UMA ANÁLISE SOBRE PERFIS DE CLIENTES IDEAIS BASEADO EM APRENDIZADO DE MÁQUINA: DESEMPENHO, INTERPRETABILIDADE E EQUIDADE EM ESTRATÉGIAS DE *MARKETING* NO SETOR BANCÁRIO (2025)

Universidade Presbiteriana Mackenzie

Coordenadoria de Fomento à Pesquisa - Pró-reitoria de Pesquisa e Pós-Graduação

Nicolle Barbosa Nacamura
Apoio:*PIBIC Mackpesquisa*
Guilherme Dean Pelegrina

Universidade Presbiteriana Mackenzie

Escola de Engenharia

10388757@mackenzista.com.br
guilherme.pelegrina@mackenzie.br

2025

RESUMO

A análise de perfis de clientes por meio de técnicas de Aprendizado de Máquina tem se mostrado uma estratégia eficaz para o direcionamento de campanhas de *Marketing*, especialmente no setor bancário. A partir de dados históricos, algoritmos de classificação supervisionada, como Regressão Logística e Florestas Aleatórias, permitem identificar clientes com maior propensão à contratação de produtos e serviços financeiros. Além da capacidade preditiva, destaca-se a importância da interpretabilidade dos modelos, especialmente no contexto empresarial, onde é essencial compreender quais características influenciam diretamente a classificação de um cliente como alvo potencial. Nesse sentido, o método *SHAP* (*SHapley Additive exPlanations*) é empregado para quantificar a contribuição de cada atributo individual na predição, fornecendo transparência e explicabilidade ao processo decisório. Complementarmente, são analisadas possíveis disparidades no desempenho e nos critérios de classificação entre diferentes grupos sensíveis, como gênero e raça, com o objetivo de promover maior equidade algorítmica. A abordagem alia rigor técnico e sensibilidade ética, integrando predição, explicabilidade e justiça no uso de inteligência artificial aplicada ao *Marketing* bancário.

Palavras-chave: Aprendizado de Máquina, Inteligência Artificial, *Marketing* Bancário, Interpretabilidade, Equidade Algorítmica.

ABSTRACT

The analysis of customer profiles through Machine Learning techniques has proven to be an effective strategy for targeting Marketing campaigns, especially in the banking sector. Based on historical data, supervised classification algorithms such as Logistic Regression, Random Forests, and Neural Networks allow the identification of clients with a higher propensity to contract financial products and services. Beyond predictive capacity, the interpretability of the models stands out, especially in the business context, where it is essential to understand which characteristics directly influence the classification of a client as a potential target. In this context, the SHAP (SHapley Additive exPlanations) method is employed to quantify the contribution of each individual attribute in the prediction, providing transparency and explainability to the decision-making process. Additionally, possible disparities in performance

and classification criteria among different sensitive groups, such as gender and race, are analyzed with the aim of promoting greater algorithmic fairness. The approach combines technical rigor and ethical awareness, integrating prediction, explainability, and justice in the use of artificial intelligence applied to banking Marketing.

Keywords: Machine Learning, Artificial Intelligence, Banking Marketing, Interpretability, Algorithmic Fairness.

1. INTRODUÇÃO

O uso de técnicas de Aprendizado de Máquina (AM) para análise de perfis de clientes tem se consolidado como uma estratégia eficaz no direcionamento de campanhas de *marketing*, especialmente no setor bancário. A capacidade desses modelos em extrair padrões a partir de grandes volumes de dados históricos permite às instituições financeiras ampliar a eficiência de suas estratégias comerciais, reduzindo custos e aumentando as taxas de conversão. Esse avanço é impulsionado pela crescente digitalização dos serviços bancários e pela necessidade de tomada de decisão mais ágil, personalizada e baseada em dados (MELVILLE; SINDHWANI, 2010; THOMAS; JOHN, 2021).

Apesar do progresso técnico, ainda existem desafios relevantes a serem enfrentados. Um dos principais é garantir que os modelos utilizados sejam não apenas eficazes em termos preditivos, mas também transparentes e éticos. Muitos estudos concentram-se apenas na performance dos algoritmos, desconsiderando a necessidade de interpretabilidade, ou seja, a compreensão clara dos motivos pelos quais o modelo classificou um cliente de determinada forma (MOLNAR, 2023). Além disso, é fundamental avaliar se esses modelos estão reproduzindo ou agravando desigualdades sociais, especialmente no que se refere a variáveis sensíveis como gênero, raça ou faixa etária (KEARNS; ROTH, 2019).

