks* e panorama

A busca de hiperparâmetros consiste em selecionar, em um espaço de configurações Λ , o vetor $\lambda \in \Lambda$ capaz de maximizar o desempenho preditivo de um modelo, respeitando simultaneamente um protocolo de validação e uma restrição de custo computacional (Bergstra; Bengio, 2012; Feurer; Hutter, 2019). Em termos gerais, considera-se um algoritmo de aprendizado \mathcal{A} que, dada uma base de treinamento $\mathcal{D}_{\text{treino}}$ e um vetor de hiperparâmetros λ , produz um preditor $\hat{f}(\cdot; \lambda)$. A avaliação do desempenho do preditor depende de uma função de perda $\mathcal{L}(\cdot, \cdot)$ e de um esquema de validação \mathcal{V} , frequentemente instanciado por validação cruzada K -fold estratificada (Kohavi, 1995).

Uma formulação usual para o problema de otimização de hiperparâmetros é dada por:

$$\hat{\lambda} = \underset{\lambda \in \Lambda}{\operatorname{argmin}} f(\lambda) \quad (2.1)$$

Em que $f(\lambda)$ representa o risco estimado, ou seja, uma estimativa do erro de generalização condicionada ao protocolo \mathcal{V} . Para validação cruzada K -fold, um estimador típico assume a forma:

$$f(\lambda) = \frac{1}{K} \sum_{k=1}^K \mathcal{L}(\hat{f}^{(k)}(\cdot; \lambda), \mathcal{D}_{\text{val}}^{(k)}) \quad (2.2)$$

Em que $\mathcal{D}_{\text{val}}^{(k)}$ denota o conjunto de validação do k -ésimo fold e $\hat{f}^{(k)}$ é treinado no complemento $\mathcal{D}_{\text{treino}}^{(k)}$. Em muitos contextos, f é tratada como uma função “caixa-preta”, pois seu cálculo requer o treinamento e a validação do modelo e, portanto, possui custo elevado e ausência de gradientes analíticos, especialmente quando o treinamento inclui procedimentos não diferenciáveis ou aleatoriedade controlada por sementes (Shahriari *et al.*, 2016; Frazier, 2018).

Além da dimensão preditiva, a busca de hiperparâmetros envolve uma dimensão econômica. Seja $c(\lambda)$ o custo computacional de avaliar $f(\lambda)$, medido, por exemplo, em tempo de processamento. Define-se um orçamento de avaliações $B \in \mathbb{N}$, frequentemente imposto por

limitações de recursos, de modo que a estratégia de busca deve selecionar uma sequência $\{\lambda_t\}_{t=1}^B$ e retornar a melhor configuração observada:

$$\hat{\lambda} \in \underset{t \in \{1, \dots, B\}}{\operatorname{argmin}} f(\lambda_t) \quad (2.3)$$

Essa perspectiva explicita que comparações justas entre métodos requerem controle simultâneo do orçamento B e do protocolo \mathcal{V} , dado que ambos impactam diretamente custo e variabilidade (Bergstra; Bengio, 2012). Em particular, métodos de busca que avaliam a mesma quantidade de configurações, sob o mesmo esquema de validação, tornam comparáveis suas trajetórias de custo computacional e suas capacidades de explorar Λ .

No que se segue, apresentam-se quatro abordagens recorrentes em *benchmarks* de otimização de hiperparâmetros, a saber, busca em grade, busca aleatória, otimização bayesiana e TPE. Embora existam extensões relevantes, como estratégias multi-fidelidade e métodos de parada antecipada, as abordagens discutidas constituem referências clássicas e amplamente usadas como linhas de base, além de servirem como alicerce conceitual para métodos mais recentes (Feurer; Hutter, 2019; Akiba *et al.*, 2019).

2.2.1 Busca em grade

A busca em grade é um procedimento determinístico que discretiza o espaço Λ e avalia todas as combinações possíveis sobre uma grade. Considere-se d hiperparâmetros e um espaço fatorável $\Lambda = \Lambda_1 \times \dots \times \Lambda_d$, no qual Λ_j é o domínio do j -ésimo hiperparâmetro. Define-se, para cada dimensão, um conjunto finito de níveis:

$$\mathcal{G}_j = \{\lambda_{j,1}, \lambda_{j,2}, \dots, \lambda_{j,n_j}\}, \quad j = 1, \dots, d \quad (2.4)$$

Com n_j pontos. A grade completa é o produto cartesiano:

$$\mathcal{G} = \mathcal{G}_1 \times \dots \times \mathcal{G}_d, \quad |\mathcal{G}| = \prod_{j=1}^d n_j \quad (2.5)$$

A estratégia avalia $f(\lambda)$ para todo $\lambda \in \mathcal{G}$ e retorna o minimizador observado, conforme (2.3) com $B = |\mathcal{G}|$. Quando $n_j = n$ para todas as dimensões, a cardinalidade da grade cresce como n^d , caracterizando o efeito conhecido como maldição da dimensionalidade (Bergstra; Bengio, 2012).

Uma leitura útil da busca em grade, quando aplicada a hiperparâmetros, é interpretá-la como um delineamento fatorial completo sobre d fatores, em que cada hiperparâmetro é tratado como um fator e os níveis correspondem aos valores discretizados. Do ponto de vista de

planejamento de experimentos, tal delineamento maximiza a cobertura determinística, porém impõe custo proibitivo para d moderado e para resoluções finas (Montgomery, 2017). Em cenários de ML, nos quais cada avaliação envolve validação cruzada, o custo torna-se ainda mais sensível a d e a n .

Do ponto de vista operacional, a busca em grade possui três atributos que explicam sua recorrência em *benchmarks*. Em primeiro lugar, é simples de implementar, inclusive com paralelização direta, pois as avaliações são independentes. Em segundo lugar, é reprodutível, dado que a lista de pontos a avaliar é fixa e não depende de resultados intermediários, exceto pelo desempate de mínimos. Em terceiro lugar, quando o conjunto \mathcal{G} é pequeno, a busca exaustiva fornece uma referência empírica útil para validar modelos mais sofisticados.

Entretanto, suas limitações são igualmente bem estabelecidas. A primeira limitação é o crescimento combinatório de $|\mathcal{G}|$, que torna a abordagem impraticável para espaços com múltiplos hiperparâmetros contínuos. A segunda limitação é a alocação ineficiente do orçamento: se apenas um subconjunto dos hiperparâmetros possui grande influência sobre f , a grade desperdiça avaliações em dimensões pouco relevantes, pois distribui pontos uniformemente em todas as direções (Bergstra; Bengio, 2012). A terceira limitação é a dependência da parametrização: alterações na escala ou nos limites dos hiperparâmetros podem induzir grades com densidades efetivas diferentes, influenciando resultados sem alterar o método subjacente.

Como consequência, lacunas comuns em comparações baseadas em busca em grade incluem a sensibilidade a escolhas *ad hoc* de limites e resoluções, além da ausência de um mecanismo interno de adaptação. Em particular, a busca em grade não incorpora informação proveniente das avaliações realizadas para priorizar regiões promissoras, sendo, portanto, um método de exploração sem orientação por desempenho. Em ambientes em que o custo computacional é um critério relevante, a busca em grade pode tornar-se uma linha de base pouco competitiva, ainda que apresente desempenho preditivo adequado quando o orçamento permite cobertura ampla.

2.2.2 Busca aleatória

A busca aleatória substitui a discretização exaustiva por amostragem aleatória no espaço Λ e, sob um orçamento B , avalia B configurações independentes. Seja $p(\lambda)$ uma distribuição

de probabilidade definida em Λ , escolhida para refletir conhecimento prévio ou hipóteses sobre escalas e regiões plausíveis. A estratégia amostra:

$$\lambda_t \sim p(\lambda), \quad t = 1, \dots, B \quad (2.6)$$

Avalia $y_t = f(\lambda_t)$ e retorna a melhor configuração observada, conforme (2.3). A escolha de $p(\lambda)$ é parte integrante do método, pois define a priorização estatística de regiões e pode incorporar distribuições log-uniformes, truncadas ou hierárquicas, particularmente adequadas quando hiperparâmetros assumem ordens de magnitude distintas (Bergstra; Bengio, 2012).

Um argumento matemático central para a busca aleatória pode ser expresso por estatísticas de ordem. Seja $Y = f(\Lambda)$ uma variável aleatória definida pela amostragem $\Lambda \sim p$, com função de distribuição acumulada F_Y . Assumindo amostras independentes, o melhor valor observado após B avaliações, para um problema de minimização, é $Y_{(1)} = \min\{Y_1, \dots, Y_B\}$. Sua distribuição é dada por:

$$\mathbb{P}(Y_{(1)} \leq y) = 1 - (1 - F_Y(y))^B \quad (2.7)$$

A Equação (2.7) evidencia que o ganho marginal de novas avaliações depende da massa de probabilidade de F_Y na região de valores baixos: se existe uma região de Λ com probabilidade não desprezível de produzir perdas pequenas, a probabilidade de observá-la cresce rapidamente com B . Em contraste com a busca em grade, a busca aleatória não desperdiça avaliações por impor cobertura regular em dimensões irrelevantes; em particular, quando apenas $d' \ll d$ hiperparâmetros controlam a maior parte da variabilidade de f , a amostragem aleatória tende a alocar mais avaliações a combinações efetivamente distintas nesses hiperparâmetros, ao passo que a grade aloca avaliações redundantes em direções pouco influentes (Bergstra; Bengio, 2012).

Entre as vantagens, destacam-se simplicidade, baixo custo de coordenação e paralelização imediata. Além disso, a busca aleatória é naturalmente compatível com espaços mistos, nos quais parte dos hiperparâmetros é discreta ou categórica, desde que p seja definido adequadamente. Por fim, pode ser vista como um *benchmark* de exploração estocástica, útil para caracterizar o efeito de variabilidade entre réplicas e sementes.

Todavia, suas limitações derivam da ausência de um mecanismo de aprendizado a partir das avaliações. A busca aleatória explora Λ sem construir uma representação explícita da relação entre hiperparâmetros e desempenho, o que a torna ineficiente quando as avaliações são muito custosas e o orçamento é restrito. Além disso, a qualidade dos resultados é sensível à escolha de p e, particularmente, à parametrização de escalas: amostragem uniforme em

intervalos lineares pode ser inadequada para hiperparâmetros cujo efeito ocorre em escala logarítmica, a exemplo de taxas de aprendizado ou regularizações (Bergstra; Bengio, 2012). Em síntese, a busca aleatória melhora a eficiência frente à busca em grade sob orçamentos fixos, mas não ataca a lacuna de orientação, permanecendo essencialmente um método de exploração não adaptativa.

2.2.3 Otimização bayesiana

A otimização bayesiana, no contexto de busca de hiperparâmetros, pode ser compreendida como uma instância de otimização sequencial baseada em modelos (*sequential model-based optimization*, SMBO), na qual se constrói um modelo probabilístico substituto para a função “caixa-preta” f e se escolhe iterativamente o próximo ponto a ser avaliado por meio de uma função de aquisição (Jones; Schonlau; Welch, 1998; Snoek; Larochelle; Adams, 2012). Denote-se por:

$$\mathcal{D}_t = \{(\boldsymbol{\lambda}_i, y_i)\}_{i=1}^t, \quad y_i = f(\boldsymbol{\lambda}_i) \quad (2.8)$$

O conjunto de observações após t avaliações. A ideia central é inferir uma distribuição a posteriori para f condicionada a \mathcal{D}_t e, em seguida, selecionar $\boldsymbol{\lambda}_{t+1}$ maximizando um critério de utilidade que balanceia exploração e exploração dirigida.

Uma escolha clássica para o modelo substituto é o processo gaussiano (Rasmussen; Williams, 2006; Snoek; Larochelle; Adams, 2012). Assume-se um prior:

$$f(\boldsymbol{\lambda}) \sim \mathcal{GP}(m(\boldsymbol{\lambda}), k(\boldsymbol{\lambda}, \boldsymbol{\lambda}')) \quad (2.9)$$

Em que m é a função média e k é o núcleo de covariância. Considerando observações com ruído gaussiano $\varepsilon \sim \mathcal{N}(0, \sigma_n^2)$, tem-se $y_i = f(\boldsymbol{\lambda}_i) + \varepsilon_i$. Definindo $\mathbf{y}_t = [y_1, \dots, y_t]^\top$ e a matriz de Gram $\mathbf{K}_t \in \mathbb{R}^{t \times t}$ com entradas $(\mathbf{K}_t)_{ij} = k(\boldsymbol{\lambda}_i, \boldsymbol{\lambda}_j) + \sigma_n^2 \delta_{ij}$, a distribuição preditiva a posteriori em um ponto candidato $\boldsymbol{\lambda}$ é gaussiana, com média e variância:

$$\mu_t(\boldsymbol{\lambda}) = m(\boldsymbol{\lambda}) + \mathbf{k}_t(\boldsymbol{\lambda})^\top \mathbf{K}_t^{-1} (\mathbf{y}_t - \mathbf{m}_t) \quad (2.10)$$

$$\sigma_t^2(\boldsymbol{\lambda}) = k(\boldsymbol{\lambda}, \boldsymbol{\lambda}) - \mathbf{k}_t(\boldsymbol{\lambda})^\top \mathbf{K}_t^{-1} \mathbf{k}_t(\boldsymbol{\lambda}) \quad (2.11)$$

Em que $\mathbf{k}_t(\boldsymbol{\lambda}) = [k(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}), \dots, k(\boldsymbol{\lambda}_t, \boldsymbol{\lambda})]^\top$ e $\mathbf{m}_t = [m(\boldsymbol{\lambda}_1), \dots, m(\boldsymbol{\lambda}_t)]^\top$.

A função de aquisição $a_t(\boldsymbol{\lambda})$ é construída a partir de μ_t e σ_t e determina o próximo ponto a avaliar. Três escolhas canônicas são probabilidade de melhoria (*Probability of Improvement*, PI), melhoria esperada (*Expected Improvement*, EI) e limite inferior e superior de confiança (Jones; Schonlau; Welch, 1998; Shahriari *et al.*, 2016; Frazier, 2018). Para um problema de minimização e definindo $y_t^* = \min_{i \leq t} y_i$, obtém-se:

$$PI_t(\boldsymbol{\lambda}) = \mathbb{P}(f(\boldsymbol{\lambda}) \leq y_t^* - \xi \mid \mathcal{D}_t) = \Phi\left(\frac{y_t^* - \xi - \mu_t(\boldsymbol{\lambda})}{\sigma_t(\boldsymbol{\lambda})}\right) \quad (2.12)$$

Em que Φ é a função de distribuição acumulada da normal padrão e $\xi \geq 0$ é um parâmetro de exploração. A melhoria esperada, por sua vez, é:

$$EI_t(\boldsymbol{\lambda}) = \mathbb{E}[\max(0, y_t^* - \xi - f(\boldsymbol{\lambda})) \mid \mathcal{D}_t] \quad (2.13)$$

Sob a hipótese gaussiana, (2.13) admite forma fechada:

$$EI_t(\boldsymbol{\lambda}) = (y_t^* - \xi - \mu_t) \Phi(z) + \sigma_t \phi(z), \quad z = \frac{y_t^* - \xi - \mu_t}{\sigma_t} \quad (2.14)$$

Em que ϕ é a densidade da normal padrão e, para simplicidade, omite-se a dependência explícita de μ_t e σ_t em $\boldsymbol{\lambda}$. Por fim, a estratégia de limite de confiança inferior (*Lower Control Boundary*, LCB), utilizada em problemas de minimização, é escrita como:

$$LCB_t(\boldsymbol{\lambda}) = \mu_t(\boldsymbol{\lambda}) - \kappa \sigma_t(\boldsymbol{\lambda}) \quad (2.15)$$

Em que $\kappa > 0$ controla o compromisso exploração-exploração dirigida. A regra de seleção pode ser expressa genericamente como:

$$\boldsymbol{\lambda}_{t+1} = \operatorname{argmax}_{\boldsymbol{\lambda} \in \Lambda} a_t(\boldsymbol{\lambda}) \quad (2.16)$$

Os benefícios da otimização bayesiana decorrem de sua eficiência amostral: ao modelar explicitamente f , o método tende a concentrar avaliações em regiões promissoras, o que é particularmente vantajoso quando cada avaliação é custosa (Snoek; Larochelle; Adams, 2012). Além disso, a estrutura probabilística permite lidar com ruído de avaliação e com incerteza, incorporando o papel da variabilidade introduzida por validação cruzada e por aleatoriedade controlada (Shahriari *et al.*, 2016). Revisões recentes destacam ainda o amadurecimento de estratégias de aquisição, mecanismos de paralelização e extensões para múltiplos objetivos e restrições (Frazier, 2018; Wistuba; Schilling; Schmidt-Thieme, 2022).

Em contrapartida, a otimização bayesiana apresenta limitações relevantes. A primeira é o custo computacional do modelo substituto, que pode tornar-se significativo à medida que t cresce, sobretudo em processos gaussianos, cuja inversão matricial, em formulações diretas, escala como $\mathcal{O}(t^3)$ (Rasmussen; Williams, 2006). A segunda limitação é a adequação do modelo aos dados: a qualidade das recomendações depende de hipóteses sobre suavidade e estrutura de f , e tais hipóteses podem ser violadas em espaços com hiperparâmetros discretos, condicionais ou com regiões não estacionárias. A terceira é a dificuldade de escalabilidade para alta dimensionalidade e para espaços heterogêneos, nos quais a otimização da função de aquisição se torna um problema não convexo de dimensão comparável à original.

Em síntese, a otimização bayesiana é frequentemente superior a estratégias puramente exploratórias sob orçamentos restritos, mas seu desempenho pode degradar em espaços

complexos e mistos. Tais limitações motivam variantes que substituem processos gaussianos por modelos mais robustos a variáveis categóricas e condicionais, como florestas aleatórias ou estimadores de densidade do tipo *Parzen* (Feurer; Hutter, 2019).

2.2.4 Hyperopt e o estimador de Parzen em árvore

O TPE é uma abordagem de otimização bayesiana desenvolvida para lidar com espaços de busca hierárquicos e com hiperparâmetros condicionais, típicos de *pipelines* de ML (Bergstra *et al.*, 2011). Diferentemente da formulação clássica em que se modela $p(y | \lambda)$ com processos gaussianos, o TPE modela a densidade $p(\lambda | y)$ e utiliza a decomposição de Bayes:

$$p(y | \lambda) = \frac{p(\lambda | y) p(y)}{p(\lambda)} \quad (2.17)$$

Dadas observações \mathcal{D}_t conforme (2.8), define-se um limiar y^* como um quantil $\gamma \in (0,1)$ da amostra $\{y_i\}_{i=1}^t$, de modo que:

$$\mathbb{P}(y < y^*) = \gamma \quad (2.18)$$

A partir desse limiar, particiona-se o conjunto de observações em duas regiões e define-se:

$$\ell(\lambda) = p(\lambda | y < y^*), \quad g(\lambda) = p(\lambda | y \geq y^*) \quad (2.19)$$

Por construção, tem-se:

$$p(\lambda) = \gamma \ell(\lambda) + (1 - \gamma) g(\lambda) \quad (2.20)$$

A conexão com a melhoria esperada pode ser demonstrada ao se escrever EI_t em termos de ℓ e g . Em particular, sob a modelagem proposta em Bergstra *et al.* (2011), maximizar a melhoria esperada é equivalente a maximizar a razão $\ell(\lambda)/g(\lambda)$ em um conjunto de candidatos. De forma intuitiva, ℓ concentra massa de probabilidade em regiões associadas a perdas menores, enquanto g concentra massa nas demais regiões. Assim, valores altos de ℓ/g indicam configurações com alta densidade sob observações boas e baixa densidade sob observações ruins.

Na prática, o procedimento itera entre duas etapas: estimar ℓ e g por métodos de estimação de densidade de Parzen (frequentemente por estimadores do tipo *kernel* ou histogramas adaptativos) e, em seguida, amostrar um conjunto de candidatos $\{\lambda^{(s)}\}_{s=1}^S \sim \ell$ e selecionar aquele que maximiza $\ell(\lambda^{(s)})/g(\lambda^{(s)})$ (Bergstra *et al.*, 2011). O adjetivo *tree-structured* decorre do fato de que o espaço Λ pode ser descrito por uma estrutura em árvore, na qual escolhas de hiperparâmetros ativam ou desativam subespaços, de modo que a distribuição $p(\lambda)$ é construída de forma condicional e composicional.

A biblioteca *Hyperopt* implementa o TPE e fornece uma linguagem de especificação de espaços hierárquicos, além de rotinas de otimização e paralelização (Bergstra *et al.*, 2013). No contexto de busca de hiperparâmetros, o *Hyperopt* é particularmente atrativo por permitir hiperparâmetros discretos, contínuos e condicionais em uma mesma especificação, evitando adaptações artificiais que seriam necessárias em implementações puramente baseadas em processos gaussianos.

Entre os aspectos positivos, destacam-se a flexibilidade para espaços mistos e hierárquicos e a eficiência em orçamentos moderados, quando comparado a estratégias sem modelo. Adicionalmente, a estratégia de modelar $p(\lambda | y)$ tende a ser robusta quando f é não suave e quando hiperparâmetros são discretos ou possuem restrições condicionais fortes.

Entretanto, o TPE também possui limitações. A qualidade do método depende da estimação de densidades, que pode degradar em alta dimensionalidade ou quando hiperparâmetros são fortemente correlacionados. Além disso, escolhas como o quantil γ , o número de candidatos S e a forma do estimador de densidade impactam o comportamento de exploração e podem demandar calibração. Por fim, embora o método explore informação acumulada, permanece no regime de otimização “caixa-preta” (*black-box*), no sentido de que não produz, por si só, um modelo explicativo interpretável da relação entre hiperparâmetros e respostas, aspecto particularmente relevante quando se deseja compreender mecanismos e realizar inferências além da seleção do melhor ponto.

2.2.5 Síntese e lacunas

A busca em grade e a busca aleatória constituem linhas de base fundamentais por sua simplicidade, baixo acoplamento e facilidade de paralelização. Contudo, ambas carecem de mecanismos explícitos de aprendizagem sobre f : a primeira por ser exaustiva em uma discretização fixa e a segunda por ser puramente amostral. A otimização bayesiana, por sua vez, introduz orientação por meio de um modelo substituto e de funções de aquisição, alcançando eficiência amostral superior em muitos cenários, mas com restrições de escalabilidade e de adequação a espaços mistos. O TPE pode ser interpretado como uma adaptação da otimização bayesiana para espaços hierárquicos, favorecendo aplicações com hiperparâmetros condicionais, ao custo de dependência de estimação de densidades.

Em termos de lacunas, observa-se que tais abordagens são predominantemente formuladas como problemas mono-objetivo, nos quais se minimiza uma perda escalar (ou se

maximiza uma métrica única), tratando custo computacional apenas como restrição indireta via orçamento B . Além disso, mesmo quando eficientes para seleção de pontos, tais métodos tendem a produzir explicabilidade limitada no sentido estatístico-experimental, pois não estruturam deliberadamente o conjunto de avaliações para inferir efeitos principais, interações e relações funcionais entre hiperparâmetros e respostas. Esse ponto é particularmente relevante em cenários nos quais o objetivo não é apenas obter um bom ponto, mas também justificar a escolha com base em princípios de planejamento e inferência.

2.3 Extreme Gradient Boosting

O XGBoost é um algoritmo de *boosting* baseado em árvores de decisão, proposto com ênfase em escalabilidade e eficiência computacional, mantendo rigor estatístico por meio de uma função objetivo regularizada. A formulação do XGBoost pode ser interpretada como um caso particular de *gradient boosting* para modelos aditivos, em que cada estimador base é uma árvore de regressão (Friedman, 2001; Chen; Guestrin, 2016). Nesta dissertação, o XGBoost é empregado como modelo preditivo central, cujos hiperparâmetros são otimizados nos experimentos descritos nos capítulos subsequentes.

2.3.1 Notação e modelo aditivo por árvores

Considere um conjunto de dados supervisionados:

$$D = \{(\mathbf{x}_i, y_i)\}_{i=1}^n, \quad \mathbf{x}_i \in \mathbb{R}^p, \quad y_i \in \mathcal{Y}. \quad (2.21)$$

No caso binário, tem-se tipicamente $\mathcal{Y} = \{0,1\}$. O XGBoost constrói um preditor aditivo composto por T árvores de regressão, de modo que a predição no espaço do *score* (ou margem) é dada por:

$$\hat{y}_i = \sum_{t=1}^T f_t(\mathbf{x}_i), \quad f_t \in \mathcal{F} \quad (2.22)$$

O conjunto \mathcal{F} representa o espaço de árvores de regressão. Seguindo Chen e Guestrin (2016), uma árvore f pode ser descrita por (i) uma função de estrutura $q(\cdot)$, que mapeia cada instância para uma das folhas, e (ii) um vetor de pesos (valores preditivos) das folhas. Assim:

$$f(\mathbf{x}) = w_{q(\mathbf{x})}, \quad q: \mathbb{R}^p \rightarrow \{1, 2, \dots, T_f\}, \quad \mathbf{w} \in \mathbb{R}^{T_f} \quad (2.23)$$

Em que T_f denota o número de folhas da árvore f e w_j é o valor predito pela folha j .

Em problemas de classificação binária, é comum converter a margem \hat{y}_i em probabilidade via função logística (sigmoide):

$$\hat{p}_i = \sigma(\hat{y}_i) = \frac{1}{1 + \exp(-\hat{y}_i)} \quad (2.24)$$

2.3.2 Função objetivo regularizada

O treinamento do XGBoost busca minimizar uma função objetivo composta por um termo de ajuste aos dados (derivado de uma perda) e por um termo de regularização de complexidade do conjunto de árvores (Chen; Guestrin, 2016). Denotando por $\ell(\cdot, \cdot)$ a função de perda, a função objetivo global é:

$$\mathcal{L}(\{f_t\}) = \sum_{i=1}^n \ell(y_i, \hat{y}_i) + \sum_{t=1}^T \Omega(f_t) \quad (2.25)$$

O termo $\Omega(f)$ penaliza a complexidade da árvore, controlando sobreajuste e favorecendo soluções mais parcimoniosas. Uma forma amplamente utilizada (com penalização L_2) é:

$$\Omega(f) = \gamma T_f + \frac{1}{2} \lambda \sum_{j=1}^{T_f} w_j^2, \quad \gamma \geq 0, \lambda \geq 0 \quad (2.26)$$

Em implementações modernas, pode-se ainda adicionar um termo L_1 ($\alpha \geq 0$) para promover esparsidade nos pesos das folhas:

$$\Omega(f) = \gamma T_f + \frac{1}{2} \lambda \sum_{j=1}^{T_f} w_j^2 + \alpha \sum_{j=1}^{T_f} |w_j| \quad (2.27)$$

A presença explícita de γ e λ na função objetivo diferencia o XGBoost de variantes clássicas de *boosting* com árvores, ao incorporar penalizações diretas sobre a estrutura e sobre a magnitude das predições nas folhas (Chen; Guestrin, 2016).

Para a classificação binária, uma escolha frequente de perda é a *log-loss* (entropia cruzada), definida em função da probabilidade prevista \hat{p}_i por:

$$\ell(y_i, \hat{p}_i) = -[y_i \log(\hat{p}_i) + (1 - y_i) \log(1 - \hat{p}_i)] \quad (2.28)$$

2.3.3 Otimização incremental via aproximação de segunda ordem

O XGBoost constrói o conjunto de árvores de forma incremental. Suponha que, após $(t - 1)$ iterações, a predição seja $\hat{y}_i^{(t-1)}$. Ao adicionar uma nova árvore f_t , obtém-se:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta f_t(\mathbf{x}_i), \quad 0 < \eta \leq 1 \quad (2.29)$$

Em que η é a taxa de aprendizado (*learning rate*), responsável por “encolher” a contribuição de cada árvore e, em geral, melhorar a capacidade de generalização às custas de um maior número de iterações (Friedman, 2001; Chen; Guestrin, 2016).

A estratégia central do XGBoost consiste em aproximar a variação da função de perda ao incluir f_t por meio de uma expansão de Taylor de segunda ordem ao redor de $\hat{y}_i^{(t-1)}$. Definindo:

$$g_i = \frac{\partial \ell(y_i, \hat{y}_i)}{\partial \hat{y}_i} \Big|_{\hat{y}_i = \hat{y}_i^{(t-1)}} \quad h_i = \frac{\partial^2 \ell(y_i, \hat{y}_i)}{\partial \hat{y}_i^2} \Big|_{\hat{y}_i = \hat{y}_i^{(t-1)}} \quad (2.30)$$

Pode-se escrever o objetivo aproximado para a iteração t como:

$$\tilde{\mathcal{L}}^{(t)} = \sum_{i=1}^n \left[g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t(\mathbf{x}_i)^2 \right] + \Omega(f_t) + \text{cte} \quad (2.31)$$

O termo “cte” não depende de f_t e, portanto, não afeta a minimização em relação à nova árvore.

2.3.4 Pesos ótimos de folhas e ganho de divisão

Como $f_t(\mathbf{x}) = w_{q(\mathbf{x})}$, pode-se agrupar as instâncias por folha. Seja:

$$I_j = \{i \in \{1, \dots, n\} \mid q(\mathbf{x}_i) = j\}, \quad j = 1, \dots, T_f \quad (2.32)$$

E defina os somatórios de gradientes e hessianas na folha j por:

$$G_j = \sum_{i \in I_j} g_i, \quad H_j = \sum_{i \in I_j} h_i \quad (2.33)$$

Substituindo em (2.31) e considerando, para fins de derivação fechada, a regularização quadrática (sem L_1), obtém-se:

$$\tilde{\mathcal{L}}^{(t)} = \sum_{j=1}^{T_f} \left[G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2 \right] + \gamma T_f + \text{cte} \quad (2.34)$$

A minimização de (2.34) em relação a cada w_j fornece o peso ótimo de folha:

$$w_j^* = -\frac{G_j}{H_j + \lambda} \quad (2.35)$$

Ao substituir (2.35) em (2.34), obtém-se a melhor (menor) contribuição possível da estrutura $q(\cdot)$ da árvore, expressa por:

$$\tilde{\mathcal{L}}^{(t)}(q) = -\frac{1}{2} \sum_{j=1}^{T_f} \frac{G_j^2}{H_j + \lambda} + \gamma T_f + \text{cte} \quad (2.36)$$

Conseqüentemente, a construção da árvore pode ser realizada por busca gulosa de divisões que maximizem a redução de perda aproximada (*gain*). Para uma divisão candidata que separa um nó pai em dois nós filhos (esquerda e direita), define-se o ganho:

$$\text{Gain} = \frac{1}{2} \left(\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{G_P^2}{H_P + \lambda} \right) - \gamma \quad (2.37)$$

Em que P denota o nó pai e L e R os nós filhos, com (G, H) definidos por somas de (g_i, h_i) sobre as instâncias pertencentes a cada região. A penalização γ reduz o ganho e atua como mecanismo de poda (*prunning*) baseado em complexidade: divisões que não geram redução suficiente no objetivo são descartadas (Chen; Guestrin, 2016).

2.3.5 Considerações práticas e relação com hiperparâmetros

A formulação anterior evidencia três componentes fundamentais do XGBoost:

- Função de perda e derivadas: a escolha de ℓ determina g_i e h_i em (2.30) e, portanto, influencia diretamente o critério de divisão e os pesos ótimos das folhas;
- Regularização explícita: γ controla a complexidade estrutural (número de folhas), enquanto λ (e, opcionalmente, α) controla a magnitude e esparsidade das predições nas folhas;
- Atualização incremental com encolhimento: a taxa η modula o tamanho do passo em (2.39), estabelecendo um *trade-off* entre capacidade do modelo e estabilidade da generalização.

Além disso, o XGBoost incorpora estratégias computacionais que permitem escalabilidade em bases volumosas, incluindo paralelismo no cálculo de estatísticas de divisão, tratamento explícito de valores ausentes e algoritmos aproximados para encontrar divisões sob restrições de memória (Chen; Guestrin, 2016). Tais mecanismos, embora relevantes para desempenho computacional, não alteram a estrutura matemático-estatística do objetivo em (2.35) - (2.37).

Finalmente, observa-se que diversos parâmetros operacionais do XGBoost afetam diretamente: (i) a estrutura da árvore (profundidade e critérios de divisão), (ii) o número de folhas, (iii) a estabilidade do par (T, η) e (iv) a qualidade da solução por meio da estocasticidade de amostragem e da regularização dos pesos de folha. Esses aspectos são retomados na Seção 2.4 e nos experimentos conduzidos nesta pesquisa.

2.4 Hiperparâmetros do XGBoost: papel, efeitos e compromissos (*trade-offs*)

A construção do modelo no algoritmo XGBoost é controlada por hiperparâmetros que regulam, simultaneamente, (i) a capacidade representacional do conjunto de árvores e (ii) os mecanismos de regularização estatística e computacional do processo de treinamento. Em termos formais, considere-se um conjunto de treinamento $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$, com $\mathbf{x}_i \in \mathbb{R}^p$ e $y_i \in \{0,1\}$ no caso binário. O XGBoost ajusta um preditor aditivo composto por T árvores de regressão $f_t \in \mathcal{F}$:

$$\hat{y}_i^{(T)} = \sum_{t=1}^T f_t(\mathbf{x}_i), \quad f_t \in \mathcal{F} \quad (2.38)$$

Em que cada f_t corresponde a uma árvore com estrutura $q_t(\cdot)$ (mapeando instâncias para folhas) e pesos de folha \mathbf{w}_t . O ajuste por iteração é obtido pela minimização de uma função objetivo regularizada (Chen; Guestrin, 2016; Friedman, 2001):

$$\mathcal{L}^{(T)} = \sum_{i=1}^n \ell(y_i, \hat{y}_i^{(T)}) + \sum_{t=1}^T \Omega(f_t) \quad (2.39)$$

Onde $\ell(\cdot)$ é uma função de perda apropriada ao tipo de tarefa (por exemplo, logística no caso binário) e $\Omega(f)$ penaliza a complexidade da árvore. Na implementação do XGBoost, $\Omega(f)$ costuma assumir forma do tipo (Chen; Guestrin, 2016):

$$\Omega(f) = \gamma |\mathcal{T}| + \frac{\lambda}{2} \|\mathbf{w}\|_2^2 + \alpha \|\mathbf{w}\|_1 \quad (2.40)$$

Em que $|\mathcal{T}|$ é o número de folhas, \mathbf{w} o vetor de pesos das folhas, e $(\gamma, \lambda, \alpha)$ hiperparâmetros de regularização. Como o modelo é construído de forma sequencial, a atualização do preditor em cada iteração pode ser escrita com fator de contração (η) $\eta \in (0,1]$:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta f_t(\mathbf{x}_i) \quad (2.41)$$

O que evidencia que, mesmo com uma mesma estrutura de árvore, a intensidade do ajuste é controlada por η . Assim, os hiperparâmetros do XGBoost podem ser interpretados como controles que afetam diretamente (2.38) - (2.41): o número de iterações T (número de árvores), a escala do passo η , a complexidade estrutural das árvores (profundidade e critérios de divisão), a estocasticidade da amostragem (linhas e colunas) e a regularização dos pesos de folha.

A seguir, detalham-se os principais hiperparâmetros adotados em problemas de classificação binária, enfatizando seu efeito matemático, seus benefícios e suas limitações, bem

como as interações que conduzem a compromissos (*trade-offs*) de desempenho e custo computacional.

2.4.1 Número de estimadores (árvores)

O hiperparâmetro número de estimadores (frequentemente denotado por T ou *number of estimators*) define o número total T de árvores no modelo aditivo (2.38). Um aumento em T amplia o espaço funcional acessível ao preditor, reduzindo viés, porém elevando a variância e o risco de sobreajuste, sobretudo quando a regularização (2.40) e/ou a estocasticidade (subamostragem) são fracas (Friedman, 2001; Natekin; Knoll, 2013).

Do ponto de vista computacional, o custo de treinamento tende a crescer aproximadamente de forma linear com T , pois cada iteração demanda (i) o cálculo de gradientes e hessianas da perda e (ii) a construção de uma nova árvore. Assim, para um algoritmo de busca de hiperparâmetros com orçamento B avaliações, a contribuição de T para o custo total pode ser aproximada por:

$$\text{Custo total} \propto \sum_{b=1}^B \sum_{t=1}^{T_b} \text{custo_árvore}(t, b) \quad (2.42)$$

Onde T_b é o número de árvores utilizado na b -ésima configuração avaliada.

Em problemas práticos, T raramente é escolhido isoladamente: sua interpretação deve ser conjunta com o η . Em particular, a estratégia de contração (2.41) sugere que valores menores de η exigem valores maiores de T para atingir o mesmo nível de ajuste do treinamento, em troca de maior robustez à variabilidade amostral (Friedman, 2001).

2.4.2 Taxa de aprendizado

O hiperparâmetro taxa de aprendizado η (*learning rate*) regula a magnitude do incremento adicionado ao modelo em cada iteração (2.41). Interpretando $f_t(\mathbf{x})$ como uma direção de descida no espaço funcional, η atua como um fator de passo, controlando a velocidade de convergência e a propensão ao sobreajuste (Friedman, 2001; Natekin; Knoll, 2013).

Uma consequência imediata de (2.41) é que, para uma sequência fixa de árvores $\{f_t\}$, reduzir η reduz a norma do preditor final $\hat{y}^{(T)}$ e, portanto, tende a impor regularização implícita por encolhimento (*shrinkage*). Todavia, valores muito pequenos de η podem demandar T muito

grande, implicando aumento do custo computacional e maior sensibilidade à escolha de critérios de parada (por exemplo, pelo parâmetro *early stopping rounds* de parada antecipada).

A interação entre (T, η) pode ser explicitada observando que, para $\eta \ll 1$, o acúmulo de atualizações requer mais iterações para produzir variação equivalente em $\hat{y}^{(T)}$:

$$\|\hat{y}^{(T)}\| \approx \left\| \sum_{t=1}^T \eta f_t \right\| \leq \eta \sum_{t=1}^T \|f_t\| \quad (2.43)$$

Consequentemente, η e T formam um *trade-off* estrutural: reduzir η pode melhorar a generalização, mas eleva o custo e pode exigir mecanismos auxiliares para controlar a convergência.

2.4.3 Profundidade máxima da árvore

O hiperparâmetro profundidade máxima da árvore (*maximal depth*) limita a profundidade máxima d_{\max} de cada árvore. Para árvores binárias, o número máximo de folhas é limitado por:

$$|\mathcal{T}| \leq 2^{d_{\max}} \quad (2.44)$$

O que implica que o termo de penalização $\gamma|\mathcal{T}|$ em (2.40) pode crescer rapidamente com d_{\max} caso a árvore expanda até a profundidade máxima.

Do ponto de vista estatístico, d_{\max} controla a complexidade de interações entre variáveis capturáveis por cada árvore: árvores rasas tendem a modelar efeitos mais globais (maior viés, menor variância), ao passo que árvores profundas capturam interações de alta ordem, porém com maior risco de sobreajuste, sobretudo em bases pequenas ou com ruído (Friedman, 2001; Hastie; Tibshirani; Friedman, 2009).

Do ponto de vista computacional, o número de nós internos cresce aproximadamente como 2^d , elevando o custo de busca por divisões e o consumo de memória. Assim, em cenários de otimização de hiperparâmetros, a profundidade máxima também participa do compromisso (*trade-off*) qualidade-custo.

2.4.4 Peso mínimo do nó-filho

O hiperparâmetro peso mínimo do nó-filho (*minimal child weight*) impõe um limiar mínimo para que um nó seja dividido. No XGBoost, a construção das árvores é guiada por

estatísticas de primeira e segunda ordem da perda. Denotando por g_i e h_i o gradiente e a hessiana da perda em relação à predição corrente:

$$g_i = \frac{\partial \ell(y_i, \hat{y}_i)}{\partial \hat{y}_i}, \quad h_i = \frac{\partial^2 \ell(y_i, \hat{y}_i)}{\partial \hat{y}_i^2} \quad (2.45)$$

A estatística de hessiana agregada em uma folha j é $H_j = \sum_{i \in I_j} h_i$, em que I_j é o conjunto de instâncias atribuídas à folha. O critério de peso mínimo pode ser visto como uma restrição do tipo:

$$H_j \geq \omega_{\min} \quad (2.46)$$

Onde $\omega_{\min} > 0$ é o valor do hiperparâmetro. Como, em perdas convexas típicas, h_i relaciona-se à incerteza local do ajuste, (2.46) desencoraja divisões que produzam folhas com pouca “massa” informacional, estabilizando o cálculo de pesos ótimos de folha e reduzindo sobreajuste (Chen; Guestrin, 2016).

Em problemas binários com perda logística, h_i assume a forma $h_i = p_i(1 - p_i)$, com p_i probabilidade predita; assim, H_j depende tanto do número de amostras quanto do grau de incerteza nas previsões, reforçando a interpretação do parâmetro como estabilizador de divisões.

2.4.5 Perda mínima para divisão

O hiperparâmetro γ (também chamado *minimal split loss* ou parâmetro *gamma*) implementa uma penalização explícita para a criação de novas folhas, aparecendo diretamente no ganho de divisão. Para um nó com estatísticas agregadas (G, H) e uma proposta de divisão em nós esquerdo e direito com estatísticas (G_L, H_L) e (G_R, H_R) , o ganho de divisão no XGBoost (ignorando, por simplicidade, o termo $\alpha\alpha$) pode ser escrito como (Chen; Guestrin, 2016):

$$\text{Gain} = \frac{1}{2} \left(\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{G^2}{H + \lambda} \right) - \gamma \quad (2.47)$$

A regra operacional de expansão exige $\text{Gain} > 0$. Portanto, aumentar γ eleva o limiar mínimo de melhora na função objetivo necessário para que uma divisão seja aceita. Como consequência, valores maiores de γ tendem a produzir árvores mais simples (menor número de folhas), diminuindo variância e custo computacional; em contrapartida, podem aumentar viés ao impedir divisões relevantes, sobretudo quando combinados com limites estritos de profundidade.

2.4.6 Fração de amostragem de instâncias

$$|\mathcal{D}_t| = \lfloor \rho n \rfloor \quad (2.48)$$

O treinamento da árvore f_t utiliza apenas \mathcal{D}_t , de modo que as estatísticas G e H em (2.47) passam a ser computadas sobre o subconjunto (*subsample*) amostrado. Esse mecanismo introduz estocasticidade análoga ao *bagging*, reduzindo correlação entre árvores e mitigando sobreajuste (Friedman, 2001; Natekin; Knoll, 2013).

Do ponto de vista do custo, reduzir ρ reduz o número de instâncias processadas por iteração, o que pode diminuir o tempo de treinamento por árvore. Contudo, ρ muito pequeno aumenta a variância das estimativas de gradiente/hessiana e pode degradar o desempenho ou exigir aumento de T .

2.4.7 Fração de colunas por árvore; e variantes por nível e por nó

Os hiperparâmetros κ_{tree} , κ_{level} e κ_{node} (*Column sample by tree, level, node*) controlam a subamostragem de colunas (atributos) em diferentes granularidades (árvore, nível e nó), promovendo diversidade estrutural entre árvores (Breiman, 2001; Chen; Guestrin, 2016):

$$|\mathcal{P}_t| = \lfloor \kappa_{\text{tree}} p \rfloor \quad (2.49)$$

Onde $\kappa_{\text{tree}} \in (0,1]$ é o hiperparâmetro fração de colunas por árvore. Analogamente, no nível de profundidade ℓ ou de nó v , definem-se subconjuntos adicionais $\mathcal{P}_{t,\ell}$ e $\mathcal{P}_{t,v}$, com cardinalidades controladas por κ_{level} e κ_{node} , respectivamente. A subamostragem de colunas reduz a probabilidade de que o modelo dependa excessivamente de um pequeno conjunto de atributos e diminui o custo de avaliação de candidatos de divisão.

Como $\kappa_{\text{tree}} < 1$, a subamostragem agressiva pode impedir que determinadas variáveis relevantes sejam consideradas em divisões críticas, sobretudo em bases com baixo número de atributos ou quando há atributos fortemente determinantes.

2.4.8 Regularizações Alpha e Lambda

Regularizações λ e α ("*lambda*" e "*alpha*") são definidas por:

$$w_j^* = -\frac{G_j}{H_j + \lambda} \quad (2.50)$$

Onde $G_j = \sum_{i \in I_j} g_i$ e $H_j = \sum_{i \in I_j} h_i$. A equação (2.50) evidencia que λ atua como termo de estabilização no denominador, reduzindo a magnitude de w_j^* e, portanto, impondo contração

nos incrementos do modelo. Assim, aumentar λ reduz variância e tende a melhorar generalização em troca de maior viés.

Já α implementa regularização ℓ_1 sobre pesos de folhas, favorecendo soluções com alguns pesos próximos de zero. No contexto de árvores, essa esparsidade não se manifesta como seleção direta de atributos, mas como contração e possível anulação de incrementos locais em folhas com fraca evidência estatística. Em geral, α pode aumentar robustez a ruído e reduzir complexidade efetiva, mas também pode levar a subajuste (*underfitting*) quando aplicado de forma intensa.

2.4.9 Peso da classe positiva

Em classificação binária, o desbalanceamento de classes pode ser tratado por ponderação da perda. O hiperparâmetro de peso da classe positiva (*scale positive weight*) define um peso $w_+ > 0$ aplicado às instâncias positivas, de modo que a função objetivo (2.39) pode ser reescrita como:

$$\mathcal{L}^{(T)} = \sum_{i=1}^n w_i \ell(y_i, \hat{y}_i^{(T)}) + \sum_{t=1}^T \Omega(f_t) \quad (2.51)$$

Em que $w_i = w_+$ quando $y_i = 1$ e $w_i = 1$ quando $y_i = 0$; $w_+ > 0$ corresponde ao hiperparâmetro. A ponderação altera gradientes e hessianas e, portanto, os ganhos de divisão (2.48) e os pesos ótimos (2.50). Em bases desbalanceadas, escolher w_+ de modo proporcional à razão entre classes (por exemplo, $w_+ \approx n_-/n_+$) pode reduzir viés de decisão em favor da classe majoritária; entretanto, pesos excessivos podem aumentar falsos positivos e degradar métricas como precisão, sendo necessário alinhar a ponderação às prioridades do problema e às métricas de avaliação (He; Garcia, 2009; Saito; Rehmsmeier, 2015).

Em bases balanceadas por amostragem (por exemplo, por sobreamostragem ou subamostragem), w_+ tende a ser próximo de 1, e seu papel torna-se secundário.

2.4.10 Critérios auxiliares: parada antecipada, métrica de avaliação e função objetivo

Além de hiperparâmetros estruturais, a prática de ajuste de XGBoost frequentemente utiliza critérios auxiliares que afetam o processo de treinamento e a seleção efetiva de T :

$$\text{parar em } t \text{ se } \max_{s \in \{t-R, \dots, t-1\}} m(s) - m(t) < \Delta \quad (2.52)$$

Esse mecanismo de parada antecipada (*early stopping rounds*) em (2.52) atua como regularização adicional e reduz custo computacional (Prechelt, 1998), mas sua eficácia depende de uma partição de validação representativa e da escolha coerente de $m(\cdot)$.

A escolha da métrica de avaliação (*evaluation metric*) e da função objetivo (*objective*) deve ser consistente com o problema. Em classificação binária, é comum empregar objetivos logísticos (saída probabilística) e métricas como a medida F1 (*F1-score*), Área sob a curva ROC (*Area Under the ROC Curve*, AUC-ROC), Área sob a curva precisão-revocação (*Area Under the Precision-Recall Curve*, AUC-PR), acurácia ou funções derivadas da matriz de confusão (*confusion matrix*) (Fawcett, 2006; Saito; Rehmsmeier, 2015). A métrica de monitoramento influencia a seleção de \hat{T} sob parada antecipada e pode induzir diferentes compromissos entre sensibilidade e especificidade. Assim, a seleção da métrica de avaliação e da função objetivo não é apenas operacional: ela define a noção de “melhor” durante o treinamento.

2.5 Métricas de avaliação e custo computacional

Em problemas de classificação binária, a avaliação de desempenho deriva de estatísticas da matriz de confusão, que relaciona as predições \hat{y} com os rótulos verdadeiros y . Denotando por TP (verdadeiros positivos), TN (verdadeiros negativos), FP (falsos positivos) e FN (falsos negativos), tem-se a partição do conjunto de teste $\mathcal{D}_{\text{test}}$:

$$\begin{aligned}
 TP &= \sum_{i \in \mathcal{D}_{\text{test}}} \mathbb{I}(y_i = 1 \wedge \hat{y}_i = 1) \\
 TN &= \sum_{i \in \mathcal{D}_{\text{test}}} \mathbb{I}(y_i = 0 \wedge \hat{y}_i = 0) \\
 FP &= \sum_{i \in \mathcal{D}_{\text{test}}} \mathbb{I}(y_i = 0 \wedge \hat{y}_i = 1) \\
 FN &= \sum_{i \in \mathcal{D}_{\text{test}}} \mathbb{I}(y_i = 1 \wedge \hat{y}_i = 0)
 \end{aligned} \tag{2.53}$$

Em que $\mathbb{I}(\cdot)$ é a função indicadora. A partir dessas quantidades, definem-se métricas clássicas de qualidade (Powers, 2011):

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.54}$$

$$\text{Precisão} = \frac{TP}{TP + FP} \tag{2.55}$$

$$\text{Revocação (Sensibilidade)} = \frac{TP}{TP + FN} \tag{2.56}$$

$$\text{Especificidade} = \frac{TN}{TN + FP} \tag{2.57}$$

A acurácia em (2.54) resume o erro global, mas pode ser enganosa em cenários desbalanceados, pois um classificador trivial pode obter acurácia elevada ao privilegiar a classe majoritária (He; Garcia, 2009). Nesses casos, métricas orientadas a erros de classe específica (precisão, revocação e especificidade) tornam-se mais informativas. Mesmo em bases balanceadas por reamostragem, é comum reportar o conjunto (2.54) - (2.57) para explicitar o padrão de erros do classificador.

Métricas compostas também são frequentes. O *F1-score*, por exemplo, é a média harmônica de precisão e revocação:

$$F_1 = 2 \frac{\text{Precisão} \cdot \text{Revocação}}{\text{Precisão} + \text{Revocação}} \quad (2.58)$$

E por sua vez, generaliza para F_β ao ponderar a importância relativa de revocação. Outra métrica de interesse em problemas binários é o coeficiente de correlação de Matthews (MCC), especialmente útil em classes desbalanceadas (Matthews, 1975):

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (2.59)$$

No contexto de classificadores probabilísticos, métricas baseadas em ranking e calibração podem ser consideradas. A AUC-ROC sintetiza o desempenho ao variar o limiar de decisão, sendo amplamente discutida na literatura (Fawcett, 2006; Bradley, 1997). Entretanto, quando as classes são desbalanceadas, a AUC-PR tende a ser mais informativa (Saito; Rehmsmeier, 2015). Já a calibração probabilística pode ser avaliada por medidas como o escore de Brier, também conhecido como Brier score (Brier, 1950).

