

AVALIAÇÃO DO IMPACTO DO RETREINO PERIÓDICO NA PREDIÇÃO DE NÃO- COMPARECIMENTOS DE PACIENTES EM CONSULTAS MÉDICAS UTILIZANDO BALANCED RANDOM FOREST

Ciências Exatas e da Terra, Volume 29 - Edição 146/MAI 2025 / 14/05/2025

EVALUATING THE IMPACT OF PERIODIC RETRAINING ON PATIENT NO-SHOW PREDICTION IN MEDICAL APPOINTMENTS USING BALANCED RANDOM FOREST

REGISTRO DOI: 10.69849/revistaft/ni10202505142025

Carlos Eduardo Gonçalves de Oliveira¹; Ana Beatriz Marinho de Jesus
Teixeira²; Mateus Simão Oliveira³; Aldo André Diaz Salazar⁴; Emerson
Nobuyuki Itikawa⁵

Resumo

Este estudo avalia o impacto do retreino periódico na performance preditiva de modelos de aprendizado de máquina no contexto de predição de não-comparecimentos de pacientes a consultas médicas previamente agendadas, fenômeno conhecido por gerar desperdício de recursos e redução da eficiência em instituições de saúde. Utilizando dados reais da Clínica Radiológica de Anápolis (CRA), referentes ao período de 2015 a 2023, compararam-se cenários em que o modelo Balanced Random Forest (BRF) foi treinado apenas uma vez com cenários nos quais o treinamento foi realizado trimestralmente, considerando janelas históricas de 1, 2 e 3 anos. O cenário com retreino periódico apresentou desempenho superior, especialmente em métricas críticas para classes desbalanceadas, como F1-score e Área sob a Curva Precision-Recall (AUC-PR), destacando-se as janelas de 1 e 2 anos. A análise comparativa das janelas temporais não indicou diferenças estatisticamente significativas entre elas, mas sugeriu-se o uso da janela de 2 anos devido ao bom balanço entre desempenho e custo computacional. Os resultados reforçam a importância do retreino periódico para melhorar a capacidade preditiva de modelos no contexto dinâmico das instituições de saúde. Estudos futuros poderiam explorar técnicas de aprendizado incremental, bem como otimização avançada de hiperparâmetros, com potencial de melhora das métricas de performance.

Palavras-chave: Não-comparecimento. Machine Learning. Balanced Random Forest. Retreino periódico. Consultas médicas.

1 INTRODUÇÃO

Os casos de não-comparecimento ocorrem quando um paciente falta a uma consulta ou a um procedimento agendado sem notificação prévia (Marbouh *et al.*, 2020). Mais especificamente, eles representam horários e recursos subutilizados no contexto das instituições de saúde (Marbouh *et al.*, 2020). Num cenário de agendamento prévio, os pacientes em potencial são impactados, haja vista que se tornam incapazes de agendar

seus procedimentos em tempo hábil, enquanto o paciente que não comparece pode ter a saúde afetada, tendo em vista que o diagnóstico ou o tratamento são adiados (Marbouh *et al.*, 2020). Há também o impacto financeiro para as instituições de saúde, pois os não-comparecimentos levam a uma redução na eficiência operacional, o que pode ser inviável, por exemplo, para departamentos de radiologia, onde as máquinas de diagnóstico tendem a ser muito caras e um alto volume de pacientes é exigido (Marbouh *et al.*, 2020).

A dimensão do problema pode ser evidenciada em estudos que relatam sua prevalência em diferentes continentes e especialidades médicas. Particularmente, a prevalência do nãocomparecimento de pacientes na América do Sul foi reportada como sendo de 27.8% de acordo com uma revisão sistemática sobre esse tópico (Dantas *et al.*, 2018), estando atrás apenas da África, com prevalência de 43.0%. Em relação a como essa prevalência varia por modalidade de procedimento, os valores foram maiores para Fisioterapia (57.3%), Endocrinologia (36.0%) e Cardiologia (30.0%) (Dantas *et al.*, 2018).

Além da alta prevalência, os impactos financeiros dos não-comparecimentos não carecem de exemplos. Em uma clínica de endoscopia, uma taxa de absenteísmo de 18% foi associada a uma redução de 16.4% na receita final (Berg *et al.*, 2013). Em um laboratório vascular, uma taxa de 12% resultou em perdas anuais estimadas em US\$89107,00, desconsiderando os custos operacionais relacionados à ineficiência do agendamento (Satiani; Miller; Patel, 2009). O mesmo estudo indicou que a redução dessa taxa em apenas 5% poderia elevar os lucros em até US\$51769,00 (Satiani; Miller; Patel, 2009). Em contextos de saúde pública, mesmo pequenas reduções nas taxas de não-comparecimento podem representar economias significativas (Satiani; Miller; Patel, 2009), e isso é relevante para países como o Brasil.

Apesar da ampla gama de estudos acerca dos impactos dos não-comparecimentos de pacientes, ainda não há consenso quanto aos

fatores mais preditivos desses eventos, que incluem variáveis demográficas do paciente, histórico prévio de faltas, distância até a instituição e aspectos temporais do agendamento, como hora, dia da semana e mês (CarrerasGarcía *et al.*, 2020; Dantas *et al.*, 2018). Entretanto, há o consenso de que o comportamento de não-comparecimento não é aleatório, e isso tem motivado a aplicação de modelos de aprendizado de máquina, com potencial para apoiar estratégias como o overbooking e o uso otimizado de recursos (Carreras-García *et al.*, 2020).