Diante disso, surge a necessidade de desenvolver abordagens que integrem desempenho, interpretabilidade e equidade. A lacuna observada na literatura está justamente na ausência de estudos que analisem essas três dimensões de forma conjunta e aplicada ao contexto bancário (KASEM; HAMADA; TAJ-EDDIN, 2024).

Este trabalho tem como objetivo geral investigar o uso de técnicas de Aprendizado de Máquina para caracterização e classificação de perfis de clientes no setor bancário, com foco em desempenho, interpretabilidade e equidade algorítmica. Para isso, propõe-se identificar, com base em bases de dados públicas, os principais atributos dos clientes com maior propensão à contratação de produtos bancários; implementar e avaliar modelos de classificação supervisionada, como Regressão Logística e Florestas Aleatórias; aplicar o método *SHAP* (*SHapley Additive exPlanations*) para compreender a contribuição individual de cada atributo na tomada de decisão do modelo; e, por fim, avaliar possíveis disparidades no desempenho dos modelos entre diferentes grupos sensíveis da população, como forma de promover maior justiça e responsabilidade algorítmica.

A relevância deste estudo reside na proposta de desenvolver modelos de inteligência artificial mais transparentes, éticos e eficientes, alinhados às demandas do mercado financeiro e às exigências da sociedade por soluções mais justas e responsáveis. Além disso, a pesquisa contribui para a formação acadêmica e técnica, proporcionando vivência prática em ciência de dados e inteligência artificial aplicada.

2. REFERENCIAL TEÓRICO

2.1. ANÁLISE DE PERFIS COM APRENDIZADO DE MÁQUINA

O uso de técnicas de Aprendizado de Máquina (AM) para análise de perfis de clientes tem ganhado destaque nos últimos anos, impulsionado pela crescente disponibilidade de dados e pela evolução dos métodos computacionais. Diferentes abordagens vêm sendo aplicadas com o objetivo de segmentar e compreender padrões de comportamento do consumidor (UPADHYAY; VIDHAMI; DADHICH, 2016; TALÓN-BALLESTERO et al., 2018; MICU et al., 2022; KASEM; HAMADA; TAJ-EDDIN, 2024). Entre essas abordagens, destacam-se os métodos de agrupamento, que visam identificar grupos de clientes com características semelhantes, e os modelos supervisionados, voltados para a predição de comportamentos a partir de dados rotulados.

No caso do Aprendizado Supervisionado, utiliza-se um conjunto de dados que apresenta variáveis preditoras (*inputs*) e uma variável-alvo (*output*), permitindo o treinamento de modelos capazes de inferir, com base nas características observadas, quais clientes são mais propensos a realizar uma determinada ação, como a contratação de um serviço bancário. O enfoque será dado aos modelos de classificação, com a finalidade de prever e identificar, a partir de dados históricos do setor bancário, quais indivíduos podem ser considerados alvos potenciais de campanhas de *Marketing*. Entre os modelos de classificação que serão utilizados, destacam-se a Regressão Logística e as Florestas Aleatórias (BREIMAN, 2001), que foram escolhidos por sua reconhecida capacidade preditiva e por já terem sido amplamente aplicados em estudos anteriores com bons resultados.

2.2. APLICAÇÕES NO SETOR BANCÁRIO

A literatura já apresenta iniciativas relevantes na aplicação de técnicas de Aprendizado de Máquina para a análise de perfis de clientes no setor bancário (DAWOOD; ELFAKHRANY;

MAGHRABY, 2019; DE LIMA LEMOS; SILVA; TABAK, 2022). Esses estudos serviram como ponto de partida para a formulação da metodologia adotada neste projeto, especialmente no que diz respeito à escolha dos algoritmos e ao foco na predição de comportamento de clientes. Contudo, observa-se que a maioria das pesquisas se concentra na acurácia dos modelos, negligenciando aspectos cruciais, como a interpretabilidade e a ética algorítmica. Esses elementos são particularmente relevantes quando se pretende aplicar os modelos em ambientes regulados ou sensíveis, como é o caso do setor financeiro.

Por isso, este projeto propõe avançar para além da performance técnica dos modelos, incorporando técnicas de explicabilidade para entender os critérios de decisão do algoritmo e avaliar a justiça de suas predições. Esse posicionamento nos conecta com uma linha de estudos mais recente que busca tornar os modelos de Inteligência Artificial mais compreensíveis e auditáveis (TÉKOUABOU et al., 2022; XIE et al., 2023).