Em experimentos com validação cruzada estratificada em k partes, cada configuração de hiperparâmetros é avaliada em k partições, produzindo estimativas $m^{(1)}, \dots, m^{(k)}$ para uma métrica m . O estimador por média é:

$$\bar{m} = \frac{1}{k} \sum_{j=1}^k m^{(j)} \quad (2.60)$$

E a variabilidade entre dobras pode ser sumarizada por desvio-padrão amostral:

$$s_m = \sqrt{\frac{1}{k-1} \sum_{j=1}^k (m^{(j)} - \bar{m})^2} \quad (2.61)$$

A estratificação assegura que a proporção de classes seja preservada em cada dobra, reduzindo viés na estimativa em problemas binários (Kohavi, 1995; Arlot; Celisse, 2010).

Além de métricas de qualidade, otimizações de hiperparâmetros requerem mensurar custo computacional. Seja t_j o tempo de treinamento e avaliação na j -ésima dobra. Define-se o tempo médio por dobra como:

$$\bar{t}_{\text{fold}} = \frac{1}{k} \sum_{j=1}^k t_j \quad (2.62)$$

Que, por construção, incorpora o custo de treino + predição na validação. Para uma configuração avaliada sob validação cruzada, o tempo total é aproximadamente $t_{\text{CV}} = \sum_{j=1}^k t_j = k \bar{t}_{\text{fold}}$.

No nível de um algoritmo de busca de hiperparâmetros que avalia B configurações (orçamento), o tempo de otimização pode ser modelado como:

$$t_{\text{opt}} = \sum_{b=1}^B t_{\text{CV}}^{(b)} + t_{\text{overhead}} \quad (2.63)$$

Onde t_{overhead} agrega custos de seleção de candidatos, atualização de modelos probabilísticos (no caso bayesiano) e operações de controle (por exemplo, parada antecipada). Em termos experimentais, t_{opt} constitui uma aproximação operacional do custo computacional e pode ser utilizado como proxy de custo em formulações multiobjetivo, ao lado de métricas de qualidade.

Por fim, é importante notar que a mensuração de tempo depende de hardware e ambiente de execução. Assim, para comparações justas, recomenda-se controlar o cenário computacional (mesmo equipamento, mesmas bibliotecas e versões) e adotar critérios de comparação por orçamento (B fixo) ou por tempo de parede (*wall-clock time*) fixo. Essa preocupação é central em experimentos de otimização de hiperparâmetros, nos quais diferenças aparentes de desempenho podem ser consequência de diferentes regimes de custo.

2.6 Planejamento de Experimentos: princípios, arranjos e análise

O DoE é um conjunto de princípios e técnicas estatísticas destinado a estruturar a coleta de dados, de modo que seja possível identificar, estimar e interpretar os efeitos de fatores controláveis sobre uma ou mais variáveis resposta, com eficiência no uso do orçamento experimental (Fisher, 1935; Montgomery, 2017). Em termos práticos, o DoE responde simultaneamente a duas necessidades: (i) garantir validade inferencial, reduzindo vieses e confundimentos; e (ii) maximizar a informação extraída por execução experimental, algo particularmente relevante quando cada execução possui custo elevado.

Em problemas de aprendizagem de máquina, o ajuste de hiperparâmetros pode ser formulado como um experimento computacional no qual fatores são hiperparâmetros e respostas são métricas de desempenho e de custo. Nesse contexto, o DoE permite substituir varreduras exaustivas por delineamentos parcimoniosos que investigam o espaço de busca de forma estruturada e interpretável, preservando a capacidade de quantificar efeitos principais, interações e curvaturas, especialmente quando combinado com RSM (Box; Wilson, 1951; Myers; Montgomery; Anderson-Cook, 2016).

Do ponto de vista estatístico, considera-se um conjunto de k fatores controláveis reunidos no vetor $\mathbf{x} = (x_1, \dots, x_k)^\top$ e uma variável resposta y , observada em N execuções experimentais. Um modelo genérico para a i -ésima execução assume a forma:

$$y_i = f(\mathbf{x}_i) + \varepsilon_i, \quad i = 1, \dots, N \quad (2.64)$$

Em que $f(\cdot)$ denota a relação desconhecida entre fatores e resposta, e ε_i representa a variação não explicada pelo modelo. Para sustentar testes clássicos e intervalos de confiança em modelos lineares, admite-se frequentemente que:

$$\mathbb{E}[\varepsilon_i] = 0, \quad \text{Var}(\varepsilon_i) = \sigma^2, \quad \varepsilon_i \text{ independentes} \quad (2.65)$$

Com distribuição aproximadamente normal, ao menos no regime em que o Teorema Central do Limite torne adequada a aproximação. Em experimentos computacionais, essa formulação permanece útil, embora a origem do ruído possa incluir estocasticidade algorítmica, variabilidade de amostragem e flutuações de desempenho computacional; por essa razão, recomenda-se atenção explícita a reprodutibilidade (semeioamento), randomização de ordem e replicação por múltiplas sementes e partições.

2.6.1 Princípios gerais

O DoE estrutura-se em torno de princípios que asseguram validade inferencial e eficiência do delineamento. A literatura clássica enfatiza randomização, replicação e bloqueio, complementados por propriedades desejáveis do delineamento, tais como balanceamento, ortogonalidade e possibilidade de execução sequencial (Fisher, 1935; Montgomery, 2017).

A randomização consiste em executar as combinações de níveis em ordem aleatória, reduzindo o risco de confundimento entre efeitos de fatores e tendências externas ao experimento. Em experimentos físicos, essa prática combate viés associado a deriva instrumental, efeitos de aquecimento, fadiga e variabilidade ambiental. Em experimentos computacionais, ainda que não haja instrumentação física, existem fontes de deriva e

heterogeneidade, como variação de carga do sistema, efeitos de cache, escalonamento e condições de execução; nesse cenário, a randomização protege o delineamento contra correlação espúria entre condições e tempo de execução.

A replicação introduz repetições planejadas de condições experimentais, permitindo separar variabilidade inerente da variação explicada por fatores. Seja y_r a r -ésima repetição de uma condição fixa. A média amostral $\bar{y} = \frac{1}{n_r} \sum_{r=1}^{n_r} y_r$ possui variância:

$$\text{Var}(\bar{y}) = \frac{\sigma^2}{n_r} \quad (2.66)$$

O que evidencia o ganho de precisão com aumento de n_r . Em DoE, as repetições no ponto central desempenham papel adicional, pois viabilizam estimativa de erro puro e testes de falta de ajuste em modelos de regressão para superfície de resposta (Myers; Montgomery; Anderson-Cook, 2016). Em ajuste de hiperparâmetros, replicação costuma ser implementada por múltiplas sementes e, quando aplicável, por múltiplas partições de validação cruzada.

O bloqueio permite controlar fatores de perturbação (*nuisance factors*) agrupando execuções em blocos relativamente homogêneos. Em um modelo aditivo simples, com tratamentos indexados por t e blocos por b , tem-se:

$$y_{tb} = \mu + \tau_t + \gamma_b + \varepsilon_{tb} \quad (2.67)$$

Em que τ_t representa o efeito do tratamento e γ_b representa o efeito do bloco. Quando variações sistemáticas entre blocos existem, esse artifício reduz a variância residual e aumenta o poder de detecção de efeitos de interesse (Montgomery, 2017). Em experimentos computacionais, blocos podem ser definidos por janelas de execução, configurações de hardware, versões de dependências ou nós de cluster, quando tais fatores não podem ser completamente fixados.

Além desses princípios, propriedades estruturais do delineamento são desejáveis para favorecer estimação e interpretação. Em termos de matriz de planejamento \mathbf{X} , a ortogonalidade entre colunas reduz colinearidade e melhora o condicionamento de $\mathbf{X}^T \mathbf{X}$, diminuindo variâncias dos estimadores de mínimos quadrados e simplificando interpretações. O balanceamento, por sua vez, assegura representação simétrica de níveis, evitando viés na estimação de efeitos principais.

Por fim, a experimentação sequencial é central na RSM. A estratégia proposta por Box e Wilson (1951) enfatiza uma sequência: triagem para identificar fatores ativos, ajuste de modelos de primeira ordem para direção de melhoria e, em seguida, ajuste de modelos de segunda ordem em uma região promissora para refinar e localizar condições ótimas. Essa

orientação é particularmente compatível com otimização de hiperparâmetros, na qual o orçamento de avaliações é limitado e o interesse inclui tanto desempenho quanto compreensão do papel de cada hiperparâmetro.

2.6.2 Arranjos clássicos e superfícies de resposta

Os principais delineamentos podem ser organizados segundo objetivos e estrutura algébrica. Nesta seção, apresentam-se delineamentos clássicos e suas propriedades, com foco em aplicações em modelagem e otimização de respostas.

Em delineamentos fatoriais completos de dois níveis, os fatores codificados assumem $x_j \in \{-1, +1\}$ e o número de execuções na parte fatorial é $N_F = 2^k$. Esses delineamentos são adequados para estimar efeitos principais e interações e servem como base para estratégias de triagem e para a construção de delineamentos compostos em RSM. Um modelo com interações até segunda ordem pode ser representado por:

$$y = \beta_0 + \sum_{j=1}^k \beta_j x_j + \sum_{1 \leq j < \ell \leq k} \beta_{j\ell} x_j x_\ell + \varepsilon \quad (2.68)$$

Em delineamentos ortogonais, os coeficientes associados a efeitos podem ser estimados com interpretações diretas. A limitação fundamental é o crescimento exponencial de N_F com k , tornando o delineamento impraticável quando o número de fatores é elevado.

Delineamentos fatoriais fracionados reduzem esse custo ao selecionar uma fração 2^{k-p} das combinações fatoriais, com $p \geq 1$, introduzindo aliasing controlado pela resolução do delineamento. Em aplicações de triagem, assume-se frequentemente que interações de alta ordem são desprezíveis, o que torna aceitável o confundimento entre efeitos de alta ordem e efeitos principais ou de baixa ordem (Box; Hunter; Hunter, 2005; Montgomery, 2017). A escolha da resolução determina o grau de proteção contra confundimentos, devendo ser coerente com o conhecimento de domínio e o objetivo experimental.

Para triagem com número muito reduzido de execuções, delineamentos de Plackett e Burman são amplamente utilizados. Esses delineamentos visam estimar efeitos principais em matrizes quase ortogonais com número de execuções múltiplo de quatro, sendo eficientes sob a premissa de que interações são pequenas na região estudada (Plackett; Burman, 1946). A vantagem reside no custo, mas a limitação é que, em presença de interações relevantes, efeitos principais podem ser distorcidos por confundimento.

Em experimentação robusta, a abordagem associada a Taguchi (1986) emprega arranjos ortogonais e funções de perda para incorporar variabilidade, buscando configurações que maximizem desempenho e minimizem sensibilidade a ruído. Apesar de sua relevância histórica e adoção industrial, recomenda-se análise criteriosa das suposições e do tratamento de interações, frequentemente reinterpretando resultados em modelos lineares e ANOVA para fornecer fundamentação inferencial (Montgomery, 2017).

2.6.3 Metodologia de Superfície de Resposta

A RSM busca aproximar a relação $f(\mathbf{x})$ por um polinômio de baixa ordem em uma vizinhança de interesse, com o objetivo de otimizar a resposta e compreender efeitos de curvatura. Um modelo quadrático padrão pode ser escrito como:

$$y = \beta_0 + \sum_{j=1}^k \beta_j x_j + \sum_{j=1}^k \beta_{jj} x_j^2 + \sum_{1 \leq j < \ell \leq k} \beta_{j\ell} x_j x_\ell + \varepsilon \quad (2.69)$$

Em notação matricial, denotando $\mathbf{y} \in \mathbb{R}^N$ o vetor de respostas, $\mathbf{X} \in \mathbb{R}^{N \times p}$ a matriz de regressão construída a partir dos termos do modelo, $\boldsymbol{\beta} \in \mathbb{R}^p$ o vetor de parâmetros e $\boldsymbol{\varepsilon} \in \mathbb{R}^N$ o vetor de resíduos, obtém-se:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.70)$$

O estimador de mínimos quadrados ordinários é:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (2.71)$$

Assumindo $\mathbf{X}^T \mathbf{X}$ inversível e bem condicionado. A inferência sobre a adequação do modelo e a significância de termos pode ser conduzida via decomposição de somas de quadrados e estatísticas F . Uma decomposição clássica considera soma de quadrados total SS_T , de regressão SS_R e residual SS_E :

$$SS_T = \sum_{i=1}^N (y_i - \bar{y})^2, \quad SS_R = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2, \quad SS_E = \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (2.72)$$

Em que \bar{y} é a média amostral e \hat{y}_i é a predição do modelo na i -ésima execução. O teste global do modelo pode ser realizado por:

$$F = \frac{MS_R}{MS_E} = \frac{SS_R/\nu_R}{SS_E/\nu_E} \quad (2.73)$$

Com ν_R e ν_E representando graus de liberdade de regressão e erro, respectivamente (Montgomery, 2017; Myers; Montgomery; Anderson-Cook, 2016). Quando há replicação no ponto central ou em outros pontos, pode-se decompor SS_E em erro puro e falta de ajuste,

fornecendo teste específico para avaliar se o modelo quadrático é compatível com o comportamento observado na região experimental.

Entre delineamentos para ajuste de modelos quadráticos, destacam-se os Delineamentos Compostos Centrais (*Central Composite Designs*, CCD) e os Planejamentos de Box e Behnken. No CCD, o conjunto de pontos é composto por uma parte fatorial, pontos axiais e pontos centrais. Para k fatores, o número de execuções é:

$$N = 2^k + 2k + n_0 \quad (2.74)$$

Em que n_0 representa o número de repetições no ponto central. Em cenários com muitos fatores, utiliza-se um fatorial fracionado de 2^{k-p} pontos, obtendo-se:

$$N = 2^{k-p} + 2k + n_0 \quad (2.75)$$

Viabilizando ajuste quadrático com custo reduzido, ainda que com possíveis estruturas de *aliasing* que precisam ser controladas (Box; Hunter; Hunter, 2005; Montgomery, 2017). O planejamento de Box e Behnken, proposto para evitar pontos nos vértices do hipercubo, requer $N = 2k(k - 1) + n_0$ execuções e costuma ser atrativo quando combinações extremas de fatores são indesejáveis (Box; Behnken, 1960). O planejamento de Doehlert distribui pontos de forma aproximadamente uniforme em uma hiperesfera e oferece flexibilidade sequencial em algumas configurações (Doehlert, 1970).

Além desses delineamentos, delineamentos ótimos (por exemplo, delineamentos *D*-ótimos) formulam a escolha de pontos como um problema de otimização combinatória, frequentemente maximizando $\det(\mathbf{X}^T \mathbf{X})$ para reduzir o volume da elipsoide de confiança dos estimadores (Atkinson; Donev; Tobias, 2007). Em problemas com restrições severas ou regiões não retangulares, delineamentos ótimos podem ser mais adequados do que delineamentos clássicos, embora sua interpretabilidade e simplicidade operacional dependam do contexto.

Em experimentos computacionais e calibração de modelos, é comum empregar delineamentos de preenchimento espacial, como hipercubo latino, *maximin* e outros, com o objetivo de cobrir o domínio uniformemente, sobretudo quando não se pressupõe estrutura polinomial simples. Esses delineamentos são particularmente relevantes quando a resposta é altamente não linear ou quando se deseja treinar metamodelos de alta flexibilidade, mas tendem a oferecer menor transparência paramétrica quando comparados à RSM.

No problema de otimização de hiperparâmetros, delineamentos de RSM são atraentes por equilibrar custo, capacidade explicativa e estrutura de análise. Há evidências na literatura de que DoE e RSM podem reduzir o número de avaliações necessárias para calibrar modelos, mantendo desempenho competitivo e oferecendo interpretações sobre sensibilidade a

hiperparâmetros (Balestrassi *et al.*, 2009; Pontes *et al.*, 2016; Arboretti; Olivieri; Lanteri, 2021; Freiesleben *et al.*, 2020). As principais lacunas consistem em: (i) crescimento do custo com muitos fatores; (ii) hiperparâmetros discretos e categóricos que exigem tratamentos específicos; (iii) superfícies com descontinuidades, platôs e ruído elevado que dificultam ajuste quadrático; e (iv) dependência do pré-processamento e do protocolo de validação, que podem reduzir a sensibilidade do desempenho a variações de hiperparâmetros, como discutido posteriormente.

2.7 Metodologia de Superfície de Resposta e Delineamento Central

Composto Face-Centrado

A RSM constitui um arcabouço estatístico e matemático destinado à modelagem empírica, à análise de sensibilidade e à otimização de sistemas em que uma ou mais respostas dependem de um conjunto finito de fatores controláveis. Em sua formulação clássica, a RSM parte do pressuposto de que a relação funcional exata entre fatores e respostas é desconhecida ou demasiado complexa para ser derivada por modelagem mecânica completa. Nesses casos, é conveniente e metodologicamente defensável substituir a relação desconhecida por uma aproximação local, ajustada em uma região experimental delimitada, com propriedades analíticas suficientes para suportar inferência estatística e procedimentos de otimização.

A RSM foi formalizada no contexto de experimentação industrial por Box e Wilson (1951), no estudo seminal sobre a obtenção experimental de condições ótimas, no qual se estabelece a conexão entre delineamentos planejados, modelos polinomiais e busca sistemática do ponto de melhor desempenho sob restrições operacionais. Essa perspectiva permanece atual em aplicações contemporâneas nas quais as respostas são obtidas por simulação ou por execução computacional repetida, uma vez que cada execução do sistema pode ser interpretada como um ensaio experimental, desde que se mantenha o controle de variáveis de contexto e se adote replicação para quantificação de variabilidade. A escolha de delineamentos específicos, como o CCD, decorre precisamente dessa necessidade de obter informação suficiente para ajustar modelos quadráticos com eficiência amostral e com capacidade de interpretação.

Em aplicações com potencial de não linearidade e com possibilidade de interação entre fatores, o modelo quadrático completo é frequentemente adotado por equilibrar capacidade de aproximação local e parcimônia. Além disso, esse modelo fornece base direta para testes formais de significância de efeitos, para avaliação de ajuste e capacidade preditiva e para construção de superfícies e contornos, os quais permitem interpretar o comportamento das

respostas no entorno de uma região de interesse. Em termos práticos, essa escolha também se justifica pelo fato de que o modelo quadrático é o de menor ordem capaz de representar curvatura, o que é essencial para a identificação de máximos e mínimos locais em problemas de otimização.

2.7.1 Formulação do modelo quadrático completo e interpretação dos termos

Considere uma resposta escalar y , observada em função de k fatores quantitativos codificados (x_1, x_2, \dots, x_k) . Um modelo quadrático completo pode ser expresso como:

$$\hat{y}(\mathbf{x}) = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x_i^2 + \sum_{i < j} \beta_{ij} x_i x_j \quad (2.76)$$

Em que β_0 representa o intercepto, β_i representam efeitos lineares, β_{ii} capturam curvatura e β_{ij} descrevem interações bilaterais entre fatores. A interpretação dos termos é importante porque conecta diretamente a estrutura estatística à lógica do sistema em estudo. Os termos lineares descrevem tendências direcionais, isto é, variações médias na resposta quando um fator se desloca ao longo do intervalo experimental, mantendo os demais fixos. Os termos quadráticos introduzem a possibilidade de concavidade ou convexidade, permitindo representar regiões de saturação, pontos de retorno e condições em que incrementos adicionais em um fator deixam de produzir melhoria, ou passam a degradar a resposta. Já os termos de interação modelam dependências entre fatores, caracterizando situações em que o efeito de um fator depende do nível assumido por outro. Em problemas de calibração de algoritmos, essa componente é relevante, pois hiperparâmetros frequentemente operam em regime de compensação, como ocorre, por exemplo, quando a agressividade de regularização altera o benefício marginal do aumento de complexidade do modelo.

O modelo (2.76) é, em geral, ajustado por Mínimos Quadrados Ordinários (MQO), a partir de um conjunto de n observações coletadas conforme um delineamento experimental. Denotando por \mathbf{y} o vetor $n \times 1$ de respostas observadas e por \mathbf{X} a matriz de projeto $n \times p$, em que p representa o número de coeficientes do modelo, o modelo linear nos parâmetros pode ser escrito como:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.77)$$

Onde $\boldsymbol{\beta}$ é o vetor de coeficientes e $\boldsymbol{\varepsilon}$ é o vetor de erros aleatórios. A estimativa por MQO é obtida por:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (2.78)$$

A partir de $\hat{\beta}$, obtém-se a resposta predita $\hat{y} = X\hat{\beta}$ e os resíduos $e = y - \hat{y}$, os quais são fundamentais para diagnóstico do modelo e verificação das hipóteses subjacentes, tais como normalidade aproximada, homoscedasticidade e independência condicional. Em experimentos computacionais, a independência é usualmente garantida por controle de *seeds* e aleatorização da ordem de execução; a homoscedasticidade deve ser verificada, pois é comum que variância do erro aumente em regiões de maior custo computacional ou maior instabilidade numérica.

2.7.2 Codificação de fatores, escalas e região experimental

A RSM assume, de forma operacional, a definição de uma região experimental em torno de valores de referência. Para fatores quantitativos, é prática padrão operar com variáveis codificadas, de modo que níveis baixos e altos sejam mapeados para -1 e $+1$, respectivamente, enquanto o ponto central é mapeado para 0 . Seja z_i a variável na escala natural e x_i a variável codificada. A transformação linear clássica pode ser escrita como:

$$x_i = \frac{z_i - z_{0i}}{\Delta_i} \quad (2.79)$$

Onde z_{0i} é o nível central do fator e Δ_i é o semi-intervalo entre níveis alto e baixo. Essa codificação torna os coeficientes mais comparáveis entre fatores, melhora condicionamento numérico do ajuste e simplifica a interpretação geométrica de pontos fatoriais, axiais e centrais no espaço experimental.

A definição da região experimental é uma escolha de projeto com implicações diretas sobre validade local do modelo. Em particular, modelos quadráticos ajustados por RSM são aproximações locais, de modo que extrapolações além do domínio experimental não são recomendadas sem validação adicional. Em problemas de calibração de hiperparâmetros, isso é ainda mais relevante, pois regiões fora dos limites podem corresponder a configurações que geram modelos degenerados, convergência inadequada ou custos computacionais desproporcionais.

2.7.3 Delineamento Central Composto: estrutura, estimação e replicação no centro

O CCD é um delineamento projetado para suportar a estimação do modelo quadrático completo com eficiência amostral. Sua estrutura é formada por três conjuntos de pontos. O primeiro conjunto é composto por pontos fatoriais, que podem ser completos ou fracionários.

O segundo conjunto inclui pontos axiais, também denominados pontos estrela, que permitem estimar termos quadráticos de forma independente dos termos lineares e de interação. O terceiro conjunto contém pontos centrais, frequentemente replicados, os quais permitem estimar erro puro e testar falta de ajuste.

Para k fatores, um CCD completo com componente fatorial 2^k contém 2^k pontos fatoriais, $2k$ pontos axiais e n_0 pontos centrais, totalizando:

$$n = 2^k + 2k + n_0 \quad (2.80)$$

A presença de replicação no centro é essencial por três razões: fornece uma estimativa direta do erro puro, permite testar a hipótese de ausência de curvatura de forma global e melhora a estabilidade do ajuste ao aumentar o grau de liberdade associado ao erro. Em contextos computacionais, a replicação é também uma forma de capturar variabilidade estocástica intrínseca, decorrente de aleatoriedade na inicialização e em operações internas do algoritmo.

A análise de variância aplicada ao modelo quadrático permite decompor a soma de quadrados total em componentes associadas ao modelo e ao erro. Seja SST a soma de quadrados total, SSR a soma de quadrados explicada pelo modelo e SSE a soma de quadrados dos erros. A decomposição clássica é:

$$SST = SSR + SSE \quad (2.81)$$

O teste global de significância do modelo pode ser conduzido pela estatística F , construída a partir das médias quadráticas do modelo e do erro. Além disso, a falta de ajuste pode ser avaliada comparando a variabilidade residual explicada por erro puro, estimada por replicações, com a variabilidade residual devida à inadequação do modelo.

2.7.4 Delineamento Central Composto Face-Centrado: definição, posição dos pontos axiais e implicações

No CCDFC, os pontos axiais são posicionados nas faces do hipercubo experimental. Isso equivale a fixar o parâmetro axial em $\alpha = 1$. Para cada fator i , existem dois pontos axiais, um em $+1$ e outro em -1 , mantendo os demais fatores em zero. Assim, os pontos axiais podem ser representados como:

$$\mathbf{x}_{\text{axial}}^{(i,\pm)} = (0, \dots, 0, \pm 1, 0, \dots, 0) \quad (2.82)$$

Em que o elemento não nulo ocupa a i -ésima posição. Em consequência, todos os níveis experimentais permanecem no intervalo $[-1, +1]$ para cada fator, o que evita extrapolação operacional. Em aplicações nas quais os limites físico-operacionais são rígidos, ou em que

configurações extremas geram instabilidade, essa propriedade é particularmente desejável. Em problemas de calibração de hiperparâmetros, a adoção do CCDFC reduz a probabilidade de avaliações inválidas, melhora a taxa de sucesso dos ensaios e evita que o delineamento explore pontos com comportamento numérico atípico, não representativo do regime operacional de interesse.

A principal limitação teórica do CCDFC é que ele não é rotacionável no sentido estrito, pois a rotação exige α ajustado em função de k e do componente fatorial. A rotatabilidade é uma propriedade segundo a qual a variância de predição depende apenas da distância ao centro, e não da direção. Em estudos em que se busca homogeneidade direcional da precisão de predição, delineamentos rotacionáveis podem ser preferidos. Todavia, em situações práticas em que a prioridade é manter os experimentos dentro dos limites operacionais e reduzir risco de extrapolação, a perda de rotatabilidade é frequentemente aceitável, sobretudo quando a região de interesse é relativamente compacta e centrada.

2.7.5 Comparação conceitual entre Delineamento Central Composto de Face Centrada e Box-Behnken: critério de escolha

O delineamento Box-Behnken também suporta ajuste de modelos quadráticos, mas difere do CCDFC ao evitar pontos em que todos os fatores estejam simultaneamente em níveis extremos. Em termos geométricos, o Box-Behnken amostra combinações intermediárias e reduz a exposição do experimento aos cantos do hipercubo. Em contrapartida, o CCDFC inclui explicitamente o componente fatorial, o que facilita a interpretação de interações em níveis extremos codificados e fornece uma base direta para análise de efeitos principais e interações no padrão tradicional de engenharia. Quando o objetivo inclui, de forma explícita, a discussão de efeitos principais, interações e termos quadráticos com apoio em gráficos de Pareto, gráficos de efeitos e superfícies de resposta, o CCDFC tende a oferecer uma estrutura mais direta e alinhada à tradição de RSM.

No caso deste trabalho, a escolha do CCDFC está alinhada ao objetivo de produzir um planejamento com interpretação clara e com cobertura adequada para estimação de curvatura, mantendo todos os pontos dentro de uma região operacional previamente delimitada. Essa escolha também favorece a comunicação dos resultados, uma vez que os deslocamentos axiais no CCDFC correspondem a variações controladas de um fator por vez, o que facilita a

argumentação sobre causalidade local e sobre mecanismos de influência dos fatores nas respostas.

2.7.6 Metodologia de Superfície de Resposta como base para otimização mono-objetivo e multiobjetivo e conexão com o Método de Interseção Normal à Fronteira

Uma vez ajustadas as superfícies de resposta para cada métrica de interesse, a RSM fornece uma descrição analítica e diferenciável do comportamento esperado das respostas no espaço de fatores. Isso permite formular problemas de otimização diretamente sobre modelos empíricos, reduzindo a necessidade de avaliação exaustiva do sistema real em toda a região de interesse. Em problemas com múltiplas respostas, é comum que critérios de desempenho preditivo e de custo computacional sejam conflitantes, o que caracteriza um cenário de otimização multiobjetivo. Nessa situação, torna-se inadequado reduzir o problema a uma única função escalar sem justificativa formal, pois diferentes escalas podem induzir soluções distintas e potencialmente enviesadas em relação às preferências do decisor.

Métodos de geração de soluções Pareto-ótimas, como o NBI, oferecem uma forma sistemática de mapear o conjunto de compromissos entre objetivos. O NBI foi proposta para contornar limitações de métodos de soma ponderada, especialmente em frentes não convexas, e pode ser aplicada diretamente quando as funções objetivo são dadas por modelos analíticos, como as superfícies quadráticas da RSM. Dessa forma, a integração RSM-NBI fornece uma abordagem metodologicamente coerente, na qual o delineamento experimental provê dados suficientes para ajuste de modelos, e a etapa de otimização multiobjetivo utiliza esses modelos como aproximações para explorar *trade-offs* entre respostas. Essa integração é particularmente conveniente quando se deseja, além de identificar soluções viáveis, discutir o papel de cada fator e de suas interações na configuração das soluções obtidas.

Em síntese, o emprego do CCDFC, aliado ao ajuste de modelos quadráticos pela RSM, fornece as condições necessárias para uma etapa subsequente de decisão por otimização, seja ela mono ou multiobjetivo. Em particular, em problemas de calibração de algoritmos com múltiplos hiperparâmetros, a existência de diversos graus de liberdade e de critérios concorrentes torna necessária uma estratégia sistemática de modelagem e de otimização. A combinação de delineamento estruturado, modelagem quadrática e métodos multiobjetivo fornece uma base formal para decisões tecnicamente justificáveis, com transparência estatística e com possibilidade de verificação por experimentos de confirmação.

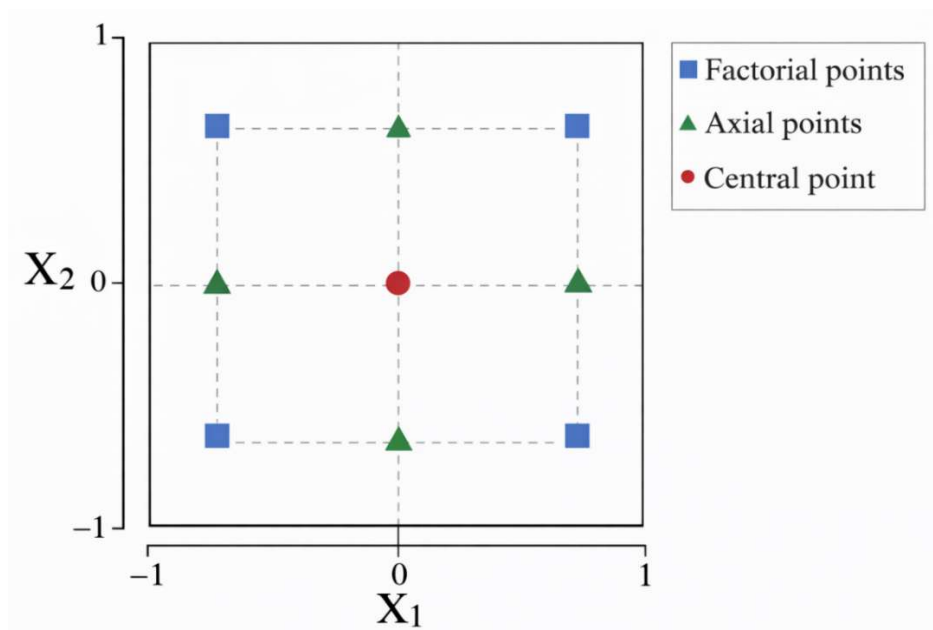


Figura 1 - Representação do CCDFC no hiper-cubo codificado e decomposição em pontos fatoriais, axiais e centrais.

2.8 Análise fatorial, componentes principais e rotação *Varimax* para síntese de métricas

Esta subseção apresenta a fundamentação matemática de análise fatorial exploratória, PCA e rotação *Varimax* com nível de formalismo compatível com publicações de alto impacto, tomando como referência a formulação empregada por Pereira *et al.* (2025).

Esta subseção consolida os fundamentos matemáticos necessários para (i) reduzir um conjunto de métricas observadas a um número menor de variáveis latentes (fatores), (ii) aumentar a interpretabilidade desses fatores por meio de rotação ortogonal do tipo *Varimax* e (iii) construir índices compostos adequados a cenários de comprometimento (*trade-off*) entre qualidade preditiva e custo temporal. Em termos metodológicos, tais conceitos permitem transformar múltiplas respostas (por exemplo, acurácia, precisão, sensibilidade e tempo de execução) em um espaço latente de baixa dimensão, no qual se tornam viáveis estratégias sistemáticas de busca e comparação de soluções.

2.8.1 Notação e padronização das variáveis

Considere um conjunto de n observações (execuções experimentais) e p variáveis observadas (métricas), organizadas na matriz de dados:

$$\mathbf{Y} = [y_{ij}] \in \mathbb{R}^{n \times p}, \quad i = 1, \dots, n, j = 1, \dots, p \quad (2.83)$$

No contexto desta dissertação, cada linha de \mathbf{Y} corresponde a uma configuração avaliada (por exemplo, uma combinação de hiperparâmetros), enquanto cada coluna representa uma métrica de desempenho (por exemplo, Acurácia, Precisão, Revocação, Especificidade e Tempo). Como tais métricas podem apresentar escalas distintas e variâncias heterogêneas, a padronização é etapa crítica para evitar que variáveis com maior amplitude numérica dominem as componentes extraídas.

Denotando por \bar{y}_j a média amostral e por s_j o desvio-padrão amostral da j -ésima variável, define-se a padronização *z-score* por:

$$z_{ij} = \frac{y_{ij} - \bar{y}_j}{s_j}, \quad \bar{y}_j = \frac{1}{n} \sum_{i=1}^n y_{ij}, \quad s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_{ij} - \bar{y}_j)^2} \quad (2.84)$$

A matriz padronizada é, então, $\mathbf{Z} = [z_{ij}] \in \mathbb{R}^{n \times p}$. A partir de \mathbf{Z} , obtém-se a matriz de correlações amostrais (equivalente a covariância quando as variáveis são padronizadas):

$$\mathbf{R} = \frac{1}{n-1} \mathbf{Z}^T \mathbf{Z} \in \mathbb{R}^{p \times p} \quad (2.85)$$

A utilização de \mathbf{R} , em vez da covariância não padronizada, é recomendada quando as unidades de medida e magnitudes variam substancialmente entre as variáveis observadas.

2.8.2 Modelo de análise fatorial comum

A análise fatorial comum (*common factor analysis*) parte da hipótese de que a variabilidade conjunta das variáveis observadas pode ser explicada por um conjunto menor de variáveis latentes não observáveis, chamadas *fatores comuns*. Em sua forma matricial, o modelo fatorial pode ser escrito como:

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda} \mathbf{f} + \boldsymbol{\varepsilon} \quad (2.86)$$

Nessa expressão: (i) $\mathbf{x} \in \mathbb{R}^p$ é o vetor aleatório de variáveis observadas; (ii) $\boldsymbol{\mu} \in \mathbb{R}^p$ é o vetor de médias; (iii) $\mathbf{f} \in \mathbb{R}^m$ é o vetor de fatores comuns, com $m \ll p$; (iv) $\boldsymbol{\Lambda} \in \mathbb{R}^{p \times m}$ é a matriz de cargas fatoriais (*factor loadings*); e (v) $\boldsymbol{\varepsilon} \in \mathbb{R}^p$ é o vetor de termos específicos (ruído/variância única), assumido não explicado pelos fatores comuns.

Para fins de estimação e interpretação, são usualmente adotadas as seguintes hipóteses:

1. $\mathbb{E}[\mathbf{f}] = \mathbf{0}$ e $\mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}$;
2. $\text{Cov}(\mathbf{f}) = \boldsymbol{\Phi}$ (matriz de covariância dos fatores);
3. $\text{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}$, tipicamente diagonal (especificidades não correlacionadas);
4. $\text{Cov}(\mathbf{f}, \boldsymbol{\varepsilon}) = \mathbf{0}$.

Sob tais hipóteses, a matriz de covariância das variáveis observadas decompõe-se como:

$$\boldsymbol{\Sigma} = \text{Cov}(\mathbf{x}) = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi} \quad (2.87)$$

Em particular, quando se assume ortogonalidade entre fatores ($\boldsymbol{\Phi} = \mathbf{I}_m$), obtém-se:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi} \quad (2.88)$$

No modelo fatorial comum, a variância total de cada variável observada x_j é particionada em (i) parcela comum, explicada pelos fatores (h_j^2 , comunalidade), e (ii) parcela específica (ψ_j , unicidade), de modo que:

$$\text{Var}(x_j) = h_j^2 + \psi_j, \quad h_j^2 = \sum_{k=1}^m \lambda_{jk}^2, \quad \psi_j = \text{Var}(\varepsilon_j) \quad (2.89)$$

Do ponto de vista interpretativo, a matriz $\boldsymbol{\Lambda}$ é central: cada elemento λ_{jk} representa a associação linear entre a variável observada j e o fator k . Cargas fatoriais elevadas em módulo indicam que a variável contribui fortemente para a definição do fator.

Embora a análise fatorial comum modele explicitamente variância comum e variância específica Equações (2.87) a (2.89), a PCA é originalmente um procedimento de decomposição da variância total, sem distinção explícita entre componentes comuns e específicos. Ainda assim, é recorrente a utilização da PCA como método de extração inicial de fatores, sobretudo quando o objetivo é obter uma representação de baixa dimensão estável e numericamente bem condicionada.

Referenciais clássicos do modelo fatorial comum e de sua interpretação incluem Spearman (1904), Thurstone (1935) e Harman (1976).

2.8.3 Análise de Componentes Principais: formulação e propriedades

A PCA busca encontrar combinações lineares ortogonais das variáveis originais que maximizem a variância explicada. Partindo da matriz padronizada \mathbf{Z} Equação (2.84), considera-se a matriz de correlações \mathbf{R} Equação (2.85) e resolve-se o problema espectral:

$$\mathbf{R}\mathbf{v}_k = \lambda_k \mathbf{v}_k, \quad k = 1, \dots, p \quad (2.90)$$

Onde $\lambda_k \geq 0$ é o k -ésimo autovalor e $\mathbf{v}_k \in \mathbb{R}^p$ é o autovetor correspondente, com normalização $\|\mathbf{v}_k\|_2 = 1$. Ordenam-se os autovalores de forma não crescente, isto é:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0 \quad (2.91)$$

A k -ésima componente principal é definida pela combinação linear:

$$\mathbf{t}_k = \mathbf{Z}\mathbf{v}_k \in \mathbb{R}^n \quad (2.92)$$

Onde \mathbf{t}_k é o vetor de escores da componente principal k ao longo das n observações. A variância explicada por essa componente é igual ao autovalor associado:

$$\text{Var}(\mathbf{t}_k) = \lambda_k \quad (2.93)$$

A proporção de variância total explicada pela k -ésima componente é:

$$\rho_k = \frac{\lambda_k}{\sum_{j=1}^p \lambda_j} \quad (2.94)$$

Consequentemente, a variância acumulada explicada pelas m primeiras componentes é:

$$\rho_{\text{acum}}(m) = \sum_{k=1}^m \rho_k = \frac{\sum_{k=1}^m \lambda_k}{\sum_{j=1}^p \lambda_j} \quad (2.95)$$

A escolha de m (número de componentes/fatores) pode ser guiada por critérios como: (i) variância acumulada mínima desejada; (ii) critério de Kaiser ($\lambda_k > 1$ quando se usa \mathbf{R}); e (iii) inspeção do *scree plot* (ponto de inflexão da curva de autovalores). Independentemente do critério adotado, a PCA oferece uma base objetiva para reduzir dimensionalidade preservando, em sentido quadrático, a maior parcela possível da variância dos dados.

Definindo $\mathbf{V}_m = [\mathbf{v}_1, \dots, \mathbf{v}_m] \in \mathbb{R}^{p \times m}$ e $\mathbf{\Lambda}_m = \text{diag}(\lambda_1, \dots, \lambda_m)$, uma matriz de cargas associada às m componentes pode ser construída por:

$$\mathbf{L} = \mathbf{V}_m \mathbf{\Lambda}_m^{1/2} \in \mathbb{R}^{p \times m} \quad (2.96)$$

Cada elemento ℓ_{jk} quantifica a correlação entre a variável j e a componente k quando \mathbf{R} é utilizada.

Os escores das m primeiras componentes são:

$$\mathbf{T} = \mathbf{Z}\mathbf{V}_m \in \mathbb{R}^{n \times m} \quad (2.97)$$

Nessa definição, a k -ésima coluna de \mathbf{T} corresponde ao vetor \mathbf{t}_k da Equação (2.92).

A formulação fundacional da PCA remonta a Pearson (1901) e Hotelling (1933), sendo sistematizada e amplamente discutida em Jolliffe (2002).

2.8.4 Rotação *Varimax*: motivação e formulação

Um aspecto recorrente em aplicações práticas de PCA e análise fatorial é que as cargas fatoriais obtidas diretamente pela decomposição espectral tendem a ser densas: cada

componente pode apresentar contribuições não desprezíveis de muitas variáveis, o que dificulta interpretação semântica. Para mitigar esse problema, aplica-se uma rotação de fatores, cujo propósito é produzir uma estrutura simples (*simple structure*), em que cargas altas concentram-se em poucas variáveis por fator, enquanto cargas próximas de zero predominam nas demais.

Quando se busca preservar a ortogonalidade (isto é, fatores não correlacionados), utiliza-se uma matriz de rotação ortogonal $\mathbf{Q} \in \mathbb{R}^{m \times m}$ tal que:

$$\mathbf{Q}^T \mathbf{Q} = \mathbf{I}_m \quad (2.98)$$

A rotação atua sobre as cargas iniciais \mathbf{L} Equação (2.96) produzindo cargas rotacionadas:

$$\mathbf{A} = \mathbf{LQ} \quad (2.99)$$

No caso da rotação *Varimax*, \mathbf{Q} é escolhida para maximizar a variância das cargas ao quadrado dentro de cada fator, incentivando que cada variável apresente ou uma carga alta ou uma carga próxima de zero em um determinado fator. Um critério clássico de *Varimax* pode ser expresso por:

$$V(\mathbf{A}) = \sum_{k=1}^m \left[\sum_{j=1}^p a_{jk}^4 - \frac{1}{p} \left(\sum_{j=1}^p a_{jk}^2 \right)^2 \right] \quad (2.100)$$

A rotação *Varimax* busca, portanto:

$$\max_{\mathbf{Q}} V(\mathbf{LQ}) \quad \text{sujeito a} \quad \mathbf{Q}^T \mathbf{Q} = \mathbf{I}_m \quad (2.101)$$

Uma propriedade relevante é que rotações ortogonais preservam a variância total explicada pelas m componentes, uma vez que $\text{tr}(\mathbf{AA}^T) = \text{tr}(\mathbf{LL}^T)$. Assim, a rotação altera a distribuição das cargas (e, portanto, a interpretabilidade), mas não altera o subespaço gerado pelas componentes selecionadas.

O critério *Varimax* foi proposto por Kaiser (1958) e permanece como uma das rotações ortogonais mais utilizadas para obtenção de estrutura simples.

2.8.5 Escores fatoriais rotacionados e indeterminação de sinal

A rotação aplicada às cargas pode ser estendida aos escores. Se \mathbf{T} são os escores não rotacionados Equação (2.97), então os escores rotacionados podem ser obtidos por:

$$\mathbf{S} = \mathbf{TQ} \in \mathbb{R}^{n \times m} \quad (2.102)$$

O vetor linha $\mathbf{s}_i \in \mathbb{R}^m$ representa, para a i -ésima observação, a coordenada no espaço fatorado rotacionado. Em aplicações empíricas, é comum reportar: (i) a matriz de cargas rotacionadas \mathbf{A} Equação (2.99); e (ii) a matriz de escores rotacionados \mathbf{S} Equação (2.101).

Deve-se notar, entretanto, que fatores (e componentes) possuem indeterminação de sinal: se (\mathbf{A}, \mathbf{S}) é solução, então $(-\mathbf{A}, -\mathbf{S})$ descreve o mesmo subespaço e explica a mesma variância. Por essa razão, a orientação do sinal pode ser escolhida de modo a facilitar interpretação, por exemplo impondo que o fator associado a “qualidade” apresente cargas positivas nas métricas de desempenho (acurácia, precisão etc.), enquanto o fator associado a “tempo/custo” apresente cargas positivas no tempo de execução (ou, alternativamente, cargas negativas, dependendo da convenção adotada na etapa de otimização).

2.8.6 Construção de índices compostos: “Qualidade” e “Tempo”

Uma vez obtidos m fatores rotacionados, é possível construir índices compostos que resumem múltiplas métricas em objetivos de alto nível, com vistas a problemas multiobjetivo. Seja $\mathbf{s}_i = [s_{i1}, \dots, s_{im}]$ o vetor de escores rotacionados da observação i . Defina dois índices escalares:

1. um índice de Q_i , que agrega fatores associados às métricas de acerto preditivo;
2. um índice de C_i , que agrega fatores associados ao custo computacional (por exemplo, tempo médio por *fold*).

A construção mais direta consiste em selecionar subconjuntos de fatores $\mathcal{J}_Q \subseteq \{1, \dots, m\}$ e $\mathcal{J}_C \subseteq \{1, \dots, m\}$ com base na interpretação das cargas (isto é, observando quais métricas apresentam cargas dominantes em cada fator). Em seguida, definem-se agregações lineares ponderadas:

$$Q_i = \sum_{k \in \mathcal{J}_Q} w_k s_{ik}, \quad C_i = \sum_{k \in \mathcal{J}_C} u_k s_{ik} \quad (2.103)$$

Onde w_k e u_k são pesos não negativos. Uma escolha frequente consiste em ponderar por variância explicada relativa, por exemplo:

$$w_k = \frac{\lambda_k}{\sum_{j \in \mathcal{J}_Q} \lambda_j}, \quad u_k = \frac{\lambda_k}{\sum_{j \in \mathcal{J}_C} \lambda_j} \quad (2.104)$$

De modo que fatores com maior contribuição à variância do sistema tenham maior influência no índice. Alternativamente, pode-se atribuir pesos uniformes quando se deseja evitar indução de preferências implícitas.

Como Q_i e C_i são definidos em escala padronizada (devido à construção a partir de \mathbf{Z}), pode ser conveniente reescalá-los para um intervalo unitário, especialmente para comparações visuais e algoritmos de otimização. Um reescalonamento *min-max* em uma amostra de referência pode ser escrito como:

$$\tilde{Q}_i = \frac{Q_i - \min(Q)}{\max(Q) - \min(Q)}, \quad \tilde{C}_i = \frac{C_i - \min(C)}{\max(C) - \min(C)} \quad (2.105)$$

Nessa convenção, valores maiores de \tilde{Q}_i indicam melhor desempenho de qualidade, enquanto valores maiores de \tilde{C}_i indicam maior custo temporal (a ser minimizado). Caso a convenção desejada seja “maior é melhor” para ambos, basta definir $\tilde{C}_i^* = 1 - \tilde{C}_i$.

Sintetizando o fluxo: (i) matriz \mathbf{Y} de métricas, com dimensões $n \times p$; (ii) padronização \mathbf{Z} e construção de \mathbf{R} ; (iii) autodecomposição de \mathbf{R} e seleção de m componentes; (iv) estimação de \mathbf{Q} por *Varimax* e obtenção de \mathbf{A} e \mathbf{S} ; (v) cargas rotacionadas (interpretação) e escores rotacionados (insumo para índices Q_i e C_i).

No contexto de otimização multiobjetivo assistida por redução de dimensionalidade, a utilização de escores rotacionados como variáveis-resposta latentes é compatível com a estratégia de regressão por escores e pós-otimização discutida por Pereira *et al.* (2025).

2.8.7 Discussão crítica: explicabilidade, vantagens e limitações

Do ponto de vista da explicabilidade, a PCA e a análise fatorial têm caráter eminentemente linear, o que permite rastrear, de modo determinístico, a relação entre variáveis originais e fatores. A rotação *Varimax* reforça essa interpretabilidade ao promover estrutura simples, aproximando cada fator de um “construto” semântico, por exemplo: um fator associado à qualidade preditiva (cargas elevadas em acurácia, precisão, especificidade e/ou sensibilidade) e outro fator associado ao custo temporal (carga dominante em tempo de execução). Portanto, em comparação a agregações arbitrárias de métricas (por exemplo, somas não justificadas), o uso de PCA/*Varimax* oferece uma fundamentação mais rigorosa e auditável.

Entretanto, devem ser reconhecidas limitações importantes:

- Fatores e componentes são combinações lineares; relações não lineares entre métricas não são capturadas explicitamente.
- Escolhas distintas de escalonamento (covariância versus correlação; escores-z versus escalonamento robusto) podem alterar cargas e interpretação.
- A rotação ortogonal impõe fatores não correlacionados; em alguns domínios, fatores oblíquos podem ser mais realistas, porém com custo interpretativo e estatístico.
- A escolha de m , a orientação do sinal e o agrupamento de fatores em “qualidade” e “tempo” requerem critérios explícitos para evitar arbitrariedade.

É pertinente a intuição de que, ao empregar uma estratégia de “meta-otimização” (isto é, otimizar hiperparâmetros por meio de um procedimento de busca e comparação de candidatos), não se perde necessariamente explicabilidade, desde que a metodologia preserve rastreabilidade. Contudo, essa afirmação exige uma nuance crítica: a meta-otimização, por si só, não garante interpretabilidade; ela apenas redefine o processo de busca no espaço de decisões. A preservação da explicabilidade decorre do fato de que o processo é documentado e validado.

No presente trabalho, a rastreabilidade é mantida quando:

1. As cargas fatoriais (antes e depois da rotação) e os escores fatoriais são exportados e versionados como artefatos do experimento;
2. Os índices compostos (Q e C) são computados a partir de transformações explícitas Equações (2.103) a (2.105);
3. Os candidatos gerados na etapa de otimização são avaliados novamente nas métricas originais (não apenas nos índices), produzindo um conjunto final de resultados comparáveis e auditáveis.

Em outras palavras, a “comprovação” experimental (por exemplo, via execução de confirmação e reavaliação das métricas originais) não substitui a explicabilidade; ela a complementa. A explicabilidade reside na capacidade de explicar por que um candidato foi preferido (em termos de fatores e métricas), enquanto a comprovação reside em demonstrar que o candidato mantém desempenho sob o mesmo protocolo experimental. Assim, uma boa prática metodológica consiste em reportar simultaneamente: (i) os índices fatoriais que orientam a busca, e (ii) as métricas originais que sustentam a interpretação substantiva dos resultados.

No tocante às métricas de classificação comumente empregadas, precisão e revocação são tradicionais na literatura de recuperação de informação (Van Rijsbergen, 1979), enquanto sensibilidade/especificidade são clássicas em avaliação de diagnóstico (Yerushalmy, 1947). Discussões contemporâneas sobre relações e vieses dessas métricas podem ser encontradas em Powers (2011).

2.9 Método da Interseção Normal à Fronteira

O NBI foi proposto por Das e Dennis (1998) como uma estratégia geométrica para gerar pontos Pareto-ótimos em problemas de otimização com múltiplos objetivos. A ideia central é substituir a escolha direta de “pesos” em somas ponderadas por um procedimento que explora,

de forma sistemática, a geometria do espaço de objetivos, construindo interseções entre a fronteira factível e retas normais a uma superfície de referência definida pelos ótimos individuais.

2.9.1 Formulação do problema multiobjetivo

Considere um vetor de variáveis de decisão $\mathbf{x} \in \Omega \subset \mathbb{R}^d$ e n funções objetivo a minimizar, reunidas no vetor:

$$\mathbf{F}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{bmatrix}, \quad \text{com } \mathbf{x} \in \Omega, \quad g(\mathbf{x}) \leq 0, \quad h(\mathbf{x}) = 0 \quad (2.106)$$

Diz-se que $\mathbf{x}^a \in \Omega$ domina $\mathbf{x}^b \in \Omega$ se $f_i(\mathbf{x}^a) \leq f_i(\mathbf{x}^b)$ para todo i e existe pelo menos um índice j tal que $f_j(\mathbf{x}^a) < f_j(\mathbf{x}^b)$. Uma solução \mathbf{x}^* é Pareto-ótima quando não existe outra solução factível que a domine. O conjunto imagem dessas soluções no espaço de objetivos define a fronteira de Pareto.

Métodos baseados em soma ponderada, embora simples, tendem a apresentar limitações conhecidas (por exemplo, sensibilidade à escala entre objetivos e dificuldade de produzir amostragens uniformes ao longo da fronteira). O NBI foi concebido justamente como alternativa para mitigar esses problemas, usando uma construção geométrica fundamentada em um “*simplex*” no espaço de objetivos.

2.9.2 Elementos geométricos do NBI: pontos âncora, linha utopia e CHIM

O primeiro passo consiste em obter os ótimos individuais (também chamados de *anchor points* no espaço de decisão). Para cada objetivo $f_i(\mathbf{x})$, resolve-se o problema mono-objetivo:

$$\mathbf{x}_i^* = \operatorname{argmin}_{\mathbf{x} \in \Omega} f_i(\mathbf{x}), \quad i = 1, \dots, n \quad (2.107)$$

Os vetores $\mathbf{A}_i = \mathbf{F}(\mathbf{x}_i^*)$ são denominados pontos âncora no espaço de objetivos. A partir deles constrói-se a matriz *payoff*:

$$\Phi = [\mathbf{A}_1 \ \mathbf{A}_2 \ \dots \ \mathbf{A}_n] = [\mathbf{F}(\mathbf{x}_1^*) \ \mathbf{F}(\mathbf{x}_2^*) \ \dots \ \mathbf{F}(\mathbf{x}_n^*)] \quad (2.108)$$

O *convex hull of individual minima* (CHIM) é o conjunto convexo gerado por combinações convexas das colunas de Φ , isto é:

$$\left\{ \Phi \boldsymbol{\beta}: \boldsymbol{\beta} \in \mathbb{R}^n, \sum_{i=1}^n \beta_i = 1, \beta_i \geq 0 \right\} \quad (2.109)$$

No caso bidimensional ($n = 2$), o CHIM se reduz ao segmento de reta ligando \mathbf{A}_1 e \mathbf{A}_2 , frequentemente chamado de linha utopia. O ponto utópico (ou ideal) pode ser escrito como o vetor que reúne os melhores valores individuais:

$$\mathbf{F}^U = [f_1^*(\mathbf{x}_1^*), f_2^*(\mathbf{x}_2^*), \dots, f_n^*(\mathbf{x}_n^*)]^T \quad (2.110)$$

Em contraste, o ponto nadir corresponde a uma referência de “pior caso” na região factível do espaço de objetivos. Para $n = 2$, uma forma direta de obtê-lo a partir dos pontos âncora é:

$$\mathbf{F}^N = [f_1(\mathbf{x}_2^*), f_2(\mathbf{x}_1^*)]^T \quad (2.111)$$

Em problemas com $n > 2$, o nadir pode ser aproximado por combinações das avaliações cruzadas dos pontos âncora (ou por estimativas baseadas no conjunto Pareto amostrado), desde que mantenha consistência com a escala dos objetivos.

2.9.3 Normalização e formulação básica do subproblema NBI

Como objetivos podem possuir unidades e ordens de grandeza muito distintas, a formulação do NBI é usualmente escrita em termos de objetivos escalonados (adimensionais). Para cada objetivo, define-se a normalização:

$$\bar{f}_i(\mathbf{x}) = \frac{f_i(\mathbf{x}) - f_i^U}{f_i^N - f_i^U} \quad (2.112)$$

Onde f_i^U e f_i^N são, respectivamente, as componentes de \mathbf{F}^U e \mathbf{F}^N . Essa transformação mapeia cada objetivo para a faixa $[0,1]$ quando f_i^U e f_i^N são referências consistentes.

Denotando por $\bar{\mathbf{F}}(\mathbf{x})$ o vetor de objetivos escalonados e por $\bar{\Phi}$ a matriz *payoff* escalonada, cada vetor de pesos β_k (com $\sum_i \beta_i = 1$ e $\beta_i \geq 0$) identifica um ponto no CHIM, $\bar{\Phi}\beta_k$. O NBI então procura o ponto sobre a fronteira factível obtido ao “caminhar” a partir desse ponto na direção normal ao CHIM, orientada para a origem. Em muitas implementações, a normal unitária é substituída por um vetor quasi-normal proporcional à geometria do CHIM, dado por:

$$\hat{\mathbf{n}} = -\bar{\Phi}\mathbf{e} \quad (2.113)$$

Onde \mathbf{e} é o vetor-coluna de uns. Com isso, o subproblema NBI associado a β_k pode ser escrito como o seguinte sistema equivalente:

$$\left. \begin{array}{l} \text{Max}_{(x,t)} t \\ \text{sujeito a: } \overline{\Phi}\beta_k + t\hat{\mathbf{n}} = \overline{\mathbf{F}}(\mathbf{x}) \\ g(\mathbf{x}) \leq 0, \quad h(\mathbf{x}) = 0 \\ \mathbf{x} \in \Omega \end{array} \right\} \Rightarrow \left. \begin{array}{l} \text{Min}_{(x,t)} (-t) \\ \text{sujeito a: } \overline{\mathbf{F}}(\mathbf{x}) - \overline{\Phi}\beta_k - t\hat{\mathbf{n}} = 0 \\ g(\mathbf{x}) \leq 0, \quad h(\mathbf{x}) = 0 \\ \mathbf{x} \in \Omega \end{array} \right\} \quad (2.114)$$

O escalar t controla a distância percorrida sobre a normal. Intuitivamente, maximizar t empurra a solução ao longo da normal até tocar a fronteira factível, gerando um ponto Pareto-ótimo associado àquela escolha de β_k .

2.9.4 Caso bidimensional: relações geométricas e redução algébrica

No caso $n = 2$, é possível obter uma interpretação particularmente clara. A linha utopia é o segmento entre os pontos âncora \mathbf{A}_1 e \mathbf{A}_2 . Um ponto genérico nessa linha pode ser parametrizado por pesos (β_1, β_2) , com $\beta_1 + \beta_2 = 1$.

A partir de relações de semelhança de triângulos na construção geométrica do NBI (Das e Dennis, 1998), obtêm-se as proporções:

$$\frac{\Delta A_2 B_2 C_2}{\Delta A_1 A_2 C_1} = \frac{\overline{B_1 C_1}}{A_1 C_1} = \frac{f_2(\mathbf{x}) - f_2^*(\mathbf{x}_2^*)}{f_2(\mathbf{x}_1^*) - f_2^*(\mathbf{x}_2^*)} = \frac{\beta_1}{\beta_1 + \beta_2} \quad (2.115)$$

$$\frac{\Delta A_1 B_1 C_2}{\Delta A_1 A_2 C_1} = \frac{\overline{C_1 B_2}}{A_1 C_1} = \frac{f_1(\mathbf{x}) - f_1^*(\mathbf{x}_1^*)}{f_1(\mathbf{x}_2^*) - f_1^*(\mathbf{x}_1^*)} = \frac{\beta_2}{\beta_1 + \beta_2} = \frac{1 - \beta_1}{\beta_1 + \beta_2} \quad (2.116)$$

Com as definições anteriores, também é possível escrever explicitamente a relação $\mathbf{F}(\mathbf{x}) = \Phi\beta$ para duas funções objetivo:

$$\mathbf{F}(\mathbf{x}) = \Phi\beta \begin{cases} f_1(\mathbf{x}) = \beta_1 f_1^*(\mathbf{x}_1^*) + \beta_2 f_1(\mathbf{x}_2^*) \\ f_2(\mathbf{x}) = \beta_1 f_2(\mathbf{x}_1^*) + \beta_2 f_2^*(\mathbf{x}_2^*) \end{cases} \Rightarrow \underbrace{\begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \end{bmatrix}}_{\check{\mathbf{F}}(\mathbf{x})} = \underbrace{\begin{bmatrix} f_1^*(\mathbf{x}_1^*) & f_1(\mathbf{x}_2^*) \\ f_2(\mathbf{x}_1^*) & f_2^*(\mathbf{x}_2^*) \end{bmatrix}}_{\check{\Phi}} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}}_{\check{\beta}} \quad (2.117)$$

Ao normalizar o sistema, obtêm-se $\overline{\mathbf{F}}(\mathbf{x}) = \overline{\Phi}\beta$ com:

$$\overline{\mathbf{F}}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) - f_1^*(\mathbf{x}_1^*) \\ f_1(\mathbf{x}_2^*) - f_1^*(\mathbf{x}_1^*) \\ f_2(\mathbf{x}) - f_2^*(\mathbf{x}_2^*) \\ f_2(\mathbf{x}_1^*) - f_2^*(\mathbf{x}_2^*) \end{bmatrix} \quad \overline{\Phi} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad (2.118)$$

Substituindo $\hat{\mathbf{n}} = -\overline{\Phi}\mathbf{e}$ na Eq. (2.106) e reorganizando os termos, chega-se a:

$$\overline{\mathbf{F}}(\mathbf{x}) - \overline{\Phi}(\beta - t\mathbf{e}) = \begin{bmatrix} \overline{f}_1(\mathbf{x}) - \beta_2 + t \\ \overline{f}_2(\mathbf{x}) - \beta_1 + t \end{bmatrix} \quad (2.119)$$

Impondo-se a igualdade a zero, obtêm-se $-t = \overline{f}_1(\mathbf{x}) - \beta_2$ e, analogamente, $-t = \overline{f}_2(\mathbf{x}) - \beta_1$.

Combinando essas relações, elimina-se t e obtêm-se uma restrição escalar que relaciona diretamente as duas funções normalizadas:

$$\bar{f}_1(\mathbf{x}) - \bar{f}_2(\mathbf{x}) + 2\beta_1 - 1 = 0 \quad (2.120)$$

Além disso, do fato de $-t = \bar{f}_1(\mathbf{x}) - \beta_2$ e $\beta_2 \in [0,1]$, segue que minimizar $(-t)$ equivale a minimizar $\bar{f}_1(\mathbf{x})$ (pois o melhor caso para o deslocamento é $\beta_2 = 0$):

$$\text{Min}_{\mathbf{x} \in \Omega} (-t) = \text{Min}_{\mathbf{x} \in \Omega} \bar{f}_1(\mathbf{x}), \quad \text{com } \beta \in [0,1] \quad (2.121)$$

Com base nas Eqs. (2.112) e (2.113), é possível escrever um esquema algébrico do NBI bidimensional destacando que o subproblema multiobjetivo pode ser interpretado como um problema mono-objetivo com uma restrição de acoplamento entre objetivos:

$$\text{Min}_{(\mathbf{x}, t)} (-t) \quad \text{sujeito a: } \bar{\mathbf{F}}(\mathbf{x}) - \bar{\Phi}\boldsymbol{\beta}_k - t\hat{\mathbf{n}} = 0, \quad g(\mathbf{x}) \leq 0, \quad h(\mathbf{x}) = 0 \quad (2.122)$$

$$\text{Min}_{\mathbf{x} \in \Omega} \bar{f}_1(\mathbf{x}) \quad \text{sujeito a: } \bar{f}_1(\mathbf{x}) - \bar{f}_2(\mathbf{x}) + 2\beta_1 - 1 = 0, \quad g(\mathbf{x}) \leq 0, \quad h(\mathbf{x}) = 0$$

2.9.5 Discretização do *simplex* de pesos e número de subproblemas

Uma etapa prática do NBI é definir quantos vetores de pesos $\boldsymbol{\beta}_k$ serão avaliados. Assumindo espaçamento uniforme δ entre valores sucessivos de peso em cada dimensão do simplex e definindo $p = 1/\delta$, o número de subproblemas (isto é, o número de vetores $\boldsymbol{\beta}_k$ no grid) pode ser calculado por:

$$N_{\text{Sub}} = \binom{n+p-1}{p} \quad (2.123)$$

Esse resultado é estruturalmente equivalente à contagem de pontos de um *Simplex-Lattice Design* $\{q, m\}$, amplamente utilizado em experimentos de misturas. Para esse tipo de planejamento, a quantidade de pontos é dada por:

$$N_{\text{runs}} = \binom{q+m-1}{m} = \frac{(q+m-1)!}{m!(q-1)!} \quad (2.124)$$

Comparando as equações (2.123) e (2.124), identifica-se a correspondência $q = n$ e $m = p = 1/\delta$. Essa equivalência é útil porque conecta a geração de pesos do NBI a um planejamento estatístico bem estabelecido, permitindo, por exemplo, randomização de sequências, controle direto do orçamento computacional (via escolha de δ) e análises de sensibilidade sobre como os pesos influenciam a distribuição e a qualidade dos pontos Pareto-ótimos.

2.9.6 Procedimento resumido para aplicação do NBI

Em termos operacionais, uma implementação típica do NBI pode ser descrita pelos passos:

1. Otimizações individuais: resolver n problemas mono-objetivo para obter \mathbf{x}_i^* e montar Φ .
2. Escalonamento: definir \mathbf{F}^U e \mathbf{F}^N e construir $\bar{\mathbf{F}}(\mathbf{x})$ e $\bar{\Phi}$ (2.108 - 2.113).
3. Grade de pesos: escolher o número de divisões q e gerar o conjunto de vetores de pesos $\{\beta_k\}$ sobre o simplex (Eqs. 2.123 - 2.124).
4. Subproblemas NBI: para cada β_k , resolver o sistema (2.122) e registrar o ponto Pareto-ótimo correspondente.
5. Pós-análise: avaliar a distribuição do conjunto Pareto amostrado e, quando aplicável, aplicar métricas de apoio à decisão para selecionar uma solução final conforme critérios do estudo.

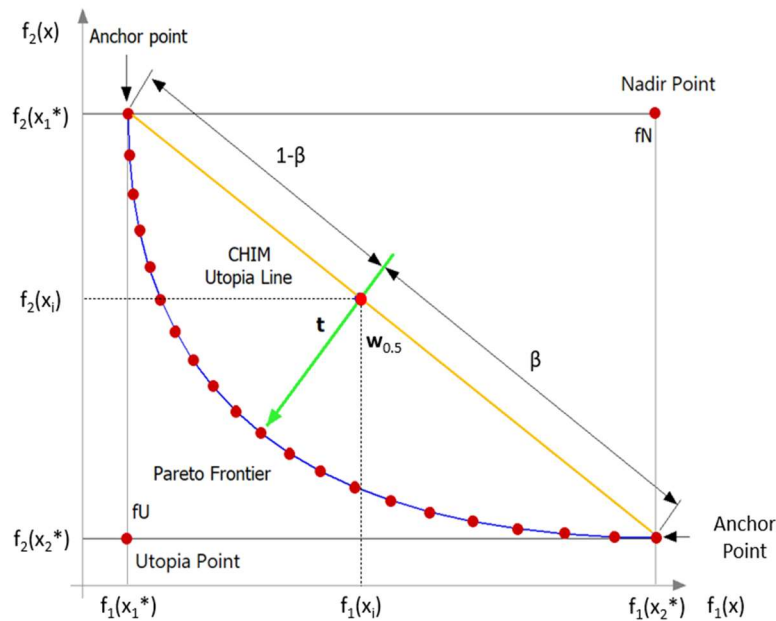


Figura 2 - Fronteira de Pareto obtida via *Normal Boundary Intersection* (NBI). Adaptado de Brito *et al.* (2014), com base em Das e Dennis (1998).

2.9.7 Conexões com a literatura e implicações

O método NBI vem sendo estendido e aplicado de forma recorrente a problemas de otimização multiobjetivo com respostas correlacionadas. Lopes *et al.* introduziram uma formulação multivariada robusta baseada em NBI que integra Análise de Componentes

Principais e projeto robusto de parâmetros, produzindo frentes de Pareto uniformemente espaçadas e estáveis (Lopes *et al.*, 2016). Em seguida, Costa *et al.* propuseram o acoplamento NBI-MMSE para otimização multivariada em usinagem a seco, evidenciando vantagens frente a somas ponderadas e melhor adequação em cenários com respostas correlacionadas (Costa *et al.*, 2016). Evoluções subsequentes incluem o uso de escores fatoriais rotacionados para reduzir dimensionalidade mantendo interpretabilidade, o que reforça a viabilidade de NBI em contextos com múltiplas métricas de desempenho (Naves *et al.*, 2017).

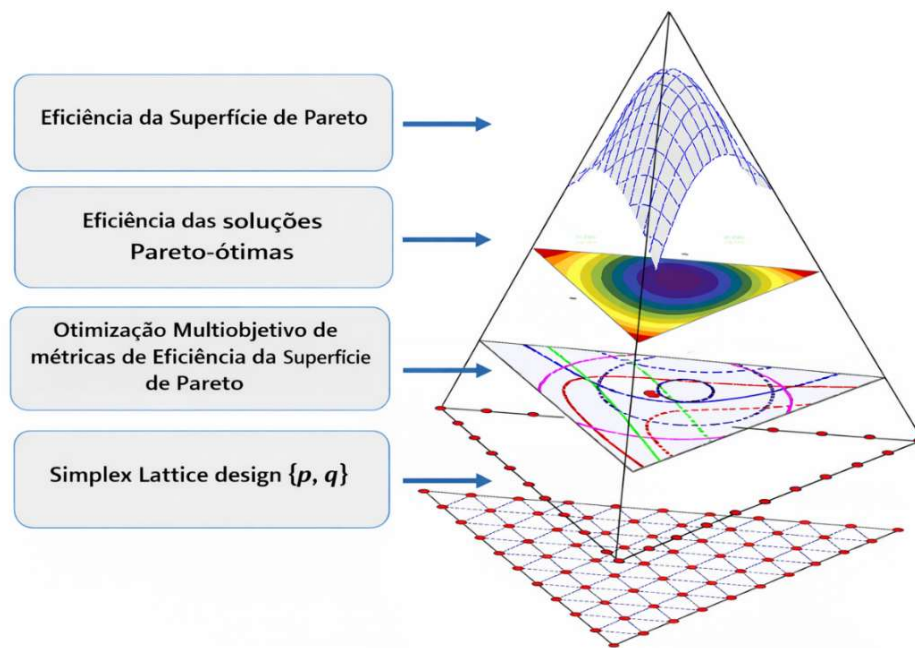


Figura 3 - Encadeamento conceitual: *Simplex Lattice* $\{p, q\}$, Otimização Multiobjetivo, Eficiência MODM, Superfície de Pareto.

O protocolo experimental de calibração de hiperparâmetros segue a linha de trabalhos do grupo que empregam DoE para reduzir custo exploratório e garantir reprodutibilidade estatística. Em particular, a estratégia de definição de níveis e blocos na etapa de treinamento de redes neurais artificiais é consistente com o procedimento proposto por Balestrassi *et al.* para previsão de séries temporais (Balestrassi; Popova; Paiva; Lima, 2009), e com o fluxo sistemático de projeto e busca focalizada de Pontes *et al.* (2016). Assim, as escolhas de fatores controlados, réplicas e critérios de análise encontram suporte em evidências já publicadas. Além de aplicações em redes neurais, a linha metodológica DoE+NBI foi expandida para problemas multivariados com respostas correlacionadas em manufatura e previsão, demonstrando generalidade do arcabouço. Destacam-se a otimização multivariada robusta via NBI com PCA (Lopes *et al.*, 2016); o NBI acoplado ao MMSE em fresamento a seco (Costa *et al.*, 2016); e a formulação baseada em escores fatoriais rotacionados para reduzir dimensionalidade sem

perder propriedades de Pareto (Naves *et al.*, 2017). Aplicações correlatas incluem a redução de dimensionalidade em problemas de engenharia computacional via NBI com PCA (Belinato *et al.*, 2019) e a modelagem multivariada robusta em forças de usinagem (Pereira *et al.*, 2018). Em previsão, ver ainda o acoplamento NBI com delineamentos de mistura e escores rotacionados (Bacci *et al.*, 2019). Essas contribuições consolidam a base metodológica adotada neste trabalho.

2.10 Otimização de Hiperparâmetros em XGBoost

A discussão dos hiperparâmetros do XGBoost (Seções 2.4.1 - 2.4.10) evidencia que a escolha desses parâmetros não é um detalhe de implementação, mas parte constitutiva do desempenho e da viabilidade computacional do modelo. Em algoritmos de *gradient boosting* com árvores, pequenas variações em parâmetros de complexidade, regularização e amostragem podem produzir diferenças substanciais tanto em métricas de generalização quanto no tempo de treinamento, sobretudo em bases tabulares de alta dimensionalidade e/ou com ruído e colinearidade. Essa sensibilidade é coerente com a própria motivação do XGBoost: combinar um procedimento de *boosting* eficiente com mecanismos explícitos de regularização e de controle de complexidade para produzir modelos competitivos em termos de acurácia e custo computacional (Chen; Guestrin, 2016).

Do ponto de vista conceitual, a otimização de hiperparâmetros (*hyperparameter optimization*, HPO, também chamada de *hyperparameter tuning*) consiste em automatizar a busca por um vetor de configurações λ que maximize (ou minimize) uma função objetivo baseada em desempenho fora da amostra. Considere um conjunto de dados \mathcal{D} particionado em subconjuntos de treinamento e validação, respectivamente $\mathcal{D}_{\text{train}}$ e \mathcal{D}_{val} , um algoritmo de treinamento \mathcal{A} (no caso, o procedimento de boosting do XGBoost) e um vetor de hiperparâmetros $\lambda \in \Lambda$, onde Λ representa o espaço de busca. Para uma métrica de avaliação $\mathcal{M}(\cdot)$ (por exemplo, AUC, *logloss*, F_1 ou *balanced accuracy*), pode-se definir a função objetivo como:

$$f(\lambda) = \mathcal{M}(\mathcal{A}(\mathcal{D}_{\text{train}}, \lambda), \mathcal{D}_{\text{val}}) \quad (2.125)$$

E, assim, escrever o problema como:

$$\lambda^* = \operatorname{argmax}_{\lambda \in \Lambda} f(\lambda) \quad (2.126)$$

Em termos mais gerais, a literatura contemporânea descreve HPO como um problema em dois níveis: no nível interno, ajustam-se os parâmetros do modelo (isto é, o ensemble de

árvores e seus parâmetros estruturais) condicionados a λ , enquanto no nível externo seleciona-se λ com base em generalização (Franceschi *et al.*, 2024). Essa visão é particularmente útil para diferenciar (i) parâmetros “aprendidos” a partir dos dados durante o treinamento (por exemplo, a estrutura específica de cada árvore) de (ii) hiperparâmetros que governam o processo de aprendizagem, como taxa de aprendizagem, profundidade e regularização (Chen; Guestrin, 2016; Franceschi *et al.*, 2024).

2.10.1 Natureza de “caixa-preta”, custo e ruído

Na prática, a função $f(\lambda)$ raramente é conhecida em forma fechada, podendo ser não convexa, não suave e definida em espaço misto (contínuo, inteiro e categórico). Além disso, sua avaliação é custosa, pois envolve treinar e validar um modelo completo para cada ponto no espaço Λ (Franceschi *et al.*, 2024). No caso do XGBoost, embora o procedimento de treinamento possa ser determinístico dado um conjunto fixo de dados e uma semente aleatória, a avaliação típica introduz estocasticidade por múltiplas vias: partições de validação (*holdout* ou *k-fold*), reamostragem interna (subamostragem de linhas e colunas), além de variações do *pipeline* (por exemplo, diferentes estratégias de imputação, padronização ou seleção de atributos quando incorporadas ao fluxo de modelagem). Assim, é frequente modelar $f(\lambda)$ como uma variável aleatória, onde cada avaliação é uma realização sujeita a variabilidade estatística. Essa interpretação conecta HPO a problemas clássicos de otimização sob incerteza e, mais especificamente, à otimização por simulação (*simulation optimization*), em que a função objetivo é observada via simulações ruidosas em vez de calculada diretamente (Ramamohan *et al.*, 2024).

2.10.2 HPO como configuração de algoritmo e implicações para XGBoost.

A literatura frequentemente enquadra a escolha de hiperparâmetros como um caso particular de configuração de algoritmos, na qual se deseja encontrar uma parametrização que minimize um custo sobre um conjunto de instâncias (Hutter; Hoos; Leyton-Brown, 2011). Nesse enquadramento, a “instância” pode representar uma tarefa (conjunto de dados, em inglês, *dataset*), uma partição de validação ou um cenário de execução (por exemplo, limite de tempo), e a avaliação de uma configuração resulta em um custo observado (erro, perda, tempo). Mesmo quando o objetivo é otimizar para uma única base de dados, a analogia é útil porque explicita

que o desempenho observado depende do protocolo de validação e pode variar por ruído amostral, tornando decisões sobre replicação e agregação parte do problema.

Em particular, SMBOs foram concebidos para cenários de avaliação cara e espaços mistos, nos quais cada execução do algoritmo é tratada como uma “caixa-preta” e se busca maximizar eficiência amostral por meio de escolhas informadas das próximas configurações (Hutter; Hoos; Leyton-Brown, 2011; Franceschi *et al.*, 2024). No caso do XGBoost, essa conexão é especialmente relevante porque (i) há hiperparâmetros categóricos e condicionais (por exemplo, escolhas de *booster* que ativam conjuntos distintos de parâmetros) e (ii) o custo de treinamento pode variar por ordens de magnitude entre configurações. Assim, abordagens SMBO e correlatas tendem a ser naturais quando se busca reduzir o número de tentativas necessárias para atingir níveis elevados de desempenho sob orçamento restrito.

2.10.3 Estrutura do espaço de busca em XGBoost

A escolha de Λ é uma decisão metodológica com impacto direto na eficiência da busca. No XGBoost, Λ é tipicamente composto por: (i) hiperparâmetros inteiros, como profundidade máxima e número de estimadores; (ii) hiperparâmetros contínuos, como frações de amostragem e taxas de regularização (α e λ); e (iii) escolhas categóricas, como *booster* e função objetivo. Além disso, há restrições de domínio (por exemplo, *subsample* em $(0,1]$) e hiperparâmetros condicionais, isto é, parâmetros que só fazem sentido sob determinado *booster*. Esse caráter condicional e de tipagem mista aproxima a configuração do XGBoost do problema de “configuração automática de algoritmos” (*algorithm configuration*), cujo objetivo é escolher parâmetros de controle de um algoritmo para otimizar desempenho empírico sob orçamento limitado (Hutter; Hoos; Leyton-Brown, 2011; Franceschi *et al.*, 2024). Em espaços estruturados, métodos capazes de lidar com condicionais e com diferentes escalas, como abordagens baseadas em modelos probabilísticos e surrogadas (*surrogates*) que suportam variáveis mistas, tendem a apresentar vantagens sobre estratégias enumerativas ingênuas, principalmente à medida que a dimensionalidade cresce (Bergstra *et al.*, 2011; Shahriari *et al.*, 2016; Franceschi *et al.*, 2024).

2.10.4 Interações e regimes de equivalência entre hiperparâmetros.

Um traço marcante do XGBoost é que muitos hiperparâmetros não atuam de forma isolada, mas em conjunto, formando “regimes” de comportamento que podem ser aproximadamente equivalentes em desempenho, porém distintos em custo. Entre as interações mais recorrentes, destacam-se:

1. A relação entre taxa de aprendizagem e número de estimadores. Em geral, valores menores de taxa de aprendizagem exigem mais iterações para atingir desempenho comparável, o que tende a aumentar o tempo de treinamento. Por outro lado, valores muito altos podem levar a atualizações agressivas e pior generalização, especialmente em presença de ruído e variáveis correlacionadas (Chen; Guestrin, 2016);
2. A interação entre profundidade, peso mínimo por nó e penalização para dividir um nó. Profundidades maiores aumentam expressividade, mas também risco de sobreajuste; peso mínimo do nó filho e perda mínima por divisão podem atuar como contrapesos, restringindo divisões com baixo ganho ou suportes pequenos;
3. O papel combinado da fração de amostragem por instância, fração de colunas por árvore e da regularização explícita. Subamostragem pode reduzir variância e custo, funcionando como regularizador estocástico; entretanto, subamostragem excessiva pode elevar viés e reduzir estabilidade, o que é particularmente relevante quando $f(\lambda)$ é avaliado via partições com variância elevada;
4. A influência de critérios auxiliares como parada antecipada, que altera a natureza do problema: ao invés de fixar o número de estimadores, a parada antecipada define um mecanismo adaptativo de seleção do número efetivo de iterações em função de um critério em validação. Assim, a parada antecipada pode ser visto como um “hiperparâmetro procedural” ou como uma forma de alocação adaptativa de recursos dentro de cada avaliação.

Essas interações reforçam a limitação de análises univariadas e justificam abordagens experimentais que explorem efeitos principais e interações (via planejamento de experimentos) ou que modelem explicitamente dependências entre variáveis (via modelos substitutos e estratégias sequenciais) (Montgomery, 2017; Franceschi *et al.*, 2024). Além disso, a existência de regimes equivalentes sugere que “melhor configuração” pode depender não apenas da métrica, mas do orçamento computacional e da estabilidade desejada.

2.10.5 Importância relativa e estratégia de alocação do orçamento

Um resultado recorrente em HPO é que nem todos os hiperparâmetros contribuem igualmente para o desempenho: frequentemente, apenas um subconjunto domina a variação explicada, enquanto outros atuam como ajustes finos. Probst; Boulesteix; Bischl (2019) discutem esse fenômeno por meio do conceito de *tunability*, enfatizando que (i) algoritmos diferem quanto ao quanto vale a pena tunar e (ii) a sensibilidade pode depender do conjunto de dados e da métrica. Em modelos de árvores impulsionadas, hiperparâmetros relacionados à complexidade do aprendiz base (profundidade e restrições de divisão), à regularização e à taxa de aprendizagem tendem a exercer impacto central, ao passo que outros hiperparâmetros podem ter efeito secundário ou se tornar relevantes apenas em regimes específicos.

Esse fato tem implicações diretas para eficiência: sob orçamento restrito, a escolha de quais dimensões explorar com maior resolução pode ser tão importante quanto o algoritmo de busca em si. Além disso, a força de baselines simples como a busca aleatória pode ser explicada quando as dimensões realmente relevantes são poucas e o espaço é grande; nesse cenário, amostrar combinações tende a ser mais eficiente do que varrer grades regulares em dimensões pouco influentes (Bergstra; Bengio, 2012). Em XGBoost, onde existem interações fortes e escalas muito diferentes entre hiperparâmetros, o desenho do espaço (limites, escalas e condicionais) e a adoção de estratégias sistemáticas de amostragem (incluindo planejamento de experimentos) tornam-se parte essencial do próprio problema de otimização (Franceschi *et al.*, 2024).

2.10.6 Métrica e protocolo de validação: risco de sobreajuste na seleção de métricas-alvo

Em HPO, a métrica \mathcal{M} e o protocolo de validação definem o “alvo” da busca. Para problemas de classificação, diferentes métricas privilegiam compromissos distintos entre sensibilidade e especificidade, podendo levar a ótimos conflitantes. A situação se agrava quando múltiplos ciclos de busca reutilizam repetidamente o mesmo conjunto de validação: ainda que cada modelo individual não sobreajuste diretamente ao conjunto de validação (pois ele não é usado para ajustar parâmetros internos), o procedimento de seleção pode sobreajustar em nível de escolha de hiperparâmetros, gerando viés otimista na estimativa de desempenho. Esse fenômeno é conhecido como sobreajuste em seleção de modelo (*model selection overfitting*) e está associado ao chamado viés de seleção *selection bias* em avaliação

comparativa (Cawley; Talbot, 2010; Franceschi *et al.*, 2024). Uma consequência prática é que avaliações reportadas no conjunto de validação podem superestimar a performance esperada em dados novos, sobretudo quando o orçamento de busca é grande, quando o espaço Λ é amplo ou quando o ruído na avaliação é elevado.

A literatura recomenda, como contramedidas, o uso de estratégias como validação cruzada aninhada (*nested cross-validation*) para estimativa imparcial do desempenho do procedimento completo (incluindo HPO), ou a separação explícita de três subconjuntos: treino, validação para HPO e teste final “intocado” para estimativa final. Em problemas onde custo computacional é crítico, variantes pragmáticas incluem: (i) reduzir a agressividade do HPO (menor orçamento) para mitigar seleção oportunista; (ii) usar reamostragens repetidas e agregação (médias e intervalos) para reduzir variância; ou (iii) empregar critérios de decisão que considerem incerteza e não apenas o valor pontual da métrica (Cawley; Talbot, 2010; Ramamohan *et al.*, 2024).

2.10.7 Custo computacional, multi-fidelidade e alocação adaptativa de recursos

Um obstáculo central em HPO é o custo de avaliação: se cada tentativa exige treinar centenas ou milhares de árvores profundas, a busca torna-se rapidamente proibitiva. Em resposta, emergiu uma classe de métodos chamada multi-fidelidade (*multi-fidelity* HPO), que utiliza aproximações baratas para priorizar configurações promissoras e investir mais recursos apenas onde há evidências de ganho (Franceschi *et al.*, 2024; Won; Lee; Lee, 2025). Em XGBoost, há fidelidades naturais e operacionalizáveis: (i) reduzir o número de estimadores; (ii) limitar profundidade máxima; (iii) treinar em subconjuntos de instâncias ou de variáveis; e (iv) utilizar *early stopping* como mecanismo interno de interrupção. Em essência, essas estratégias constroem uma curva de aprendizagem parcial para cada configuração, permitindo descartar precocemente alternativas fracas.

Do ponto de vista algorítmico, uma referência importante é o *Hyperband*, que formaliza a alocação adaptativa de orçamento como um problema de exploração e eliminação sucessiva de configurações, combinando amostragem aleatória com paradas antecipadas e redistribuição de recursos (Li *et al.*, 2018). O BOHB, por sua vez, combina o princípio multi-fidelidade (inspirado no *Hyperband*) com um componente de otimização bayesiana baseado em TPE, buscando melhorar a amostragem de configurações em função de desempenho observado e mantendo bons resultados sob orçamentos variados (Falkner; Klein; Hutter, 2018). Para

XGBoost, esses métodos são conceitualmente compatíveis com o uso de *early stopping* e com o fato de que o número de iterações/árvores é um recurso ajustável. Entretanto, a adoção prática requer cuidado: fidelidades muito “agressivas” (por exemplo, treinar com pouquíssimas iterações) podem inverter a ordem de mérito entre configurações, pois algumas combinações de hiperparâmetros (por exemplo, taxas de aprendizagem pequenas) só se tornam competitivas em regimes de maior orçamento.

2.10.8 Estocasticidade e perspectiva de otimização por simulação

Mesmo em ambientes controlados, HPO pode ser entendido como uma otimização estocástica, pois cada avaliação envolve variabilidade amostral e, frequentemente, aleatoriedade operacional (subamostragem, *shuffle*, inicialização, paralelismo). Nessa perspectiva, comparar duas configurações λ_1 e λ_2 significa comparar médias (ou quantis) de distribuições de desempenho, e não valores determinísticos. Ramamohan *et al.* (2024) argumentam que a analogia com otimização por simulação é natural: o avaliador de desempenho do modelo desempenha o papel de um “simulador” que, dado um ponto no espaço de decisão (hiperparâmetros), retorna uma observação ruidosa da função objetivo. A partir desse enquadramento, os autores investigam métodos de otimização discretos e estatísticos, como técnicas de *ranking and selection* (R&S) e procedimentos de busca aleatória com garantias assintóticas, além de estabelecer bases teóricas para sua aplicação em *tuning*.

Esse enquadramento é útil por duas razões em XGBoost. Primeiro, reforça a necessidade de replicação: ao repetir a avaliação de uma configuração sob diferentes partições e/ou sementes, pode-se estimar não só a média de desempenho, mas também sua variância, reduzindo a probabilidade de selecionar uma configuração “sortuda” por ruído. Segundo, abre espaço para critérios de decisão baseados em significância ou garantias estatísticas, em contraste com a seleção puramente por melhor valor pontual. Em trabalhos como o de Ramamohan *et al.* (2024), métodos como o procedimento de Kim e Nelson e variações do *stochastic ruler* são comparados a bibliotecas de HPO amplamente usadas (por exemplo, estratégias aleatórias e TPE), ilustrando que, sob certos regimes de ruído e custo, abordagens estatísticas podem ser competitivas e, em alguns casos, superiores.

2.10.9 Multiobjetividade: desempenho versus custo (e outros critérios)

Para aplicações reais, “melhor” configuração raramente significa apenas maximizar uma métrica de acerto. Em XGBoost, configurações com maior capacidade de ajuste tendem a aumentar tempo de treinamento e, por vezes, custo de inferência. Assim, HPO pode ser formulada como um problema multiobjetivo, em que se deseja simultaneamente maximizar desempenho preditivo e minimizar custo computacional (tempo, memória, energia), ou ainda controlar outros atributos como estabilidade ou complexidade do modelo. Um exemplo de formulação biobjetivo é:

$$\min_{\lambda \in \Lambda} \mathbf{F}(\lambda) = \min_{\lambda \in \Lambda} [1 - f(\lambda), T(\lambda)] \quad (2.127)$$

Onde $T(\lambda)$ representa uma medida de custo como tempo de execução. Nessa classe de problemas, busca-se tipicamente o conjunto de soluções não-dominadas (fronteira de Pareto), do qual uma solução de compromisso é escolhida segundo preferências do decisor (Miettinen, 1999). Franceschi *et al.* (2024) discutem explicitamente extensões multiobjetivo em HPO, apontando que a inclusão de custo e/ou restrições se torna cada vez mais relevante à medida que modelos e bases crescem.

Em XGBoost, a multiobjetividade também pode ser interpretada em termos de orçamento computacional: para um orçamento de tempo fixo, algumas configurações são inviáveis; já para um orçamento flexível, pode-se aceitar custo maior em troca de desempenho marginalmente superior. Portanto, a discussão multiobjetivo não é apenas uma extensão teórica, mas uma forma de explicitar escolhas implícitas que já ocorrem na prática.

2.10.10 Implicações práticas: definição de espaços, escalas e orçamentos

Uma decisão frequentemente subestimada em HPO é a parametrização do espaço de busca: além de escolher limites, deve-se escolher escalas (linear ou logarítmica), distribuições de amostragem (uniforme, log-uniforme) e relações condicionais. Em XGBoost, é comum que hiperparâmetros de regularização e taxa de aprendizagem operem em ordens de grandeza distintas, o que favorece escalas logarítmicas para exploração eficiente. Além disso, evidências empíricas sugerem que nem todos os hiperparâmetros têm o mesmo impacto: alguns são fortemente “ajustáveis” (*tunable*) e dominam o ganho de desempenho, enquanto outros exercem efeito marginal ou dependente de contexto. Probst; Boulesteix; Bischl (2019) formalizam e quantificam a importância de hiperparâmetros em diferentes algoritmos,

reforçando a ideia de que estratégias de *tuning* devem priorizar variáveis mais influentes quando o orçamento é limitado.

Por fim, comparar metodologias de HPO (ou reportar resultados) requer explicitar o orçamento e o critério de custo adotado (número de avaliações, tempo de parede, custo computacional total) e relatar o protocolo de validação. Franceschi *et al.* (2024) salientam que conclusões podem se inverter quando se altera o orçamento, a dimensionalidade e o regime de ruído; de modo análogo, Won; Lee; Lee (2025) destacam que ganhos de multi-fidelidade dependem do quão bem a fidelidade reduzida preserva a ordenação entre configurações. Em síntese, no XGBoost, a otimização de hiperparâmetros deve ser entendida como um problema de otimização de caixa-preta em espaço misto, potencialmente multiobjetivo e ruidoso, cuja solução depende tanto do algoritmo de busca quanto de decisões experimentais (métricas, validação, replicação e orçamento). Essa caracterização fundamenta a adoção de estratégias sistemáticas de experimentação e tomada de decisão no capítulo metodológico, particularmente quando se deseja conciliar desempenho e custo computacional.

2.10.11 Tendências contemporâneas e avanços metodológicos em otimização de hiperparâmetros

Nos últimos anos, a literatura em HPO tem evoluído além das abordagens clássicas baseadas em busca exaustiva, amostragem aleatória, otimização bayesiana tradicional e metaheurísticas populacionais. Observa-se um movimento consistente em direção a estratégias híbridas, adaptativas e orientadas por dados (*data-driven*), cujo objetivo central é reduzir o custo experimental associado à exploração do espaço de busca, mantendo ou ampliando a qualidade das soluções encontradas.

Um dos eixos recentes dessa evolução envolve o uso de Modelos de Linguagem de Grande Porte (*Large Language Models*, LLMs) como componentes auxiliares no processo de configuração de algoritmos. Em particular, investiga-se a capacidade desses modelos de sugerir regiões promissoras do espaço de busca ou de ajustar dinamicamente parâmetros de metaheurísticas, com base em descrições estruturadas do problema e histórico de execuções. Martinek *et al.* (2026) exploram essa direção ao utilizar LLMs para guiar a sintonia de parâmetros de métodos metaheurísticos, reduzindo dependência de tentativa e erro manual e acelerando a convergência prática. Embora ainda emergente, essa linha sugere um

deslocamento do paradigma puramente exploratório para um paradigma assistido por conhecimento contextual e aprendizado prévio.

Em paralelo, revisões recentes consolidam o papel das estratégias multi-fidelity e mecanismos de early stopping como instrumentos centrais para eficiência computacional em HPO (Won; Lee; Lee, 2025). A premissa dessas abordagens é que avaliações de baixo custo, como treinos parciais, subconjuntos de dados ou estimativas intermediárias de desempenho, podem fornecer informação suficiente para descartar configurações subótimas precocemente. Essa lógica formaliza o *trade-off* entre profundidade de avaliação e cobertura do espaço de busca, permitindo melhor alocação do orçamento computacional. Tais estratégias são particularmente relevantes em cenários onde o treinamento completo do modelo representa o principal gargalo de custo.

No contexto específico de modelos baseados em árvores, estudos recentes têm apontado para a integração entre técnicas clássicas de DoE, RSM e estratégias de busca como alternativa competitiva frente a métodos puramente estocásticos (Vasquez-Ramos *et al.*, 2025). Essas abordagens enfatizam a interpretabilidade dos efeitos principais e interações entre hiperparâmetros, oferecendo não apenas soluções numéricas, mas também compreensão estrutural da paisagem de resposta. Em espaços de busca moderadamente dimensionados e sob restrição de orçamento, tais métodos têm se mostrado capazes de alcançar soluções de alto desempenho com menor variabilidade e maior transparência decisional.

Adicionalmente, problemas de ajuste de hiperparâmetros em ambientes altamente não estacionários, como aprendizado por reforço ou sistemas dinâmicos complexos; têm motivado investigações sobre robustez, estabilidade e importância relativa de hiperparâmetros (Weller; Moll, 2026). Esses estudos indicam que nem todos os hiperparâmetros exercem influência equivalente sobre o desempenho final, sugerindo oportunidades para estratégias de priorização e redução dimensional no espaço de busca. Tal constatação dialoga diretamente com abordagens baseadas em síntese multivariada e análise fatorial, que buscam estruturar e hierarquizar informações em contextos com múltiplas métricas correlacionadas.

De forma geral, as tendências contemporâneas em HPO apontam para três direções convergentes: (i) maior eficiência amostral e computacional por meio de estratégias adaptativas; (ii) incorporação de conhecimento prévio ou assistido na exploração do espaço de busca; e (iii) valorização da interpretabilidade e da robustez das soluções encontradas. Nesse cenário, abordagens que integram planejamento experimental estruturado, modelagem aproximadora e formulação explícita multiobjetivo mantêm relevância metodológica,

sobretudo quando o objetivo não se restringe à obtenção de um único ponto ótimo, mas à caracterização sistemática de compromissos entre desempenho e custo.

3. METODOLOGIA

3.1 Visão geral do *pipeline* e hipóteses operacionais

Este capítulo descreve a metodologia proposta para seleção de hiperparâmetros do XGBoost em problemas de classificação binária sob uma perspectiva explicitamente multiobjetivo, em que se busca, de modo simultâneo, maximizar desempenho preditivo e reduzir custo computacional. O custo é tratado como tempo de execução medido em ambiente controlado e comparável entre métodos.

O *pipeline* possui cinco blocos principais: (i) planejamento de experimentos no espaço de hiperparâmetros; (ii) avaliação reprodutível das configurações por validação cruzada estratificada; (iii) síntese das respostas em escores operacionais de qualidade e custo; (iv) modelagem por RSM para aproximação das funções objetivo; e (v) geração orientada de soluções de compromisso via NBI, com confirmação no modelo real.

A Figura 4 sintetiza as etapas e os artefatos produzidos a cada fase. A Figura 5 apresenta um pseudocódigo equivalente, utilizado como referência operacional para a implementação e para garantir rastreabilidade do protocolo.

3.2 Definição do problema e notação

Seja o conjunto de dados $\mathcal{D} = (\mathbf{X}, \mathbf{y})$, em que $\mathbf{X} \in \mathbb{R}^{n \times p}$ é a matriz de atributos e $\mathbf{y} \in \{0,1\}^n$ é o vetor de rótulos. Denota-se por $\mathbf{h} \in \mathcal{H} \subset \mathbb{R}^d$ o vetor de hiperparâmetros do XGBoost, com $d = 7$ fatores especificados na Seção 3.4.

A seleção de hiperparâmetros é formulada como um problema multiobjetivo com dois objetivos:

1. Maximizar uma medida de qualidade $\mathcal{Q}(\mathbf{h})$, construída a partir de métricas derivadas da matriz de confusão;
2. Minimizar um custo $\mathcal{C}(\mathbf{h})$, associado ao tempo de execução do treinamento/validação de um modelo com hiperparâmetros \mathbf{h} .

Na prática, o método estima e por RSM a partir de um conjunto experimental gerado por DoE (Seção 3.5), e emprega NBI para gerar soluções candidatas Pareto-ótimas (Seção 3.10), posteriormente confirmadas no modelo real.

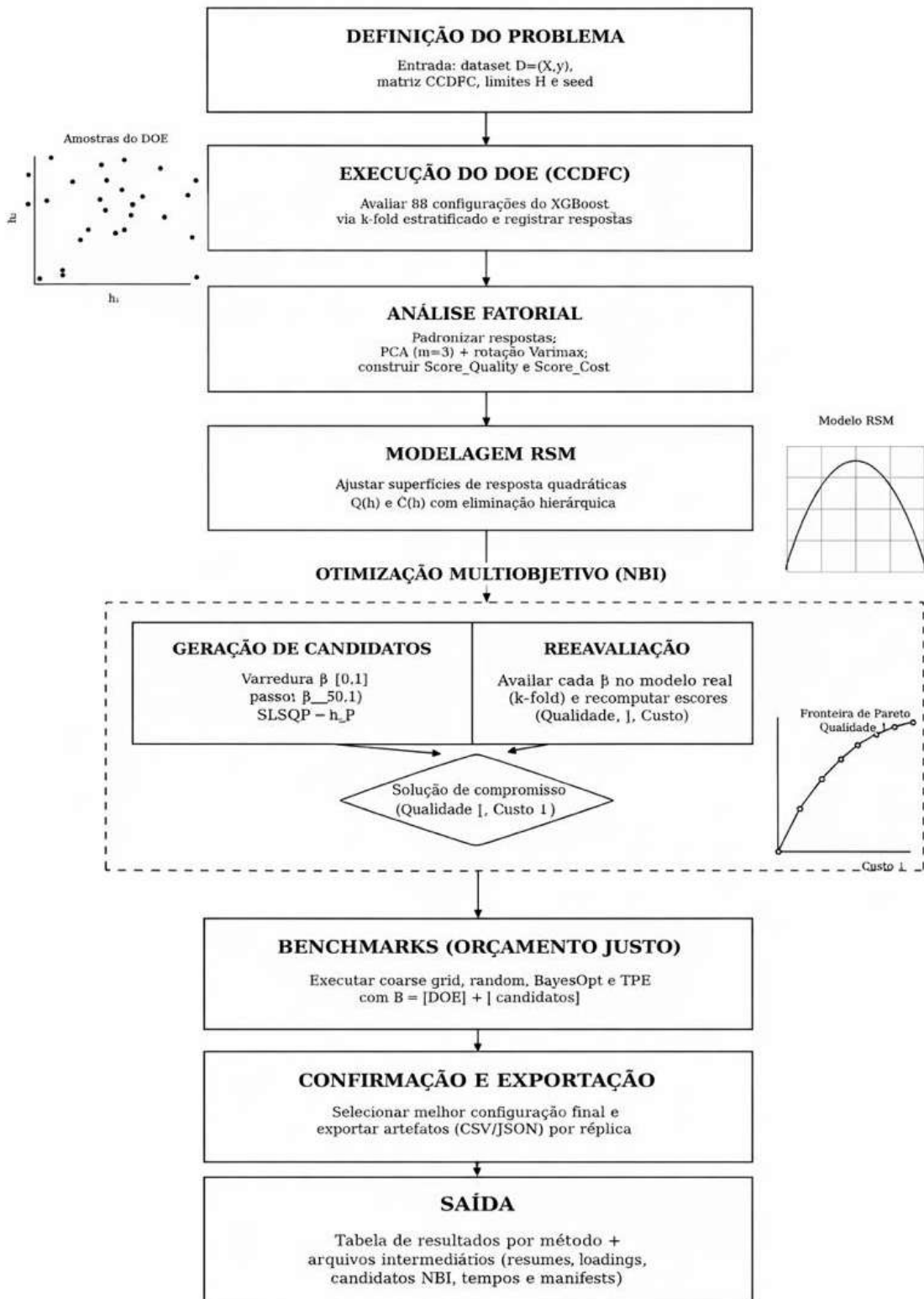


Figura 4 - Fluxograma do *pipeline* proposto (DoE + RSM + PCA/Varimax + NBI).