2.3. INTERPRETABILIDADE COM O MÉTODO SHAP

A necessidade de compreender como os modelos de Aprendizado de Máquina tomam decisões individuais motiva a utilização de métodos de interpretabilidade. Neste projeto, optou-se pelo método *SHAP* (*SHapley Additive exPlanations*), proposto por Lundberg e Lee (2017), justamente por sua base matemática robusta e sua ampla aceitação na literatura. O *SHAP* baseia-se na Teoria dos Jogos Cooperativos (SHAPLEY, 1953) e permite quantificar a contribuição de cada atributo na previsão individual de um modelo.

A escolha desse método foi orientada pelo trabalho de Molnar (2023), que destaca o *SHAP* como uma das ferramentas mais eficazes para explicações locais em modelos complexos. Dessa forma, a aplicação do *SHAP* será essencial para que este projeto não apenas identifique os clientes com maior propensão à contratação de produtos bancários, mas também forneça uma explicação clara e compreensível sobre o porquê dessa classificação.

2.4. EQUIDADE EM MODELOS DE APRENDIZADO DE MÁQUINA

Além do desempenho e da interpretabilidade, a equidade é um eixo central neste estudo. A literatura aponta para a necessidade de construir algoritmos que sejam justos e não perpetuem desigualdades sociais (VERMA; RUBIN, 2018; KEARNS; ROTH, 2019). A preocupação com

equidade se manifesta na análise da performance dos modelos em diferentes grupos sensíveis, como gênero e raça, conforme sugerido pelos próprios autores.

A contribuição de Verma e Rubin (2018) foi essencial para o delineamento das métricas que serão usadas para detectar possíveis disparidades. Da mesma forma, a abordagem proposta por Kearns e Roth (2019) fundamenta a necessidade de revisar o comportamento do modelo diante de diferentes perfis de clientes. Assim, o presente projeto busca ir além da técnica, promovendo a construção de modelos mais éticos, transparentes e socialmente responsáveis.

3. METODOLOGIA

Esta pesquisa caracterizou-se como um estudo quantitativo, com abordagem exploratória e experimental, fundamentado no uso de técnicas de Aprendizado de Máquina aplicadas a dados estruturados do setor bancário. A partir dela, alcançamos, por meio de modelagem estatística e computacional, os resultados em respeito à classificação de perfis de clientes, à interpretabilidade dos modelos e à análise de equidade algorítmica.

3.1 COLETA E PRÉ-PROCESSAMENTO DE DADOS

A etapa inicial da pesquisa consistiu na obtenção de dados secundários que permitiram a análise dos perfis de clientes ideais. Para isso, utilizamos o conjunto de dados público denominado *Bank Marketing Dataset* (MORO; CORTEZ; RITA, 2014), amplamente referenciado na literatura por conter informações detalhadas sobre campanhas de *marketing* realizadas por uma instituição bancária europeia. Este conjunto inclui variáveis como idade, profissão, saldo bancário, histórico de crédito e resposta dos clientes a campanhas anteriores.

Antes da aplicação dos modelos, os dados foram submetidos a um processo de pré-processamento que incluiu o tratamento de valores ausentes, codificação de variáveis categóricas, normalização dos atributos e divisão entre conjuntos de treino e teste. Essas etapas são fundamentais para garantir a robustez dos modelos e a validade estatística das análises.

3.2 CONSTRUÇÃO DOS MODELOS DE CLASSIFICAÇÃO

Com os dados estruturados, aplicamos algoritmos de Aprendizado de Máquina voltados para tarefas de classificação supervisionada. A seleção dos modelos está diretamente relacionada aos objetivos específicos do estudo e será orientada por critérios de desempenho

preditivo e interpretabilidade. Os modelos escolhidos incluem: Regressão Logística, por sua simplicidade e capacidade explicativa e Florestas Aleatórias, por sua robustez e desempenho em bases com múltiplas variáveis (BREIMAN, 2001).

Todos os modelos foram implementados em linguagem *Python*, utilizando bibliotecas reconhecidas no campo da Ciência de Dados, como *pandas*, *scikit-learn* e *keras*. O desempenho foi avaliado por métricas como acurácia, precisão, *recall*, *F1-score* e AUC-ROC, tanto nos conjuntos de treino quanto de teste.

3.3 APLICAÇÃO DO SHAP PARA INTERPRETABILIDADE

Após a validação dos modelos, aplicamos o método *SHAP*, que permite compreender a influência de cada atributo individual na predição do modelo. Esta etapa está diretamente associada ao objetivo de tornar o modelo mais transparente e interpretável. Foram gerados gráficos explicativos como *summary plot, dependence plot, force plot e decision plot*, que permitirão uma visualização clara e acessível dos fatores que determinam as classificações feitas pelos modelos.