Begin

```
Inicializar seed para módulos pseudoaleatórios (numpy, random, xgboost);
Carregar dataset  $D = (X,y)$ ;
If  $y$  not in  $\{0,1\}$  do
  | Mapear classes do alvo  $y$  para  $\{0,1\}$ ;
End If "If  $y$  not in  $\{0,1\}$ ";
Ler matriz de delineamento CCDFC  $H$  ( $|DOE| = 88$ );
Definir  $k$ -fold  $\leftarrow 5$ ; Inicializar tabela  $R\_DOE \leftarrow NULL$ ;
For  $i \leftarrow 1$  up to  $|DOE|$  repeat
  | Selecionar configuração  $h \leftarrow H[i]$ ;
  | Avaliar  $XGBoost(h)$  via  $k$ -fold estratificado;
  | Armazenar respostas em  $R\_DOE$ ;
End For "For  $i \leftarrow 1$  up to  $|DOE|$ ";
Padronizar respostas do DOE e calcular PCA ( $m = 3$ );
Aplicar rotação Varimax e calcular escores fatoriais rotacionados;
Construir  $Score\_Quality$  (maximizar) e  $Score\_Cost$  (minimizar);
Ajustar modelos quadráticos de RSM  $Q(h)$  e  $\hat{C}(h)$  com eliminação hierárquica;
Calcular utopia  $u$  e nadir  $n$  a partir dos escores observados no DOE;
Inicializar conjunto de candidatos  $P \leftarrow NULL$ ;
For  $\beta \leftarrow 0$  to 1 step  $\beta\_step$  repeat
  | Resolver:  $\max \beta \cdot Q\_norm(h) + (1-\beta) \cdot C\_norm(h)$  sujeito aos limites  $H$ ;
  | If solução é viável do
  | | Armazenar  $h_\beta$  em  $P$ ;
  | Else
  | | Descartar  $h_\beta$ ;
  | End If "If solução é viável";
End For "For  $\beta \leftarrow 0$  to 1 step  $\beta\_step$ ";
Reavaliar cada  $h$  em  $P$  no modelo real ( $k$ -fold);
Selecionar solução de compromisso  $h^*$  (Qualidade  $\uparrow$ , Custo  $\downarrow$ );
If  $benchmark\_budget \leq 0$  do
  | Definir  $B \leftarrow |DOE| + |P|$ ; (auto-fairness)
Else
  | Definir  $B \leftarrow benchmark\_budget$ ;
End If "If  $benchmark\_budget \leq 0$ ";
Executar benchmarks sob orçamento  $B$ : coarse grid, random, BayesOpt, TPE;
Exportar artefatos (CSV/JSON) e resumo desta réplica;
End Algorithm.
```

Figura 5 - Pseudocódigo do pipeline proposto.

3.3 Pré-processamento, balanceamento e particionamento

O pré-processamento é aplicado antes da avaliação por validação cruzada, de modo idêntico em todos os métodos comparados. As etapas incluem: (i) codificação binária consistente dos rótulos; (ii) balanceamento por subamostragem (*undersampling*) da classe

majoritária; e (iii) validação cruzada estratificada em k dobras (k-fold), preservando aproximadamente as proporções de classe em cada partição.

Seja $\mathcal{F} = \{1, \dots, k\}$ o conjunto de dobras e $m(\mathbf{h}; j)$ uma métrica avaliada na dobra j . A estimativa agregada por validação cruzada é:

$$\bar{m}(\mathbf{h}) = \frac{1}{k} \sum_{j=1}^k m(\mathbf{h}; j) \quad (3.1)$$

Sendo análogo para o tempo, isto é, $\bar{t}(\mathbf{h}) = \frac{1}{k} \sum_{j=1}^k t(\mathbf{h}; j)$.

3.3.1 Caracterização resumida da base de dados canônica

A base principal utilizada neste estudo é o conjunto *MAGIC Gamma Telescope*, que reúne instâncias geradas por simulação de Monte Carlo para modelar o registro de partículas gama de alta energia em um telescópio *Cherenkov* atmosférico. O objetivo é classificar eventos entre gama (*signal*) e *hadrons* (*background*) a partir de um conjunto de descritores contínuos do formato da imagem do chuveiro atmosférico (Bock, 2004).

A Tabela 1 a seguir resume a caracterização e a estrutura da base de dados utilizada.

Tabela 1 - Caracterização descritiva do conjunto de dados *MAGIC Gamma Telescope* utilizado nos experimentos.

Item	Valor
Nº de instâncias	19.020
Nº de variáveis preditoras	10
Tipo das variáveis	Contínuas (reais)
Valores faltantes	Não
Classes	g (<i>gama</i>) / h (<i>hadron</i>)
Nº de instâncias	19.020
Distribuição de classes	g = 12.332; h = 6.688

3.3.2 Pré-processamento e descrição de atributos

Para fins técnicos, o número de eventos da classe h é subestimado no conjunto disponibilizado, embora em dados reais o *background* costume ser majoritário. Ainda assim, o conjunto apresenta desbalanceamento relevante ($\approx 65\%$ vs. 35%), o que justifica a etapa de balanceamento descrita na metodologia (Bock, 2004). A Tabela 2 apresenta a especificação das

variáveis preditoras do conjunto de dados, com resumo conceitual e respectivas unidades físicas.

Tabela 2 - Especificação das variáveis preditoras do conjunto de dados *MAGIC Gamma Telescope* utilizado nos experimentos, com resumo conceitual e respectivas unidades físicas.

Variável	Descrição (resumo)	Unidade
f_{Length}	eixo maior da elipse (parâmetro de <i>Hillas</i>)	mm
f_{Width}	eixo menor da elipse (parâmetro de <i>Hillas</i>)	mm
f_{Size}	\log_{10} da soma do conteúdo de todos os pixels	fotóelétrons (phe)
f_{Conc}	razão (soma dos 2 maiores pixels) / f_{Size}	adimensional
f_{Conc1}	razão (maior pixel) / f_{Size}	adimensional
f_{Asym}	distância do maior pixel ao centro, projetada no eixo maior	mm
f_{M3Long}	raiz cúbica do 3º momento ao longo do eixo maior	mm
$f_{M3Trans}$	raiz cúbica do 3º momento ao longo do eixo menor	mm
f_{Alpha}	ângulo do eixo maior com o vetor até a origem	graus (°)
f_{Dist}	distância da origem ao centro da elipse	mm

3.3.2.1 Normalização, valores discrepantes e balanceamento

Como as variáveis são contínuas e o modelo base (XGBoost) é baseado em árvores, não é estritamente necessária a normalização das *features* para fins de aprendizado do modelo. Assim, optou-se por manter as variáveis no formato original do conjunto, priorizando comparabilidade entre métodos de otimização sob o mesmo pipeline de pré-processamento. O desbalanceamento entre classes foi aplicado estritamente aos dados de treino em cada partição.

3.4 Espaço de busca de hiperparâmetros

Define-se o espaço admissível de busca de hiperparâmetros, denotado por \mathcal{H} , a partir do estabelecimento de limites inferiores e superiores para cada variável de decisão considerada no processo de calibração do modelo. Neste capítulo, os hiperparâmetros são identificados por suas denominações originais entre aspas duplas e em *snake case*, conforme especificadas na implementação computacional do algoritmo XGBoost. Tal escolha é adotada com o objetivo de garantir consistência terminológica entre a formulação matemática, o planejamento experimental, os resultados numéricos e as representações gráficas subsequentes, nas quais esses nomes aparecem de forma explícita. Os fatores investigados são definidos como segue:

1. h_1 : fração de instâncias por árvore (parâmetro “subsample”);
2. h_2 : fração de colunas por árvore (parâmetro “colsample_bytree”);
3. h_3 : fração de colunas por nível (parâmetro “colsample_bylevel”);
4. h_4 : taxa de aprendizado (parâmetro “learning_rate”);
5. h_5 : profundidade máxima (parâmetro “max_depth”);
6. h_6 : perda mínima para divisão (parâmetro “gamma”);
7. h_7 : número de estimadores (parâmetro “n_estimators”).

Os limites numéricos adotados para cada fator, bem como os pontos centrais utilizados no CCDFC, são apresentados a seguir. Para fatores discretos (profundidade máxima (“max_depth”) e número de estimadores (“n_estimators”)), valores centrais eventualmente não inteiros são operacionalizados por projeção e arredondamento para o inteiro admissível mais próximo, mantendo-se os limites estabelecidos.

$$\begin{array}{llll}
\text{subsample} & \in [0,05, 1,00], & z_0 = 0,525, \\
\text{colsample_bytree} & \in [0,05, 1,00], & z_0 = 0,525, \\
\text{colsample_bylevel} & \in [0,05, 1,00], & z_0 = 0,525, \\
\text{learning_rate} & \in [0,01, 0,30], & z_0 = 0,155, \\
\text{max_depth} & \in \{3, \dots, 18\}, & z_0 = 10,5, \\
\text{gamma} & \in [0,05, 5,00], & z_0 = 2,525, \\
\text{n_estimators} & \in \{50, \dots, 700\}, & z_0 = 375.
\end{array} \tag{3.2}$$

3.5 Planejamento experimental CCDFC

3.5.1 Construção do delineamento e codificação

O planejamento experimental baseia-se em um delineamento do tipo CCDFC com componente fatorial fracionário 2^{7-1} (64 ensaios), 14 pontos axiais e 10 pontos centrais, totalizando $N = 88$ configurações avaliadas. Denota-se por $\mathbf{x}_i \in [-1,1]^d$ o vetor de variáveis codificadas no ensaio i .

A transformação entre a variável real z_ℓ e o código x_ℓ é dada por:

$$x_\ell = \frac{z_\ell - z_{\ell,0}}{\Delta z_\ell}, \quad z_{\ell,0} = \frac{z_{\ell,-1} + z_{\ell,+1}}{2}, \quad \Delta z_\ell = \frac{z_{\ell,+1} - z_{\ell,-1}}{2} \tag{3.3}$$

Em que $z_{\ell,-1}$ e $z_{\ell,+1}$ representam, respectivamente, os níveis baixo e alto do fator ℓ , e $z_{\ell,0}$ é o ponto central. A inversão para obtenção do valor real a partir do código é:

$$z_\ell = z_{\ell,0} + x_\ell \Delta z_\ell \tag{3.4}$$

3.5.2 Definição da região experimental e restrições experimentais

Em problemas de calibração de hiperparâmetros, determinadas combinações extremas podem conduzir a comportamento numérico atípico, instabilidade no processo de treinamento ou custo computacional desproporcional. Durante experimentos preliminares exploratórios, observou-se que configurações em que os hiperparâmetros de amostragem (“subsample”, “colsample_bytree” e “colsample_bylevel”) assumiam valores próximos de zero resultavam em ausência de convergência ou em desempenho degenerado do modelo, decorrente da insuficiência de amostras e atributos disponíveis para construção das árvores.

Para mitigar tal comportamento e preservar a validade operacional das execuções experimentais, estabeleceu-se um limite inferior prático (*lower bound*) de 0,05 para esses hiperparâmetros, aplicado antes da execução do delineamento. Assim, embora o CCDFC opere com níveis codificados no intervalo $[-1, +1]$, a transformação para a escala real respeitou esse limite mínimo, evitando regiões inviáveis do espaço de busca.

Ressalta-se que o objetivo do planejamento experimental não é “garantir convergência” do algoritmo em sentido absoluto, mas concentrar a amostragem em uma região operacional realista e informativa, reduzindo a probabilidade de avaliações inválidas e aumentando a taxa de sucesso e interpretabilidade das respostas obtidas.

3.5.3 Execução do DoE e registro das respostas

Cada configuração \mathbf{h}_i definida pelo delineamento é avaliada com validação cruzada estratificada e os seguintes vetores de respostas são registrados:

1. $\mathbf{q}_i \in \mathbb{R}^4$: métricas de qualidade (Acurácia, Precisão, Revocação e Especificidade);
2. $c_i \in \mathbb{R}_+$: tempo médio por *fold* (dobra), em segundos, como *proxy* de custo.

A matriz de respostas $\mathbf{R} \in \mathbb{R}^{N \times 5}$ é formada empilhando-se as respostas em cada ensaio. O delineamento e a ordem de execução são registrados para rastreabilidade e auditoria experimental.

3.6 Métricas de avaliação e medida de custo

3.6.1 Matriz de confusão e métricas de qualidade

Sejam \hat{y}_i a predição do classificador e y_i o rótulo verdadeiro associado à i -ésima observação. Os termos TP , TN , FP e FN , previamente definidos na Seção 2.5 por meio das Equações (2.54) a (2.57), compõem a matriz de confusão e são utilizados como base para a definição das métricas de qualidade. A acurácia do classificador é definida conforme a Equação (3.6), enquanto as métricas de precisão, revocação e especificidade são dadas pela Equação (3.7):

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.6)$$

$$\text{Prec} = \frac{TP}{TP + FP} \quad (3.7.1)$$

$$\text{Rec} = \frac{TP}{TP + FN} \quad (3.7.2)$$

$$\text{Esp} = \frac{TN}{TN + FP} \quad (3.7.3)$$

Para fins de síntese e visualização comparativa do desempenho, adota-se adicionalmente um índice escalar de qualidade, definido como a média aritmética simples das métricas derivadas da matriz de confusão, conforme formalizado:

$$J_Q(\mathbf{h}) = \frac{\text{Acc}(\mathbf{h}) + \text{Prec}(\mathbf{h}) + \text{Rec}(\mathbf{h}) + \text{Esp}(\mathbf{h})}{4} \quad (3.8)$$

3.6.2 Tempo de execução como *proxy* de custo computacional

O custo computacional de uma configuração \mathbf{h} é quantificado por tempo de execução medido em segundos. Para cada dobra j , registra-se o tempo $t(\mathbf{h}; j)$ consumido pela etapa de treinamento e avaliação. O custo utilizado no delineamento é o tempo médio por dobra (*fold*) $\bar{t}(\mathbf{h})$.

Em análises agregadas, também se registra o tempo total do processo de otimização como a soma dos tempos associados a todas as avaliações executadas por um método, permitindo comparação sob o mesmo orçamento de avaliações.

3.7 Etapa mono-objetivo: modelagem das respostas individuais

Nesta etapa, cada resposta de desempenho do Estudo I (Acurácia, Precisão, Revocação, Especificidade e Tempo de execução) é tratada individualmente, com o objetivo de (i) quantificar os efeitos principais e de interação dos hiperparâmetros, (ii) obter modelos preditivos via RSM e (iii) verificar premissas de modelagem por meio de análise residual. Os resultados desta etapa subsidiam a interpretação do comportamento do modelo e a definição de variáveis sintetizadas utilizadas na etapa multiobjetivo.

3.7.1 Procedimento iterativo de ajuste ponderado e diagnóstico dos resíduos

Para garantir reprodutibilidade e padronização entre as respostas do Estudo I, adotou-se o procedimento iterativo a seguir (baseado em ajuste por Superfície de Resposta e tratamento de heterocedasticidade por ponderação):

1. Ajuste inicial de modelos de RSM por seleção *Stepwise (Forward Selection)* para cada variável-resposta de interesse (Acurácia, Precisão, Revocação, Especificidade e Tempo de Execução);
2. Armazenamento dos resíduos do ajuste, $e_i = y_i - \hat{y}_i$, para cada execução experimental i ;
3. Cálculo do valor absoluto dos resíduos $|e_i|$, como medida de magnitude do erro local;
4. Ajuste de um modelo RSM tomando $|e_i|$ como resposta, de modo a modelar a variabilidade em função dos fatores;
5. Estimativa da variância local pelo princípio do POE (*Propagation of Error*), obtendo-se uma aproximação $\sigma^2(x_i)$ para cada ponto experimental;
6. Construção da matriz de pesos $W = \text{diag}(1/\sigma^2(x_i))$ e reajuste do modelo RSM original via mínimos quadrados ponderados;
7. Reavaliação da qualidade do ajuste (R^2 adj. e R^2 pred.) e inspeção diagnóstica dos resíduos para validar as premissas do modelo;
8. Caso os critérios de ajuste não sejam atendidos (por exemplo, R^2 adj. e R^2 pred. > 90%), repete-se os passos 2. Até 7. adotando-se, quando necessário, uma ponderação empírica alternativa $W_i = 1/e_i^2$, até estabilização dos indicadores.

Concluída a modelagem e a validação dos ajustes mono-objetivo, procede-se à síntese das múltiplas respostas em VRFs, de modo a reduzir colinearidade entre métricas e preparar o espaço de objetivos empregado na otimização multiobjetivo. As etapas seguintes descrevem a

construção dos VRFs (Seção 3.8), o ajuste de metamodelos para cada VRF (Seção 3.9) e, por fim, a otimização multiobjetivo via NBI (Seção 3.10).

3.8 Síntese das respostas e construção de escores

3.8.1 Padronização das respostas

As respostas são padronizadas por variável para remover efeito de escala. Seja r_{ij} o valor da resposta j no ensaio i . A padronização é:

$$z_{ij} = \frac{r_{ij} - \mu_j}{\sigma_j}, \quad \mu_j = \frac{1}{N} \sum_{i=1}^N r_{ij}, \quad \sigma_j = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (r_{ij} - \mu_j)^2} \quad (3.9)$$

Denota-se por $\mathbf{Z} \in \mathbb{R}^{N \times 5}$ a matriz padronizada.

3.8.2 PCA, rotação *Varimax* e escores operacionais

Aplica-se PCA sobre \mathbf{Z} , obtendo-se componentes principais e escores associados. Em notação matricial, a matriz de correlações é:

$$\mathbf{S} = \frac{1}{N-1} \mathbf{Z}^T \mathbf{Z} \quad (3.10)$$

E sua decomposição espectral $\mathbf{S} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T$ fornece autovetores \mathbf{V} e autovalores $\mathbf{\Lambda}$. Os escores não rotacionados são $\mathbf{T} = \mathbf{ZV}$.

Para maximizar interpretabilidade, aplica-se rotação ortogonal *Varimax*. Seja \mathbf{L} a matriz de cargas fatoriais retida e \mathbf{T} uma matriz ortogonal; então $\tilde{\mathbf{L}} = \mathbf{LT}$. A rotação *Varimax* busca maximizar o critério definido em:

$$V(\mathbf{T}) = \sum_{k=1}^m \left[\frac{1}{p} \sum_{j=1}^p \tilde{\ell}_{jk}^4 - \left(\frac{1}{p} \sum_{j=1}^p \tilde{\ell}_{jk}^2 \right)^2 \right] \quad (3.11)$$

Em que p é o número de variáveis observadas, m é o número de componentes retidos e $\tilde{\ell}_{jk}$ é o elemento (j, k) de $\tilde{\mathbf{L}}$. Os escores rotacionados são obtidos por $\tilde{\mathbf{T}} = \mathbf{ZVT}$.

Ressalta-se que, após a obtenção dos fatores latentes por meio da Análise de Componentes Principais e da rotação *Varimax*, admite-se a construção de escores sintéticos a partir da combinação dos fatores extraídos, com o objetivo de apoiar análises subsequentes e facilitar a interpretação dos resultados. Em particular, esses escores podem ser associados, de

forma conceitual, a dimensões relacionadas à qualidade do modelo e ao custo computacional. Contudo, a definição concreta dessa agregação não é estabelecida a priori, sendo dependente da análise empírica das cargas fatoriais e dos escores observados, a qual é apresentada no Capítulo 4.

3.9 Modelagem por Superfícies de Resposta

Para aproximar o comportamento dos escores em função dos fatores do delineamento, ajustam-se superfícies quadráticas. Seja $\mathbf{x} \in [-1,1]^d$ o vetor de fatores codificados e $y(\mathbf{x})$ a resposta (por exemplo, Score_Q ou Score_C). A forma geral do modelo quadrático com a substituição das variáveis pelos escores obtidos fica:

$$\hat{y}(\mathbf{x}) = \beta_0 + \sum_{i=1}^d \beta_i x_i + \sum_{i=1}^d \beta_{ii} x_i^2 + \sum_{i=1}^{d-1} \sum_{j=i+1}^d \beta_{ij} x_i x_j \quad (3.12)$$

Com termo de erro $\varepsilon(\mathbf{x})$ assumido aproximadamente homocedástico no domínio experimental.

Em notação matricial, para N observações, tem-se $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, em que \mathbf{X} é a matriz de regressão (incluindo termos lineares, quadráticos e interações). O estimador de mínimos quadrados ordinários é o mesmo tal qual definido na Seção (2.6.3) na Eq. (2.71).

Embora o ajuste de superfícies de resposta seja naturalmente contínuo, parte do espaço de busca de hiperparâmetros em modelos como XGBoost é discreto (por exemplo, “max_depth” e “n_estimators”). Assim, a superfície estimada é uma aproximação contínua de um problema misto. Na etapa de otimização, valores não inteiros sugeridos pelo modelo contínuo são operacionalizados por projeção e arredondamento para o conjunto admissível. Essa discretização pode induzir pequenas descontinuidades na função objetivo real; por isso, todas as soluções candidatas são confirmadas por avaliação direta do modelo (no domínio discreto), mantendo a coerência entre a recomendação da RSM e a performance observada.

Para reduzir sobreajuste (*overfitting*) e preservar hierarquia do modelo, aplica-se eliminação hierárquica (*backward elimination*), removendo termos não significativos enquanto se mantém o princípio de hierarquia: interações e termos quadráticos são preservados apenas quando os termos lineares correspondentes permanecem no modelo.

3.10 Otimização Multiobjetivo por NBI

3.10.1 Normalização e Pontos de referência (Utopia x Nadir)

A NBI é aplicada sobre objetivos normalizados para evitar dominância por escala. Definem-se os pontos de utopia e nadir a partir das respostas observadas no DoE: $Q_{\text{utopia}} = \max_i \text{Score}_{Q,i}$, $Q_{\text{nadir}} = \min_i \text{Score}_{Q,i}$ e, para custo, $C_{\text{utopia}} = \min_i \bar{t}_i$, $C_{\text{nadir}} = \max_i \bar{t}_i$.

A normalização é definida por:

$$Q_{\text{norm}}(\mathbf{h}) = \frac{\hat{Q}(\mathbf{h}) - Q_{\text{nadir}}}{Q_{\text{utopia}} - Q_{\text{nadir}}} \quad (3.13)$$

$$C_{\text{norm}}(\mathbf{h}) = \frac{C_{\text{nadir}} - \hat{C}(\mathbf{h})}{C_{\text{nadir}} - C_{\text{utopia}}} \quad (3.14)$$

Por construção, Q_{norm} e C_{norm} variam aproximadamente em $[0,1]$ e a convenção “maior é melhor” é preservada em ambos os objetivos.

3.10.2 Geração de candidatos na fronteira de Pareto

Para $\beta \in [0,1]$, a NBI é operacionalizada por varredura discreta com passo $\Delta\beta = 0,02$, resolvendo-se, para cada β , o subproblema escalar:

$$\max_{\mathbf{h} \in \mathcal{H}} \beta Q_{\text{norm}}(\mathbf{h}) + (1 - \beta) C_{\text{norm}}(\mathbf{h}) \quad (3.15)$$

O subproblema é resolvido por Programação Quadrática Sequencial (SLSQP), com projeção das soluções no domínio admissível \mathcal{H} e múltiplas inicializações (*multi-start*), mitigando mínimos locais induzidos pela aproximação via RSM.

O conjunto de soluções viáveis obtidas é denotado por $\mathcal{P} = \{\mathbf{h}_\beta\}$ e representa uma aproximação discreta da fronteira de Pareto.

3.10.3 Confirmação no modelo real e seleção de compromisso

Como os candidatos \mathbf{h}_β são gerados a partir de modelos aproximados, realiza-se confirmação por reavaliação direta do XGBoost (*k-fold*) para cada candidato, recomputando-se $J_Q(\mathbf{h})$ e o tempo médio por dobra.

A seleção de uma solução de compromisso é realizada minimizando a distância Euclidiana ao ponto ideal (1,1) no espaço normalizado:

$$\mathbf{h}^* = \operatorname{argmin}_{\mathbf{h} \in \mathcal{P}} \left\| \begin{bmatrix} 1 - Q_{\text{norm}}(\mathbf{h}) \\ 1 - C_{\text{norm}}(\mathbf{h}) \end{bmatrix} \right\|_2 \quad (3.16)$$

3.11 Benchmarks e critério de justiça por orçamento

Para comparação, quatro estratégias de otimização de hiperparâmetros são consideradas: busca em grade, busca aleatória, otimização bayesiana e TPE (*Hyperopt*). Para evitar comparações enviesadas por diferentes custos de avaliação, define-se um orçamento comum B (número de avaliações do modelo) e cada método realiza exatamente B avaliações.

A baseline XGBoost default é tratada separadamente por não envolver busca; nesse caso reporta-se apenas uma avaliação ($B = 1$) para referência.

No método proposto, o orçamento é dado por:

$$B = N + N_\beta \quad (3.17)$$

Em que N é o número de ensaios do DoE e N_β é o número de candidatos confirmados na etapa multiobjetivo (um por valor de β). Como cada avaliação utiliza validação cruzada com k dobras, o número total de ajustes de modelos (treinos) por método é $B \cdot k$.

3.12 Replicações, controle de aleatoriedade e reprodutibilidade experimental

A aleatoriedade afeta (i) o particionamento em dobras (*folds*); (ii) operações internas de amostragem do XGBoost; e (iii) algoritmos de busca estocásticos (por exemplo, busca aleatória e TPE). Para estimar estabilidade e reduzir risco de conclusões dependentes de uma única realização, o protocolo é replicado em R execuções independentes, indexadas por $r \in \{1, \dots, R\}$, cada uma com semente pseudoaleatória distinta.

As sementes são geradas a priori e registradas em manifesto experimental, juntamente com a parametrização do experimento (delineamento, *k-fold*, passo $\Delta\beta$, orçamento e parâmetros de cada *benchmark*). A execução, o registro de artefatos (CSV/JSON) e a reprodutibilidade do *pipeline* são suportados por repositório público, estruturado para permitir reprises (*reruns*) determinísticos a partir do manifesto e das sementes.

3.13 Ambiente Computacional

Com o objetivo de assegurar reprodutibilidade experimental e adequada interpretação dos resultados relativos ao custo computacional, todos os experimentos foram conduzidos em ambiente controlado e fixo, conforme especificações detalhadas a seguir.

3.13.1 Hardware

Os experimentos foram executados em um computador Apple MacBook Pro™ (*Model Identifier*: Mac16,8), equipado com processador Apple Silicon M4 Pro™, arquitetura ARM64, contendo 12 núcleos de CPU (8 de alto desempenho e 4 de eficiência), 24 GB de memória RAM unificada e armazenamento em unidade de estado sólido (SSD NVMe).

Ressalta-se que não foi utilizada aceleração por GPU ou processamento via CUDA, sendo todas as execuções realizadas exclusivamente em CPU.

3.13.2 Sistema Operacional

Sistema operacional macOS™ versão 26.3 (Build 25D125), arquitetura ARM64.

3.13.3 Ambiente de Software

O pipeline experimental foi implementado em Python 3.9.6, executado em ambiente virtual isolado, com gerenciamento de dependências realizado por meio do gerenciador oficial de pacotes da linguagem Python “*Pip Installs Packages*” (pip 26.0.1).

As versões efetivamente utilizadas das principais bibliotecas foram:

- NumPy 2.0.2
- Pandas 2.3.3
- SciPy 1.13.1
- Scikit-learn 1.6.1
- XGBoost 2.1.4
- Statsmodels 0.14.6
- factor_analyzer 0.5.1
- scikit-optimize 0.10.2

- Hyperopt 0.2.7
- tqdm 4.66.2
- python-dotenv 1.2.1
- OpenPyXL 3.1.5

A listagem completa das dependências encontra-se registrada por meio do comando “pip freeze” e disponibilizada no repositório experimental, permitindo reconstrução integral do ambiente.

3.13.4 Controle de Execução

Para reduzir variabilidade associada à paralelização automática e garantir comparabilidade justa entre métodos, manteve-se constante o número de *threads* utilizadas nas execuções (por exemplo, “n_jobs = 1”, quando aplicável).

Adicionalmente, as sementes pseudoaleatórias foram fixadas e registradas em cada réplica experimental como mencionado na subseção anterior, assegurando consistência estatística e reprodutibilidade dos resultados.

3.14 Ameaças à validade

As principais ameaças à validade do estudo incluem:

1. Validade externa: o experimento principal é conduzido em uma base, em um tipo de problema e em um modelo específico (XGBoost para classificação binária). Como extensão de validade externa, o protocolo foi replicado em duas bases adicionais (Seção 4.9). Outrossim, a generalização para outras bases, algoritmos, problemas em aprendizado supervisionado e regimes de dimensionalidade deve ser investigada em estudos futuros;
2. Validade de construto: o tempo de execução é tratado como *proxy* de custo computacional. Embora adequado para comparação relativa em ambiente controlado, variações de hardware e carga do sistema podem afetar medidas absolutas;
3. Validade interna: a aproximação por RSM pode introduzir erro de modelo, sobretudo em regiões com comportamento não linear pronunciado. A etapa de confirmação por reavaliação direta dos candidatos foi incluída para reduzir este risco.

4. Risco de sobreajuste ao conjunto de validação: quando a avaliação do desempenho é repetida muitas vezes durante a busca (*overfitting by optimization*). Esse fenômeno ocorre quando a estratégia de busca explora flutuações aleatórias dos dados/splits e passa a selecionar configurações com desempenho aparentemente superior apenas por efeito de múltiplas tentativas (Cawley; Talbot, 2010). Em termos práticos, isso reforça a necessidade de separar adequadamente dados de seleção e dados de teste, bem como controlar orçamento experimental e aleatoriedade (Franceschi *et al.*, 2024).

4. RESULTADOS E DISCUSSÕES

Este capítulo apresenta e discute os resultados obtidos na avaliação comparativa do método proposto (DoE+RSM+PCA/*Varimax*+NBI) frente a quatro estratégias de ajuste de hiperparâmetros utilizadas como referência: busca em grade, busca aleatória, otimização bayesiana e TPE (Hyperopt). Nas tabelas e figuras, para fins de legibilidade, o método proposto é identificado como "DoE + NBI (proposto)", em alusão às etapas de otimização multiobjetivo e seleção de solução de compromisso descritas no Capítulo 3. As análises foram conduzidas a partir de 30 réplicas independentes, variando-se apenas a semente pseudoaleatória, de modo a capturar a variabilidade inerente ao particionamento e à validação cruzada. Os detalhes do *pipeline* de experimentação, do delineamento CCDFC e dos procedimentos de execução são descritos no Capítulo 3.

4.1 Visão geral da avaliação

4.1.1 Condições de comparação e orçamento computacional

Para assegurar comparabilidade entre os métodos, adotou-se um orçamento computacional uniforme, definido pelo número de avaliações completas do modelo (*k-fold* estratificado). Em todas as réplicas e para todos os métodos, o orçamento foi fixado em $B = 138$ avaliações. No método proposto, esse orçamento corresponde à soma das 88 execuções do delineamento CCDFC com as 50 reavaliações dos candidatos gerados na etapa de otimização multiobjetivo. Nos métodos de referência, o mesmo orçamento foi alocado integralmente ao procedimento de busca e seleção de hiperparâmetros, com posterior confirmação da melhor configuração sob o mesmo protocolo de validação. Adicionalmente, avaliou-se uma baseline de referência com o XGBoost em parâmetros *default* (sem *tuning*), executada sob o mesmo protocolo de validação cruzada e as mesmas sementes das réplicas; por não envolver processo de busca, essa baseline foi avaliada com uma única execução ($B = 1$).

4.1.2 Métricas avaliadas

A comparação entre os métodos foi realizada a partir de métricas de qualidade preditiva e de custo computacional. No conjunto de métricas de qualidade, consideraram-se acurácia,

precisão, recall e especificidade, calculadas a partir das matrizes de confusão obtidas em cada uma das dobras da validação cruzada e posteriormente agregadas por média. Além disso, definiu-se um índice sintético de qualidade como a média aritmética dessas quatro métricas, com o objetivo de facilitar a análise conjunta de desempenho.

Neste trabalho, diferencia-se (i) tempo total de otimização, entendido como o tempo dedicado à etapa de busca e seleção de configurações até consumir o orçamento experimental, e (ii) tempo total de execução, entendido como o tempo completo do procedimento experimental (incluindo execução do protocolo de avaliação para todas as configurações testadas dentro do orçamento, além de eventuais etapas finais de confirmação/seleção). Adicionalmente, reporta-se o tempo médio por dobra (*fold*) como uma *proxy* do custo marginal de avaliação de uma configuração, permitindo comparar métodos independentemente do número total de avaliações realizadas. Essas métricas permitem analisar, de forma integrada, o compromisso entre qualidade e custo computacional.

Este capítulo é desenvolvido em duas etapas complementares. Primeiramente, conduz-se uma análise mono-objetivo baseada CCDFC, com modelagens individuais das respostas e interpretação estatística de efeitos principais, interações e termos quadráticos. Na sequência, tais evidências motivam e subsidiam a etapa multiobjetivo, na qual métricas de qualidade e custo computacional são tratadas de forma simultânea, culminando na comparação do método proposto com abordagens de referência.

4.2 Estudo mono-objetivo com CCDFC: modelagens individuais das respostas

Esta seção apresenta uma análise mono-objetivo dos resultados obtidos no CCDFC, com ênfase na interpretação estatística dos efeitos dos hiperparâmetros do XGBoost sobre as respostas avaliadas. O objetivo é (i) caracterizar a sensibilidade de cada métrica em função de efeitos principais, interações e termos quadráticos; (ii) identificar regiões promissoras do espaço de busca em termos de desempenho; e (iii) evidenciar, de forma interpretável, a natureza de compromisso entre qualidade preditiva e custo computacional, o que motiva a formulação multiobjetivo tratada nas seções subsequentes.

4.2.1 Matriz experimental e codificação dos fatores

O CCDFC foi conduzido no espaço de sete hiperparâmetros do XGBoost, definidos como fatores quantitativos. Para fins de modelagem por Superfície de Resposta, os níveis reais de cada fator foram codificados em uma escala adimensional, usualmente no intervalo $[-1, 1]$, de modo a facilitar a estimação e a comparação entre efeitos. Denotando por h_i o valor real do i -ésimo hiperparâmetro e por x_i seu respectivo nível codificado, utiliza-se a transformação linear:

$$x_i = \frac{h_i - h_{i,0}}{\Delta_i}, \quad i = 1, \dots, k \quad (4.1)$$

Em que $h_{i,0}$ representa o nível central e Δ_i a semi-amplitude do intervalo experimental do fator i . No caso específico do CCDFC, a distância axial é $\alpha = 1$, de modo que os pontos axiais coincidem com as faces do hipercubo codificado.

Nos gráficos de efeitos principais (gerados no *software* Minitab™), os fatores quantitativos são identificados como x_1, x_2, \dots, x_7 e correspondem aos fatores A, B, ..., G da matriz CCDFC, conforme: $x_1=A$ (fração de subamostragem de instâncias, parâmetro “subsample”); $x_2=B$ (fração de colunas por árvore, parâmetro “colsample_bytree”); $x_3=C$ (fração de colunas por nível, parâmetro “colsample_bylevel”); $x_4=D$ (taxa de aprendizado, parâmetro “learning_rate”); $x_5=E$ (profundidade máxima, parâmetro “max_depth”); $x_6=F$ (perda mínima para divisão, parâmetro “gamma”); e $x_7=G$ (número de estimadores, parâmetro “n_estimators”).

A Figura 6 apresenta um extrato da matriz experimental utilizada, contendo os fatores associados aos hiperparâmetros do XGBoost: fração de subamostragem de instâncias (parâmetro “subsample”), fração de colunas por árvore (parâmetro “colsample_bytree”), fração de colunas por nível (parâmetro “colsample_bylevel”), taxa de aprendizado (parâmetro “learning_rate”), profundidade máxima (parâmetro “max_depth”), perda mínima para divisão (parâmetro “gamma”) e número de estimadores (parâmetro “n_estimators”) e as respostas originais: acurácia (resposta “Accuracy_Mean”), precisão (resposta “Accuracy_Mean”), revocação (resposta “Accuracy_Mean”), especificidade (resposta “Accuracy_Mean”), e Tempo de Execução (resposta “Tempo_execucao”), calculadas sobre a média por *fold* na validação cruzada estratificada.

subsample	colsample_bytree	colsample_bylevel	learning_rate	max_depth	gamma	n_estimators	Accuracy_Mean	Precision_Mean	Recall_Mean	Specificity_Mean	Tempo_Execucao
0.525	0.525	1.000	0.155	10.500	2.525	375.000	0.938	0.950	0.925	0.951	1.289
1.000	1.000	0.050	0.300	3.000	0.050	50.000	0.943	0.965	0.920	0.966	0.331
0.050	1.000	0.050	0.010	18.000	0.050	700.000	0.938	0.956	0.918	0.957	4.034
0.525	0.525	0.525	0.155	10.500	5.000	375.000	0.937	0.953	0.920	0.954	0.869
0.050	1.000	0.050	0.010	3.000	0.050	50.000	0.914	0.939	0.885	0.942	0.268
1.000	1.000	0.050	0.010	18.000	0.050	50.000	0.937	0.942	0.932	0.942	0.835
0.525	0.525	0.525	0.155	10.500	2.525	50.000	0.941	0.956	0.926	0.957	0.386
0.525	0.525	0.525	0.155	10.500	2.525	375.000	0.942	0.951	0.933	0.951	1.296
1.000	1.000	0.050	0.300	18.000	0.050	700.000	0.940	0.953	0.927	0.953	2.184
0.525	0.525	0.525	0.155	3.000	2.525	375.000	0.941	0.954	0.928	0.955	0.907
0.050	1.000	0.050	0.010	18.000	5.000	50.000	0.908	0.943	0.870	0.947	0.210
0.050	0.050	0.050	0.010	18.000	0.050	50.000	0.905	0.949	0.855	0.954	0.365
0.050	1.000	1.000	0.010	3.000	0.050	700.000	0.927	0.947	0.904	0.949	3.072
0.525	0.050	0.525	0.155	10.500	2.525	375.000	0.939	0.960	0.916	0.962	1.044
0.050	1.000	0.050	0.300	3.000	5.000	50.000	0.915	0.947	0.879	0.951	0.183
1.000	0.050	0.050	0.010	3.000	0.050	50.000	0.920	0.967	0.869	0.970	0.264
1.000	0.050	1.000	0.300	18.000	0.050	700.000	0.920	0.937	0.901	0.939	2.586
0.525	0.525	0.525	0.155	10.500	2.525	375.000	0.942	0.951	0.933	0.951	0.948
1.000	1.000	1.000	0.300	3.000	5.000	50.000	0.934	0.957	0.908	0.959	0.231

Figura 6 - Extrato da matriz experimental do CCDFC, com fatores (hiperparâmetros) e respostas originais.

4.2.2 Modelo de Superfície de Resposta e significância dos efeitos

Para cada resposta y observada no CCDFC, ajustou-se um modelo quadrático em variáveis codificadas, com termos de primeira ordem, interações de segunda ordem e termos quadráticos, conforme o termo geral definido na Seção (2.6.3) na Equação (2.69), em que $\mathbf{x} = (x_1, \dots, x_k)^\top$ e β denota o vetor de parâmetros a serem estimados por mínimos quadrados. A significância estatística dos termos do modelo foi investigada por meio de ANOVA e de gráficos de Pareto de efeitos padronizados. No gráfico de Pareto, cada barra corresponde ao valor absoluto do efeito padronizado (associado ao teste t) de um termo do modelo; termos que ultrapassam o limiar de significância indicado no gráfico são interpretados como estatisticamente relevantes ao nível de confiança adotado.

Nos itens a seguir, apresentam-se os resultados para cada variável-resposta de interesse com ênfase na interpretação dos termos significativos, nos efeitos de interação e nos níveis recomendados pelos modelos ajustados.

O procedimento iterativo de ponderação e diagnóstico dos resíduos, utilizado para tratar heterocedasticidade e fortalecer a robustez dos ajustes, é descrito na Seção 3.7.1.

4.2.3 Acurácia

A Figura 7 sintetiza os efeitos padronizados associados à Acurácia e ilustra, por meio de um gráfico de contorno, a variação predita da resposta em função da fração de subamostragem de instâncias (x_1) e do número de estimadores (x_7), mantendo-se os demais

fatores em seus níveis centrais. Observa-se que fração de subamostragem de instâncias (x_1) e número de estimadores (x_7) se destacam entre os fatores com maior influência na Acurácia, o que é coerente com a dinâmica de redução de viés e variância em métodos de *boosting*. No contorno, nota-se uma tendência de aumento da Acurácia à medida que se eleva o número de árvores e a fração de instâncias amostradas por árvore, com sinais de saturação em regiões de valores elevados, característica típica de retornos marginais decrescentes. A presença de interações envolvendo taxa de aprendizado (x_4) e número de estimadores (x_7) sugere que o ganho de Acurácia não depende apenas do aumento isolado de árvores, mas do equilíbrio entre a taxa de aprendizado e o orçamento de iterações. Do ponto de vista interpretativo, tais evidências reforçam que o desempenho é governado por combinações de fatores, justificando o uso de termos de interação e de modelos quadráticos.

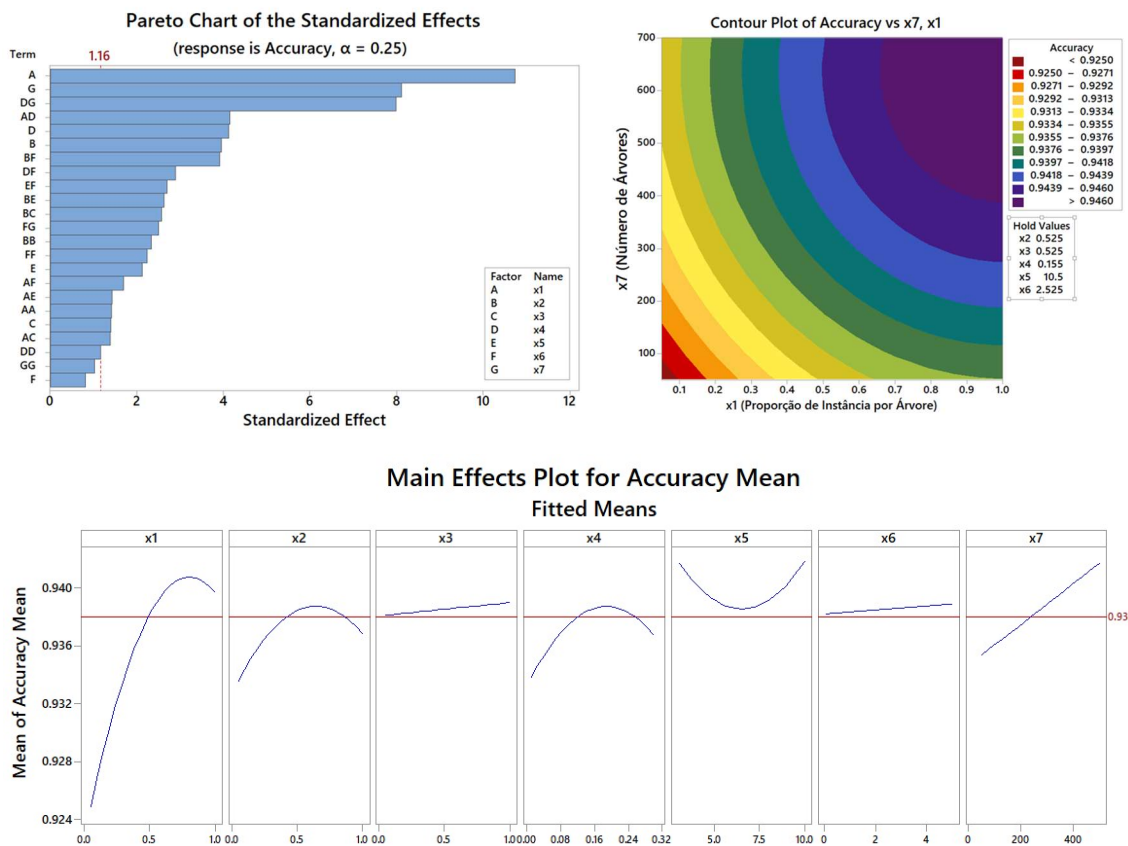


Figura 7 - Gráfico de Pareto de efeitos padronizados, contorno e efeitos principais para Acurácia.

4.2.4 Precisão

A Figura 8 apresenta os efeitos padronizados e o contorno para a resposta Precisão, com destaque para a interação entre a fração de subamostragem de instâncias (x_1) e a taxa de aprendizado (x_4).

O gráfico de Pareto evidencia que a interação x_1x_4 e o efeito principal de taxa de aprendizado (x_4) figuram entre os termos mais influentes, seguidos por fração de subamostragem de instâncias (x_1) e perda mínima para divisão (x_6). No contorno, a região de maior Precisão concentra-se em valores mais baixos de taxa de aprendizado associados a valores mais altos de fração de subamostragem de instâncias. Em termos práticos, a tendência observada indica que atualizações mais conservadoras (menor taxa de aprendizado), aliadas a maior fração de instâncias por árvore, favorecem a redução de falsos positivos e, conseqüentemente, aumentam a Precisão. A interação x_1x_4 indica que a escolha de taxa de aprendizado não deve ser analisada isoladamente: taxas mais elevadas podem ser viáveis em determinados regimes de amostragem, mas tornam-se desfavoráveis quando combinadas com baixa fração de instâncias.

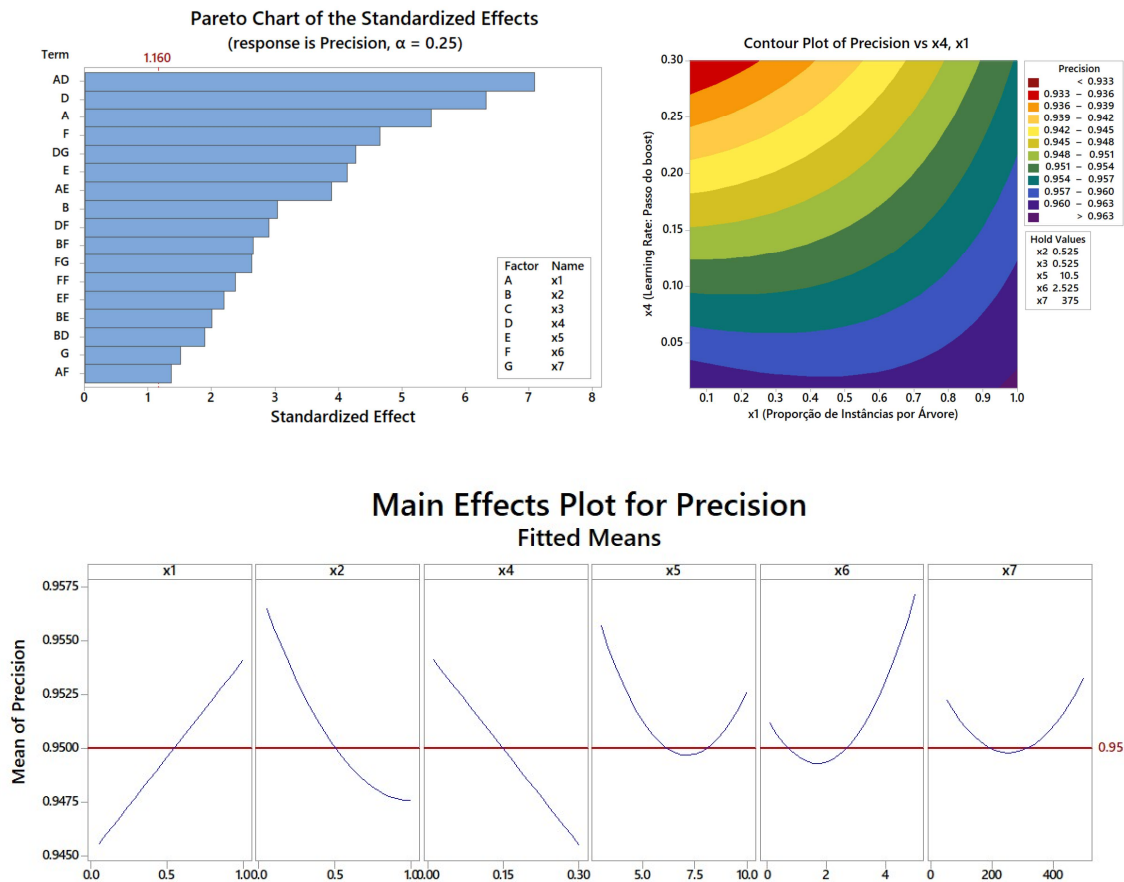


Figura 8 - Gráfico de Pareto de efeitos padronizados, contorno e efeitos principais para Precisão.

4.2.5 Revocação

A Figura 9 sintetiza os efeitos padronizados associados a Revocação e ilustra, por meio de um gráfico de contorno, a variação predita da resposta em função da taxa de aprendizado (x_4) e do número de estimadores (x_7), mantendo-se os demais fatores em seus níveis centrais.

Os resultados indicam que taxa de aprendizado (x_4) e número de estimadores (x_7) são fatores dominantes para a Revocação, com contribuição adicional de fração de subamostragem de instâncias (x_1) e fração de colunas por árvore (x_2), bem como de interações envolvendo x_4 e x_7 . O contorno sugere uma região de maior Revocação associada a valores mais altos de número de árvores e valores intermediários de taxa de aprendizado, evidenciando que a métrica não é monotônica com relação à taxa de aprendizado. Essa característica é compatível com a presença de termos quadráticos e com a existência de um ponto ótimo local no domínio experimental. Do ponto de vista de aplicação, a Revocação reflete a capacidade de identificar corretamente a classe positiva; a forma da superfície indica que escolhas agressivas ou excessivamente conservadoras de taxa de aprendizado podem degradar a sensibilidade, tornando necessária a busca orientada por modelos para localizar faixas mais apropriadas de parâmetros.

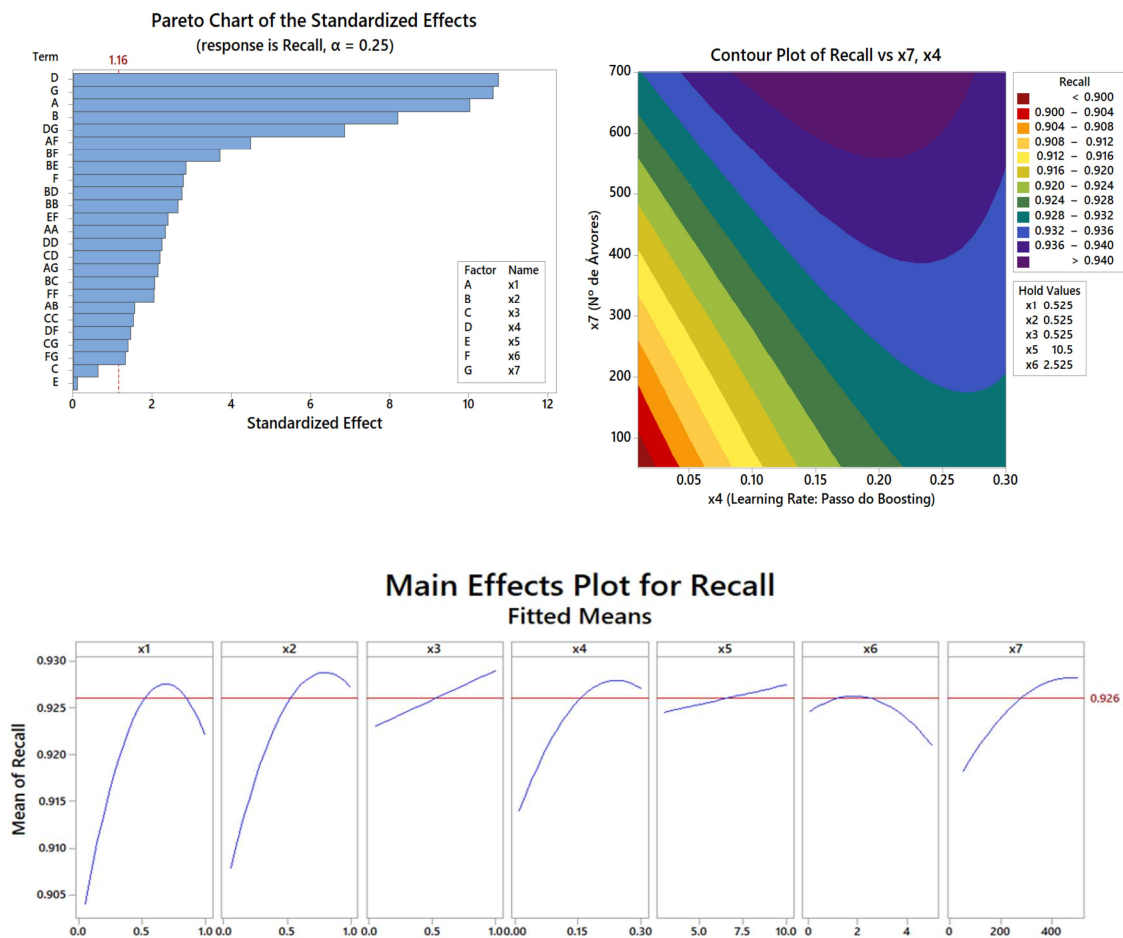
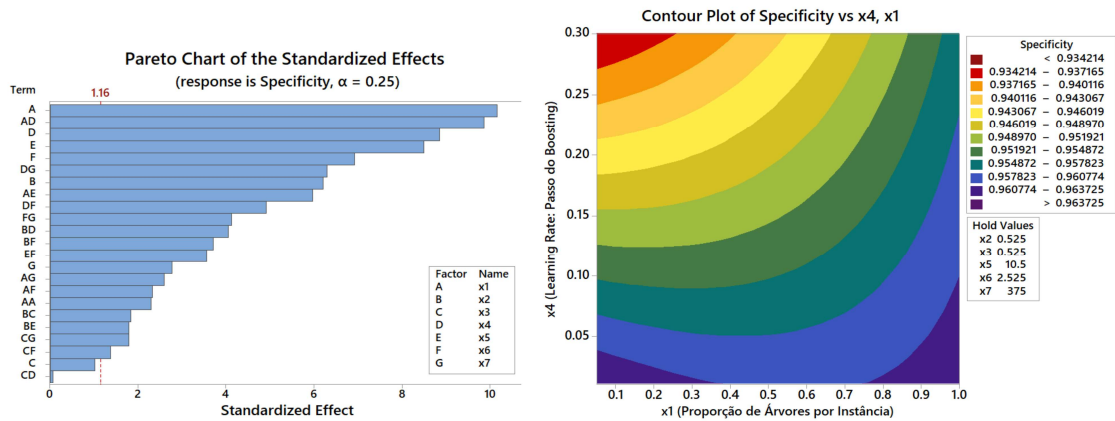


Figura 9 - Gráfico de Pareto de efeitos padronizados, contorno e efeitos principais para Revocação.

4.2.6 Especificidade

A Figura 10 apresenta os efeitos padronizados e o contorno para a Especificidade. A Especificidade apresenta comportamento qualitativamente semelhante ao da Precisão, com destaque para fração de subamostragem de instâncias (x_1), taxa de aprendizado (x_4) e sua interação, além de profundidade máxima (x_5) e perda mínima para divisão (x_6). O contorno evidencia uma região de maior Especificidade associada a valores mais baixos de taxa de aprendizado e maiores valores de fração de subamostragem de instâncias. A interpretação é consistente com o papel dessas escolhas na contenção de complexidade efetiva e na estabilidade do processo de *boosting*, reduzindo a taxa de falsos positivos. A presença de interações e termos quadráticos indica que, mesmo para uma métrica orientada à classe negativa, a relação entre hiperparâmetros não é estritamente linear.



Main Effects Plot for Specificity Fitted Means

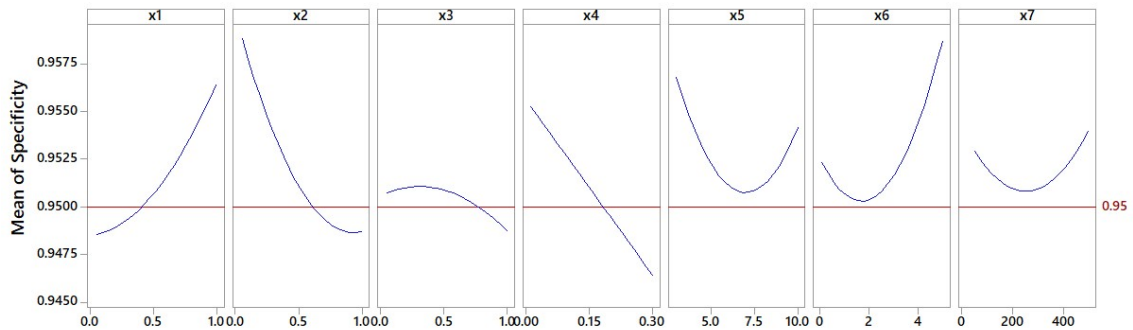
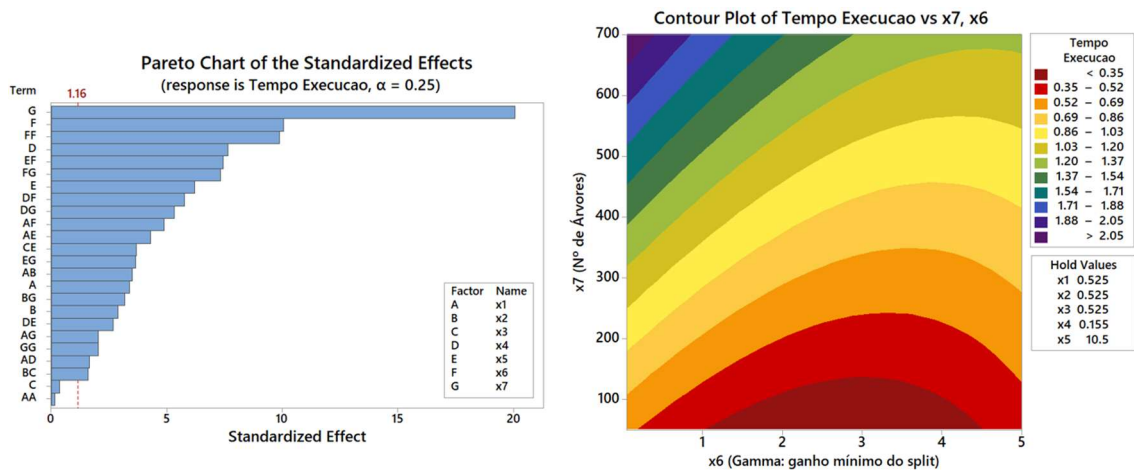


Figura 10 - Gráfico de Pareto de efeitos padronizados, contorno e efeitos principais para Especificidade.

4.2.7 Tempo de Execução

A Figura 11 apresenta os efeitos padronizados e o contorno para o Tempo de Execução. O gráfico de Pareto evidencia o predomínio de número de estimadores (x_7) como fator de maior efeito no Tempo de Execução, resultado esperado, dado que o custo de treinamento e inferência cresce com o número de árvores. Observa-se, adicionalmente, a influência de perda mínima para divisão (x_6) e de termos quadráticos associados, o que sugere que a penalização de divisão (*split*) também afeta o custo, ao controlar a quantidade de nós efetivamente expandidos. No contorno, a menor região de tempo concentra-se em valores reduzidos de número de estimadores, enquanto o aumento do número de árvores desloca o sistema para faixas mais elevadas de custo computacional. Esse resultado estabelece, de forma empírica e interpretável, o principal mecanismo de conflito com as respostas de qualidade: parte das regiões favoráveis em Acurácia e Revocação ocorre em valores elevados de número de estimadores, o que implica maior tempo.



Main Effects Plot for Tempo de Execucao Fitted Means

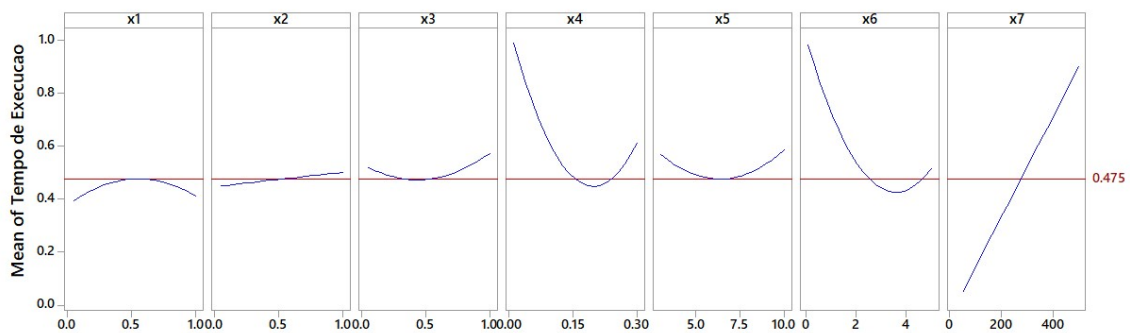


Figura 11 - Gráfico de Pareto de efeitos padronizados, contorno e efeitos principais para Tempo de execução.

4.2.8 Síntese e implicações para a formulação multiobjetivo

A análise mono-objetivo evidencia três aspectos centrais. Primeiro, a presença recorrente de interações e termos quadráticos nas respostas demonstra que o espaço de hiperparâmetros apresenta não linearidades relevantes, o que torna abordagens puramente exaustivas ou aleatórias menos eficientes, sobretudo quando se deseja inferência interpretável. Segundo, há evidência de que diferentes métricas são favorecidas em regiões distintas do domínio experimental: por exemplo, combinações que aumentam Revocação podem não coincidir com as que maximizam Precisão e Especificidade. Terceiro, a resposta Tempo de Execução é fortemente impactada por número de estimadores, estabelecendo um mecanismo direto de compromisso entre qualidade e custo.

Dessa forma, embora o estudo mono-objetivo forneça interpretabilidade e indique direções de ajuste, ele não resolve a seleção de uma configuração final quando existem objetivos concorrentes. Esse diagnóstico sustenta a necessidade de uma estratégia multiobjetivo, capaz de explorar o espaço de busca de maneira orientada por planejamento e

por modelos (Superfície de Resposta), culminando na identificação de soluções de compromisso sob restrições de custo computacional.

A análise mono-objetivo apresentada anteriormente evidencia que, embora seja possível identificar configurações favoráveis a cada resposta isoladamente, as condições que maximizam métricas de qualidade nem sempre coincidem com aquelas que minimizam o tempo de execução. Essa característica reforça a necessidade de uma formulação multiobjetivo, na qual os objetivos são considerados conjuntamente para a obtenção de soluções de compromisso com interpretabilidade e eficiência computacional. As figuras 12 e 13 representam uma síntese gráfica dos principais elementos do estudo mono-objetivo de forma sumarizada, bem como resultados individuais intra métodos.

4.3 Estudo multiobjetivo - aplicação da metodologia proposta

A etapa multiobjetivo parte dos resultados do planejamento experimental e da calibração dos modelos, em que cada configuração de hiperparâmetros produz um vetor de desempenho preditivo e custo computacional. Essas variáveis observadas são sintetizadas em fatores rotacionados (VRF₁-VRF₃) via PCA/Varimax, reduzindo redundâncias e tornando o problema tratável com três objetivos interpretáveis. As três VRFs são aproximadas por modelos quadráticos com RSM, que funcionam como meta-modelos contínuos das respostas no espaço dos hiperparâmetros codificados. Em seguida, o método NBI é aplicado sobre as três superfícies, gerando uma família de soluções Pareto-ótimas. Cada solução fornece um vetor de hiperparâmetros candidatos, que é validado por uma rodada de confirmação (treino e avaliação) para composição do conjunto final e seleção do ponto otimizado.

4.3.1 Obtenção de VRF₁, VRF₂ e VRF₃

Considere que, para cada execução experimental i ($i = 1, \dots, N$), são computadas p variáveis observadas: métricas de classificação e medidas de custo/tempo. Denotando por $\mathbf{M} \in \mathbb{R}^{N \times p}$ a matriz que empilha esses vetores por linha, procede-se à padronização por z-score para eliminar efeitos de escala e tornar comparáveis grandezas heterogêneas:

$$z_{ij} = \frac{m_{ij} - \mu_j}{\sigma_j}, \quad i = 1, \dots, N, \quad j = 1, \dots, p \quad (4.2)$$

A partir de $\mathbf{Z} = (z_{ij})$, calcula-se a matriz de correlação amostral \mathbf{R} , sobre a qual a PCA é aplicada. A PCA fornece autovalores e autovetores que quantificam a variância explicada e a direção principal de variação no espaço de métricas:

$$\mathbf{R} = \frac{1}{N-1} \mathbf{Z}^T \mathbf{Z} \quad (4.3)$$

Seja a decomposição espectral $\mathbf{R} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T$, com $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_p)$ e $\lambda_1 \geq \dots \geq \lambda_p$. Seleccionam-se r componentes ($r = 3$ neste trabalho), produzindo a matriz de cargas fatoriais não-rotacionadas \mathbf{L} :

$$\mathbf{L} = \mathbf{V}_r \mathbf{\Lambda}_r^{1/2} \quad (4.4)$$

Para facilitar interpretação e obter fatores mais 'puros' (cargas mais concentradas por fator), aplica-se uma rotação ortogonal do tipo Varimax. Seja \mathbf{T} uma matriz ortogonal ($\mathbf{T}^T \mathbf{T} = \mathbf{I}$) que maximiza o critério Varimax; as cargas rotacionadas são dadas por $\bar{\mathbf{L}}$, tal que:

$$\bar{\mathbf{L}} = \mathbf{L} \mathbf{T}, \quad \mathbf{T}^T \mathbf{T} = \mathbf{I} \quad (4.5)$$

A comunalidade h_j^2 de cada variável observada j mede a fração de variância explicada pelo subespaço fatorado e é útil para avaliar se a variável está adequadamente representada pelos r fatores:

$$h_j^2 = \sum_{k=1}^r \bar{\ell}_{jk}^2, \quad j = 1, \dots, p \quad (4.6)$$

Em seguida, calculam-se os escores fatoriais rotacionados para cada execução i . Adota-se a forma matricial do método de regressão (Thurstone), em que a matriz de coeficientes \mathbf{A} é obtida a partir de \mathbf{R} e das cargas rotacionadas $\bar{\mathbf{L}}$:

$$\mathbf{A} = \mathbf{R}^{-1} \bar{\mathbf{L}} (\bar{\mathbf{L}}^T \mathbf{R}^{-1} \bar{\mathbf{L}})^{-1} \quad (4.7)$$

Os escores fatoriais rotacionados, doravante denominados VRFs, são computados como combinações lineares das variáveis padronizadas. Seja $\mathbf{V} \times \mathbf{R} \times \mathbf{F} \in \mathbb{R}^{N \times r}$ a matriz de escores ($r = 3$). A i -ésima linha fornece $(\text{VRF}_{1i}, \text{VRF}_{2i}, \text{VRF}_{3i})$ para a execução i :

$$\mathbf{VRF} = \mathbf{Z} \mathbf{A} \Rightarrow \text{VRF}_{i,k} = \sum_{j=1}^p z_{ij} a_{jk} \quad (4.8)$$

A interpretação de cada VRF decorre do padrão de cargas em $\bar{\mathbf{L}}$. Em termos práticos, as três dimensões sintetizam: (i) um fator associado à qualidade global de classificação (maior contribuição de métricas como acurácia e precisão); (ii) um fator associado à capacidade de recuperação e robustez (maior contribuição de métricas como revocação e especificidade, dependendo do problema); e (iii) um fator associado ao custo computacional (dominância de variáveis de tempo/complexidade). Essas interpretações devem ser verificadas à luz das cargas rotacionadas observadas no estudo.

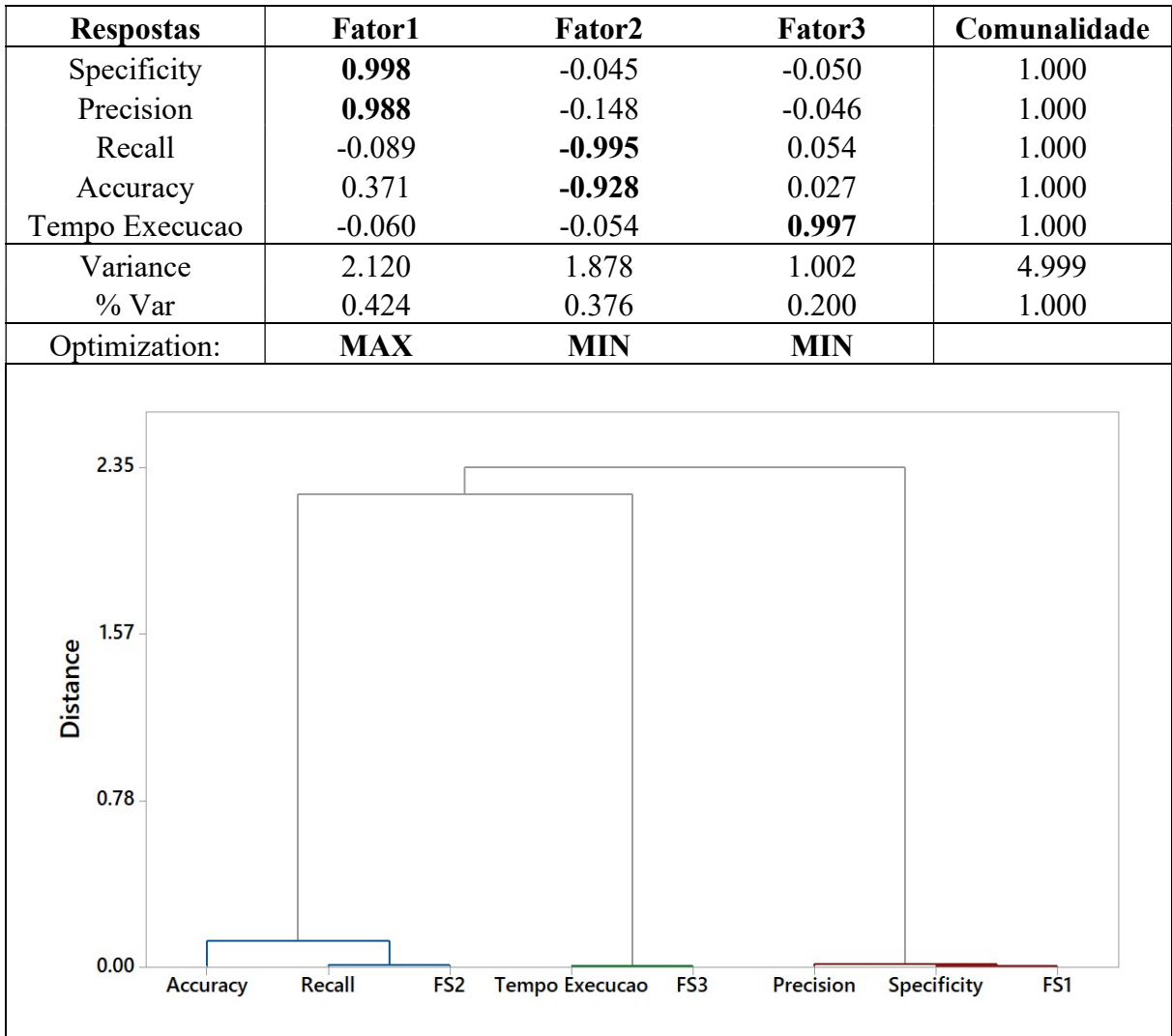


Figura 12 - Análise Fatorial com rotação Varimax e Cluster⁽¹⁾ de métricas de performance e escores rotacionados de fator.

Nota: ⁽¹⁾ Análise Hierárquica de Cluster baseada no Método de Ward e no valor absoluto da correlação entre as variáveis.

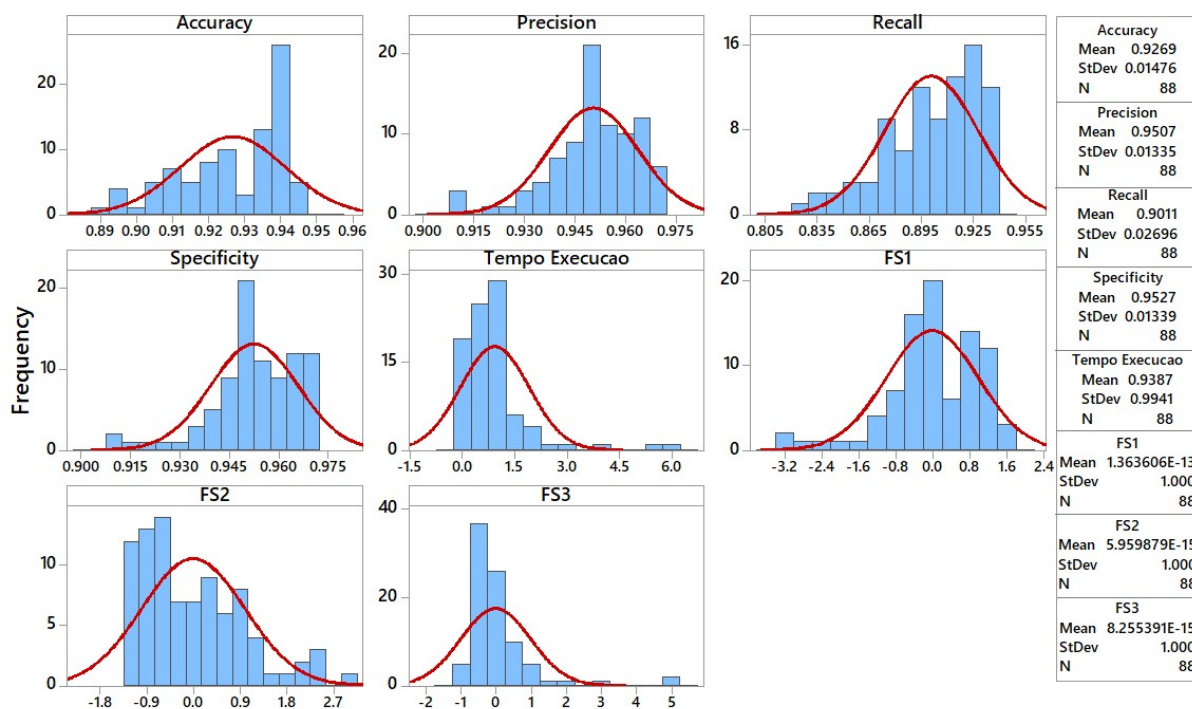


Figura 13 - Histogramas das métricas de performance do XGBoost e dos Escores rotacionados de fator.

A partir da análise das cargas fatoriais dos componentes rotacionados, observa-se que os fatores VRF_1 e VRF_2 apresentam associação predominante com métricas de desempenho preditivo do modelo, enquanto o fator VRF_3 concentra majoritariamente métricas relacionadas ao custo computacional, em especial o tempo médio de treinamento. Com base nessa evidência empírica, os fatores VRF_1 e VRF_2 passam a ser conjuntamente interpretados como uma dimensão denominada “qualidade”, ao passo que o fator VRF_3 é interpretado como uma dimensão de “Custo”. Essa agregação é adotada para fins de análise e discussão dos resultados, em consonância com a implementação computacional do estudo, mantendo-se a estrutura completa dos fatores nos procedimentos de otimização multiobjetivo.

4.3.2 Da VRF ao espaço contínuo: modelagem por Superfície de Resposta

A otimização multiobjetivo requer avaliar, de forma contínua, o impacto dos hiperparâmetros sobre cada VRF. Como as VRFs são observadas apenas nos pontos do planejamento experimental, ajusta-se um meta-modelo para cada VRF com base na Metodologia de Superfície de Resposta (RSM). Seja $\mathbf{x} = (x_1, \dots, x_K)$ o vetor de fatores codificados (K hiperparâmetros). Para cada VRF_k ($k = 1, 2, 3$), utiliza-se um modelo quadrático completo:

$$\widehat{\text{VRF}}_k(\mathbf{x}) = \beta_{0k} + \sum_{i=1}^K \beta_{ik} x_i + \sum_{i=1}^K \beta_{iik} x_i^2 + \sum_{i<j} \beta_{ijk} x_i x_j \quad (4.9)$$

A codificação linear utilizada para os hiperparâmetros segue a transformação padrão do planejamento de experimentos, em que h_i é o valor real do hiperparâmetro, $h_{i,0}$ é o ponto central e Δ_i é a semi-amplitude do intervalo estudado:

$$x_i = \frac{h_i - h_{i,0}}{\Delta_i}, \quad i = 1, \dots, K \quad (4.10)$$

Para hiperparâmetros naturalmente discretos (por exemplo, profundidade máxima), a solução contínua fornecida pela otimização deve ser posteriormente ajustada ao domínio admissível (arredondamento e truncamento), conforme a Seção 4.4.3.

Para cada resposta latente (FS_1 , FS_2 e FS_3), apresentam-se a seguir (i) o gráfico de Pareto dos efeitos padronizados e um corte em contorno no espaço de fatores, e (ii) o resumo do modelo com os coeficientes codificados selecionados via *Stepwise*. Esses resultados fundamentam a interpretação dos efeitos e servem de base para a construção das superfícies de resposta apresentadas na sequência.

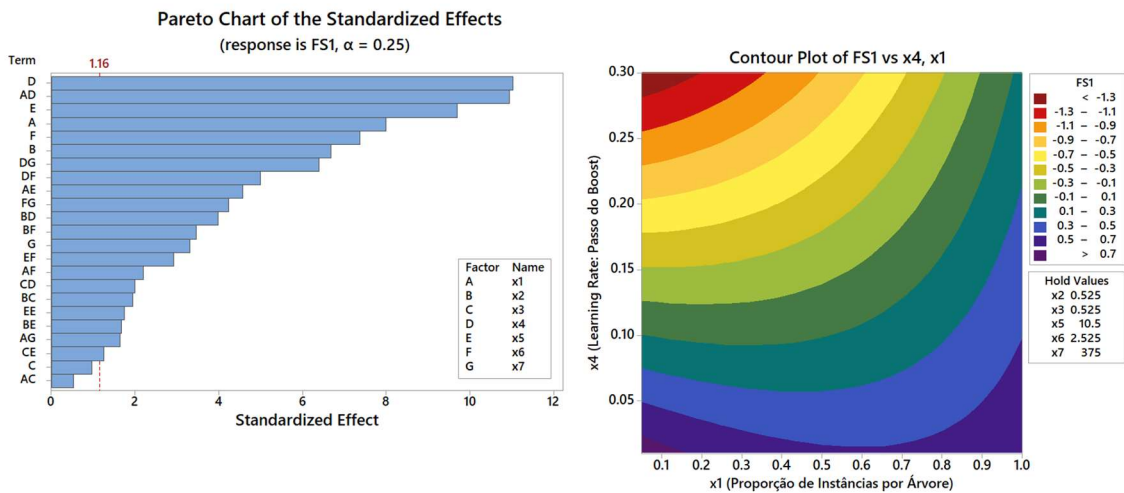


Figura 14 - Gráfico de Pareto dos efeitos padronizados e contorno para FS_1 (VRF_1).

Para o escore FS_1 (VRF_1), podemos concluir com as evidências das Figuras 14 e 15 observa-se predominância do fator D (x_4 : “learning_rate”) e da interação $A \times D$ ($x_1 \times x_4$: “subsample” \times “learning_rate”), além de contribuições relevantes de E (x_5 : “max_depth”) e F (x_6 : “gamma”), conforme o gráfico de Pareto. O contorno em função de (x_4 , x_1) apresentado na Figura 14 indica que combinações com “subsample” elevado e valores moderados de “learning_rate” conduzem a regiões mais favoráveis do VRF_1 , enquanto valores extremos de

“learning_rate” tendem a degradar o escore. Os coeficientes codificados corroboram esse comportamento, com efeito linear negativo de x_4 e interação positiva $x_1 \times x_4$, caracterizando um *trade-off* entre intensidade de aprendizagem e amostragem em cada iteração do *boosting*.

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.961056	98.84%	98.42%	92.56%

Coded Coefficients

Term	Coef	SE Coef	T-Value	P-Value
Constant	-0.0672	0.0331	-2.03	0.046
x1	0.2641	0.0330	7.99	0.000
x2	-0.2282	0.0342	-6.68	0.000
x3	-0.0332	0.0343	-0.97	0.337
x4	-0.3977	0.0360	-11.03	0.000
x5	-0.2119	0.0218	-9.70	0.000
x6	0.2695	0.0365	7.38	0.000
x7	0.1119	0.0338	3.31	0.002
x5*x5	0.0721	0.0414	1.74	0.086
x1*x3	0.0204	0.0386	0.53	0.600
x1*x4	0.4363	0.0398	10.96	0.000
x1*x5	-0.1846	0.0404	-4.57	0.000
x1*x6	-0.0906	0.0413	-2.19	0.032
x1*x7	0.0647	0.0392	1.65	0.104
x2*x3	-0.0811	0.0417	-1.95	0.056
x2*x4	0.1451	0.0364	3.99	0.000
x2*x5	0.0627	0.0375	1.67	0.100
x2*x6	-0.1381	0.0398	-3.47	0.001
x3*x4	-0.0691	0.0347	-1.99	0.051
x3*x5	0.0487	0.0389	1.25	0.215
x4*x6	0.1952	0.0390	5.00	0.000
x4*x7	-0.2581	0.0404	-6.39	0.000
x5*x6	0.1099	0.0376	2.92	0.005
x6*x7	0.1519	0.0359	4.24	0.000

Figura 15 - Resumo do modelo e coeficientes codificados para FS_1 (VRF_1).

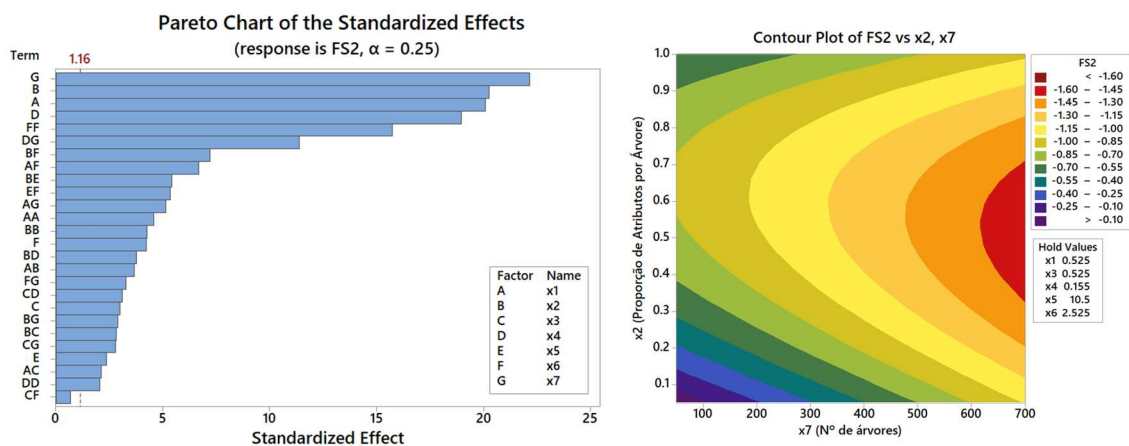


Figura 16 - Gráfico de Pareto dos efeitos padronizados e contorno para FS_2 (VRF_2).

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.960669	99.87%	99.81%	97.78%

Coded Coefficients

Term	Coef	SE Coef	T-Value	P-Value
Constant	-1.1743	0.0154	-76.37	0.000
x1	-0.3868	0.0193	-20.09	0.000
x2	-0.2933	0.0145	-20.25	0.000
x3	0.0515	0.0172	3.00	0.004
x4	-0.3611	0.0190	-18.96	0.000
x5	0.0420	0.0179	2.35	0.022
x6	0.0776	0.0183	4.24	0.000
x7	-0.3327	0.0150	-22.17	0.000
x1*x1	0.4063	0.0884	4.60	0.000
x2*x2	0.481	0.113	4.26	0.000
x4*x4	0.269	0.130	2.07	0.043
x6*x6	0.3793	0.0241	15.74	0.000
x1*x2	-0.0867	0.0236	-3.67	0.001
x1*x3	-0.0343	0.0162	-2.11	0.039
x1*x6	-0.1399	0.0209	-6.69	0.000
x1*x7	0.0837	0.0163	5.13	0.000
x2*x3	0.0600	0.0213	2.81	0.007
x2*x4	0.0804	0.0212	3.79	0.000
x2*x5	-0.1130	0.0209	-5.40	0.000
x2*x6	0.1332	0.0185	7.20	0.000
x2*x7	0.0680	0.0236	2.88	0.005
x3*x4	-0.0695	0.0225	-3.09	0.003
x3*x6	-0.0151	0.0218	-0.69	0.491
x3*x7	-0.0500	0.0180	-2.78	0.007
x4*x7	0.2127	0.0187	11.40	0.000
x5*x6	-0.1228	0.0230	-5.34	0.000
x6*x7	-0.0669	0.0203	-3.29	0.002

Figura 17 - Resumo do modelo e coeficientes codificados para FS₂ (VRF₂).

No caso de FS₂ (VRF₂), podemos concluir com as evidências das Figuras 16 e 17 que o termo G (x₇: “n_estimators”) assume a maior relevância estatística, seguido por B (x₂: “colsample_bytree”) e A (x₁: “subsample”), o que evidencia a sensibilidade do escore ao tamanho do ensemble e às frações de amostragem de instâncias e atributos. A presença de termos quadráticos (por exemplo, x₁² e x₂²) e de interações envolvendo x₇ sugere curvatura significativa na superfície, o que se reflete no contorno (x₂, x₇) pela existência de uma região ótima interna, em vez de um ótimo em fronteira. Em termos práticos, isso implica que aumentos indiscriminados de “n_estimators” não garantem melhoria monotônica do VRF₂, sendo necessário balancear o orçamento de árvores com frações de amostragem para evitar regimes de sobreajuste e/ou saturação de ganhos.

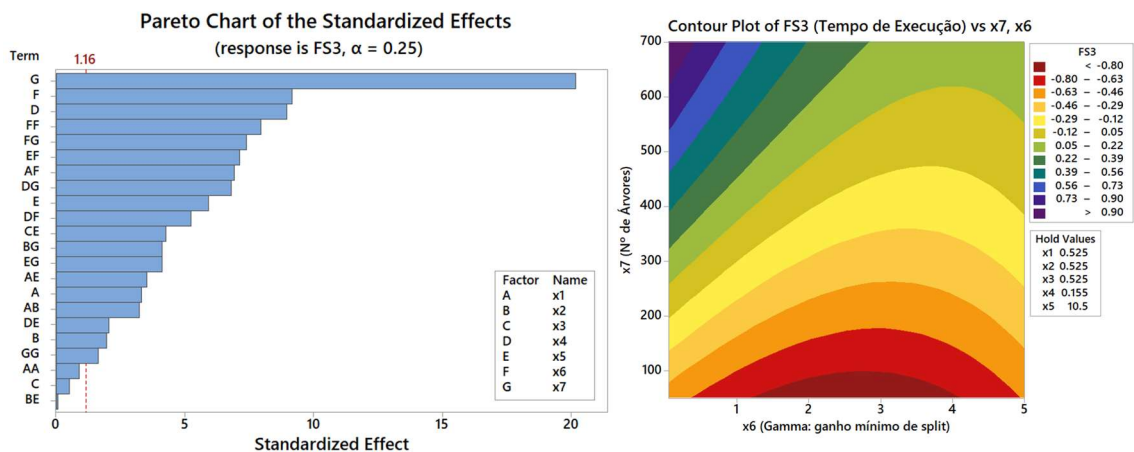


Figura 18 - Gráfico de Pareto dos efeitos padronizados e contorno para FS₃ (VRF₃).

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.956443	98.48%	97.97%	90.48%

Coded Coefficients

Term	Coef	SE Coef	T-Value	P-Value
Constant	-0.2207	0.0189	-11.66	0.000
x1	0.0708	0.0213	3.33	0.001
x2	0.0330	0.0167	1.97	0.053
x3	0.0083	0.0159	0.52	0.605
x4	-0.2548	0.0284	-8.97	0.000
x5	0.1729	0.0293	5.91	0.000
x6	-0.2428	0.0265	-9.16	0.000
x7	0.5948	0.0295	20.19	0.000
x1*x1	-0.0276	0.0310	-0.89	0.377
x6*x6	0.3329	0.0418	7.96	0.000
x7*x7	-0.1006	0.0609	-1.65	0.103
x1*x2	-0.0663	0.0205	-3.24	0.002
x1*x5	0.0815	0.0230	3.54	0.001
x1*x6	-0.1299	0.0188	-6.92	0.000
x2*x5	-0.0014	0.0217	-0.07	0.948
x2*x7	0.0759	0.0184	4.12	0.000
x3*x5	-0.0774	0.0181	-4.28	0.000
x4*x5	-0.0582	0.0284	-2.05	0.045
x4*x6	0.1442	0.0276	5.23	0.000
x4*x7	-0.1814	0.0266	-6.81	0.000
x5*x6	-0.1844	0.0259	-7.12	0.000
x5*x7	0.1159	0.0282	4.11	0.000
x6*x7	-0.2082	0.0282	-7.40	0.000

Figura 19 - Resumo do modelo e coeficientes codificados para FS₃ (VRF₃).

Por fim, FS₃ (VRF₃), pode-se verificar nas Figuras 18 e 19 que sua associação ao tempo de execução, é fortemente governado por G (x₇: “n_estimators”), bem como por F (x₆: “gamma”) e D (x₄: “learning_rate”), o que é consistente com o custo computacional do *boosting: ensembles* maiores demandam mais iterações de treinamento, e parâmetros de regularização influenciam a complexidade das partições. O contorno em (x₇, x₆) mostra que a penalização por *split* (“gamma”) e o número de árvores modulam diretamente o VRF₃, com evidências de não linearidade (termo x₆²) e interações (por exemplo, x₆×x₇). Assim, na etapa multiobjetivo, a minimização de VRF₃ atua como mecanismo explícito de controle de custo

computacional, restringindo soluções que, embora melhorem VRF_1 e VRF_2 , incorrem em tempos de execução proibitivos.

Terms	Accuracy	Precision	Recall	Specificity	Tempo	FS ₁	FS ₂	FS ₃	z(x)
1	0.94225	0.95217	0.93128	0.95179	0.7761	-0.06720	-1.17430	-0.22070	1.0000
x ₁	0.00695	0.00487	0.00946	0.00407	0.0985	0.26410	-0.38680	0.07080	-1.0000
x ₂	0.00251	-0.00295	0.00780	-0.00293	0.0564	-0.22820	-0.29330	0.03300	0.5062
x ₃	-0.00086	-0.00040	-0.00059	-0.00045	0.0071	-0.03320	0.05150	0.00830	-0.2895
x ₄	0.00259	-0.00482	0.01017	-0.00468	-0.2152	0.16045	-0.36110	-0.25480	1.0000
x ₅	-0.00138	-0.00336	0.00011	-0.00318	0.1878	-0.02114	0.04200	0.17290	-0.3702
x ₆	0.00052	0.00347	-0.00258	0.00377	-0.275	0.07005	0.07760	-0.24280	-0.0933
x ₇	0.00511	0.00194	0.00992	0.00132	0.6103	0.11190	-0.33270	0.59480	-0.9369
x ₁ ²	-0.00339	0.00000	-0.00876	0.00121	-0.0049	0.00000	0.40630	-0.02760	1.0000
x ₂ ²	-0.00562	0.00000	-0.00996	0.00000	0.00000	0.00000	0.48100	0.00000	0.2562
x ₃ ²	0.00000	-0.00212	-0.00570	0.00000	0.00000	0.00000	0.00000	0.00000	0.0838
x ₄ ²	-0.00280	0.00000	-0.00841	0.00000	0.00000	0.00000	0.26900	0.00000	1.0000
x ₅ ²	0.00000	0.00000	0.00000	0.00000	0.00000	0.07210	0.00000	0.00000	0.1370
x ₆ ²	-0.00539	0.00000	-0.00768	0.00000	0.287	0.00000	0.37930	0.33290	0.0087
x ₇ ²	-0.00317	0.00000	0.00000	0.00000	-0.1147	0.00000	0.00000	-0.10060	0.8778
x ₁ x ₂	0.00000	0.00000	0.00153	0.00000	-0.0765	0.00000	-0.08670	-0.06630	-0.5062
x ₁ x ₃	0.00089	0.00000	0.00000	0.00000	0.00000	0.02040	-0.03430	0.00000	0.2895
x ₁ x ₄	0.00264	0.00566	0.00000	0.00554	-0.0501	0.43630	0.00000	0.00000	-1.0000
x ₁ x ₅	-0.00094	-0.00269	0.00000	-0.00273	0.1229	-0.18460	0.00000	0.08150	0.3702
x ₁ x ₆	0.00112	-0.00122	0.00426	-0.00137	-0.1393	-0.09060	-0.13990	-0.12990	0.0933
x ₁ x ₇	0.00000	0.00000	-0.00203	0.00142	0.064	0.06470	0.08370	0.00000	0.9369
x ₂ x ₃	-0.00162	-0.00098	-0.00202	-0.00092	-0.0466	-0.08110	0.06000	0.00000	-0.1465
x ₂ x ₄	0.00000	0.00189	-0.00271	0.00203	0.00000	0.14510	0.08040	0.00000	0.5062
x ₂ x ₅	0.00169	0.00076	0.00279	0.00094	0.00000	0.06270	-0.11300	-0.00140	-0.1874
x ₂ x ₆	-0.00255	-0.00195	-0.00355	-0.00191	0.00000	-0.13810	0.13320	0.00000	-0.0472
x ₂ x ₇	0.00000	0.00000	0.00000	0.00000	0.0677	0.00000	0.06800	0.07590	-0.4742
x ₃ x ₄	0.00000	0.00000	0.00207	0.00003	0.00000	-0.06910	-0.06950	0.00000	-0.2895
x ₃ x ₅	0.00000	0.00000	0.00000	0.00000	-0.0777	0.04870	0.00000	-0.07740	0.1072
x ₃ x ₆	0.00000	0.00056	0.00000	0.00080	0.00000	0.00000	-0.01510	0.00000	0.0270
x ₃ x ₇	0.00000	-0.00087	0.00132	-0.00098	0.00000	0.00000	-0.05000	0.00000	0.2712
x ₄ x ₅	0.00000	0.00000	0.00000	0.00000	-0.0804	0.00000	0.00000	-0.05820	-0.3702
x ₄ x ₆	0.00187	0.00251	0.00140	0.00239	0.1568	0.19520	0.00000	0.14420	-0.0933
x ₄ x ₇	-0.00503	-0.00371	-0.00648	-0.00348	-0.1877	-0.25810	0.21270	-0.18140	-0.9369
x ₅ x ₆	0.00000	0.00000	0.00233	0.00188	-0.2012	0.10990	-0.12280	-0.18440	0.0346
x ₅ x ₇	0.00181	0.00156	0.00000	0.00000	0.1063	0.00000	0.00000	0.11590	0.3468
x ₆ x ₇	0.00161	0.00227	0.00126	0.00204	-0.2141	0.15190	-0.06690	-0.20820	0.0875
f(x)	0.92735	0.93807	0.90893	0.94125	0.12000	-0.47098	-0.31059	-1.04240	
g ₁ (x)	3.364	3.36359							

Tabela 3 - Resumo dos modelos e coeficientes⁽¹⁾ codificados para as métricas de desempenho do XGBoost e suas variáveis latentes (Escores Rotacionados de Fator).

Nota⁽¹⁾: Valores destacados em azul (coeficiente igual a 0) indicam termos não significativos para a resposta correspondente, ajustados para zero no modelo final.

4.4 Otimização multiobjetivo: do simplex NBI ao resultado otimizado

Com três meta-modelos $\widehat{VRF}_1(\mathbf{x})$, $\widehat{VRF}_2(\mathbf{x})$ e $\widehat{VRF}_3(\mathbf{x})$, define-se o problema multiobjetivo no espaço dos fatores codificados. O método NBI é particularmente adequado porque constrói pontos sobre a fronteira de Pareto a partir de interseções normais ao simplex formado pelos pontos de ancoragem fixados no espaço de objetivos (Das; Dennis, 1998).

4.4.1 Formulação do problema no espaço VRF

Por convenção, o NBI é usualmente descrito para problemas de minimização. Assim, utiliza-se um vetor de sinais $\mathbf{s}^T = (s_1, s_2, s_3)$, onde $s_k = -1$ quando o objetivo k deve ser maximizado (convertendo maximização em minimização) e $s_k = +1$ quando deve ser minimizado. Denotando por $\mathbf{F}(\mathbf{x})$ o vetor de objetivos já no sentido de minimização, obtém-se:

$$\min_{\mathbf{x} \in \Omega} \mathbf{F}(\mathbf{x}) = [s_1 \widehat{VRF}_1(\mathbf{x}), s_2 \widehat{VRF}_2(\mathbf{x}), s_3 \widehat{VRF}_3(\mathbf{x})]^T \quad (4.11)$$

Em que Ω representa o domínio admissível dos fatores (tipicamente $\Omega = [-1, 1]^K$ quando o planejamento é codificado). No caso particular deste trabalho, os dois primeiros fatores (\widehat{VRF}_1 e \widehat{VRF}_2) estão associados a desempenho e devem ser maximizados ($s_1 = s_2 = -1$), enquanto o terceiro fator (\widehat{VRF}_3) representa custo e deve ser minimizado ($s_3 = +1$). Entretanto, esta definição de sentidos de otimização precisa ser alinhada com os sentidos das otimizações individuais das respostas originais e as correlações observadas entre estas e os \widehat{VRF}_i . Para o caso do XGBoost estudado neste trabalho, \widehat{VRF}_1 deverá ser maximizado enquanto \widehat{VRF}_2 e \widehat{VRF}_3 deverão ser minimizados (Observem-se os sinais *loadings* descritos na Figura 12).

4.4.2 Pontos de Ancoragem e construção do *simplex* no espaço de objetivos

O NBI é inicializado definindo os pontos de ancoragem (ou soluções individuais) associados à otimização isolada de cada objetivo. Para cada $k = 1, 2, \dots, p$, resolve-se:

$$\mathbf{x}_k^* = \operatorname{argmin}_{\mathbf{x} \in \Omega} F_k(\mathbf{x}), \quad k = 1, 2, \dots, p \quad (4.12)$$

Com essas soluções, constroem-se os vetores-âncora no espaço de objetivos, $\mathbf{F}_k^* = \mathbf{F}(\mathbf{x}_k^*)$. Define-se também o ponto utópico \mathbf{F}^U (utopia) como o vetor formado pelos melhores valores individuais de cada objetivo, tal que:

$$\mathbf{F}^U = [F_1(\mathbf{x}_1^*), F_2(\mathbf{x}_2^*), F_3(\mathbf{x}_3^*)]^\top \quad (4.13)$$

Com isso, monta-se a matriz Φ das direções do *simplex* CHIM com colunas $\mathbf{F}_k^* - \mathbf{F}^U$:

$$\Phi = [\mathbf{F}_1^* - \mathbf{F}^U \quad \mathbf{F}_2^* - \mathbf{F}^U \quad \mathbf{F}_3^* - \mathbf{F}^U] \quad (4.14)$$

Qualquer ponto sobre o *simplex* no espaço de objetivos pode ser parametrizado por um vetor de pesos $\beta^T = (\beta_1, \beta_2, \beta_3)$, pertencente ao *simplex* unitário: $\beta_k \geq 0$ e $\beta_1 + \beta_2 + \beta_3 = 1$. Dessa forma, tem-se que:

$$\mathbf{F}^c(\beta) = \mathbf{F}^U + \Phi\beta, \quad \beta_1 + \beta_2 + \beta_3 = 1, \quad \beta_k \geq 0 \quad (4.15)$$

Para o caso do XGBoost, tem-se a seguinte parametrização:

	Matriz Payoff (NBI)							
	Accuracy	Precision	Recall	Specificity	Tempo	FS ₁	FS ₂	FS ₃
Accuracy	0.9509	0.9586	0.9405	0.9591	0.9103	0.4301	-1.6503	-0.1337
Precision	0.9374	0.9689	0.9044	0.9686	1.0846	1.1244	-0.1345	0.1421
Recall	0.9447	0.9539	0.9551	0.9524	1.4826	-0.0532	-1.9873	0.4569
Specificity	0.9386	0.9685	0.9055	0.9689	1.0157	1.0640	-0.1373	0.0403
Tempo	0.9366	0.9504	0.9233	0.9507	0.1200	-0.1718	-0.9072	-0.8729
VRF ₁	0.9497	0.9666	0.9285	0.9658	0.3510	1.2021	-1.0250	-0.6254
VRF ₂	0.9445	0.9523	0.9389	0.9554	0.9752	0.2425	-1.6665	-0.0536
VRF ₃	0.9273	0.9379	0.9089	0.9412	0.1200	-0.4710	-0.3106	-1.0425

VRF ₁	VRF ₂	VRF ₃
4.919	3.716	1.0018

Payoff		
2.120	3.041	4.919
2.289	1.878	3.716
1.176	1.980	1.002

Payoff Escalonada		
0.000	0.329	1.000
0.224	0.000	1.000
0.178	1.000	0.000

$\bar{\Phi}\beta$	$t\hat{m} = t\bar{\Phi}e$	$\bar{F}(x)$
1.000	0.000	1.000
1.000	0.000	1.000
0.000	0.000	0.000

$g_1(x)$:	3.364
$g_2(x)$:	0.000
$g_3(x)$:	0.000
$g_4(x)$:	0.000
Slack(x):	0.000
T:	0.000

F^U	F^N
2.120	4.919
1.878	3.716
1.002	1.980

β	$e=-t$
0.000	0.000
0.000	0.000
1.000	0.000
1.000	

Tabela 4 - Configuração do método VRF-NBI.