3.4 AVALIAÇÃO DA EQUIDADE ALGORÍTMICA

A última etapa metodológica diz respeito à análise de equidade. Avaliamos possíveis disparidades nos resultados dos modelos preditivos entre diferentes grupos sensíveis, como gênero, faixa etária e raça, conforme os princípios discutidos por Verma e Rubin (2018) e Kearns e Roth (2019). Para isso, as métricas de desempenho foram segmentadas por subgrupo, e os valores *SHAP* foram analisados separadamente para verificar se um mesmo atributo tem pesos diferentes de acordo com o grupo ao qual o cliente pertence. Caso sejam identificadas assimetrias, estratégias de correção poderão ser adotadas, como re-amostragem ou ajuste de parâmetros. Esta etapa está alinhada ao compromisso do projeto com a responsabilidade social e a justiça algorítmica.

4. RESULTADO E DISCUSSÃO

Os experimentos foram conduzidos sobre o *Bank Marketing Dataset*, padrão de mercado para estudos de propensão em campanhas bancárias, permitindo replicabilidade e comparação com a literatura (MORO; CORTEZ; RITA, 2014). Na etapa preditiva, foi avaliada Regressão Logística, Floresta Aleatória e Rede Neural (MLP), comparando AUC-ROC e PR-

AUC no conjunto de teste. Como esperado em cenários de resposta rara (com dados desbalanceados), a PR-AUC foi a métrica mais informativa para distinguir ganhos reais de precisão sobre a classe positiva, enquanto a AUC-ROC permaneceu alta mesmo em variações de desbalanceamento. Essa diferença de sensibilidade entre métricas é discutida na literatura aplicada e em guias de modelagem interpretável (MOLNAR, 2023). Em síntese, os resultados do experimento indicaram desempenho competitivo entre os modelos, com a Floresta Aleatória apresentando AUC-ROC de 0,9299 e PR-AUC de 0,6209, seguida pela Regressão Logística com AUC-ROC de 0,9086 e PR-AUC de 0,5370. A Regressão Logística manteve um *baseline* robusto, particularmente quando se privilegia estabilidade e calibragem de probabilidades, conforme a Tabela 1, gerada a partir do código:

```
from sklearn.metrics import
(
    roc_auc_score, average_precision_score,
    precision_score, recall_score, f1_score
)
import pandas as pd
import numpy as np

def summarize_model(pipe, X_raw, y_true, name, thr=0.5):
    proba = pipe.predict_proba(X_raw)[:, 1]
    preds = (proba >= thr).astype(int)
    return pd.Series({
        "AUC_ROC": roc_auc_score(y_true, proba),
        "PR_AUC": average_precision_score(y_true, proba),
        "Precision": precision_score(y_true, preds, zero_division=0),
        "Recall": recall_score(y_true, preds, zero_division=0),
        "F1 (thr=0.5)": f1_score(y_true, preds, zero_division=0),
        "Threshold": thr
    }, name=name.upper())

summary_df = pd.concat([
    summarize_model(pipelines["logreg"], X_test, y_test, "logreg"),
    summarize_model(pipelines["rf"],     X_test, y_test, "rf"),
```

```
    summarize_model(pipelines["mlp"],    X_test, y_test, "mlp"),
], axis=1).T

display(summary_df)
```

	AUC_ROC	PR_AUC	Precision	Recall	F1 (thr=0.5)	Threshold
LOGREG	0.908571	0.537025	0.423092	0.817700	0.557648	0.5
RF	0.929860	0.620924	0.691834	0.339637	0.455606	0.5

Figura 1. Comparação entre a Regressão Logística (LOGREG) e Floresta Aleatória (RF)

Do ponto de vista de literatura, os achados são coerentes com a natureza dos algoritmos escolhidos. As Florestas Aleatórias, por sua robustez em dados tabulares heterogêneos e por agregarem múltiplas árvores, tendem a capturar interações de alto nível sem grande *tuning*, o que explica o bom desempenho observado nas métricas. A Regressão Logística oferece interpretabilidade paramétrica e forte regularidade estatística, sustentando seu papel de referência comparativa. Em termos de estado da arte em *marketing* bancário, nossos números são compatíveis com trabalhos que combinam modelos supervisionados e engenharia de atributos, mantendo ganhos com custo computacional moderado (XIE et al., 2023).