4.4.3 Do simplex aos hiperparâmetros candidatos: subproblemas NBI

Seja \mathbf{n} o vetor normal unitário ao *simplex* (orientado para a região de melhoria). Para cada vetor de pesos $\beta^{(k)}$, o NBI busca o ponto da fronteira de Pareto alcançado pela interseção entre a fronteira viável e a semirreta que parte de $\mathbf{F}^c(\beta^{(k)})$ na direção normal \mathbf{n} . Essa busca é formulada pelo subproblema:

$$\mathbf{F}^c \max_{\mathbf{x} \in \Omega, t \geq 0} t, \quad \mathbf{F}(\mathbf{x}) = \mathbf{F}^c(\beta^{(k)}) + t \mathbf{n} \quad (4.16)$$

A solução do subproblema para cada $\beta^{(k)}$ gera um ponto Pareto-ótimo no espaço codificado, denotado $\mathbf{x}^{(k)}$. Nas implementações numéricas realizadas neste trabalho, esse problema é resolvido via SLSQP, utilizando os meta-modelos quadráticos como funções avaliáveis.

A discretização do simplex define quantos subproblemas serão resolvidos. Para H subdivisões uniformes (passo $1/H$), o número de vetores de pesos β no simplex tridimensional é igual a:

$$N_\beta = \frac{(H+2)(H+1)}{2} \quad (4.17)$$

Uma vez obtido $\mathbf{x}^{(k)}$, faz-se a decodificação para o espaço real dos hiperparâmetros, utilizando a transformação inversa (4.19). Para evidenciar o caráter de pós-processamento em variáveis discretas e restritas, denota-se por $\rho_i(\cdot)$ um operador de ajuste (arredondamento e truncamento) específico de cada hiperparâmetro:

$$h_i^{(k)} = \rho_i(h_{i,0} + \Delta_i x_i^{(k)}), \quad i = 1, \dots, k \quad (4.18)$$

O conjunto de candidatos é então definido como a coleção de vetores de hiperparâmetros gerados por todos os $\beta^{(k)}$, com remoção de duplicatas e eventuais candidatos inviáveis após ajuste:

$$\mathcal{H} = \{\mathbf{h}^{(k)}\}_{k=1}^{N_\beta} \quad (4.19)$$

Observe que \mathcal{H} é um conjunto gerado no nível de meta-modelo. A etapa seguinte é essencial: validar os candidatos por execuções reais do algoritmo de aprendizagem, obtendo as métricas e VRFs observadas para cada candidato e garantindo que o ponto ótimo não seja um artefato do ajuste quadrático.

4.4.4 Confirmação experimental e seleção do resultado otimizado

Para cada candidato $\mathbf{h} \in \mathcal{H}$, executa-se uma rodada de confirmação (treino e avaliação) sob o mesmo protocolo de validação utilizado nas comparações entre métodos. A partir das

métricas observadas, recalculam-se as VRFs usando os mesmos coeficientes \mathbf{A} (Seção 4.3.1), mantendo consistência na síntese dos objetivos.

Como a solução do NBI produz um conjunto de soluções Pareto-ótimas, é necessário um critério de decisão para selecionar o resultado final. Uma estratégia prática é utilizar uma função utilidade $U(\mathbf{h})$ construída com VRFs normalizadas, permitindo comparar candidatos em uma escala comum:

$$\mathbf{h}^* = \operatorname{argmax}_{\mathbf{h} \in \mathcal{H}} U(\mathbf{h}) \quad (4.20)$$

A utilidade pode ser definida de diversas formas. Uma escolha simples, alinhada ao uso de pesos explícitos, é:

$$U(\mathbf{h}) = \omega_1 \widetilde{\text{VRF}}_1(\mathbf{h}) + \omega_2 \widetilde{\text{VRF}}_2(\mathbf{h}) + \omega_3 \widetilde{\text{VRF}}_3(\mathbf{h}), \quad \omega_1 + \omega_2 + \omega_3 = 1 \quad (4.21)$$

Onde $\widetilde{\text{VRF}}_k(\mathbf{h})$ representa a versão normalizada (e, quando necessário, invertida) da k -ésima VRF, de modo que valores maiores indiquem sempre melhor desempenho. Alternativamente, pode-se escolher o “ponto de joelho” da fronteira de Pareto, que maximiza ganho marginal em desempenho para um incremento moderado de custo.

4.5 Encadeamento final: da obtenção dos fatores ao resultado otimizado

Em síntese, a etapa multiobjetivo segue o encadeamento:

1. Calcular VRF_1 - VRF_3 a partir das métricas e custos observados;
2. Ajustar meta-modelos RSM para cada VRF;
3. Gerar um conjunto de pontos no simplex ($\boldsymbol{\beta}^{(k)}$) e resolver os subproblemas NBI;
4. Decodificar cada solução $\mathbf{x}^{(k)}$ em um vetor real de hiperparâmetros $\mathbf{h}^{(k)}$;
5. Consolidar o conjunto \mathcal{H} , remover duplicatas e realizar confirmação experimental;
6. Selecionar \mathbf{h}^* segundo um critério de decisão (utilidade, ponto de joelho ou restrição operacional).

Essa estrutura separa claramente o papel da otimização (sobre meta-modelos) do papel da validação (sobre o modelo real), minimizando o risco de se escolher uma configuração ótima apenas no nível de aproximação. Ao mesmo tempo, o uso de VRFs permite explicar o compromisso entre objetivos de forma mais interpretável do que o uso direto de um grande número de métricas correlacionadas.

Neste trabalho, propõe-se que passo 3 desse algoritmo (gerar um conjunto de pontos no simplex ($\boldsymbol{\beta}^{(k)}$) e resolver os subproblemas NBI) seja definido por um arranjo experimental de mistura do tipo *Simplex-Lattice* $\{3, 10\}$, nos qual $q = 3$ funções objetivo e $p = 10$ (sendo p igual

ao inverso do espaçamento entre os pesos $\beta^{(k)}$). Desse modo, formam-se 66 combinações convexas de pesos apresentadas na Figura 20. Na Figura 21, apresentam-se os valores ótimos de VRF obtidos com as 66 iterações do método NBI.

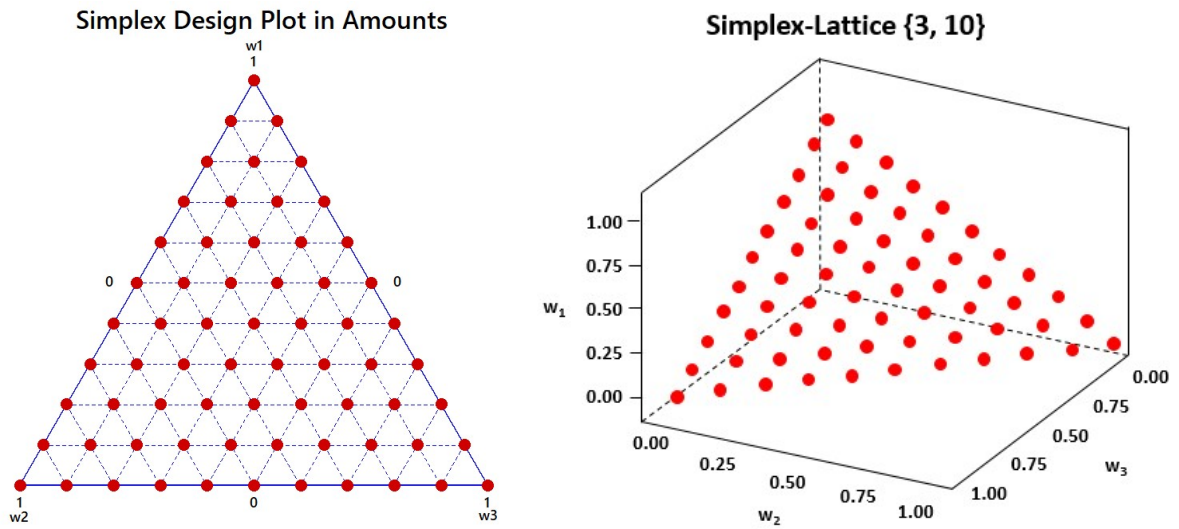


Figura 20 - Arranjo *Simplex-Lattice* {3, 10} usado no caso XGBoost.

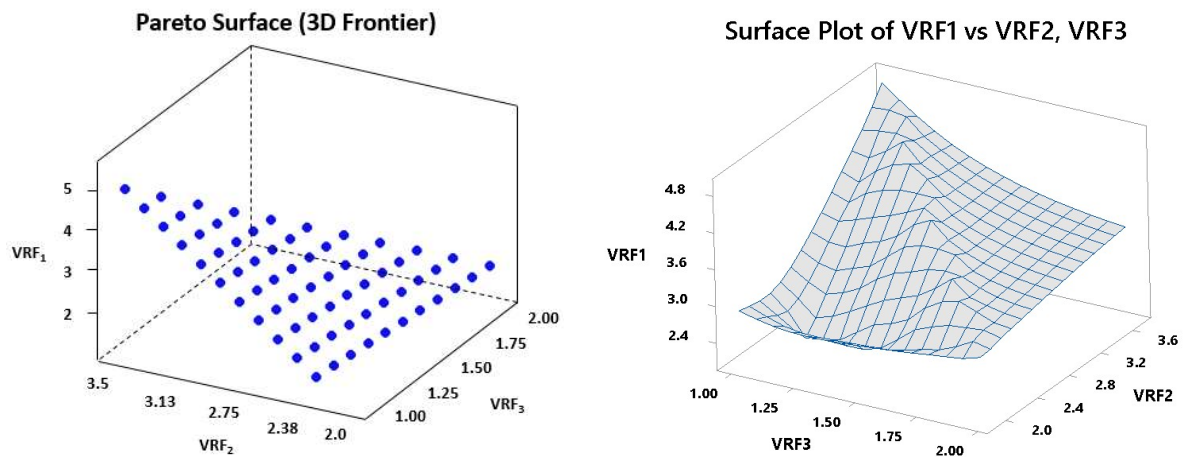


Figura 21 - Superfície de Pareto para os Índices Multivariados (VRF) relativo às métricas de desempenho do XGBoost para os pesos definidos pelo arranjo *Simplex-Lattice* {3, 10}.

Embora a formulação em *simplex* e a leitura triobjetivo (VRF₁-VRF₃) sejam apresentadas de forma informacional, para explicitar a estrutura dos fatores latentes e apoiar a interpretação do compromisso entre dimensões, o pipeline experimental efetivamente adotado operacionaliza a etapa decisória de forma biobjetivo (qualidade × custo), por meio da

condensação dos fatores de desempenho (VRF_1 e VRF_2) em um escore único de qualidade e da associação do fator ligado ao tempo (VRF_3) a um escore de custo.

A Figura 22 apresenta os gráficos de contorno para as métricas de desempenho do algoritmo XGBoost, obtidas nas 66 configurações de pesos do NBI. Observa-se claramente que os valores máximos para *Accuracy*, *Precision* e *Specificity* são alcançados quando se atribui pesos maiores a VRF_1 e VRF_2 , enquanto *Recall* é obtido com valores menores destas métricas.

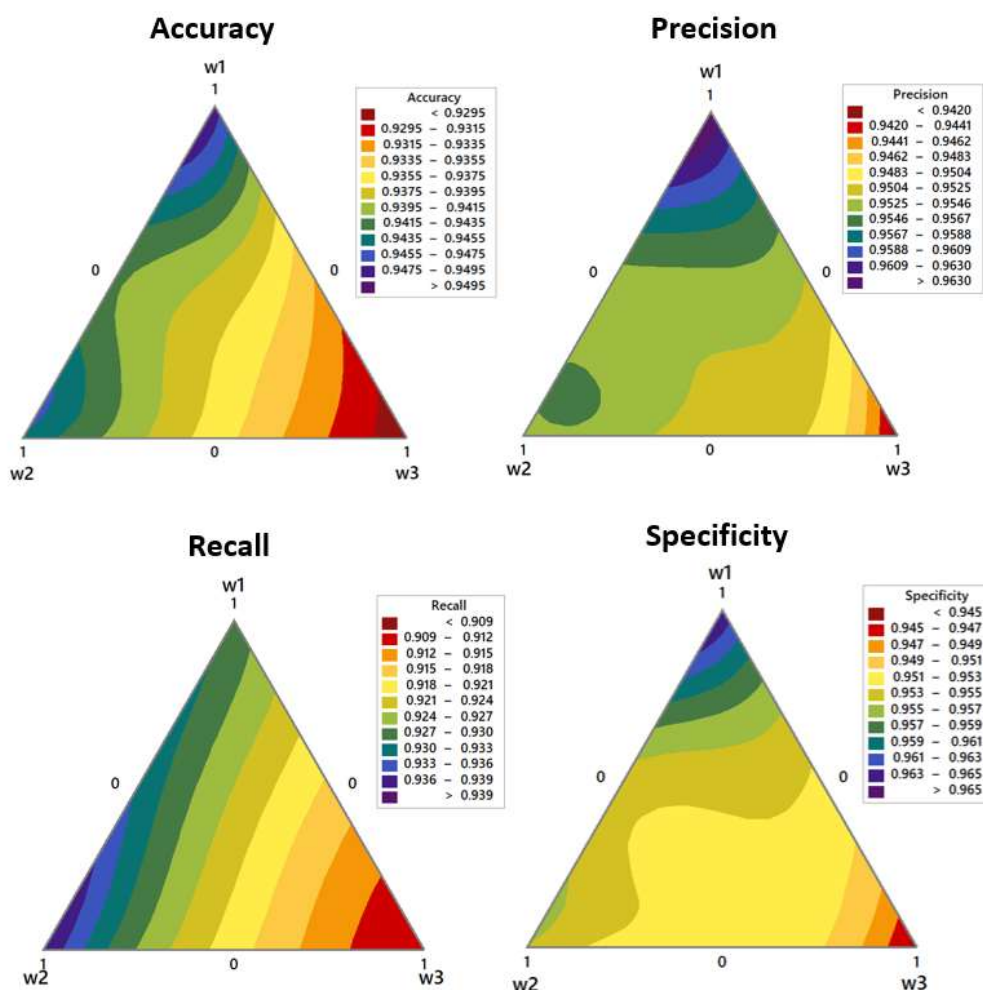


Figura 22 - Superfícies de Resposta de Misturas para as métricas de desempenho do XGBoost baseado no arranjo *Simplex-Lattice* {3, 10}.

Raciocínio análogo se aplica aos resultados ilustrados nos gráficos de contorno para as superfícies de resposta de mistura para *Tempo de Processamento* e VRFs da Figura 23. Já a Figura 24 apresenta a sobreposição de todas as superfícies de resposta de misturas no espaço do *Simplex-Lattice* {3, 10} (*Mixture Overlaid Contour Plot*). Observa-se uma região viável de solução na qual as melhores soluções Pareto-ótimas se localizam. A melhor solução Pareto-ótima é definida pelos pesos $w_1=0.67$, $w_2=0.27$ e $w_3=0.06$.

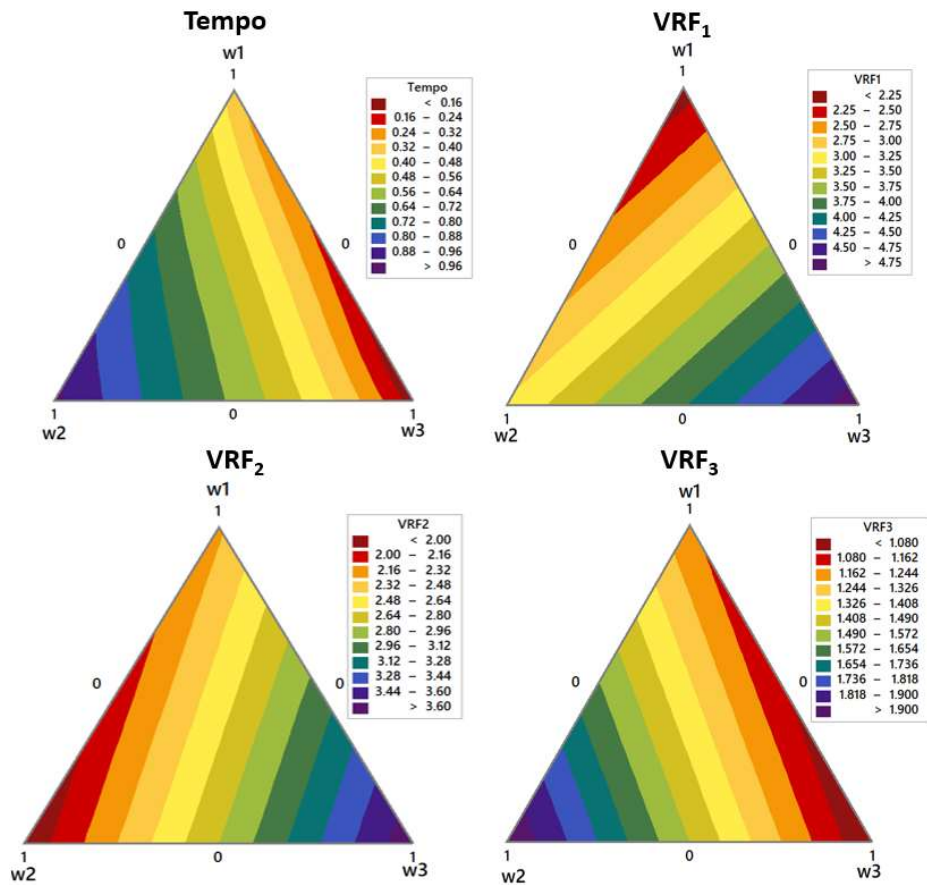


Figura 23 - Superfícies de Resposta de Misturas para o tempo de processamento do XGBoost e os indicadores multivariados das métricas de desempenho avaliadas nos pontos do arranjo *Simplex-Lattice* {3, 10}.

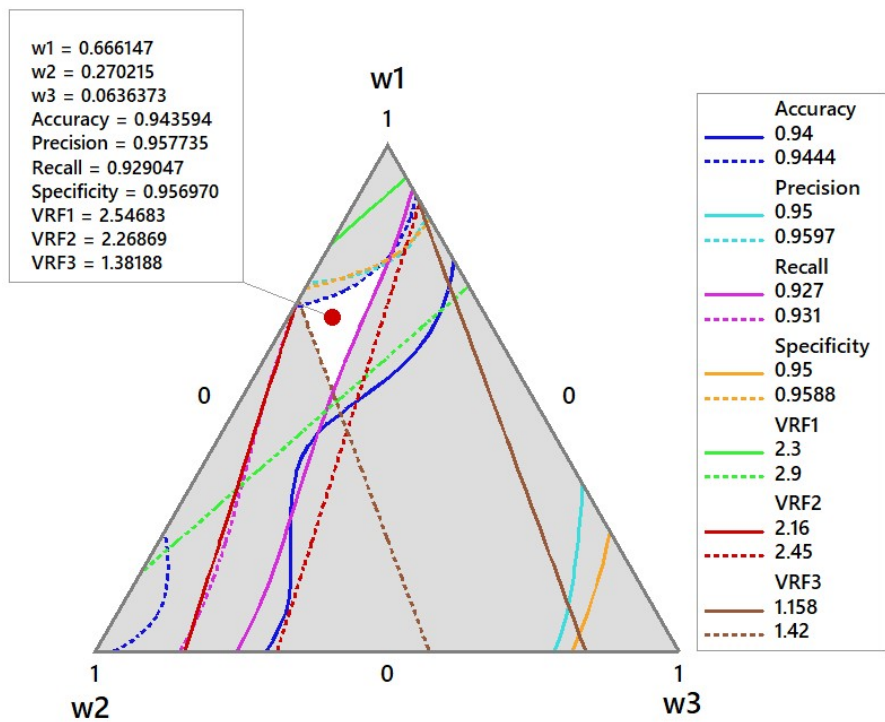


Figura 24 - *Overlaid Contour Plot* (Sobreposição das Superfícies de Resposta para o Arranjo *Simplex-Lattice* {3, 10} das métricas de desempenho do XGBoost e os respectivos indicadores multivariados (VRF).

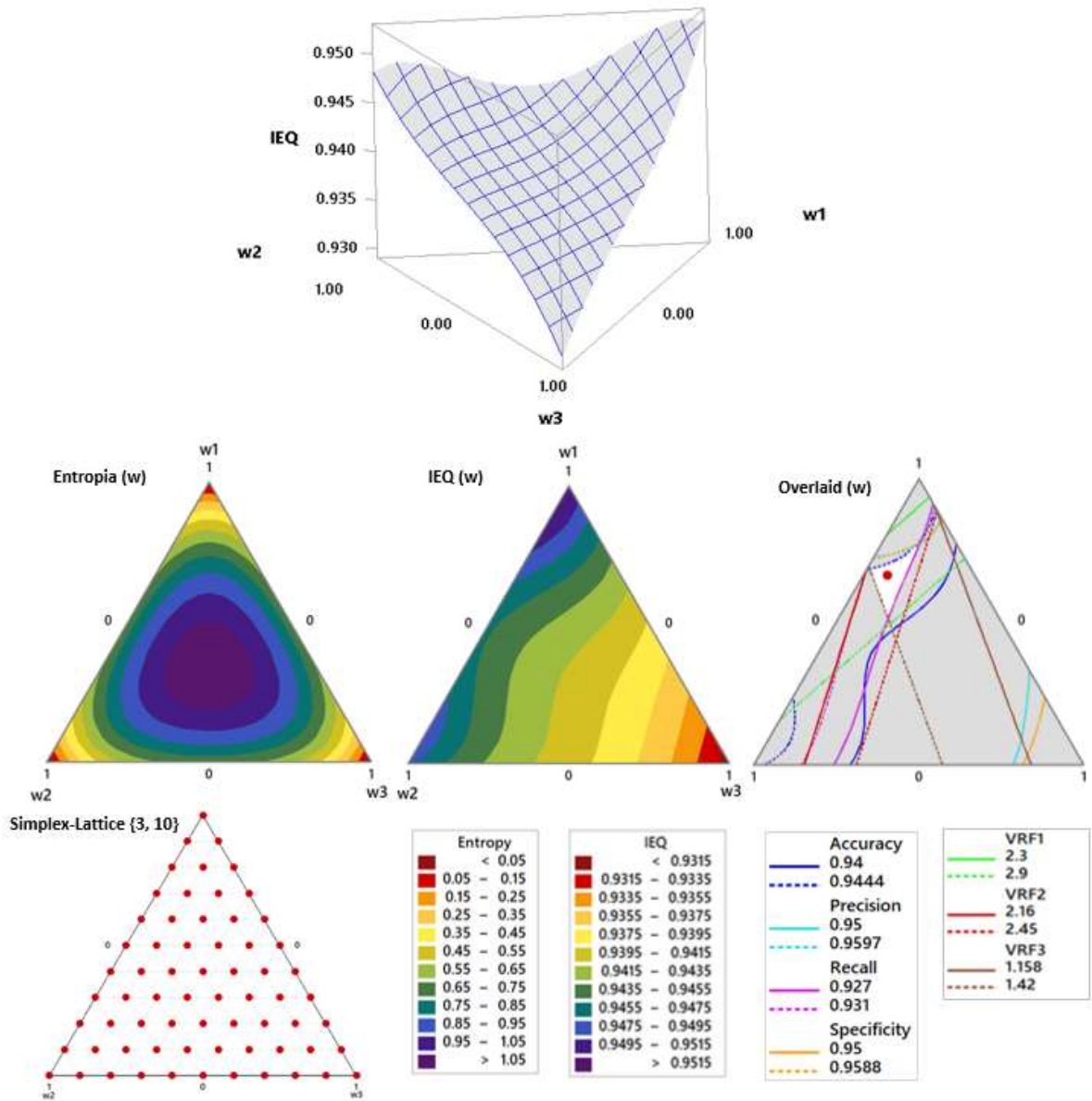


Figura 25 - Arranjo *Simplex-Lattice* {3, 10} do Índice Escalar de Qualidade do XGBoost.

Repetiu-se a análise também para o Índice Escalar de Qualidade (IEQ) e para a Entropia dos pesos utilizados no Simplex-Lattice {3, 10}. Como o IEQ é uma média aritmética simples das quatro métricas de desempenho (*Accuracy*, *Precision*, *Recall* e *Specificity*), há uma tendência de o gráfico de contorno seguir o mesmo padrão dos modelos individuais, concentrando os pesos mais em VRF₁ e VRF₂ do que em VRF₃. Isto faz bastante sentido, uma vez que é mais importante se alcançar métricas de desempenho satisfatórias do que um *Tempo de processamento* extremamente pequeno. O gráfico da *Entropia* dos pesos, por sua vez, destaca valores pequenos nas regiões de VRF₁ e VRF₂ máximo, indicando uma maior concentração de pesos nestas duas variáveis latentes. A *Entropia* é máxima quando os pesos são uniformemente distribuídos. No presente caso, quando $w_1 = w_2 = w_3 = 1/3$.

4.6 Análise de custo computacional do estudo multiobjetivo

Após a avaliação mono-objetivo e a interpretação estatística dos efeitos no espaço codificado, torna-se necessário verificar a robustez das soluções selecionadas e quantificar, sob replicação, o impacto do método de busca no custo computacional. Nesta etapa, as configurações obtidas por cada estratégia são submetidas a um estudo de confirmação com múltiplas sementes, permitindo comparar tempos de otimização e de execução total e verificar se eventuais diferenças em métricas de qualidade são estatisticamente relevantes. A análise a seguir organiza-se, primeiro, pela dimensão de custo (tempo) e, em seguida, pela dimensão de desempenho preditivo, culminando em uma discussão integrada do compromisso inerente ao problema multiobjetivo.

A Tabela 5 apresenta estatísticas descritivas do custo computacional ao longo das 30 réplicas. Observa-se que o método proposto apresenta, de forma consistente, os menores valores médios tanto no tempo total de otimização quanto no tempo total de execução. Esse comportamento é coerente com o desenho do *pipeline*, no qual o delineamento experimental CCDFC atua como mecanismo de exploração dirigida do espaço de hiperparâmetros, reduzindo a necessidade de avaliações adicionais em comparação com estratégias de busca que dependem de amostragem aleatória ou de atualizações iterativas do processo de seleção.

Tabela 5 - Estatísticas descritivas do custo computacional por método (média \pm desvio padrão; B = 138; n = 30).

Método	Tempo de otimização (s)	Tempo total (s)	Tempo médio por <i>fold</i> (s)
DOE + NBI (proposto)	157,86 \pm 6,32	157,94 \pm 6,33	0,078 \pm 0,026
Busca em Grade	208,16 \pm 14,77	209,15 \pm 14,73	0,991 \pm 0,702
Busca Aleatória	174,34 \pm 11,59	174,71 \pm 11,57	0,371 \pm 0,301
Otimização Bayesiana	553,14 \pm 53,93	554,17 \pm 54,18	1,036 \pm 0,776
TPE (Hyperopt)	192,90 \pm 38,87	193,30 \pm 39,15	0,401 \pm 0,363

Como a *baseline default* não constitui um método de otimização sob orçamento B, seus resultados são apresentados apenas como referência descritiva e não foram incluídos nas análises inferenciais (ANOVA/MANOVA e pós-testes), que permanecem restritas aos métodos comparados sob orçamento justo.

Para referência, o XGBoost *default* (sem *tuning*) apresentou tempo médio por *fold* de 0,16s (0,157 \pm 0,001, n = 30), mensurado sob o mesmo protocolo. Como não há etapa de busca, não se aplica ‘tempo de otimização’, e o ‘tempo total’ refere-se a uma única avaliação (B=1).

As Figuras 26, 27 e 28 complementam a análise, evidenciando a distribuição do tempo de otimização, do tempo total de execução e do tempo médio por *fold*, respectivamente. Em todos os gráficos, o método proposto apresenta mediana inferior e menor dispersão relativa, indicando robustez ao variar a semente da réplica.

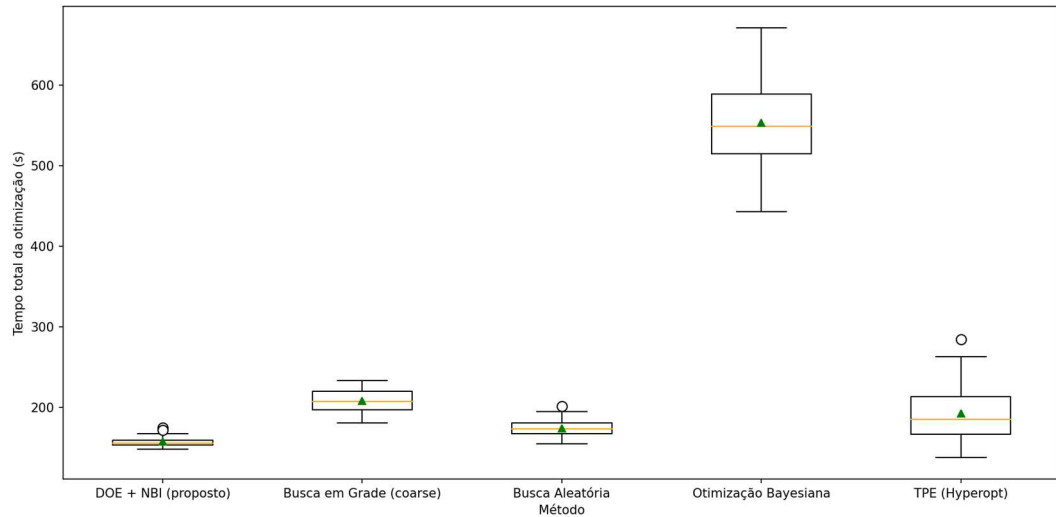


Figura 26 - *Boxplot* do tempo total da otimização por método (n = 30).

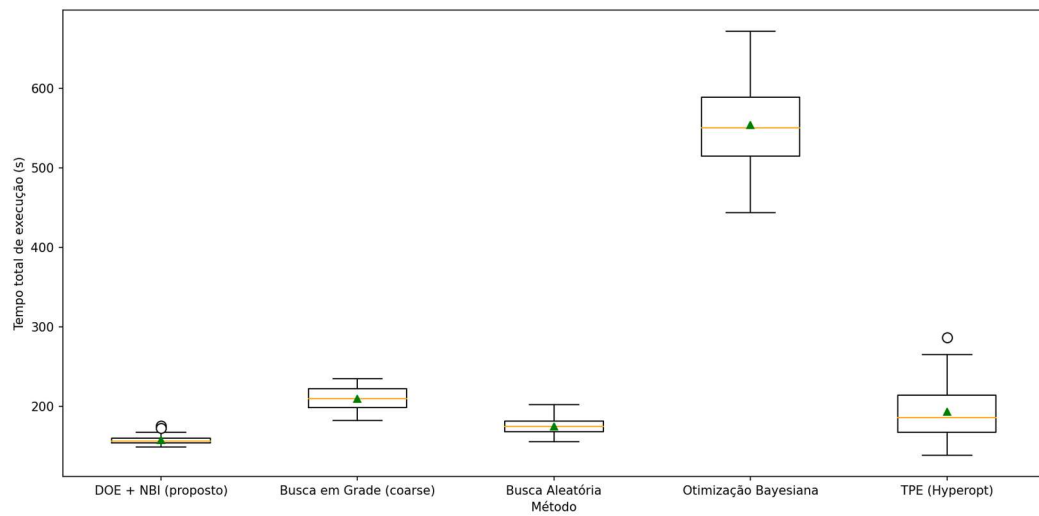


Figura 27 - *Boxplot* do tempo total de execução por método (n = 30).

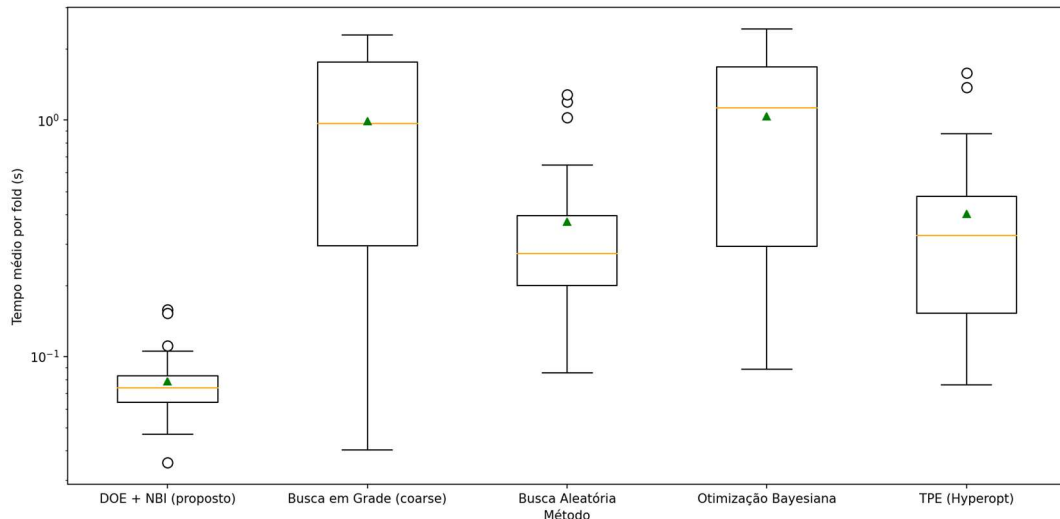


Figura 28 - *Boxplot* do tempo médio por fold por método (escala logarítmica; n = 30).

4.6.1 Análise de variância para tempos de execução

Para verificar se as diferenças observadas nos *boxplots* são estatisticamente significativas, aplicou-se análise de variância de um fator (ANOVA) considerando o método como fator e o tempo como variável resposta (Figuras 29 e 30). Como os tempos apresentaram assimetria e variâncias desiguais entre grupos, adotou-se transformação logarítmica em base 10 antes da aplicação do teste. Em ambos os casos (tempo total de otimização e tempo total de execução), rejeitou-se a hipótese nula de igualdade de médias com nível de significância de 5%. As comparações múltiplas pelo método de Tukey indicaram que o método proposto apresenta tempo médio estatisticamente inferior ao de todos os métodos de referência.

```

ANOVA de um fator: Tempo total da otimização (s)

Transformação: log10

Fonte          GL      SQ Aj.      QM Aj.      F      p-valor
Método         4       5.824268    1.456067    655.82  9.603e-92
Erro          145     0.321934    0.002220
Total         149     6.146202

Comparações múltiplas de Tukey (95%) em relação ao método DoE + NBI
(proposto)

Comparação          Dif
IC95%      p-aj.
Otimização Bayesiana vs DoE + NBI (proposto)  0.5429
[0.5093; 0.5765] 0.000e+00

Busca em Grade vs DoE + NBI (proposto)          0.1194
[0.0858; 0.1530] 0.000e+00

TPE (Hyperopt) vs DoE + NBI (proposto)          0.0792
[0.0456; 0.1128] 0.000e+00

Busca Aleatória vs DoE + NBI (proposto)          0.0425
[0.0089; 0.0761] 5.600e-03

```

Figura 29 - ANOVA ($\log_{10}T$) e comparações de Tukey para o tempo total da otimização.

ANOVA de um fator: Tempo total de execução (s)					
Transformação: log10					
Fonte	GL	SQ Aj.	QM Aj.	F	p-valor
Método	4	5.825970	1.456492	652.47	1.364e-91
Erro	145	0.323680	0.002232		
Total	149	6.149650			
Comparações múltiplas de Tukey (95%) em relação ao método DoE + NBI (proposto)					
Comparação					Dif
IC95% p-aj.					
Otimização Bayesiana vs DoE + NBI (proposto)					0.5435
[0.5098; 0.5772]					0.000e+00
Busca em Grade vs DoE + NBI (proposto)				0.1213	
[0.0876; 0.1550]					0.000e+00
TPE (Hyperopt) vs DoE + NBI (proposto)					0.0798
[0.0461; 0.1135]					0.000e+00
Busca Aleatória vs DoE + NBI (proposto)					0.0432
[0.0095; 0.0769]					4.700e-03

Figura 30 - ANOVA ($\log_{10}T$) e comparações de Tukey para o tempo total de execução.

4.7 Análise de qualidade preditiva do estudo multiobjetivo

A Tabela 6 apresenta estatísticas descritivas das métricas de qualidade preditiva ao longo das 30 réplicas. Em todos os métodos, observou-se desempenho elevado e baixa variabilidade entre réplicas, o que sugere que, no cenário controlado adotado, a tarefa apresenta separabilidade suficiente para que modelos de gradiente impulsionado atinjam valores altos mesmo com ajustes não exaustivos. Adicionalmente, o pré-processamento com balanceamento por subamostragem contribui para reduzir viés de classe, o que tende a aproximar as métricas de acurácia e especificidade.

Tabela 6 - Estatísticas descritivas de qualidade por método (média \pm desvio padrão; n = 30).

Método	Acurácia	Precisão	Recall	Especificidade
DOE + NBI (proposto)	0,9437 \pm 0,0026	0,9604 \pm 0,0042	0,9259 \pm 0,0061	0,9615 \pm 0,0044
Busca em Grade	0,9451 \pm 0,0013	0,9617 \pm 0,0045	0,9274 \pm 0,0049	0,9628 \pm 0,0047
Busca Aleatória	0,9459 \pm 0,0016	0,9633 \pm 0,0032	0,9274 \pm 0,0049	0,9644 \pm 0,0033
Otimização Bayesiana	0,9468 \pm 0,0014	0,9627 \pm 0,0036	0,9299 \pm 0,0047	0,9638 \pm 0,0038
TPE (Hyperopt)	0,9464 \pm 0,0012	0,9647 \pm 0,0042	0,9269 \pm 0,0045	0,9659 \pm 0,0043
XGBoost <i>default</i> (<i>baseline</i>)	0,9343 \pm 0,0026	0,9490 \pm 0,0044	0,9182 \pm 0,0044	0,9503 \pm 0,0047

A Figura 31 apresenta os *boxplots* de acurácia média por método. Embora existam diferenças estatisticamente detectáveis entre as distribuições, as variações ocorrem na terceira

casa decimal, indicando que os ganhos de qualidade são marginais em termos práticos para este conjunto de dados. Nesse contexto, torna-se relevante avaliar o custo associado a tais ganhos.

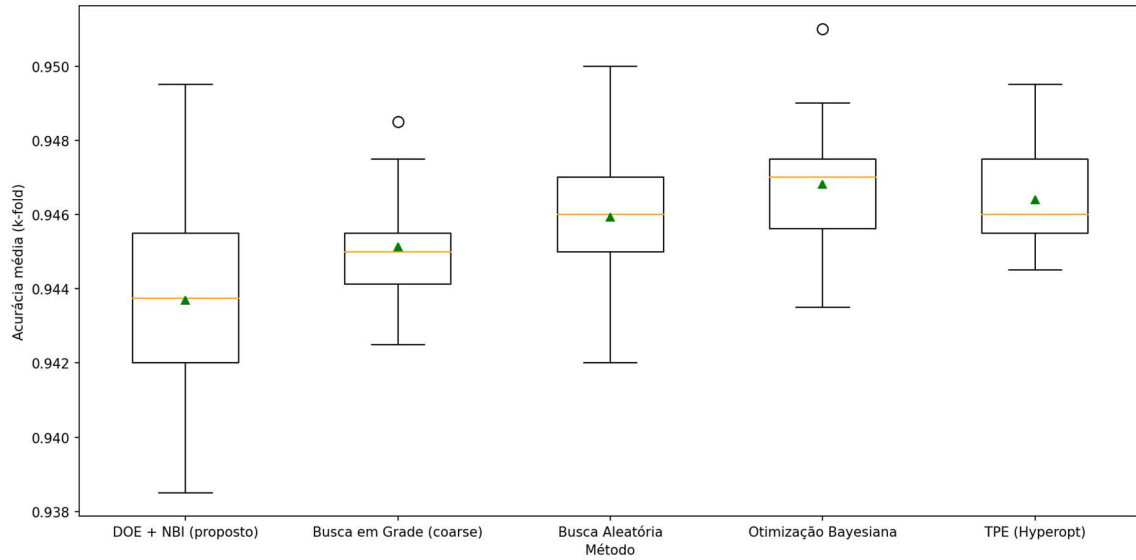


Figura 31 - *Boxplot* da acurácia média por método (n = 30).

A Figura 32 sintetiza a relação entre custo marginal e qualidade, representando o índice de qualidade em função do tempo médio por *fold*. A elipse de confiança foi traçada em torno dos pontos do método proposto para destacar a concentração das réplicas em uma região de menor custo. Observa-se que o método proposto forma um agrupamento à esquerda do gráfico, caracterizando menor tempo médio por *fold*, enquanto mantém valores de qualidade próximos aos demais métodos.

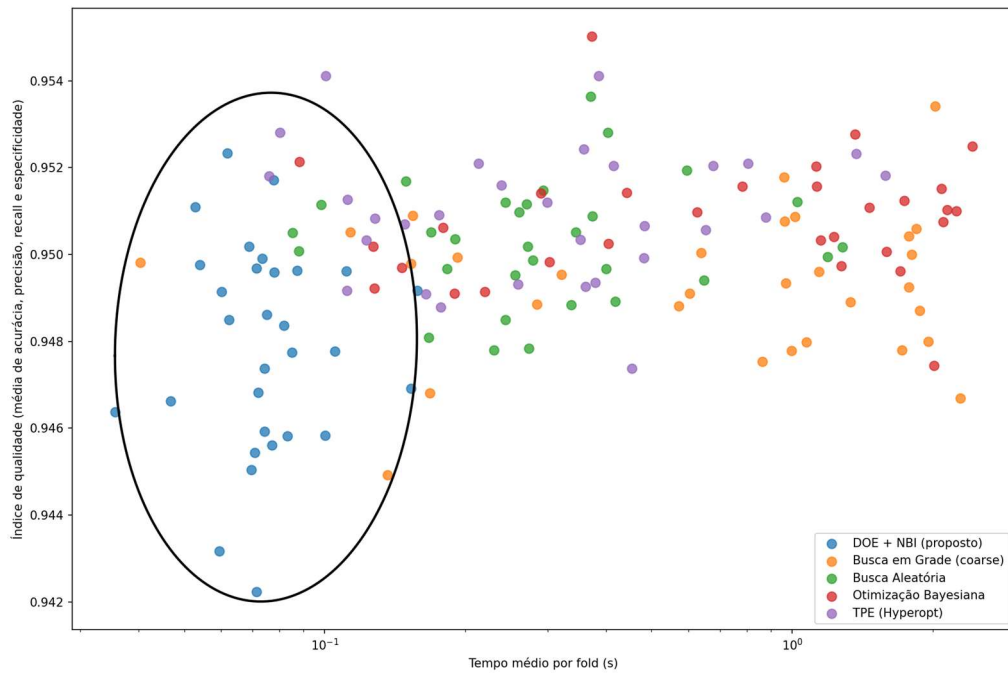


Figura 32 - Dispersão entre índice de qualidade e tempo médio por *fold*

4.7.1 Análise de variância para métricas de qualidade

Com o objetivo de avaliar se as diferenças observadas nas métricas de qualidade são estatisticamente significativas, aplicou-se ANOVA de um fator para o índice de qualidade e para a acurácia (Figura 33). Em ambos os casos, rejeitou-se a hipótese nula de igualdade de médias ($p < 0,05$). No entanto, a diferença absoluta entre a melhor média e a média do método proposto foi de aproximadamente 0,0031 no índice de qualidade, o que corresponde a cerca de 0,33% em termos relativos. Assim, apesar de estatisticamente detectáveis, os ganhos de qualidade observados nos métodos de referência apresentam pequena relevância prática frente às diferenças substanciais de custo computacional.

ANOVA de um fator: Índice de qualidade					
Transformação: nenhuma					
Fonte	GL	SQ Aj.	QM Aj.	F	p-valor
Método	4	0.000198	0.000049	16.77	2.545e-11
Erro	145	0.000427	0.000003		
Total	149	0.000625			
Comparações múltiplas de Tukey (95%) em relação ao método DoE + NBI (proposto)					
Comparação					Dif
IC95%	p-aj.				
TPE (Hyperopt) vs DoE + NBI (proposto)					0.0031
[0.0019; 0.0043] 0.000e+00					
Otimização Bayesiana vs DoE + NBI (proposto)					0.0029
[0.0017; 0.0041] 0.000e+00					
Busca Aleatória vs DoE + NBI (proposto)					0.0024
[0.0012; 0.0036] 0.000e+00					
Busca em Grade vs DoE + NBI (proposto)					0.0014
[0.0002; 0.0026] 1.470e-02					

Figura 33 - Saída da ANOVA e comparações de Tukey para o índice de qualidade.

4.8 Análises multivariadas de robustez e estabilidade

Esta subseção avalia a robustez; sendo aqui definida como sensibilidade a variações entre réplicas, e a estabilidade; sendo definida por sua vez como consistência do compromisso qualidade-custo do método proposto “DOE+NBI” em comparação com os *benchmarks* a partir de 30 réplicas por método.

4.8.1 Vetor multivariado e pré-processamento

Para cada réplica, considera-se o vetor multivariado:

$$\mathbf{y} = \left(\text{Accuracy}_{\text{Mean}}, \text{Precision}_{\text{Mean}}, \text{Recall}_{\text{Mean}}, \text{Specificity}_{\text{Mean}}, \log(\text{Time}_{\text{MeanFold}}) \right) \quad (4.22)$$

Com $p = 5$ variáveis. O termo de tempo é transformado por logaritmo natural para reduzir assimetria e tornar a escala mais comparável às métricas de desempenho (limitadas em $[0,1]$).

4.8.2 Robustez via distância de Mahalanobis (detecção de pontos atípicos)

A robustez multivariada foi analisada pela distância de Mahalanobis em relação ao “perfil típico” do método proposto. Seja $\hat{\mu}$ e $\hat{\Sigma}$ a média e a matriz de covariância amostrais

estimadas apenas com as 30 réplicas do método proposto “DoE + NBI”. Para cada observação \mathbf{y}_i , calcula-se:

$$d_i^2 = (\mathbf{y}_i - \hat{\boldsymbol{\mu}})^\top \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}) \quad (4.23)$$

Adota-se o limiar $\chi_{0,975}^2(p = 5) = 12,833$ para sinalizar observações com distância elevada (interpretação operacional de *outliers* no espaço de \mathbf{y}). A inversão de $\hat{\boldsymbol{\Sigma}}$ é feita numericamente (pseudoinversa), garantindo estabilidade mesmo sob correlação forte entre variáveis. A Tabela 7 a seguir resume a distribuição de d^2 por método.

Tabela 7 - Distâncias de Mahalanobis ao perfil do método proposto e contagem de *outliers* ($d^2 > \chi_{0,975}^2$).

Método	Média d^2	DP d^2	P5 d^2	P50 d^2	P95 d^2	<i>Outliers</i>
DOE + NBI (proposto)	3,867	2,951	1,156	3,041	8,393	1
Busca em Grade	7,067	45,545	7,172	75,551	131,324	23
Busca Aleatória	30,123	24,598	4,240	24,404	87,566	22
Otimização Bayesiana	71,693	48,051	5,496	87,914	136,850	25
TPE (Hyperopt)	33,742	28,911	6,677	27,338	92,362	20

Observa-se que o método proposto DOE + NBI concentra-se em uma região compacta do espaço multivariado (baixa média e dispersão de d^2), com apenas 1 ocorrência acima do limiar. Por outro lado, as réplicas dos *benchmarks* apresentam grande afastamento em relação à região característica do método proposto. Como as métricas de qualidade se mantêm próximas entre métodos (Tabela 3), essa separação é compatível com diferenças estruturais no componente temporal (redução pronunciada de $\text{Time}_{\text{MeanFold}}$ em DOE + NBI), e não apenas com flutuações aleatórias.

4.8.3 Normalidade multivariada (teste de Mardia) e implicações

A adequação da hipótese de normalidade multivariada foi investigada pelo teste de Mardia, separando assimetria ($b_{1,p}$) e curtose ($b_{2,p}$). A Tabela 8 reporta os principais indicadores por método.

Tabela 8 - Teste de Mardia (assimetria e curtose) por método.

Método	b1,p (assimetria)	p (assim.)	b2,p (curtose)	z (curtose)	p (curtose)	n
DOE + NBI (proposto)	5,854	0,7405	23,366	-3,808	0,0001	30
Busca em Grade	3,446	0,9949	20,806	-4,646	3,38e-06	30
Busca Aleatória	6,054	0,6959	23,673	-3,708	0,0002	30
Otimização Bayesiana	3,561	0,9930	21,810	-4,318	1.58e-05	30
(TPE) Hyperopt	4,606	0,9399	22,280	-4,164	3,13e-05	30

Em todas as abordagens, a componente de curtose rejeita a normalidade multivariada (p-valores muito pequenos), sugerindo distribuição mais concentrada (menor dispersão) do que o esperado sob normalidade, um comportamento plausível quando métricas de desempenho são limitadas e agregadas por média de *folds*. Essas evidências motivam duas escolhas nesta subseção: (i) uso de correlação por postos (Spearman) na análise de dependências internas do vetor \mathbf{y} ; e (ii) preferência por estatísticas multivariadas relativamente robustas (por exemplo, Traço de Pillai na MANOVA), acompanhadas de medidas de efeito e análises descritivas.

4.8.4 Estrutura de correlação e correção de múltiplas comparações (Holm-Bonferroni)

Para investigar dependências internas do vetor \mathbf{y} sob o método proposto DOE + NBI, foram calculadas correlações de Spearman entre todos os pares de variáveis ($\binom{5}{2} = 10$ testes). A escolha por Spearman é compatível com os resultados de normalidade (Seção 4.8.3) e com possíveis relações monotônicas não-lineares. Aplicou-se a correção de Holm-Bonferroni sobre os 10 p-valores. Os resultados são apresentados na Tabela 9.

Destacam-se duas relações robustas após correção: (i) Precision e Specificity apresentam associação muito forte e positiva, e (ii) Accuracy e Recall apresentam correlação moderada. Em contrapartida, $\log(\text{Time}_{\text{MeanFold}})$ não mostrou associação estatisticamente significativa com as métricas de qualidade nas réplicas do método proposto DOE + NBI, sugerindo que, no regime explorado, o método proposto consegue controlar o tempo sem impor uma *trade-off* monotônico forte com o desempenho preditivo médio.