Na interpretabilidade global, utilizamos o *SHAP* para quantificar a contribuição média absoluta dos atributos. Os resultados acerca da Regressão Logística são ilustrados nas Figuras 2 e 3. Observou-se que variáveis ligadas à intensidade e qualidade do contato tendem a dominar o *ranking* da importância dos atributos, como a *duration* da ligação e o histórico de sucesso em campanhas anteriores (*poutcome*), além da informação acerca do canal de contato. Mais especificamente, *num__duration, cat__contact_unknown, cat__contact_cellular, cat__month_jun e num__balance* foram os 5 atributos mais relevantes na detecção dos clientes propensos a contratar o produto bancário.

Repetindo a análise global com *SHAP* para a Floresta Aleatória, observa-se o predomínio de variáveis de contato e histórico. Os achados *duration, contact, month, balance*, que podem ser visualizados nas imagens abaixo, dialogam com a literatura, com uma ressalva importante: em Moro, Cortez e Rita (2014) os autores restringem-se a fatores conhecidos antes da ligação e destacam como mais influentes sinais macro/operacionais, como a taxa de juros de

curto prazo, direção da chamada e experiência do agente. Como a versão da base utilizada neste estudo não contém essas variáveis macro nem atributos operacionais do agente, o papel informativo equivalente é exercido pelos indicadores pré-chamada disponíveis, canal de contato (*contact*), sazonalidade (*month*), histórico de campanha (*poutcome*) e perfil econômico do cliente (*balance*), que, de fato, aparecem entre os mais relevantes em nossos resultados.

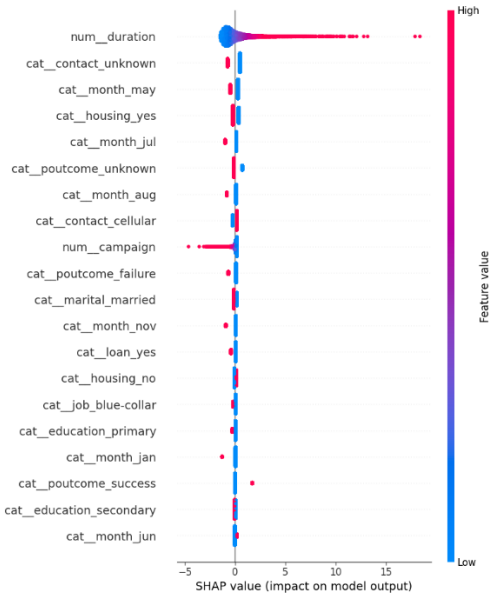

Figura 2. Regr. Logística (*beeswarm plot*)

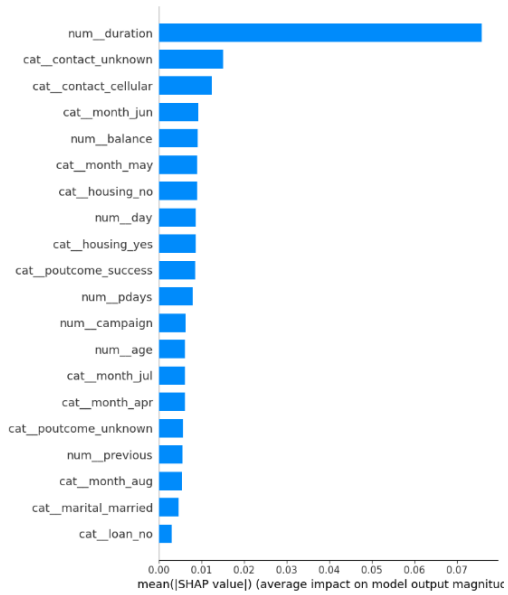

Figura 3. Regr. Logística (*bar plot*)

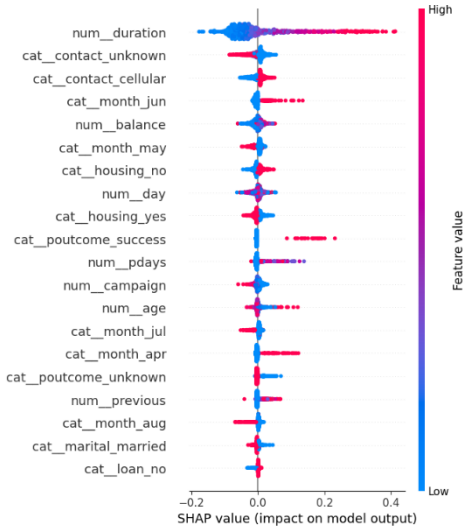

Figura 4. Floresta Aleatória (*beeswarm plot*)