Tabela 9 - Correlações (Spearman) em DoE + NBI (proposto) e p-valores com ajuste de Holm.

Variáveis	P(Spearman)	P bruto	P(Holm)
Acurácia Média × Precisão Média	0,313	0,0922	0,553
Acurácia Média × Revocação Média	0,647	0,0001	0,001
Acurácia Média × Especificidade Média	0,269	0,1507	0,753
Acurácia Média × Logaritmo Natural de Tempo Médio por Fold	0,021	0,9106	1,000
Precisão Média × Revocação Média	-0,439	0,0153	0,107
Precisão Média × Especificidade Média	0,994	3,26e ⁻²⁸	3,26e ⁻²⁷
Precisão Média × Logaritmo Natural de Tempo Médio por <i>Fold</i>	-0,100	0,5995	1,000
Revocação Média × Especificidade Média	-0,486	0,0065	0,052
Revocação Média × Logaritmo Natural de Tempo Médio por <i>Fold</i>	-0,003	0,9869	1,000
Especificidade Média × Ln de Tempo Médio por <i>Fold</i>	-0,073	0,7033	1,000

4.8.5 Comparação multivariada entre métodos (MANOVA) e pós-testes

Para comparar os cinco métodos considerando simultaneamente qualidade e custo (via y), foi conduzida uma MANOVA com fator “Método”. O efeito global foi significativo (Traço de Pillai = 0.7457, $F(16,580) = 8.307$, $p < 10^{-12}$), indicando que ao menos um método difere multivariadamente dos demais.

Tabela 10 - Pós-testes (Hotelling T^2): DoE + NBI versus *benchmarks* (com Holm).

Comparação entre métodos	T^2	F	df ₁	df ₂	p bruto	p (Holm)	D (Mahalanobis)
DoE + NBI (proposto) vs. Busca Aleatória	131,936	24,567	5	54	8,24e ⁻¹³	2,47e ⁻¹²	2,966
DoE + NBI (proposto) vs. Otimização Bayesiana	150,817	28,083	5	54	6,74e ⁻¹⁴	2,7e ⁻¹³	3,171
DoE + NBI (proposto) vs. TPE (Hyperopt)	110,994	20,668	5	54	1,78e ⁻¹¹	1,78e ⁻¹¹	2,72
DoE + NBI (proposto) vs. Busca em Grade	118,115	21,994	5	54	6,03e ⁻¹²	1,21e ⁻¹¹	2,806

Como pós-teste direcionado, comparou-se o método proposto DoE + NBI contra cada *benchmark* via estatística T^2 de Hotelling (duas amostras, $p = 5$), com correção de Holm sobre

as 4 comparações. Além do p-valor, reporta-se o efeito como a distância de Mahalanobis entre as médias (D).

Os resultados confirmam diferenças multivariadas consistentes entre método proposto DoE + NBI e todos os *benchmarks* (p ajustado muito pequeno e efeitos grandes). Em termos práticos, a separação multivariada é coerente com a redução substancial do tempo médio por *fold* no método proposto DoE + NBI (Tabela 6), sem deterioração proporcional das métricas de qualidade.

4.8.6 Métricas multiobjetivo (IGD, *Spread* e entropia de espaçamento) a partir do conjunto NBI

A qualidade geométrica dos conjuntos não-dominados obtidos na etapa multiobjetivo do método proposto “DoE + NBI” foi avaliada com base nos resultados experimentais gerados entre replicações. Para operacionalizar um espaço biobjetivo coerente com “custo/tempo versus qualidade”, adotou-se:

1. Objetivo 1 (custo/tempo): minimizar $\log(\text{Time}_{\text{MeanFold}})$;
2. Objetivo 2 (qualidade): minimizar $1 - Q$, em que Q é a média de (Accuracy, Precision, Recall, Specificity).

As métricas calculadas foram: (i) IGD - distância geracional inversa (*reverse generational distance*), menor é melhor; (ii) *Spread* Δ , uniformidade/cobertura (menor é melhor); (iii) Entropia normalizada de espaçamento H , regularidade do espaçamento (maior é melhor).

Como não há frente de Pareto visualmente analítica, a “frente de referência” foi construída como o conjunto não-dominado global obtido pela união de todas as réplicas (aproximação padrão em cenários experimentais). Para evitar dominância de escala, as distâncias foram calculadas após normalização *min-max* das duas funções objetivo.

Os valores indicam que as fronteiras não-dominadas por réplica possuem, em média, cerca de 4-5 soluções (com variabilidade moderada). O IGD médio reflete que, em várias réplicas, o conjunto não-dominado cobre a frente de referência apenas parcialmente - o que é esperado quando o orçamento por réplica é limitado. Por sua vez, quando há pelo menos 3 pontos não-dominados, a entropia H tende a valores elevados, sugerindo espaçamento relativamente regular nas melhores réplicas.

Tabela 11 - Métricas multiobjetivo (média, desvio padrão e quantis por réplica; quando aplicável).

Métrica	Média	DP	P5	P50	P95	n válido
IGD	0,18	0,092	0,081	0,163	0.332	30
Δ (Spread)	0,854	0,114	0,701	0,866	1.018	29
H (Entropia)	0,776	0,205	0,373	0,834	0.971	27
	PF	(nº pontos ND)	4,567	1,755	2	4

4.8.7 Interpretação do compromisso qualidade-custo e sensibilidade dos hiperparâmetros

(β_1/β_2 , superfícies RSM e síntese)

A interpretação do compromisso qualidade-custo do método proposto pode ser feita em dois níveis: (i) o papel dos pesos/ênfases (β_1 e β_2) na escolha de hiperparâmetros, e (ii) a sensibilidade local estimada pelas superfícies RSM na etapa multiobjetivo.

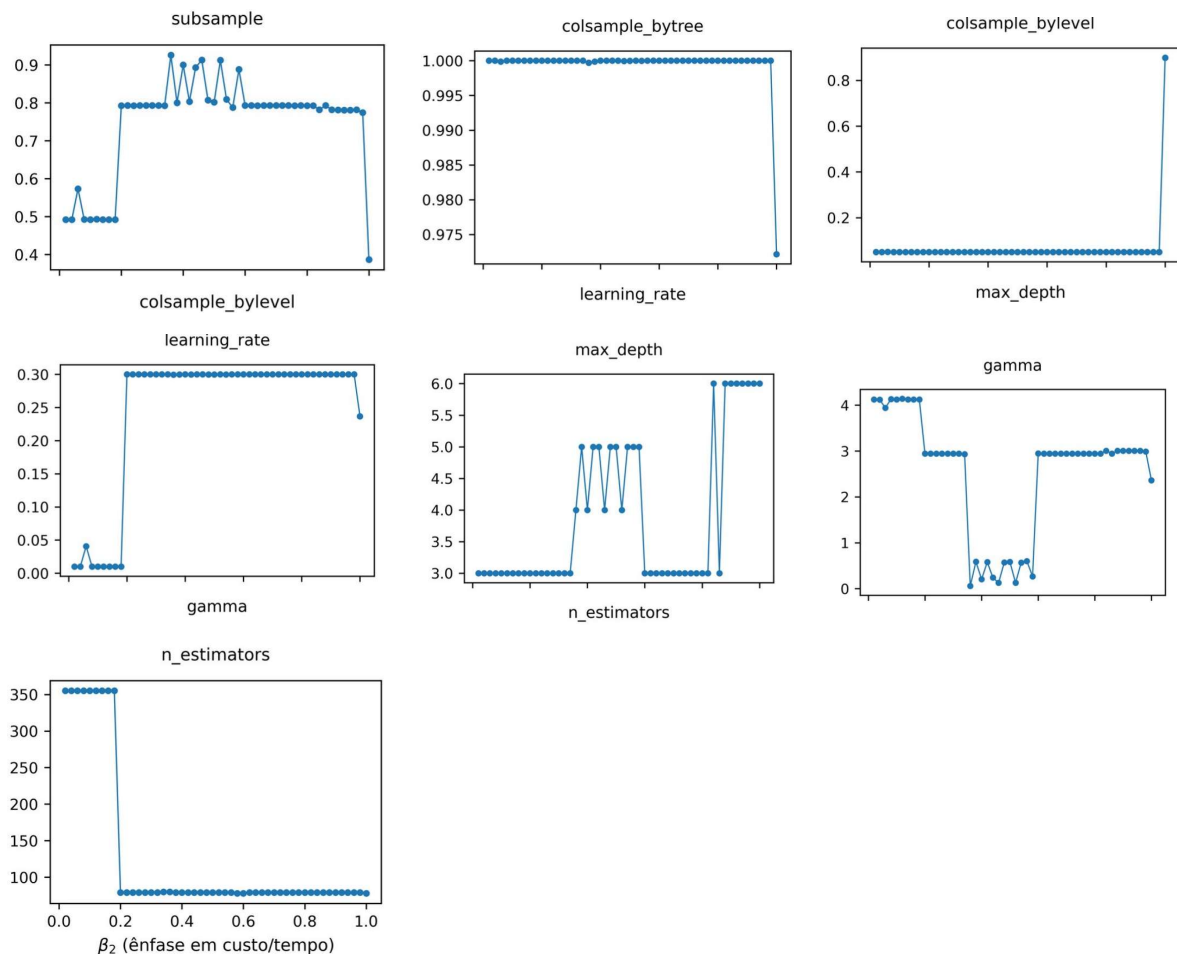


Figura 34 - Perfil dos hiperparâmetros ao longo de β_2 (ênfase em custo/tempo)

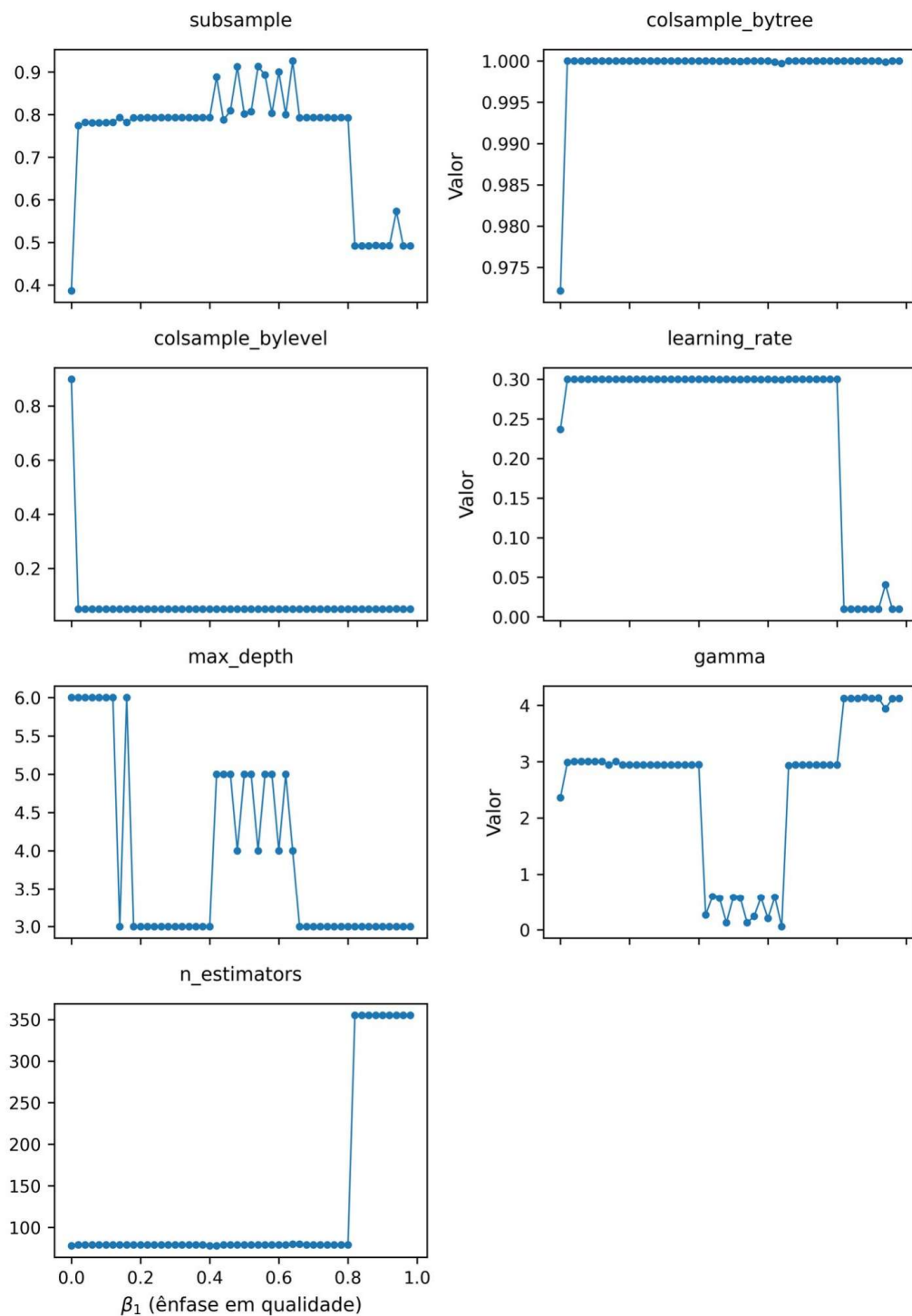


Figura 35 - Perfil dos hiperparâmetros ao longo de β_1 (ênfase em qualidade)

Nas Figuras 34 e 35, observa-se um mecanismo de controle interpretável: ao se aumentar a ênfase no custo/tempo (β_2), a tendência é deslocar a solução para configurações com menor complexidade computacional (por exemplo, uma redução no hiperparâmetro “n_estimators” e ajustes em parâmetros que afetam o esforço de treinamento). Ao aumentar a

ênfase em qualidade (β_1), ocorre o movimento inverso, privilegiando configurações de maior capacidade preditiva (por exemplo, mais estimadores e regimes de “learning_rate” compatíveis com maior refinamento). Esse comportamento é consistente com a lógica multiobjetivo ilustrada na Figura 36, em que o peso define a direção de busca ao longo da fronteira de Pareto.

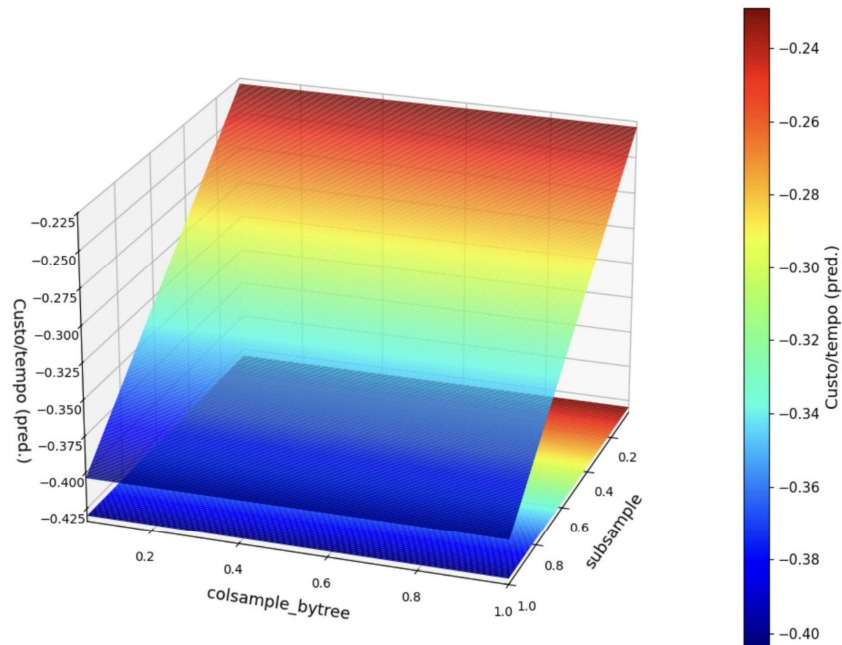


Figura 36 - Superfície RSM (custo/tempo): “subsample” × “colsample_bytree”

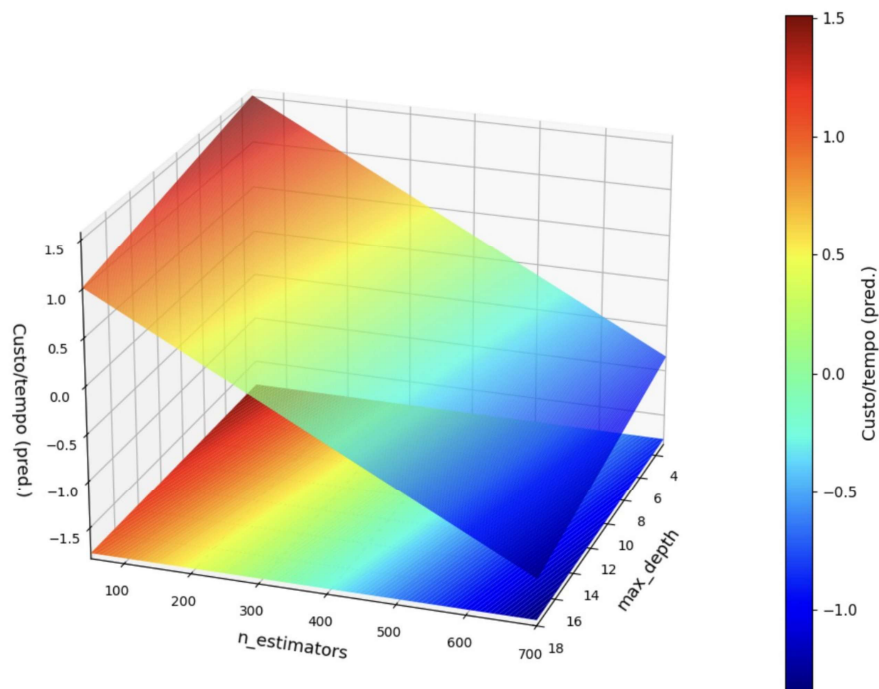


Figura 37 - Superfície RSM (custo/tempo): “n_estimators” × “max_depth”

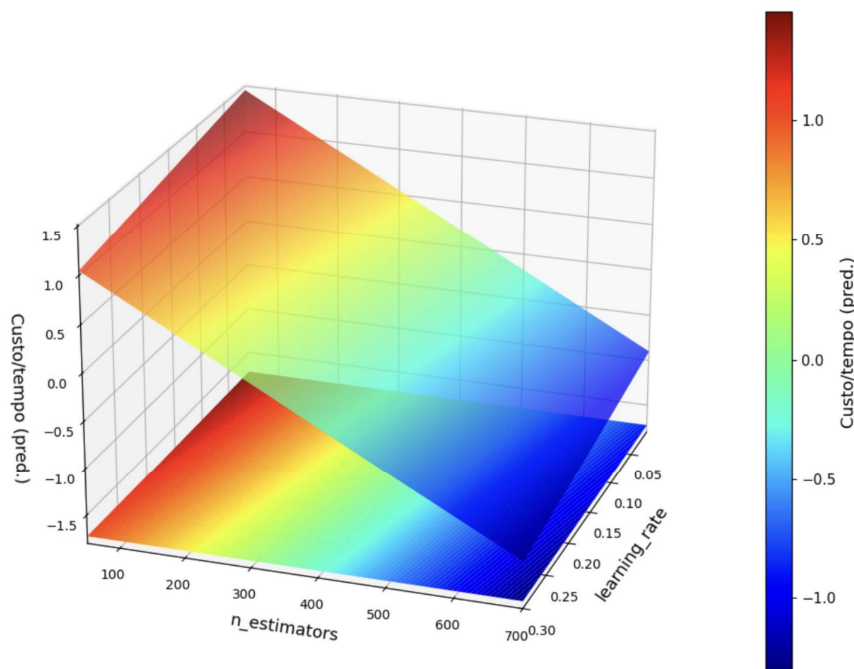


Figura 38 - Superfície RSM (custo/tempo): “n_estimators” × “learning_rate”

Nas Figuras 36 a 38, as superfícies associadas ao fator de custo/tempo exibem comportamento predominantemente quase linear (planos inclinados) e com interações fracas na região amostrada, o que é favorável à estabilidade: variações simultâneas em dois hiperparâmetros tendem a produzir alterações previsíveis. Em particular, há evidência visual de que “n_estimators” é um fator dominante para o custo/tempo (Figuras 37 e 38), enquanto “subsample” apresenta gradiente relevante na dupla (“subsample”, “colsample_bytree”) (Figura 36), sugerindo um “controle fino” adicional sobre custo ao variar a fração de amostragem. Para o fator de qualidade (Figuras 39 e 40), nota-se igualmente uma geometria suave, compatível com variações monotônicas na região explorada: “learning_rate” tende a induzir mudanças mais pronunciadas no nível predito de qualidade (Figura 39), enquanto “subsample” e “colsample_bytree” contribuem de forma mais regular (Figura 40). A ausência de picos locais abruptos nas superfícies reforça a leitura de estabilidade: a otimização opera em uma paisagem relativamente bem condicionada, o que reduz o risco de mudanças erráticas de hiperparâmetros entre réplicas para atingir níveis semelhantes de qualidade.

Em conjunto, as evidências numéricas (Mahalanobis, correlações e MANOVA) e gráficas (perfis em β_1/β_2 e RSM) suportam a interpretação de que o método proposto “DoE + NBI” produz soluções consistentes no espaço desempenho-custo: (i) concentra as réplicas em uma região multivariada compacta; (ii) preserva relações internas esperadas entre métricas de

desempenho; e (iii) oferece um mecanismo de ajuste de compromisso (via β) que se traduz em mudanças estruturadas de hiperparâmetros.

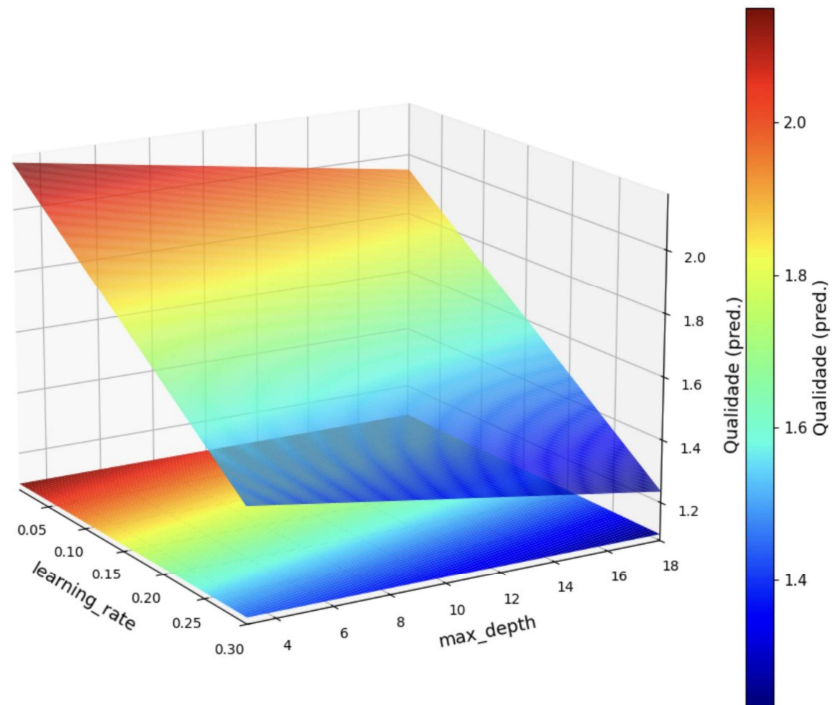


Figura 39 - Superfície RSM (qualidade): “learning_rate” × “max_depth”

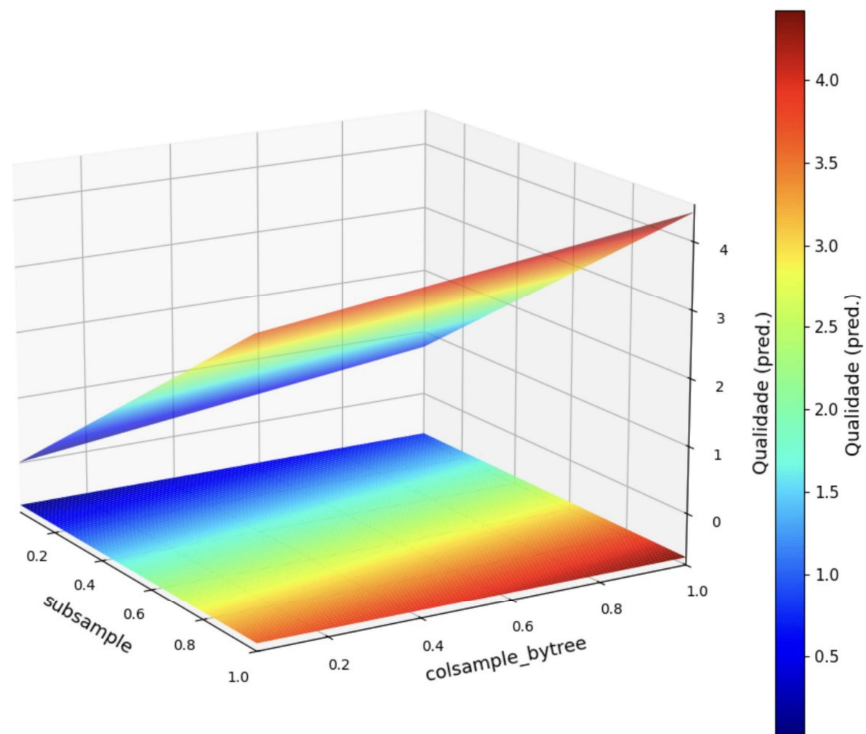


Figura 40 - Superfície RSM (qualidade): “subsample” × ”colsample_bytree”

4.9 Validação externa em bases adicionais

O Capítulo 4 analisou em profundidade o comportamento do pipeline proposto (DoE + RSM + NBI) no benchmark canônico MAGIC Gamma Telescope. A escolha dessa base como estudo principal favorece a discussão do método em um cenário mais controlado, permitindo isolar com maior clareza o efeito da metodologia frente a variáveis externas inerentes aos dados. Como validação externa, esta seção avalia a estabilidade do mesmo pipeline em duas bases adicionais, com perfis contrastantes: (i) Espécies de pimentas: “*pepper_species*”, com classes balanceadas e elevada separabilidade; e (ii) Inadimplência em Crédito: “*credit_card_default*”, base real, ruidosa e fortemente desbalanceada. Em ambos os casos, o objetivo é verificar se o padrão observado no benchmark principal (*trade-off* custo-qualidade com baixa variabilidade entre réplicas) se reproduz fora do cenário controlado.

4.9.1 Protocolo experimental preservado

Para garantir comparabilidade com os resultados das Seções 4.1-4.8, preservou-se integralmente o protocolo experimental. Mantiveram-se: (a) orçamento computacional fixo por método ($B = 138$ avaliações); (b) 30 réplicas independentes, todas derivadas de uma *seed* base 42; (c) validação cruzada estratificada; (d) métricas de qualidade (acurácia, precisão, revocação e especificidade) e métricas de custo temporal (tempo de otimização, tempo total e tempo médio por fold). As comparações incluem o método proposto (DoE + NBI) e quatro estratégias de referência: busca em grade, busca aleatória, otimização bayesiana e TPE (Hyperopt).

4.9.2 Base de Espécies de Pimentas

O presente estudo de caso encontra respaldo conceitual e empírico na tese de Freire (2019), intitulada “Diversidade Genética de *Capsicum chinense* C. *frutescens*”, desenvolvida no Programa de Pós-Graduação em Fitotecnia da Universidade Federal de Viçosa. O conjunto de dados utilizado neste experimento está diretamente relacionado às análises conduzidas naquela pesquisa, a qual investigou a estrutura genética e a variabilidade entre acessos de *Capsicum*.

Destaca-se que esta dissertação mantém vínculo acadêmico com o projeto de Pós-Doutorado da referida pesquisadora junto à Universidade Federal de Itajubá (UNIFEI), financiado pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), por meio do Projeto BPD-01045-22. Tal colaboração reforça a integração entre métodos estatísticos multivariados e aplicações em dados agronômicos reais.

Adicionalmente, ressalta-se que, embora o problema tenha sido tratado sob uma estrutura binária agregada, o conjunto original contempla 17 categorias específicas organizadas em duas classes principais, o que representa um cenário potencial de maior complexidade classificatória. Essa característica sugere a possibilidade de extensões futuras do método proposto para cenários multiclases mais granulares, ampliando ainda mais sua aplicabilidade.

A base “*pepper_species*” contém 200.000 instâncias (obtidas através de *data augmentation* de um conjunto original de 54 instâncias), 20 atributos contínuos (x_1, \dots, x_{20}) e duas classes balanceadas: $y = 0$ (*Chinense*) e $y = 1$ (*Frutescens*). As Tabelas 12 e 13 apresentam os resultados consolidados (média \pm desvio-padrão) ao longo das 30 réplicas.

Tabela 12 - Estatísticas descritivas de qualidade por método (média \pm desvio padrão; $n = 30$; $B = 138$) em “*pepper_species*”.

Método	Acurácia	Precisão	Recall	Especificidade
DOE + NBI (proposto)	0,99992 \pm 0,00003	0,99986 \pm 0,00005	0,99999 \pm 0,00001	0,99986 \pm 0,00005
Busca em Grade	0,99996 \pm 0,00001	0,99993 \pm 0,00001	1,00000 \pm 0,00000	0,99993 \pm 0,00001
Busca Aleatória	0,99997 \pm 0,00001	0,99993 \pm 0,00001	1,00000 \pm 0,00000	0,99993 \pm 0,00001
Otimização Bayesiana	0,99997 \pm 0,00000	0,99993 \pm 0,00001	1,00000 \pm 0,00000	0,99993 \pm 0,00001
TPE (Hyperopt)	0,99997 \pm 0,00000	0,99994 \pm 0,00001	1,00000 \pm 0,00000	0,99994 \pm 0,00001

Tabela 13 - Estatísticas descritivas do custo computacional por método (média \pm desvio padrão; $n = 30$; $B = 138$) em “*pepper_species*”.

Método	Tempo de otimização (s)	Tempo total (s)	Tempo médio por <i>fold</i> (s)
DOE + NBI (proposto)	316,66 \pm 8,26	316,87 \pm 8,29	0,211 \pm 0,054
Busca em Grade	423,69 \pm 27,47	424,30 \pm 27,48	0,603 \pm 0,195
Busca Aleatória	368,47 \pm 17,34	369,10 \pm 17,37	0,635 \pm 0,206
Otimização Bayesiana	610,16 \pm 21,72	610,84 \pm 21,66	0,675 \pm 0,353
TPE (Hyperopt)	403,42 \pm 54,09	404,08 \pm 54,25	0,656 \pm 0,210

Observa-se desempenho próximo ao teto em todas as abordagens, com métricas de qualidade acima de 0,9998. Nesse cenário, diferenças absolutas em qualidade são pequenas e

compatíveis com a elevada separabilidade da base. Por outro lado, o método proposto apresenta redução consistente de custo temporal: o tempo médio por fold cai para 0,211 s, contra aproximadamente 0,60-0,68 s nos benchmarks (redução entre 65% e 69%, dependendo do comparativo). Em termos de tempo de otimização, o ganho é mais evidente frente à otimização bayesiana (redução de aproximadamente 48%).

Tabela 14 - Robustez em “*pepper_species*”: distâncias de Mahalanobis ao perfil do método proposto e contagem de outliers (limiar $\chi^2_{(5; 0,975)} = 12,833$).

Método	Média d ²	DP d ²	P5 d ²	P50 d ²	P95 d ²	Outliers
DoE + NBI (proposto)	2,900	3,193	0,446	1,979	6,644	1
Busca em Grade	19,714	10,812	5,897	16,887	41,073	23
Busca Aleatória	21,473	11,319	4,330	20,830	37,167	25
Otimização Bayesiana	23,778	18,936	3,428	18,475	62,921	23
TPE (Hyperopt)	22,337	11,703	9,785	19,155	46,088	24

A robustez multivariada foi analisada com o mesmo vetor de resposta adotado na Seção 4.8: $y = (\text{Acurácia}, \text{Precisão}, \text{Revocação}, \text{Especificidade}, \ln(\text{Tempo médio por fold}))$. As distâncias de Mahalanobis foram calculadas em relação ao perfil típico do método proposto (média e covariância de suas 30 réplicas). Como na base canônica, o DoE + NBI concentra as menores distâncias (média $d^2 = 2,900$) e apresenta apenas 1 outlier no limiar χ^2 , indicando baixa variabilidade entre réplicas. Os métodos de referência exibem distâncias muito mais elevadas e alta contagem de outliers, refletindo sobretudo discrepâncias no componente temporal do vetor resposta.

Como verificação de pressupostos, o teste de Mardia aplicado ao vetor y no método proposto não rejeita assimetria ($b_{1,p} = 6,212$; $p = 0,6587$), porém rejeita normalidade pela curtose ($z = -5,477$; $p < 10^{-7}$). Esse resultado é consistente com o observado no *benchmark* principal, reforçando o uso de medidas robustas e a ênfase em efeitos de magnitude (e não apenas p-valores) na interpretação.

Tabela 15 - Correlações (Spearman) no método proposto em “*pepper_species*” e p-valores com ajuste de Holm.

Variáveis	ρ (Spearman)	P bruto	P(Holm)
Acurácia Média × Precisão Média	0,955	2,39e ⁻¹⁶	2,15e ⁻¹⁵
Acurácia Média × Revocação Média	0,503	0,0046	0,0323
Acurácia Média × Especificidade Média	0,935	3,96e ⁻¹⁴	3,16e ⁻¹³
Acurácia Média × Ln de Tempo médio por fold	0,484	0,0067	0,0402
Precisão Média × Revocação Média	0,260	0,1655	0,3310
Precisão Média × Especificidade Média	0,993	1,35e ⁻²⁷	1,35e ⁻²⁶
Precisão Média × Ln de Tempo médio por fold	0,417	0,0219	0,1020
Revocação Média × Especificidade Média	0,196	0,2994	0,3310
Revocação Média × Ln de Tempo médio por fold	0,342	0,0646	0,1938
Especificidade Média × Ln de Tempo médio por fold	0,421	0,0204	0,1020

As correlações de Spearman evidenciam dependência estrutural entre métricas: acurácia, precisão e especificidade são fortemente correlacionadas, enquanto a associação com $\ln(\text{Tempo})$ é moderada. Esse padrão é compatível com a motivação do tratamento multivariado adotado no Capítulo 4, no qual métricas não são tratadas como independentes.

4.9.3 Base de Inadimplência de Crédito

A base “*credit_card_default*” corresponde ao conjunto “Default of Credit Card Clients”, originalmente proposto por Yeh e Lien (2009), amplamente utilizado na literatura como benchmark para problemas de reputação de crédito (*credit scoring*). O conjunto contém 30.000 instâncias e 23 variáveis explicativas, sendo a variável-alvo definida como inadimplência no pagamento do cartão de crédito no mês subsequente ($y = 1$ para inadimplência; $y = 0$ caso contrário). Trata-se de um cenário realista, caracterizado por ruído e forte desbalanceamento entre classes, o que impõe maior complexidade ao problema classificatório. As Tabelas 16 e 17 resumem os resultados das 30 réplicas.

Tabela 16 - Estatísticas descritivas de qualidade por método (média ± desvio padrão; n = 30; B = 138) em “*credit_card_default*”.

Método	Acurácia	Precisão	Recall	Especificidade
DOE + NBI (proposto)	0,81810 ± 0,00364	0,67363 ± 0,00472	0,34473 ± 0,02820	0,95255 ± 0,00348
Busca em Grade	0,82211 ± 0,00031	0,68440 ± 0,00300	0,36359 ± 0,00456	0,95235 ± 0,00120
Busca Aleatória	0,82185 ± 0,00035	0,68067 ± 0,00277	0,36682 ± 0,00381	0,95108 ± 0,00105
Otimização Bayesiana	0,82219 ± 0,00028	0,68304 ± 0,00403	0,36639 ± 0,00595	0,95166 ± 0,00166
TPE (Hyperopt)	0,82220 ± 0,00031	0,68290 ± 0,00314	0,36650 ± 0,00412	0,95163 ± 0,00120

Tabela 17 - Estatísticas descritivas do custo computacional por método (média ± desvio padrão; n = 30; B = 138) em “*credit_card_default*”.

Método	Tempo de otimização (s)	Tempo total (s)	Tempo médio por <i>fold</i> (s)
DOE + NBI (proposto)	269,71 ± 6,95	269,83 ± 6,97	0,114 ± 0,038
Busca em Grade	386,01 ± 33,41	386,76 ± 33,58	0,755 ± 0,495
Busca Aleatória	385,74 ± 27,53	386,17 ± 27,52	0,431 ± 0,315
Otimização Bayesiana	547,87 ± 47,37	548,35 ± 47,42	0,482 ± 0,441
TPE (Hyperopt)	318,42 ± 46,97	318,84 ± 47,09	0,419 ± 0,283

Diferentemente da base “*pepper_species*”, a qualidade preditiva não apresenta efeito de teto, e a revocação da classe positiva permanece moderada ($\approx 0,36$) em todos os benchmarks, refletindo o caráter desbalanceado do problema. O método proposto mantém especificidade elevada ($\approx 0,953$), mas apresenta revocação média inferior (0,345) e maior variabilidade nessa métrica (DP = 0,028). Ainda assim, o ganho em custo é substancial: o tempo médio por *fold* é reduzido para 0,114 s, contra 0,42-0,76 s nos métodos de referência (redução de 73% a 85%). Em aplicações em que revocação seja requisito primário, destaca-se que o pipeline fornece um conjunto Pareto de soluções; logo, a escolha final pode ser direcionada a priorizar revocação com custo computacional maior, conforme a necessidade operacional.

Tabela 18 - Robustez em “*credit_card_default*”: distâncias de Mahalanobis ao perfil do método proposto e contagem de *outliers* (limiar $\chi^2_{(5; 0,975)} = 12,833$).

Método	Média d^2	DP d^2	P5 d^2	P50 d^2	P95 d^2	Outliers
DoE + NBI (proposto)	4,833	4,404	1,192	3,862	12,486	2
Busca em Grade	44,623	20,025	10,471	48,844	68,549	27
Busca Aleatória	23,500	15,350	9,113	17,835	49,894	20
Otimização Bayesiana	30,737	19,891	7,472	22,876	68,999	26
TPE (Hyperopt)	26,316	14,840	8,047	20,266	51,107	25

Mesmo em uma base mais ruidosa, observa-se estabilidade do método proposto: as distâncias de Mahalanobis permanecem concentradas (média $d^2 = 4,833$) e a contagem de outliers no limiar χ^2 é baixa (2 em 30 réplicas). Os benchmarks exibem distâncias significativamente maiores e maior número de outliers, coerente com a discrepância temporal e com pequenas diferenças em métricas de qualidade.

No método proposto, o teste de Mardia indica rejeição de normalidade por assimetria ($b1, p = 18,248$; $p = 6,58e^{-07}$) e curtose ($z = 2,326$; $p = 0,0200$). Assim como no *benchmark* principal, essa evidência sugere cautela na interpretação estrita de testes paramétricos, sem comprometer a conclusão geral devido à consistência direcional dos efeitos e às magnitudes observadas.

Tabela 19 - Correlações (Spearman) no método proposto em “*credit_card_default*” e p-valores com ajuste de Holm.

Variáveis	ρ (Spearman)	P bruto	P(Holm)
Acurácia Média × Precisão Média	0,520	0,0032	0,0226
Acurácia Média × Revocação Média	0,920	$6,56e^{-13}$	$6,56e^{-12}$
Acurácia Média × Especificidade Média	-0,670	$5,09e^{-05}$	0,0004
Acurácia Média × Ln de Tempo médio por fold	0,338	0,0679	0,3397
Precisão Média × Revocação Média	0,224	0,2340	0,7020
Precisão Média × Especificidade Média	0,137	0,4695	0,8790
Precisão Média × Ln de Tempo médio por fold	0,147	0,4395	0,8790
Revocação Média × Especificidade Média	-0,884	$9,16e^{-11}$	$8,24e^{-10}$
Revocação Média × Ln de Tempo médio por fold	0,367	0,0461	0,2769
Especificidade Média × Ln de Tempo médio por fold	-0,282	0,1307	0,5226

As correlações revelam *trade-offs* característicos de cenários desbalanceados: acurácia se associa fortemente à revocação, enquanto revocação e especificidade apresentam forte correlação negativa. Esse comportamento reforça a necessidade de tratamento multivariado e

de seleção de compromissos no espaço de objetivos, evitando decisões baseadas em uma única métrica.

4.9.4 Síntese multivariada e implicações para validade externa

Em ambas as bases, o padrão observado no benchmark canônico se repete: (i) o método proposto produz configurações com redução substancial de custo computacional; (ii) a variabilidade entre réplicas permanece controlada, com baixas contagens de outliers multivariados; e (iii) os perfis multivariados dos métodos diferem de forma estatisticamente significativa. A MANOVA aplicada ao vetor y indica efeito do método tanto em “*pepper_species*” (Wilks $\lambda = 0,2226$; $F = 17,247$; $p < 0,001$) quanto em “*credit_card_default*” (Wilks $\lambda = 0,2115$; $F = 14,000$; $p < 0,001$).

Tabela 20 - Pós-testes (Hotelling T^2) entre o método proposto e benchmarks no vetor y , com ajuste de Holm.

Base	Comparação entre métodos	T^2	F	df1	df2	p bruto	p (Holm)	D (Mahalanobis)
Espécies de Pimentas	DoE + NBI (proposto) vs. Busca em Grade	203,946	37,976	5	54	$1,11e^{-16}$	$4,44e^{-16}$	3,687
Espécies de Pimentas	DoE + NBI (proposto) vs. Busca Aleatória	232,204	43,238	5	54	$1,11e^{-16}$	$4,44e^{-16}$	3,934
Espécies de Pimentas	DoE + NBI (proposto) vs. Otimização Bayesiana	165,757	30,865	5	54	$1,08e^{-14}$	$1,08e^{-14}$	3,324
Espécies de Pimentas	DoE + NBI (proposto) vs. TPE (Hyperopt)	303,405	56,496	5	54	$1,11e^{-16}$	$4,44e^{-16}$	4,497
Inadimplência de Crédito	DoE + NBI (proposto) vs. Busca em Grade	308,599	57,463	5	54	$1,11e^{-16}$	$4,44e^{-16}$	4,536
Inadimplência de Crédito	DoE + NBI (proposto) vs. Busca Aleatória	216,373	40,290	5	54	$1,11e^{-16}$	$4,44e^{-16}$	3,798
Inadimplência de Crédito	DoE + NBI (proposto) vs. Otimização Bayesiana	199,543	37,156	5	54	$2,22e^{-16}$	$4,44e^{-16}$	3,647
Inadimplência de Crédito	DoE + NBI (proposto) vs. TPE (Hyperopt)	244,282	45,487	5	54	$1,11e^{-16}$	$4,44e^{-16}$	4,036

Os pós-testes de Hotelling confirmam que, para ambas as bases, o vetor de desempenho do método proposto difere do vetor dos benchmarks (p ajustado $< 10^{-12}$ em todos os casos), com distâncias de Mahalanobis entre médias variando aproximadamente de 3,3 a 4,5. No contexto desta dissertação, esses resultados sustentam uma conclusão prática: o pipeline DoE + RSM + NBI não é apenas competitivo em um único benchmark, mas tende a preservar o mesmo padrão de *trade-off* e estabilidade em bases externas com características distintas, ampliando a evidência de validade externa do método.

5. CONCLUSÕES

Este trabalho teve como objetivo propor e avaliar um procedimento sistemático e auditável para o ajuste multiobjetivo de hiperparâmetros do XGBoost em um caso de teste representativo, equilibrando desempenho preditivo, custo computacional e robustez do modelo. Para isso, foram integradas quatro camadas metodológicas: (i) DoE e modelagem por RSM, (ii) redução de dimensionalidade e interpretação das métricas por PCA com rotação *Varimax*, (iii) formulação do problema multiobjetivo no espaço de objetivos formado por VRFs e (iv) otimização multiobjetivo via NBI.

O ganho central dessa arquitetura está em separar claramente três decisões que, na prática, costumam ser tratadas de forma *ad hoc*: (a) quais métricas importam, (b) como agregá-las sem perder rastreabilidade e (c) como escolher um compromisso entre objetivos conflitantes. Ao transformar um conjunto amplo de métricas observáveis em três fatores interpretáveis (VRF₁, VRF₂ e VRF₃), o estudo reduz a complexidade do problema e permite que a seleção final de hiperparâmetros seja justificada com base em *trade-offs* explícitos, e não apenas em um único número (por exemplo, acurácia) ou em uma escolha empírica.

5.1 Síntese dos resultados e implicações

A etapa de análise fatorial evidenciou que as métricas originais não variam de forma independente: há correlações estruturais entre indicadores de qualidade preditiva e indicadores de custo/estabilidade. A rotação *Varimax* reforçou a interpretabilidade ao produzir uma estrutura simples, permitindo associar cada VRF a um construto semântico: VRF₁ como *proxy* de qualidade preditiva, VRF₂ como *proxy* de custo computacional (tempo de execução) e VRF₃ como *proxy* de robustez/estabilidade do comportamento do modelo (penalizando soluções mais instáveis).

Com as VRFs definidas, a modelagem por RSM foi utilizada para aproximar o mapeamento entre o vetor de hiperparâmetros e o espaço de objetivos, viabilizando a exploração do espaço de decisão com custo computacional menor do que uma busca exaustiva. Essa etapa é particularmente relevante porque a função objetivo real, treinar e avaliar o XGBoost, é onerosa; logo, a construção de um metamodelo atua como “atalho” para orientar a busca multiobjetivo.

Na sequência, a otimização multiobjetivo foi formulada diretamente no espaço das VRFs. O método NBI foi empregado para gerar soluções Pareto-ótimas ao longo do simplex de pesos, com três etapas-chave: (i) cálculo das soluções âncora (ótimos mono-objetivo) para cada VRF, (ii) construção do simplex no espaço de objetivos a partir do ponto utópico e das direções do CHIM e (iii) resolução, para cada vetor de pesos β , do subproblema NBI que projeta a fronteira de Pareto na direção normal ao simplex. Esse procedimento favorece uma amostragem mais uniforme do conjunto de Pareto, reduzindo o risco de concentração de soluções em regiões específicas (o que é comum em somas ponderadas tradicionais).

Em termos de desempenho, a solução selecionada como compromisso final atingiu melhora simultânea nos três eixos: houve aumento de aproximadamente 7,3% em VRF₁ (qualidade preditiva), redução de cerca de 71,5% no custo temporal médio (de 553,1 s para 157,9 s) associado a VRF₂ e redução aproximada de 22,4% em VRF₃ (indicador de robustez/estabilidade). Esses resultados sugerem que é possível deslocar a fronteira de trade-off para regiões mais favoráveis quando a busca é estruturada e guiada por modelos (DoE/RSM) e por uma formulação multiobjetivo explícita (VRFs + NBI).

Do ponto de vista prático, o principal efeito do procedimento proposto é transformar a “escolha do melhor modelo” em uma decisão informada. Em vez de produzir apenas um candidato final, o NBI entrega um conjunto de soluções Pareto-ótimas que podem ser comparadas à luz de restrições operacionais (por exemplo, tempo máximo de execução) e de preferências do decisor (por exemplo, priorizar falsos negativos em detrimento de falsos positivos). Assim, a decisão deixa de ser uma aposta em um ponto isolado e passa a ser uma seleção entre alternativas justificáveis.

Em síntese, as principais conclusões derivadas do estudo são:

- A redução de dimensionalidade via PCA/*Varimax* permitiu consolidar um conjunto amplo de métricas em três variáveis-resposta fatoriais (VRF₁-VRF₃), mantendo interpretabilidade e rastreabilidade.
- A combinação DoE + RSM fornece metamodelos úteis para explorar o espaço de hiperparâmetros com menor custo computacional, atuando como aproximação da função objetivo real.
- A formulação multiobjetivo no espaço das VRFs torna explícitos os compromissos entre qualidade preditiva, custo temporal e robustez, evitando decisões baseadas em uma única métrica.

- O método NBI, ao amostrar soluções ao longo do *simplex* de pesos, produz um conjunto Pareto mais informativo e facilita a escolha de compromissos compatíveis com requisitos operacionais.
- A solução final selecionada no caso de teste apresentou ganhos simultâneos em qualidade (VRF₁), custo (VRF₂) e estabilidade (VRF₃), indicando que o procedimento pode deslocar a fronteira de desempenho para regiões mais eficientes.
- Adicionalmente, a validação externa em “*pepper_species*” e “*credit_card_default*” reproduziu o mesmo padrão geral observado no benchmark canônico (ganhos de custo com estabilidade/baixa variabilidade entre réplicas), reforçando a validade externa do pipeline (Seção 4.9)

5.2 Evidências de validade externa e estabilidade em bases adicionais

Além do benchmark canônico utilizado como estudo principal, o pipeline proposto (DoE + RSM + NBI) foi aplicado a duas bases adicionais com perfis contrastantes, com o objetivo de avaliar se o padrão observado no cenário controlado se mantém em condições menos ideais e mais próximas de uso real. Em ambos os casos, preservou-se o mesmo protocolo experimental (30 réplicas e orçamento computacional fixo por método), permitindo interpretar as diferenças como efeitos atribuíveis predominantemente às abordagens de otimização e não a variações de configuração.

Na base “*pepper_species*”, observa-se um efeito de teto: todas as abordagens atingem métricas de qualidade muito próximas de 1,0 (por exemplo, o método proposto obtém acurácia $0,99992 \pm 0,00003$ e revocação $0,99999 \pm 0,00001$). Nesse regime, diferenças absolutas de qualidade tornam-se necessariamente pequenas e pouco informativas, e o diferencial passa a ser a eficiência computacional e a estabilidade entre réplicas. O método proposto reduz o tempo de otimização para $316,66 \pm 8,26s$, frente a $423,69 \pm 27,47s$ na busca em grade (redução de aproximadamente 25%), e reduz o tempo médio por fold para $0,211 \pm 0,054s$, frente a $0,603 \pm 0,195s$ (redução de aproximadamente 65%). No tratamento multivariado, o perfil do método proposto permanece mais concentrado: a distância de Mahalanobis ao seu próprio perfil típico apresenta média $d^2 = 2,900$ e apenas 1 outlier em 30 réplicas, enquanto os benchmarks exibem médias de d^2 entre aproximadamente 19,7 e 23,8 e 23–25 outliers, evidenciando maior variabilidade multivariada.