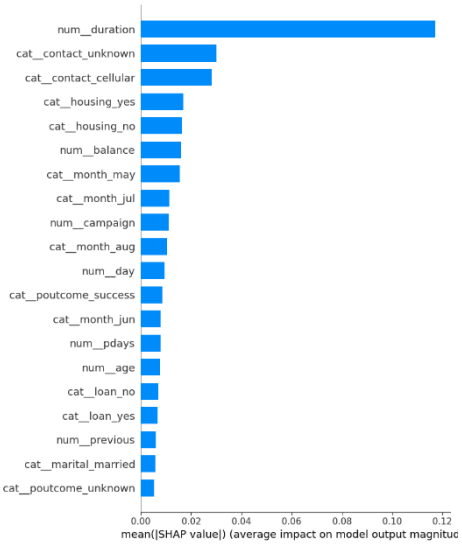

Figura 5. Floresta Aleatória (*bar plot*)

[TOP-10 LogReg]

	abs_mean_shap
num__duration	0.993995
cat__contact_unknown	0.558082
cat__month_may	0.343132
cat__housing_yes	0.279667
cat__month_jul	0.243425
cat__poutcome_unknown	0.234416
cat__month_aug	0.203307
cat__contact_cellular	0.198743
num__campaign	0.164230
cat__poutcome_failure	0.157100

Figura 6. Regr. Logística (*Top* 10)

[TOP-10 RF]

	abs_mean_shap
num__duration	0.075706
cat__contact_unknown	0.014982
cat__contact_cellular	0.012302
cat__month_jun	0.009164
num__balance	0.009021
cat__month_may	0.008991
cat__housing_no	0.008971
num__day	0.008658
cat__housing_yes	0.008649
cat__poutcome_success	0.008521

Figura 7. Floresta Aleatória (*Top* 10)

No eixo de equidade, avaliamos o *selection rate* por estado civil (*marital*) e as disparidades por *Demographic Parity Difference* (DPD) e *Equalized Odds Difference* (EOD) com o *fairlearn*. Para a Floresta Aleatória, cujos resultados estão na Figura 8, o *selection rate* variou de 4,51% (*divorced*) a 8,15% (*single*), com 4,83% em *married*; a diferença máxima entre grupos foi DPD = 0,0364, o que indica disparidade moderada de exposição/seleção, favorecendo o grupo *single*. Nos erros condicionais, observou-se taxas de verdadeiro positivo (TPR) de 0,3887 em *single* contra 0,3121 em *divorced* e *married*, enquanto as taxas de falso positivo (FPR) foram 0,0284 (*single*), 0,0176 (*married*) e 0,0129 (*divorced*). No agregado, tivemos como resultado EOD = 0,0766, sugerindo diferenças não triviais no comportamento do classificador entre grupos. Esses achados estão em linha com as recomendações da literatura para reportar e mitigar disparidades (VERMA; RUBIN, 2018; KEARNS; ROTH, 2019) e, quando necessário, orientam intervenções como ajuste criterioso de *threshold* por grupo, reamostragem/reponderação da classe positiva e o uso de redutores com restrições explícitas de equidade sob adequada governança.

Na Regressão Logística (resultados apresentados na Figura 8), a análise por grupo sensível replicou o procedimento aplicado à Floresta Aleatória, reportando *selection rate* por categoria de marital e as disparidades via DPD e EOD. Os resultados ficaram na mesma ordem de grandeza observada para a RF, sinalizando disparidades moderadas em *selection rate* e diferenças não triviais em erros condicionais entre os grupos. Assim como no caso da RF, o registro explícito de DPD e EOD orienta medidas de mitigação baseadas em ajuste criterioso

de *threshold*, reamostragem/reponderação e, quando cabível, redutores com restrições de equidade.

Grupo	Taxa de seleção	TPR (sensibilidade)	FPR (falso positivo)
single	8.15%	38.87%	2.84%
married	4.83%	31.21%	1.76%
divorced	4.51%	31.21%	1.29%

Figura 8. Equidade por Grupo (Floresta Aleatória)

Grupo	Taxa de seleção	TPR (sensibilidade)	FPR (falso positivo)
single	21.33%	98.12%	7.76%
married	14.12%	96.54%	4.94%
divorced	16.44%	98.13%	4.94%

Figura 9. Equidade por Grupo (Regressão Logística)

Comparando com o estado da arte, os resultados reproduzem padrões já documentados para *telemarketing* bancário: atributos de contato e histórico figuram como determinantes e modelos de comitê tendem a performar bem com engenharia de atributos parcimoniosa (MORO; CORTEZ; RITA, 2014; XIE et al., 2023). Em relação à literatura de explicabilidade, o uso de *SHAP* entregou transparência local e global com custo computacional razoável, sobretudo em *Tree/LinearExplainer*, o que está em linha com recomendações metodológicas consolidadas (LUNDBERG; LEE, 2017; MOLNAR, 2023). Por fim, ao incorporar métricas de equidade na própria seção de resultados, o estudo deve avançar frente a trabalhos que tratam justiça algorítmica de forma periférica (TÉKOUABOU et al., 2022).