Já na base “*credit_card_default*”, que representa um cenário mais realista (ruído e forte desbalanceamento), o efeito de teto desaparece e o problema expõe com clareza o trade-off entre critérios. A revocação da classe positiva permanece moderada em todas as abordagens (no método proposto, $0,34473 \pm 0,02820$), ao passo que a especificidade se mantém elevada ($0,95255 \pm 0,00348$), comportamento típico de problemas desbalanceados. Mesmo nesse cenário, o método proposto preserva a proposta central de eficiência: o tempo de otimização cai para $269,71 \pm 6,95$ s, frente a aproximadamente 386 s na busca em grade (redução de cerca de 30%), e o tempo médio por fold é reduzido para $0,114 \pm 0,038$ s (redução de aproximadamente 85% em relação à busca em grade). Além disso, a robustez multivariada segue superior: média $d^2 = 4,833$ com 2 outliers, em contraste com benchmarks apresentando médias de d^2 entre aproximadamente 23,5 e 44,6 e 20 a 27 outliers.

Complementarmente, os pós-testes de Hotelling T^2 no vetor multivariado confirmam que, em ambas as bases, o método proposto apresenta perfil multivariado estatisticamente distinto dos benchmarks (p-valor ajustado $< 10^{-14}$ em todas as comparações), com distâncias de Mahalanobis entre médias na ordem de 3,3 a 4,5. Em conjunto, esses resultados sustentam que o padrão identificado no benchmark canônico; redução substancial de custo computacional com variabilidade entre réplicas controlada, e com compromissos explícitos no espaço multivariado de métricas, não se restringe a um único conjunto de dados, reproduzindo-se em domínios com características significativamente diferentes. Do ponto de vista aplicado, isso reforça o pipeline como ferramenta de decisão: em bases com alta separabilidade, o ganho tende a se manifestar majoritariamente como eficiência; em bases desbalanceadas, o método fornece um mecanismo estruturado para navegar compromissos (por exemplo, custo versus revocação) de forma rastreável e auditável.

5.3 Contribuições do trabalho

Além do resultado empírico, o trabalho contribui metodologicamente ao propor um fluxo completo, do diagnóstico de métricas até a seleção final de hiperparâmetros, com justificativas formais em cada etapa. As contribuições podem ser sintetizadas em três frentes complementares:

- Contribuição metodológica: integração de DoE/RSM, PCA/*Varimax* e NBI em um *pipeline* único para ajuste multiobjetivo, com etapas explicitadas e auditáveis.

- Contribuição analítica: definição de objetivos interpretáveis (VRFs) que preservam a semântica de grupos de métricas e permitem discutir *trade-offs* de forma mais clara do que agregações arbitrárias.
- Contribuição prática: geração de um conjunto Pareto de configurações de hiperparâmetros candidatas, permitindo adaptar a escolha final às prioridades do problema e às restrições do ambiente de execução.
- Contribuição empírica: demonstração de estabilidade do pipeline em bases adicionais com perfis contrastantes (efeito de teto e cenário real desbalanceado), utilizando o mesmo protocolo experimental e tratamento multivariado, reforçando a validade externa da metodologia e a consistência do padrão custo–qualidade–robustez observado.

5.4 Limitações e trabalhos futuros

Como qualquer estudo aplicado, os resultados devem ser interpretados à luz de limitações. Primeiro, o metamodelo RSM é uma aproximação do comportamento real do treinamento/avaliação e pode degradar em regiões de maior não linearidade ou interação não capturada pela estrutura quadrática. Segundo, as VRFs dependem do conjunto de métricas adotado e da amostra experimental utilizada na análise fatorial; em outras bases, a estrutura de cargas pode mudar e exigir reestimação. Terceiro, a solução “final” reflete um critério de compromisso adotado para o caso de teste e não elimina a existência de outras soluções Pareto igualmente válidas sob preferências diferentes.

Com base nessas limitações, destacam-se como direções promissoras para continuidade:

- Ampliar o conjunto de objetivos, incluindo métricas de custo de memória, estabilidade sob mudanças de distribuição (*drift*) e medidas de calibração probabilística.
- Investigar metamodelos alternativos (por exemplo, Processos Gaussianos ou *ensembles*) para reduzir erro de aproximação em regiões altamente não lineares.
- Incorporar preferências do decisor de forma explícita (métodos de decisão multicritério ou otimização com restrições), reduzindo a subjetividade da seleção final dentro do conjunto de Pareto.
- Ampliar a aplicação do *pipeline* em outros casos de teste e domínios, verificando a estabilidade das VRFs e a generalização das conclusões.

- Automatizar a etapa de pós-análise com critérios padronizados de escolha de compromisso (por exemplo, distância ao ponto utópico, ponto de joelho, ou métricas de robustez), facilitando a adoção industrial.
- Pós-otimização local sobre a região confirmada pelo DoE+RSM+NBI, à semelhança do procedimento discutido por Costa *et al.* (2025). No presente trabalho, a varredura do vetor de pesos já foi realizada com passo relativamente fino ($\Delta\beta\approx 0,02$), o que fornece uma discretização prática do compromisso qualidade-custo. Como extensão, pode-se reduzir ainda mais esse passo ou realizar uma busca local (malha refinada, iteração multiobjetivo com NBI, ou ainda uma heurística de curto orçamento) na vizinhança promissora para refinar a solução sem reiniciar todo o processo.
- Aplicação a bases com maior variabilidade de desempenho (incluindo diferentes balanços de classes, regimes de ruído e graus de dificuldade), nas quais a fronteira multiobjetivo tende a ser mais informativa e a etapa de pós-otimização pode produzir ganhos mais visíveis.
- Ampliação do conjunto de fatores/objetivos quando justificado pelo problema. Em particular, quando a redução multivariada indicar três (ou mais) fatores relevantes, torna-se possível recuperar visualizações em *simplex* (por exemplo, um triângulo para três fatores) e explorar malhas *simplex-lattice* sem degenerescência.
- Adoção sistemática de testes multivariados robustos/permutacionais como padrão de validação, reduzindo a dependência de hipóteses paramétricas em aplicações industriais (onde normalidade e homoscedasticidade são frequentemente frágeis).
- Avaliação do VRF-NBI com outros algoritmos de aprendizado (por exemplo, SVM, redes neurais artificiais, regressão logística) e com variações do pré-processamento/seleção de atributos, investigando como a estrutura fatorial e as superfícies RSM se alteram em diferentes famílias de modelos.

Em conclusão, o estudo demonstra que a combinação de modelagem estatística, redução de dimensionalidade e otimização multiobjetivo fornece um caminho consistente para ajustar modelos de aprendizagem de máquina em cenários onde múltiplos critérios importam simultaneamente. Ao substituir escolhas empíricas por um processo formal e reproduzível, o trabalho reforça a ideia de que desempenho, custo e robustez devem ser tratados como componentes de uma mesma decisão de projeto, e não como ajustes posteriores ao treinamento.

Os resultados evidenciam que o método proposto oferece um mecanismo de busca orientada capaz de reduzir substancialmente o custo computacional associado ao ajuste de

hiperparâmetros, mantendo desempenho preditivo elevado. Diferentemente das abordagens de referência, predominantemente formuladas como otimização de objetivo único, o método proposto trata explicitamente um problema multiobjetivo, no qual qualidade e custo são otimizados de forma simultânea. Essa formulação implica que a solução selecionada corresponde a uma solução de compromisso, e não ao ponto de utopia, o que explica a ocorrência de pequenas reduções nas métricas de qualidade quando comparadas a métodos que ignoram o custo computacional.

Sob a perspectiva metodológica, a eficiência do método proposto decorre de três elementos complementares. Primeiro, o delineamento experimental CCDFC promove cobertura estruturada do espaço de hiperparâmetros com número reduzido de avaliações, evitando tanto a exploração exaustiva quanto a amostragem puramente aleatória. Segundo, a modelagem por superfícies de resposta fornece um modelo aproximado que permite navegar no espaço de decisão com baixo custo, além de manter interpretabilidade por meio dos efeitos principais e de interações identificadas. Terceiro, o *framework* de otimização multiobjetivo DoE+NBI fornece uma forma sistemática de gerar candidatos ao longo de uma fronteira de Pareto, possibilitando selecionar configurações que equilibram custo e qualidade conforme o cenário de aplicação.

REFERÊNCIAS

- AKIBA, Takuya; SANO, Shotaro; YANASE, Toshihiko; OHTA, Takeru; KOYAMA, Masanori. **Optuna: A next-generation hyperparameter optimization framework**. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2019. p. 2623-2631.
- ARBORETTI, Roberta; OLIVIERI, Marta; LANTERI, Giuseppe. **Machine learning and design of experiments with an application to industrial experiments**. Quality and Reliability Engineering International, v. 37, n. 3, p. 1042-1058, 2021. DOI: 10.1002/qre.3025.
- ARLOT, S.; CELISSE, A. A survey of cross-validation procedures for model selection. Statistics Surveys, v. 4, p. 40-79, 2010. DOI: 10.1214/09-SS054.
- ATKINSON, Anthony C.; DONEV, Alexander N.; TOBIAS, Randall D. **Optimum Experimental Designs, with SAS**. Oxford: Oxford University Press, 2007.
- BACCI, Livio Agnew; MELLO, Luiz Gustavo; INCERTI, Taynara; PAIVA, Anderson Paulo de; BALESTRASSI, Pedro Paulo; *et al.* **Optimization of combined time series methods to forecast the demand for coffee in Brazil: A new approach using Normal Boundary Intersection coupled with mixture designs of experiments and rotated factor scores**. International Journal of Production Economics, v. 212, p. 186-211, 2019. DOI: 10.1016/j.ijpe.2019.03.001.
- BALESTRASSI, Pedro Paulo; POPPI, Rodrigo C.; PAIVA, Anderson P.; FERREIRA, Jorge R.; ROMA, Wesley N. **Design of experiments on neural network's training for nonlinear time series forecasting**. Expert Systems with Applications, v. 36, n. 7, p. 10349-10356, 2009.
- BELINATO, G.; ALMEIDA, F. A.; PAIVA, A. P.; GOMES, J. H. F.; BALESTRASSI, P. P.; ROSA, P. A. R. C. **A multivariate Normal Boundary Intersection PCA-based approach to reduce dimensionality in optimization problems for LBM process**. Engineering with Computers, v. 35, p. 1533-1544, 2019. DOI: 10.1007/s00366-018-0678-3.
- BERGSTRA, James; BARDENET, Rémi; BENGIO, Yoshua; KÉGL, Balázs. **Algorithms for hyper-parameter optimization**. In: Advances in Neural Information Processing Systems (NeurIPS). [S.l.]: Curran Associates, 2011.
- BERGSTRA, James; BENGIO, Yoshua. Random search for hyper-parameter optimization. Journal of Machine Learning Research, [S.l.], v. 13, p. 281-305, 2012.

BERGSTRA, James; YAMINS, Daniel; COX, David D. **Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures.** In: **International Conference on Machine Learning.** Atlanta: Pmlr, 2013. p. 115-123.

BISHOP, Christopher M. **Pattern recognition and machine learning.** New York: Springer, 2006.

BOCK, R. **MAGIC Gamma Telescope [Dataset].** UCI Machine Learning Repository, 2004. DOI: 10.24432/C52C8B.

BOCK, Richard K. *et al.* **Methods for multivariate analysis of data from astrophysical experiments.** Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, [S.l.], v. 516, n. 2-3, p. 511-528, 2004.

BORISOV, Vadim *et al.* **Deep neural networks and tabular data: a survey.** IEEE Transactions on Neural Networks and Learning Systems, [S.l.], v. 34, n. 11, p. 8504-8527, 2023.

BOX, George E. P.; WILSON, K. B. **On the experimental attainment of optimum conditions.** Journal of the Royal Statistical Society: Series B (Methodological), v. 13, n. 1, p. 1-45, 1951.

BOX, George E. P.; BEHNKEN, Donald W. **Some new three level designs for the study of quantitative variables.** Technometrics, v. 2, n. 4, p. 455-475, 1960. DOI: 10.1080/00401706.1960.10489912.

BOX, George E. P.; HUNTER, J. Stuart; HUNTER, William G. **Statistics for Experimenters: Design, Innovation, and Discovery.** 2. ed. Hoboken: Wiley, 2005.

BRADLEY, Andrew P. **The use of the area under the ROC curve in the evaluation of machine learning algorithms.** Pattern Recognition, v. 30, n. 7, p. 1145-1159, 1997. DOI: 10.1016/S0031-3203(96)00142-2.

BREIMAN, L. **Random forests.** Machine Learning, v. 45, n. 1, p. 5-32, 2001. DOI: 10.1023/A:1010933404324.

BRIER, Glenn W. **Verification of forecasts expressed in terms of probability.** Monthly Weather Review, v. 78, n. 1, p. 1-3, 1950.

BRITO, T. G.; PAIVA, A. P.; FERREIRA, J. R.; GOMES, J. H. F.; BALESTRASSI, P. P. **A normal boundary intersection approach to multiresponse robust optimization of the surface roughness in end milling process with combined arrays.** Precision Engineering, v. 38, n. 3, p. 628-638, 2014. DOI: 10.1016/j.precisioneng.2014.02.009.

CAWLEY, Gavin C.; TALBOT, Nicola L. C. **On over-fitting in model selection and subsequent selection bias in performance evaluation.** *Journal of Machine Learning Research*, v. 11, p. 2079–2107, 2010.

CHEN, Tianqi; GUESTRIN, Carlos. **XGBoost: a scalable tree boosting system.** In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. New York: ACM, 2016. p. 785-794.

COSTA, Danielle Martins Duarte; BRITO, Tarcísio Gonçalves; PAIVA, Anderson Paulo de; LEME, Rafael Coradi; BALESTRASSI, Pedro Paulo. **A normal boundary intersection with multivariate mean square error approach for dry end milling process optimization of the AISI 1045 steel.** *Journal of Cleaner Production*, v. 135, p. 1658-1672, 2016. DOI: 10.1016/j.jclepro.2016.01.062.

DAS, Indraneel; DENNIS, John E. **Normal-boundary intersection: a new method for generating the Pareto surface in nonlinear multicriteria optimization problems.** *SIAM Journal on Optimization*, [S.l.], v. 8, n. 3, p. 631-657, 1998. DOI: 10.1137/S1052623496307510.

DEB, Kalyanmoy. **Multi-objective optimization using evolutionary algorithms.** Chichester: Wiley, 2001.

DEB, Kalyanmoy; PRATAP, Amrit; AGARWAL, Sameer; MEYARIVAN, T. **A fast and elitist multiobjective genetic algorithm: NSGA-II.** *IEEE Transactions on Evolutionary Computation*, v. 6, n. 2, p. 182–197, 2002.

DOEHLERT, David H. **Uniform shell designs.** *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, v. 19, n. 3, p. 231-239, 1970.

FALKNER, Stefan; KLEIN, Aaron; HUTTER, Frank. **BOHB: robust and efficient hyperparameter optimization at scale.** In: DY, Jennifer; KRAUSE, Andreas (ed.). *Proceedings of the 35th International Conference on Machine Learning (ICML 2018)*. *Proceedings of Machine Learning Research*, v. 80. [S.l.]: PMLR, 2018. p. 1437–1446.

FAWCETT, Tom. **An introduction to ROC analysis.** *Pattern Recognition Letters*, v. 27, n. 8, p. 861-874, 2006. DOI: 10.1016/j.patrec.2005.10.010.

FEURER, Matthias; HUTTER, Frank. **Hyperparameter optimization.** In: HUTTER, Frank; KOTTHOFF, Lars; VANSCHOREN, Joaquin (org.). *Automated machine learning: methods, systems, challenges*. Cham: Springer, 2019. p. 3-33.

FISHER, Ronald Aylmer. **The Design of Experiments.** Edinburgh: Oliver and Boyd, 1935.

FRANCESCHI, Luca; DONINI, Michele; PERRONE, Valerio; KLEIN, Aaron; ARCHAMBEAU, Cédric; SEEGER, Matthias; PONTIL, Massimiliano; FRASCONI, Paolo. **Hyperparameter optimization in machine learning**. arXiv:2410.22854 [stat.ML], 2024.

FRAZIER, Peter I. **A tutorial on Bayesian optimization**. arXiv:1807.02811, 2018.

FREIESLEBEN, Johannes; KNEIB, Thomas; STELZER, Christian; WEST, Mike. **Design of experiments: a powerful tool for machine learning model calibration**. In: AMERICAN STATISTICAL ASSOCIATION. Proceedings of the 2020 Joint Statistical Meetings. Alexandria: Asa, 2020.

FREIRE, A. I. **Diversidade genética de *Capsicum chinense* C. frutescens**. 2019. 75 f. Tese (Doutorado em Fitotecnia) - Universidade Federal de Viçosa, Viçosa, 2019.

FRIEDMAN, Jerome H. **Greedy function approximation: a gradient boosting machine**. The Annals of Statistics, [S.l.], v. 29, n. 5, p. 1189-1232, 2001.

GRAMACY, Robert. Surrogates: **Gaussian process modeling, design, and optimization for the applied sciences**. Boca Raton: CRC Press, 2020.

GRINSZTAJN, Alexandre; GABRIEL, François; LEE, Jim. **Why do tree-based models still outperform deep learning on tabular data?** In: Advances in Neural Information Processing Systems (NeurIPS). [S.l.]: Curran Associates, 2022.

HARMAN, Harry H. **Modern Factor Analysis**. 3. ed. Chicago: The University of Chicago Press, 1976.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The elements of statistical learning: data mining, inference, and prediction**. 2. ed. New York: Springer, 2009.

HE, Haibo; GARCIA, Eduardo A. **Learning from imbalanced data**. IEEE Transactions on Knowledge and Data Engineering, v. 21, n. 9, p. 1263-1284, 2009. DOI: 10.1109/TKDE.2008.239.

HOTELLING, Harold. **Analysis of a complex of statistical variables into principal components**. Journal of Educational Psychology, v. 24, n. 6, p. 417-441, 1933; v. 24, n. 7, p. 498-520, 1933.

HUTTER, Frank; HOOS, Holger H.; LEYTON-BROWN, Kevin. **Sequential model-based optimization for general algorithm configuration**. In: COELLO, Carlos A. (ed.). Learning and Intelligent Optimization (LION 2011). Lecture Notes in Computer Science, v. 6683. Berlin; Heidelberg: Springer, 2011. p. 507-523.

JOLLIFFE, Ian T. **Principal Component Analysis**. 2. ed. New York: Springer, 2002.

JONES, Donald R.; SCHONLAU, Matthias; WELCH, William J. **Efficient global optimization of expensive black-box functions.** Journal of Global Optimization, v. 13, n. 4, p. 455-492, 1998.

KAISER, Henry F. **The varimax criterion for analytic rotation in factor analysis.** Psychometrika, v. 23, n. 3, p. 187-200, 1958. DOI: 10.1007/BF02289233.

KOHAVI, Ron. **A study of cross-validation and bootstrap for accuracy estimation and model selection.** In: International Joint Conference on Artificial Intelligence. Montreal: Morgan Kaufmann, 1995. p. 1137-1143.

LI, Lisha; JAMIESON, Kevin; DESALVO, Giulia; ROSTAMIZADEH, Afshin; TALWALKAR, Ameet. **Hyperband: a novel bandit-based approach to hyperparameter optimization.** Journal of Machine Learning Research, v. 18, n. 185, p. 1-52, 2018.

LOPES, Luiz Gustavo Dias; BRITO, Tarcisio Gonçalves; PAIVA, Anderson Paulo de; PERUCHI, Rodolfo S.; BALESTRASSI, Pedro Paulo. **Robust parameter optimization based on multivariate normal boundary intersection.** Computers & Industrial Engineering, v. 93, p. 55-66, 2016. DOI: 10.1016/j.cie.2015.12.023.

MARTINEK, A.; BARTUZI-TROKIELEWICZ, E.; ŁUKASIK, S.; GANDOMI, A. H. **Tuning metaheuristic parameters with the use of Large Language Models.** Neurocomputing, v. 675, art. 132976, 2026. DOI: 10.1016/j.neucom.2026.132976.

MATTHEWS, B. W. **Comparison of the predicted and observed secondary structure of T4 phage lysozyme.** Biochimica et Biophysica Acta (BBA) - Protein Structure, v. 405, n. 2, p. 442-451, 1975.

MCELFRESH, D. *et al.* **When do neural nets beat decision trees on tabular data?** In: Advances in Neural Information Processing Systems (NeurIPS). 2023.

MIETTINEN, Kaisa. Nonlinear multiobjective optimization. Dordrecht: Kluwer Academic Publishers, 1999.

MONTGOMERY, Douglas C. **Design and analysis of experiments.** 9. ed. Hoboken: Wiley, 2017.

MORALES-HERNÁNDEZ, Alejandro *et al.* **A survey on multi-objective hyperparameter optimization in machine learning.** arXiv preprint, arXiv:2111.13755v3, 2022.

MYERS, Raymond H.; MONTGOMERY, Douglas C.; ANDERSON-COOK, Christine M. **Response surface methodology: process and product optimization using designed experiments.** 4. ed. Hoboken: Wiley, 2016.

NATEKIN, Alexey; KNOLL, Alois. **Gradient boosting machines, a tutorial**. *Frontiers in Neurorobotics*, v. 7, art. 21, 2013. DOI: 10.3389/fnbot.2013.00021.

NAVES, Fabiano Luiz; PAULA, Taynara Incerti de; BALESTRASSI, Pedro Paulo; BRAGA, Washington Luis Moreira; SAWHNEY, Rapinder Singh; PAIVA, Anderson Paulo de. **Multivariate normal boundary intersection based on rotated factor scores: A multiobjective optimization method for methyl orange treatment**. *Journal of Cleaner Production*, v. 143, p. 413-439, 2017. DOI: 10.1016/j.jclepro.2016.12.092.

PEARSON, Karl. **On lines and planes of closest fit to systems of points in space**. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, v. 2, n. 11, p. 559-572, 1901. DOI: 10.1080/14786440109462720.

PEREIRA, R. B. D.; LEITE, R. R.; ALVIM, A. C.; PAIVA, A. P.; BALESTRASSI, P. P.; FERREIRA, J. R.; DAVIM, J. P. **Multivariate robust modeling and optimization of cutting forces of the helical milling process of the aluminum alloy Al 7075**. *The International Journal of Advanced Manufacturing Technology*, v. 95, p. 2691-2715, 2018. DOI: 10.1007/s00170-017-1398-3.

PEREIRA, Matheus Costa; RIBEIRO, Caio Tertuliano; MENDES, Ronã Rinston Amaury; CAMPOS, Paulo Henrique da Silva; PAIVA, Anderson Paulo de Paiva. **A hybrid multivariate normal boundary intersection approach with post-optimization assisted by mixture design of experiments**. *Engineering Applications of Artificial Intelligence*, v. 162, p. 112510, 2025. DOI: 10.1016/j.engappai.2025.112510.

PLACKETT, Robin L.; BURMAN, J. P. **The design of optimum multifactorial experiments**. *Biometrika*, v. 33, n. 4, p. 305-325, 1946.

PONTES, Francisco J.; ALMEIDA, Cecília A. P.; BARBOSA, João R.; SOUSA, Sérgio L.; SOARES, José J.; LIMA, Fernando R.; PAIVA, Anderson P. **Design of experiments and focused grid search for neural network parameter optimization**. *Neurocomputing*, v. 186, p. 22-34, 2016.

POWERS, David M. W. **Evaluation: From Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation**. *Journal of Machine Learning Technologies*, v. 2, n. 1, p. 37-63, 2011.

PRECHELT, L. **Early stopping - but when?** In: ORR, G. B.; MÜLLER, K.-R. (org.). . Berlin: Springer, 1998. p. 55-69.

PROBST, Philipp; WRIGHT, Marvin N.; BOULESTEIX, Anne-Laure. **Hyperparameters and tuning strategies for random forest**. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, v. 9, n. 3, e1301, 2019. DOI: 10.1002/widm.1301.

PROBST, Philipp; BOULESTEIX, Anne-Laure; BISCHL, Bernd. **Tunability: importance of hyperparameters of machine learning algorithms**. Journal of Machine Learning Research, v. 20, n. 53, p. 1-32, 2019

RAMAMOCHAN, Varun *et al.* **Discrete simulation optimization for tuning machine learning method hyperparameters**. Journal of Simulation, v. 18, n. 5, p. 745-765, 2024

RASMUSSEN, Carl Edward; WILLIAMS, Christopher K. I. **Gaussian processes for machine learning**. Cambridge: MIT Press, 2006.

SAITO, Takaya; REHMSMEIER, Marc. **The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets**. PLOS ONE, v. 10, n. 3, e0118432, 2015. DOI: 10.1371/journal.pone.0118432.

SHAHRIARI, Bobak; SWERSKY, Kevin; WANG, Ziyu; ADAMS, Ryan P.; DE FREITAS, Nando. **Taking the human out of the loop: a review of Bayesian optimization**. Proceedings of the IEEE, v. 104, n. 1, p. 148-175, 2016.

SHALEV-SHWARTZ, Shai; BEN-DAVID, Shai. **Understanding machine learning: from theory to algorithms**. Cambridge: Cambridge University Press, 2014.

SNOEK, Jasper; LAROCHELLE, Hugo; ADAMS, Ryan P. **Practical Bayesian optimization of machine learning algorithms**. In: Advances in Neural Information Processing Systems. Red Hook: Curran Associates, 2012. v. 25.

SPEARMAN, Charles. **“General Intelligence”, objectively determined and measured**. The American Journal of Psychology, v. 15, n. 2, p. 201-292, 1904. DOI: 10.2307/1412107.

TAGUCHI, Genichi. **Introduction to Quality Engineering: Designing Quality into Products and Processes**. Tokyo: Asian Productivity Organization, 1986.

THURSTONE, Louis L. **The Vectors of Mind: Multiple-Factor Analysis for the Isolation of Primary Traits**. Chicago: The University of Chicago Press, 1935.

VASQUEZ-RAMOS, J.; ROUX, F. C.; DIMOU, P.; YANNIS, G.; PANOSKALTSIS, N. **Response surface-driven hyperparameter optimization for XGBoost**. The Journal of Supercomputing, v. 81, art. 1112, 2025. DOI: 10.1007/s11227-025-07600-4.

VAN RIJSBERGEN, Cornelis J. **Information Retrieval**. 2. ed. London: Butterworths, 1979.

WELLER, S.; MOLL, T. **Hyperparameter importance estimation in reinforcement learning**. Neurocomputing, v. 659, art. 131770, 2026. DOI: 10.1016/j.neucom.2025.131770.

WISTUBA, Martin; SCHILLING, Nils; SCHMIDT-THIEME, Lars. **Recent advances in Bayesian optimization.** arXiv:2206.03301, 2022.

WOLPERT, D. H.; MACREADY, W. G. **No free lunch theorems for optimization.** IEEE Transactions on Evolutionary Computation, v. 1, n. 1, p. 67-82, 1997.

WON, Jonghyeon; LEE, Hyun-Suk; LEE, Jang-Won. **A review on multi-fidelity hyperparameter optimization in machine learning.** ICT Express, v. 11, p. 245–257, 2025.

YEH, I.-C.; LIEN, C.-H. **The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients.** *Expert Systems with Applications*, v. 36, n. 2, p. 2473–2480, 2009. DOI: 10.1016/j.eswa.2007.12.020.

YERUSHALMY, Jacob. **Statistical problems in assessing methods of medical diagnosis, with special reference to X-ray techniques.** Public Health Reports, v. 62, n. 40, p. 1432-1449, 1947.

APÊNDICE A - DISPONIBILIDADE DO CÓDIGO-FONTE E ARTEFATOS EXPERIMENTAIS

O código-fonte, resultados completos desta experimentação, *scripts* de execução e instruções de reprodutibilidade desta pesquisa estão disponibilizados em repositório público, no endereço: https://github.com/caioribeiro99/doi_nbi_hpo_project

Os artefatos completos gerados pelos experimentos (arquivos CSV/JSON, *logs* e demais saídas por réplica) foram organizados em arquivo compactado (.zip) e disponibilizados como material suplementar na seção de "*Releases*" do repositório.

A baseline 'XGBoost *default*' foi obtida por execução adicional sob o mesmo manifesto e sementes e registrada nos artefatos de sumarização do experimento.

Disponível em: https://github.com/caioribeiro99/doi_nbi_hpo_project/releases. Acesso em: 30 dez. 2025.

APÊNDICE C - SUPERFÍCIE DE PARETO (*SIMPLEX-LATTICE* {3, 10} E MÉTRICAS DE PERFORMANCE)

Simplex-Lattice {3, 10}				Métricas de Desempenho do XGBoost				
n	w ₁	w ₂	w ₃	Accuracy	Precision	Recall	Specificity	Tempo
1	1.00	0.00	0.00	0.950	0.967	0.929	0.966	0.351
2	0.90	0.10	0.00	0.948	0.962	0.930	0.962	0.459
3	0.90	0.00	0.10	0.944	0.960	0.927	0.959	0.321
4	0.80	0.20	0.00	0.948	0.962	0.931	0.962	0.543
5	0.80	0.10	0.10	0.945	0.961	0.928	0.960	0.421
6	0.80	0.00	0.20	0.941	0.958	0.924	0.957	0.284
7	0.70	0.30	0.00	0.944	0.958	0.931	0.958	0.632
8	0.70	0.20	0.10	0.945	0.961	0.928	0.960	0.494
9	0.70	0.10	0.20	0.941	0.958	0.925	0.957	0.383
10	0.70	0.00	0.30	0.938	0.956	0.922	0.955	0.262
11	0.60	0.40	0.00	0.943	0.956	0.932	0.956	0.700
12	0.60	0.30	0.10	0.940	0.954	0.927	0.954	0.615
13	0.60	0.20	0.20	0.942	0.958	0.925	0.957	0.472
14	0.60	0.10	0.30	0.939	0.957	0.922	0.956	0.368
15	0.60	0.00	0.40	0.936	0.954	0.920	0.954	0.239
16	0.50	0.50	0.00	0.942	0.956	0.933	0.956	0.747
17	0.50	0.40	0.10	0.940	0.953	0.929	0.953	0.668
18	0.50	0.30	0.20	0.939	0.951	0.926	0.951	0.584
19	0.50	0.20	0.30	0.939	0.955	0.923	0.955	0.461
20	0.50	0.10	0.40	0.937	0.954	0.920	0.953	0.353
21	0.50	0.00	0.50	0.934	0.952	0.918	0.952	0.225
22	0.40	0.60	0.00	0.942	0.955	0.934	0.954	0.799
23	0.40	0.50	0.10	0.941	0.953	0.930	0.952	0.719
24	0.40	0.40	0.20	0.940	0.953	0.927	0.952	0.632
25	0.40	0.30	0.30	0.939	0.955	0.923	0.954	0.535
26	0.40	0.20	0.40	0.937	0.953	0.920	0.953	0.442
27	0.40	0.10	0.50	0.935	0.952	0.918	0.952	0.342
28	0.40	0.00	0.60	0.932	0.950	0.915	0.951	0.199
29	0.30	0.70	0.00	0.943	0.954	0.935	0.954	0.858
30	0.30	0.60	0.10	0.942	0.953	0.931	0.953	0.767
31	0.30	0.50	0.20	0.940	0.953	0.927	0.953	0.686
32	0.30	0.40	0.30	0.939	0.954	0.923	0.954	0.598
33	0.30	0.30	0.40	0.937	0.953	0.921	0.953	0.518
34	0.30	0.20	0.50	0.935	0.951	0.918	0.952	0.431
35	0.30	0.10	0.60	0.933	0.950	0.916	0.951	0.325
36	0.30	0.00	0.70	0.931	0.950	0.913	0.951	0.155
37	0.20	0.80	0.00	0.945	0.954	0.937	0.954	0.895
38	0.20	0.70	0.10	0.942	0.954	0.932	0.953	0.813
39	0.20	0.60	0.20	0.941	0.954	0.928	0.953	0.738
40	0.20	0.50	0.30	0.939	0.954	0.924	0.954	0.658
41	0.20	0.40	0.40	0.937	0.953	0.921	0.953	0.583
42	0.20	0.30	0.50	0.936	0.951	0.919	0.952	0.507
43	0.20	0.20	0.60	0.934	0.951	0.916	0.952	0.408
44	0.20	0.10	0.70	0.932	0.950	0.913	0.951	0.296
45	0.20	0.00	0.80	0.930	0.949	0.911	0.951	0.123
46	0.10	0.90	0.00	0.947	0.956	0.939	0.956	0.934
47	0.10	0.80	0.10	0.943	0.954	0.933	0.954	0.860
48	0.10	0.70	0.20	0.941	0.954	0.928	0.953	0.787
49	0.10	0.60	0.30	0.939	0.954	0.924	0.953	0.712
50	0.10	0.50	0.40	0.938	0.952	0.922	0.952	0.643
51	0.10	0.40	0.50	0.936	0.951	0.919	0.952	0.573
52	0.10	0.30	0.60	0.935	0.951	0.916	0.952	0.486
53	0.10	0.20	0.70	0.933	0.951	0.914	0.952	0.392
54	0.10	0.10	0.80	0.931	0.949	0.912	0.951	0.277
55	0.10	0.00	0.90	0.929	0.946	0.911	0.948	0.120
56	0.00	1.00	0.00	0.945	0.952	0.939	0.955	0.975
57	0.00	0.90	0.10	0.944	0.954	0.934	0.953	0.907
58	0.00	0.80	0.20	0.941	0.954	0.929	0.953	0.834
59	0.00	0.70	0.30	0.940	0.953	0.925	0.953	0.763
60	0.00	0.60	0.40	0.938	0.952	0.922	0.952	0.698
61	0.00	0.50	0.50	0.936	0.952	0.919	0.952	0.627
62	0.00	0.40	0.60	0.935	0.951	0.917	0.952	0.553
63	0.00	0.30	0.70	0.934	0.951	0.914	0.952	0.472
64	0.00	0.20	0.80	0.932	0.950	0.912	0.951	0.375
65	0.00	0.10	0.90	0.930	0.949	0.910	0.950	0.248
66	0.00	0.00	1.00	0.927	0.938	0.909	0.941	0.120

APÊNDICE D - PARETO MDE (*SETUPS* ÓTIMOS E NÃO-DOMINADOS)

n	Variáveis de Entrada Codificadas							Variáveis de Entrada Decodificadas						
	x ₁	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₁	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇
1	1.000	-0.484	-0.171	1.000	-0.994	0.322	-0.093	1.000	0.295	0.444	0.300	3.047	3.322	344.6
2	1.000	-0.438	-0.025	0.852	-0.425	0.200	-0.148	1.000	0.317	0.513	0.279	7.313	3.019	327.1
3	0.813	-0.595	-0.009	0.712	-0.227	0.206	-0.549	0.911	0.242	0.521	0.258	8.799	3.036	196.4
4	1.000	-0.412	-0.004	0.639	-0.458	0.173	-0.037	1.000	0.330	0.523	0.248	7.068	2.954	362.9
5	0.927	-0.503	0.008	0.469	-0.308	0.247	-0.390	0.965	0.286	0.529	0.223	8.192	3.137	248.4
6	0.424	-0.653	-0.001	0.606	-0.343	0.390	-0.657	0.727	0.215	0.524	0.243	7.928	3.491	161.4
7	0.615	-0.623	-0.008	0.918	-0.542	0.188	0.512	0.817	0.229	0.521	0.288	6.436	2.991	541.5
8	0.580	-0.615	0.140	0.390	-0.710	0.349	-0.169	0.801	0.233	0.592	0.212	5.178	3.389	320.2
9	0.364	-0.600	0.076	0.505	-0.456	0.484	-0.504	0.698	0.240	0.561	0.228	7.081	3.723	211.1
10	0.246	-0.665	-0.023	0.464	-0.231	0.339	-0.698	0.642	0.209	0.514	0.222	8.769	3.364	148.1
11	0.570	-0.608	-0.036	0.919	-0.436	0.162	0.735	0.796	0.236	0.508	0.288	7.227	2.926	613.7
12	0.330	-0.778	-0.036	1.000	-0.468	0.160	0.579	0.682	0.155	0.508	0.300	6.989	2.920	563.1
13	0.338	-0.624	0.109	0.351	-0.478	0.404	-0.299	0.686	0.228	0.577	0.206	6.916	3.524	277.9
14	0.213	-0.656	0.107	0.280	-0.451	0.306	-0.500	0.626	0.213	0.576	0.196	7.116	3.283	212.5
15	0.049	-0.628	-0.045	0.346	-0.171	0.306	-0.732	0.548	0.227	0.504	0.205	9.216	3.283	137.1
16	0.454	-0.573	0.000	0.883	-0.246	0.348	0.913	0.741	0.253	0.525	0.283	8.655	3.387	671.6
17	0.240	-0.694	0.062	0.957	-0.163	0.343	0.584	0.639	0.195	0.555	0.294	9.278	3.374	564.8
18	0.048	-0.696	0.003	0.988	-0.022	0.349	0.235	0.548	0.195	0.526	0.298	10.336	3.389	451.3
19	0.191	-0.649	-0.007	0.254	-0.119	0.238	-0.366	0.616	0.217	0.522	0.192	9.604	3.115	256.1
20	-0.035	-0.593	-0.034	0.279	-0.147	0.330	-0.542	0.508	0.243	0.509	0.196	9.397	3.342	198.8
21	-0.223	-0.528	-0.036	0.343	-0.190	0.315	-0.755	0.419	0.274	0.508	0.205	9.071	3.305	129.7
22	0.434	-0.495	-0.030	0.859	-0.073	0.307	0.896	0.731	0.290	0.511	0.280	9.951	3.285	666.3
23	0.218	-0.652	-0.013	0.844	0.003	0.301	0.560	0.629	0.215	0.519	0.277	10.524	3.271	556.9
24	0.050	-0.671	0.005	0.618	-0.006	0.311	0.139	0.549	0.206	0.527	0.245	10.452	3.295	420.3
25	0.134	-0.615	-0.010	0.174	-0.069	0.212	-0.255	0.589	0.233	0.520	0.180	9.979	3.050	292.2
26	-0.117	-0.558	-0.026	0.209	-0.109	0.310	-0.383	0.469	0.260	0.513	0.185	9.683	3.291	250.6
27	-0.286	-0.480	-0.026	0.253	-0.140	0.327	-0.558	0.389	0.297	0.513	0.192	9.447	3.335	193.8
28	-0.357	-0.438	-0.024	0.173	-0.052	0.163	-0.767	0.355	0.317	0.513	0.180	10.109	2.928	125.6
29	0.477	-0.424	-0.052	0.796	-0.125	0.082	0.802	0.752	0.323	0.500	0.270	9.563	2.728	635.7
30	0.214	-0.614	-0.011	0.621	0.012	0.228	0.454	0.627	0.233	0.520	0.245	10.591	3.090	522.6
31	0.058	-0.626	0.009	0.406	0.006	0.242	0.141	0.553	0.227	0.529	0.214	10.547	3.125	420.8
32	0.039	-0.554	-0.016	0.104	-0.048	0.243	-0.147	0.544	0.262	0.517	0.170	10.143	3.125	327.4
33	-0.195	-0.510	-0.018	0.135	-0.073	0.271	-0.253	0.432	0.283	0.516	0.175	9.951	3.195	292.8
34	-0.321	-0.465	-0.020	0.104	-0.032	0.173	-0.408	0.373	0.304	0.515	0.170	10.261	2.953	242.6
35	-0.445	-0.331	-0.009	0.133	-0.090	0.269	-0.586	0.314	0.368	0.521	0.174	9.826	3.191	184.6
36	-0.368	-0.271	0.042	-0.097	-0.077	0.077	-0.842	0.350	0.396	0.545	0.141	9.921	2.717	101.2
37	0.448	-0.301	-0.061	0.603	-0.027	0.138	0.654	0.738	0.382	0.496	0.242	10.297	2.867	587.4
38	0.210	-0.545	-0.010	0.438	0.011	0.175	0.400	0.625	0.266	0.520	0.219	10.584	2.958	504.9
39	0.059	-0.565	0.011	0.253	0.018	0.173	0.149	0.553	0.257	0.530	0.192	10.632	2.954	423.4
40	-0.038	-0.504	-0.013	0.044	-0.021	0.185	-0.062	0.507	0.285	0.519	0.161	10.340	2.984	354.9
41	-0.268	-0.446	-0.011	0.062	-0.040	0.223	-0.151	0.398	0.313	0.520	0.164	10.196	3.078	326.0
42	-0.401	-0.376	-0.008	0.040	-0.017	0.166	-0.290	0.334	0.346	0.521	0.161	10.369	2.936	280.7
43	-0.396	-0.265	0.044	-0.139	-0.059	0.101	-0.490	0.337	0.399	0.546	0.135	10.054	2.774	215.9
44	-0.476	-0.152	0.051	-0.141	-0.077	0.115	-0.649	0.299	0.453	0.549	0.135	9.922	2.809	163.9
45	-0.511	-0.045	0.053	-0.115	-0.102	0.127	-0.887	0.282	0.503	0.550	0.138	9.734	2.840	86.7
46	0.477	-0.110	-0.043	0.316	-0.248	0.141	0.614	0.751	0.473	0.505	0.201	8.641	2.873	574.4
47	0.217	-0.459	-0.011	0.305	0.014	0.111	0.366	0.628	0.307	0.520	0.199	10.607	2.800	494.1
48	0.055	-0.488	0.012	0.131	0.029	0.109	0.154	0.551	0.293	0.531	0.174	10.715	2.795	425.0
49	-0.108	-0.438	-0.008	-0.013	0.001	0.132	0.004	0.474	0.317	0.521	0.153	10.510	2.852	376.4
50	-0.331	-0.365	-0.004	-0.008	-0.010	0.169	-0.074	0.368	0.352	0.523	0.154	10.425	2.944	350.8
51	-0.468	-0.268	0.003	-0.025	-0.009	0.156	-0.196	0.303	0.398	0.526	0.151	10.429	2.912	311.3
52	-0.445	-0.140	0.055	-0.214	-0.041	0.098	-0.393	0.314	0.459	0.551	0.124	10.189	2.767	247.3
53	-0.510	-0.021	0.065	-0.212	-0.077	0.152	-0.525	0.283	0.515	0.556	0.124	9.925	2.900	204.5
54	-0.560	0.087	0.066	-0.155	-0.121	0.209	-0.691	0.259	0.566	0.556	0.132	9.593	3.042	150.4
55	-0.770	0.231	0.013	0.170	-0.149	0.107	-0.907	0.159	0.635	0.531	0.180	9.380	2.789	80.1
56	0.409	0.072	1.000	0.396	-1.000	-0.097	1.000	0.719	0.559	1.000	0.212	3.000	2.285	700.0
57	0.250	-0.347	-0.010	0.199	0.021	0.026	0.320	0.644	0.360	0.520	0.184	10.659	2.589	479.1
58	0.037	-0.403	0.012	0.044	0.039	0.052	0.167	0.543	0.333	0.531	0.161	10.795	2.654	429.1
59	-0.174	-0.356	-0.002	-0.062	0.020	0.083	0.058	0.442	0.356	0.524	0.146	10.650	2.731	393.9
60	-0.375	-0.270	0.005	-0.080	0.017	0.101	-0.026	0.347	0.397	0.527	0.143	10.631	2.775	366.5
61	-0.364	-0.140	0.054	-0.267	0.011	0.025	-0.230	0.352	0.458	0.550	0.116	10.579	2.587	300.4
62	-0.462	-0.006	0.067	-0.292	-0.023	0.083	-0.323	0.305	0.522	0.557	0.113	10.329	2.729	270.1
63	-0.525	0.114	0.074	-0.282	-0.060	0.142	-0.431	0.275	0.579	0.560	0.114	10.052	2.877	235.0
64	-0.550	0.212	0.081	-0.259	-0.112	0.201	-0.567	0.264	0.626	0.563	0.117	9.659	3.023	190.8
65	-0.544	0.286	0.087	-0.230	-0.174	0.245	-0.748	0.267	0.661	0.566	0.122	9.193	3.131	131.8
66	-1.000	0.506	-0.290	1.000	-0.370	-0.093	-0.937	0.050	0.765	0.387	0.300	7.724	2.294	70.5

APÊNDICE E - SUPERFÍCIE DE PARETO (MÉTRICAS MULTIVARIADAS)

n	Fatores Rotacionados			Índice Fatorial VRF			Restrições do Método NBI		
	FS ₁	FS ₂	FS ₃	VRF ₁	VRF ₂	VRF ₃	g ₁ (x)	g ₂ (x)	g ₃ (x)
1	1.201	-1.025	-0.625	2.120	2.289	1.176	0.000	0.000	0.000
2	0.899	-1.058	-0.538	2.212	2.248	1.256	0.000	0.000	0.000
3	0.673	-0.922	-0.647	2.400	2.432	1.158	0.000	0.000	0.000
4	0.773	-1.093	-0.464	2.304	2.207	1.337	0.000	0.000	0.000
5	0.592	-0.950	-0.556	2.492	2.391	1.239	0.000	0.000	0.000
6	0.454	-0.832	-0.669	2.680	2.574	1.141	0.000	0.000	0.000
7	0.677	-1.130	-0.398	2.396	2.166	1.417	0.000	0.000	0.000
8	0.521	-0.980	-0.479	2.584	2.349	1.319	0.000	0.000	0.000
9	0.395	-0.857	-0.574	2.772	2.533	1.221	0.000	0.000	0.000
10	0.286	-0.750	-0.694	2.960	2.717	1.124	0.000	0.000	0.000
11	0.595	-1.170	-0.339	2.488	2.124	1.497	0.000	0.000	0.000
12	0.456	-1.010	-0.412	2.676	2.308	1.400	0.000	0.000	0.000
13	0.340	-0.883	-0.495	2.864	2.492	1.302	0.000	0.000	0.000
14	0.237	-0.773	-0.593	3.052	2.676	1.204	0.000	0.000	0.000
15	0.144	-0.675	-0.719	3.240	2.860	1.106	0.000	0.000	0.000
16	0.524	-1.213	-0.284	2.580	2.083	1.578	0.000	0.000	0.000
17	0.397	-1.042	-0.351	2.768	2.267	1.480	0.000	0.000	0.000
18	0.288	-0.909	-0.426	2.956	2.451	1.382	0.000	0.000	0.000
19	0.190	-0.796	-0.511	3.144	2.635	1.284	0.000	0.000	0.000
20	0.101	-0.696	-0.613	3.332	2.819	1.187	0.000	0.000	0.000
21	0.019	-0.606	-0.748	3.520	3.003	1.089	0.000	0.000	0.000
22	0.459	-1.261	-0.232	2.672	2.042	1.658	0.000	0.000	0.000
23	0.342	-1.076	-0.295	2.860	2.226	1.560	0.000	0.000	0.000
24	0.239	-0.937	-0.364	3.048	2.410	1.463	0.000	0.000	0.000
25	0.146	-0.820	-0.440	3.236	2.594	1.365	0.000	0.000	0.000
26	0.060	-0.718	-0.528	3.424	2.778	1.267	0.000	0.000	0.000
27	-0.019	-0.625	-0.633	3.612	2.961	1.169	0.000	0.000	0.000
28	-0.094	-0.541	-0.779	3.799	3.145	1.071	0.000	0.000	0.000
29	0.399	-1.315	-0.184	2.765	2.001	1.738	0.000	0.000	0.000
30	0.290	-1.112	-0.243	2.952	2.185	1.641	0.000	0.000	0.000
31	0.192	-0.966	-0.307	3.140	2.369	1.543	0.000	0.000	0.000
32	0.103	-0.845	-0.377	3.328	2.553	1.445	0.000	0.000	0.000
33	0.021	-0.740	-0.455	3.516	2.736	1.347	0.000	0.000	0.000
34	-0.056	-0.645	-0.545	3.704	2.920	1.250	0.000	0.000	0.000
35	-0.129	-0.559	-0.655	3.892	3.104	1.152	0.000	0.000	0.000
36	-0.198	-0.479	-0.814	4.079	3.288	1.054	0.000	0.000	0.000
37	0.344	-1.380	-0.139	2.857	1.960	1.819	0.000	0.000	0.000
38	0.241	-1.151	-0.194	3.044	2.144	1.721	0.000	0.000	0.000
39	0.147	-0.996	-0.254	3.232	2.328	1.623	0.000	0.000	0.000
40	0.062	-0.870	-0.319	3.420	2.511	1.526	0.000	0.000	0.000
41	-0.018	-0.762	-0.390	3.608	2.695	1.428	0.000	0.000	0.000
42	-0.092	-0.666	-0.470	3.796	2.879	1.330	0.000	0.000	0.000
43	-0.163	-0.578	-0.563	3.984	3.063	1.232	0.000	0.000	0.000
44	-0.230	-0.496	-0.678	4.172	3.247	1.134	0.000	0.000	0.000
45	-0.294	-0.420	-0.856	4.359	3.431	1.037	0.000	0.000	0.000
46	0.292	-1.464	-0.095	2.949	1.919	1.899	0.000	0.000	0.000
47	0.194	-1.192	-0.148	3.137	2.103	1.801	0.000	0.000	0.000
48	0.105	-1.027	-0.205	3.324	2.286	1.704	0.000	0.000	0.000
49	0.022	-0.897	-0.265	3.512	2.470	1.606	0.000	0.000	0.000
50	-0.055	-0.785	-0.331	3.700	2.654	1.508	0.000	0.000	0.000
51	-0.128	-0.686	-0.403	3.888	2.838	1.410	0.000	0.000	0.000
52	-0.196	-0.597	-0.485	4.076	3.022	1.313	0.000	0.000	0.000
53	-0.262	-0.514	-0.581	4.264	3.206	1.215	0.000	0.000	0.000
54	-0.325	-0.437	-0.703	4.451	3.389	1.117	0.000	0.000	0.000
55	-0.385	-0.364	-0.911	4.639	3.573	1.019	0.000	0.000	0.000
56	0.242	-1.666	-0.054	3.041	1.878	1.980	0.000	0.000	0.000
57	0.149	-1.238	-0.104	3.229	2.061	1.882	0.000	0.000	0.000
58	0.063	-1.060	-0.158	3.417	2.245	1.784	0.000	0.000	0.000
59	-0.016	-0.924	-0.215	3.604	2.429	1.686	0.000	0.000	0.000
60	-0.091	-0.809	-0.277	3.792	2.613	1.589	0.000	0.000	0.000
61	-0.162	-0.708	-0.343	3.980	2.797	1.491	0.000	0.000	0.000
62	-0.229	-0.616	-0.417	4.168	2.981	1.393	0.000	0.000	0.000
63	-0.293	-0.532	-0.501	4.356	3.164	1.295	0.000	0.000	0.000
64	-0.355	-0.454	-0.600	4.544	3.348	1.197	0.000	0.000	0.000
65	-0.414	-0.380	-0.730	4.731	3.532	1.100	0.000	0.000	0.000
66	-0.471	-0.311	-1.042	4.919	3.716	1.002	0.000	0.000	0.000