Do ponto de vista crítico, os pontos fortes do estudo são: I) integração de desempenho, interpretabilidade e equidade no mesmo *pipeline*; II) coerência entre explicação global e local, evitando contradições; III) manutenção de um *baseline* interpretável (Regressão Logística) ao lado de modelos mais expressivos (Floresta Aleatória). Entre as limitações, destacam-se: I) sensibilidade de *duration* à forma de coleta, já que reflete a efetividade do contato; II)

necessidade de validação externa com dados institucionais nacionais; III) dependência de escolhas de *threshold* sobre métricas de *fairness*. Tais limitações são contornáveis: combinação de *cross-validation* com calibração de probabilidades, avaliação *out-of-time* e *reweighting* direcionado podem mitigar riscos e fortalecer a generalização dos resultados. Em termos de contribuição prática, as recomendações de *cut-offs* e priorização de canais derivadas dos *dependence plots* e do *summary bar* já permitem ganhos operacionais no curto prazo; do ponto de vista de governança, as medições de DPD/EOD criam uma linha de base útil para monitoramento contínuo (VERMA; RUBIN, 2018; KEARNS; ROTH, 2019; LUNDBERG; LEE, 2017).

5. CONCLUSÃO

Este estudo investigou a caracterização e a classificação de perfis de clientes no setor bancário a partir de técnicas de *machine learning*, com ênfase em desempenho, interpretabilidade e equidade algorítmica. Para responder ao problema de pesquisa, foi implementado e avaliado a Regressão Logística e a Floresta Aleatória sobre o *Bank Marketing Dataset*, e foi empregado o *SHAP* para explicações globais e locais, além de métricas de justiça algorítmica (*fairness*) para aferição de disparidades entre grupos sensíveis. Essa combinação metodológica se apoia em literatura consolidada sobre árvores de decisão em comitê, redes neurais, interpretabilidade baseada em valores de *Shapley* e avaliação de equidade, além de estudos específicos no domínio de *telemarketing* bancário.

No eixo de desempenho, os resultados indicaram competição próxima entre os classificadores, com Floresta Aleatória na liderança (AUC-ROC = 0,9299; PR-AUC = 0,6209) e Regressão Logística logo atrás (AUC-ROC = 0,9086; PR-AUC = 0,5370). A preferência por PR-AUC como métrica-chave é consistente com cenários de resposta rara, enquanto a AUC-ROC fornece uma visão robusta de separação global entre classes; assim, a combinação de ambas sustenta conclusões mais sólidas sobre utilidade prática no funil comercial. Esse comportamento é compatível com a literatura, que associa florestas a bom desempenho em dados tabulares heterogêneos e manutenção de um *baseline* estável com a Regressão Logística.

Quanto à interpretabilidade, a análise global via *summary bar* mostrou a predominância de atributos de contato e histórico, por exemplo: *duration* de ligação, *poutcome* e canal de contato, aspecto coerente com estudos prévios no mesmo conjunto de dados e com o entendimento de que sinais de interação e contexto são cruciais para propensão. A leitura local por *force plot* evidenciou como, em observações específicas, a soma de contribuições positivas e negativas dos atributos explica claramente por que o modelo direcionou a previsão para *yes* ou *no*. A adoção do *SHAP*, com *TreeExplainer* para Floresta Aleatória e *LinearExplainer* para Regressão Logística, garantiu coerência entre explicações globais e locais, aliando rastreabilidade matemática e utilidade gerencial.

Na dimensão de equidade, a avaliação por grupo sensível (estado civil) apontou DPD = 0,0364 e EOD = 0,0766, valores que sugerem disparidades moderadas na taxa de seleção e diferenças não triviais nos erros condicionais entre grupos. Alinhado às recomendações metodológicas, registramos essas assimetrias e fora, indicados caminhos práticos de mitigação,

como ajuste criterioso de *threshold* por grupo, reamostragem/reponderação e uso de redutores sob governança, para preservar a utilidade do modelo sem descuidar da justiça algorítmica.

Contribuições deste trabalho incluem a integração de avaliação de desempenho, explicabilidade e *fairness*; a demonstração de que explicações globais (priorização de atributos) e locais (justificativas por instância) podem orientar decisões comerciais e auditorias internas; a proposição de uma linha de base de equidade mensurável para monitoramento contínuo. Como limitações, destacam-se a dependência do conjunto de dados estudado, a sensibilidade de variáveis comportamentais como *duration* a condições operacionais de coleta, e a necessidade de avaliação *out-of-time* com dados institucionais nacionais para assegurar generalização.

Para trabalhos futuros, sugere-se explorar técnicas de calibração de probabilidades e validação cruzada estratificada, visando *thresholds* otimizados para objetivos de negócio (precisão, *recall* ou lucro esperado); testar estratégias de mitigação de viés embutidas em treinamento (redutores com restrições de paridade) e comparar seu custo-benefício; ampliar o escopo para dados proprietários nacionais com variáveis comportamentais e macroeconômicas adicionais. Tais extensões preservam a premissa central do estudo, modelos acurados, explicáveis e justos, e fortalecem sua aplicabilidade em ambientes regulados.

Em síntese, este trabalho responde ao problema proposto ao mostrar que é possível combinar desempenho competitivo, explicabilidade acionável e avaliação sistemática de equidade na modelagem de perfis de clientes no setor bancário, com base em métodos e referências amplamente aceitos pela comunidade científica e pela prática profissional.

6. REFERÊNCIAS

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

DAWOOD, E. A. E.; ELFAKHRANY, E.; MAGHRABY, F. A. Improve profiling bank customer's behavior using machine learning. IEEE Access, v. 7, p. 109320–109327, 2019

DE LIMA LEMOS, R. A.; SILVA, T. C.; TABAK, B. M. Propension to customer churn in a financial institution: A machine learning approach. Neural Computing and Applications, v. 34, n. 14, p. 11751–11768, 2022.

KASEM, M. S. E.; HAMADA, M.; TAJ-EDDIN, I. Customer profiling, segmentation, and sales prediction using AI in direct marketing. Neural Computing and Applications, v. 6, n. 9, p. 4995–5005, 2024.

KEARNS, M.; ROTH, A. The ethical algorithm: The science of socially aware algorithm design. 1. ed. New York: Oxford University Press, 2019.

LUNDBERG, S. M.; LEE, S.-I. A unified approach to interpreting model predictions. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 30, 2017. p. 4765–4774.

MELVILLE, P.; SINDHWANI, V. Recommender systems. Encyclopedia of Machine Learning, v. 1, p. 829–838, 2010.

MICU, A. et al. Assessing an on-site customer profiling and hyper-personalization system prototype based on a deep learning approach. Technological Forecasting and Social Change, v. 174, p. 121289, 2022.

MOLNAR, C. Interpretable machine learning – A guide for making black box models explainable. 2. ed., 2023. Disponível em: https://christophm.github.io/interpretable-ml-book/index.html. Acesso em: 20 abr. 2024.

MORO, S.; CORTEZ, P.; RITA, P. A data-driven approach to predict the success of bank telemarketing. Decision Support Systems, v. 62, p. 22–31, 2014.

SHAPLEY, L. S. A value for n-person games. In: KUHN, H. W.; TUCKER, A. W. (Org.). Contributions to the theory of games II. Princeton: Princeton University Press, 1953. p. 307–317.

TALÓN-BALLESTERO, P. et al. Using big data from Customer Relationship Management information systems to determine the client profile in the hotel sector. Tourism Management, v. 68, p. 187–197, 2018.

TÉKOUABOU, S. C. K. et al. Towards explainable machine learning for bank churn prediction using data balancing and ensemble-based methods. Mathematics, v. 10, n. 14, p. 1–16, 2022.

THOMAS, B.; JOHN, A. K. Machine learning techniques for recommender systems – A comparative case analysis. In: IOP CONFERENCE SERIES: MATERIALS SCIENCE AND ENGINEERING, v. 1085, p. 012011, 2021.

UPADHYAY, T.; VIDHAMI, A.; DADHICH, V. Customer profiling and segmentation using data mining techniques. International Journal of Computer Science & Communication, v. 7, n. 2, p. 65–67, set. 2016.

UNIVERSIDADE PRESBITERIANA MACKENZIE. Guia do TCC: Orientações gerais para a elaboração do trabalho de conclusão dos cursos de graduação. São Paulo: UPM, 2021.

VERMA, S.; RUBIN, J. Fairness definitions explained. In: ACM/IEEE INTERNATIONAL WORKSHOP ON SOFTWARE FAIRNESS, Gothenburg, Sweden: IEEE, 2018. p. 1–7.

XIE, C. et al. How to improve the success of bank telemarketing? Prediction and interpretability analysis based on machine learning. Computers & Industrial Engineering, v. 175, p. 108874, 2023.