Recentemente, diversos modelos preditivos têm sido propostos para prever o nãocomparecimento de pacientes. Em uma revisão abrangente realizada em (Toffaha *et al.*, 2025), constatou-se que modelos baseados em regressão logística (LaValley, 2008) são os mais utilizados, presentes em cerca de 68% dos estudos analisados. Contudo, técnicas mais sofisticadas, como Random Forest (Breiman, 2001) e XGBoost (Chen; Guestrin, 2016), demonstraram desempenho superior, com os melhores modelos alcançando valores de Área Sob a Curva ROC (AUC-ROC) entre 0,75 e 0,95 e acurácia até 99,44%.

Além da escolha dos modelos preditivos mais eficazes, existem importantes desafios relacionados à qualidade e à completude dos dados utilizados para treinamento, ao desbalanceamento entre as classes (uma vez que eventos de não-comparecimento geralmente são minoritários nos conjuntos de dados) e às particularidades organizacionais que influenciam a aplicabilidade prática e cotidiana dessas ferramentas. Para lidar com problemas relacionados à completude dos dados, estratégias como imputação de valores ausentes ou inclusão de variáveis indicadoras têm sido amplamente utilizadas (García-Laencina; Sancho-

Gómez; Figueiras-Vidal, 2010; Wujek; Hall; Güneş, 2016). Já o desbalanceamento de classes é frequentemente abordado por meio de técnicas de reamostragem, como o Synthetic Minority Over-sampling Technique (SMOTE) (Alkhaldeh; Albalkhi; Naswhan, 2023) e o Random Under-Sampling (RUS). Contudo, um aspecto ainda pouco explorado é a

importância da dinâmica temporal dos dados utilizados no treinamento, especialmente em relação à estabilidade das previsões ao longo do tempo.

Considerando que o comportamento de pacientes e as condições institucionais são dinâmicos ao longo do tempo, fenômeno conhecido como *concept drift* no contexto de aprendizado de máquina (Lu *et al.*, 2019), é essencial avaliar não somente a performance pontual dos modelos preditivos, mas também a importância da atualização periódica desses modelos. Dessa forma, o presente estudo visa avaliar o impacto do retreino periódico dos modelos de aprendizado de máquina sobre a previsão do não-comparecimento de pacientes em procedimentos previamente agendados, identificando qual janela temporal de treinamento (1, 2 ou 3 anos) é mais eficaz. Para isso, comparou-se cenários em que modelos treinados uma única vez são aplicados para prever múltiplos trimestres futuros com o cenário no qual os modelos são retreinados trimestralmente, utilizando dados reais da Clínica Radiológica de Anápolis (CRA), Goiás, Brasil.

2 METODOLOGIA

2.1 Conjunto de dados

Os dados utilizados neste estudo foram coletados diretamente do sistema de registros da CRA, localizada em Anápolis, Goiás, Brasil. A CRA é uma instituição de diagnóstico por imagem, oferecendo serviços, como Angiotomografia de Coronárias, Tomografia Computadorizada, Ressonância Magnética, Densitometria Óssea, Medicina Nuclear, Ultrassom e de consultas médicas. Devido à variedade de serviços oferecidos, a CRA ameniza a necessidade da população de Anápolis (398869 habitantes, de acordo com o Censo Demográfico de 2022) de se deslocar até Goiânia (localizada a cerca de 55 km de distância), capital de Goiás, em busca de serviços avançados de diagnóstico.

Para este estudo, foram considerados somente os dados referentes ao serviço de consultas médicas. Identificadores dos pacientes foram remapeados para números aleatórios visando assegurar o anonimato e evitar o rastreamento dos pacientes.

Foram incluídas somente instâncias envolvendo procedimentos agendados entre os anos de 2015 e 2023, nas quais os pacientes compareceram ou deixaram de comparecer sem cancelamento ou notificação prévia. Instâncias agendadas para o mesmo dia foram excluídas, uma vez que este estudo tem como objetivo prever não-comparecimentos com pelo menos um dia de antecedência. As informações coletadas dos pacientes incluíram data de nascimento, sexo, CEP, cidade, estado, bairro e plano de saúde. Além disso, foram registrados o status do procedimento (realizado ou não-comparecimento), o tipo de procedimento (modalidade específica e tipo de exame), a data agendada para o procedimento e a data em que o agendamento foi realizado.

Assim, o conjunto de dados incluiu um total de 64367 instâncias, dentre as quais 11282 houve não-comparecimento de pacientes, resultando em uma proporção de não-comparecimentos igual a 17.5%.

2.2 Pré-processamento dos dados

Os dados crus foram processados de modo a servirem apropriadamente como input para os modelos de aprendizado de máquina.

Resumidamente, foi feito o seguinte:

– Idade dos pacientes: calculada com base na data de nascimento e na data agendada; os valores ausentes foram imputados por K-Nearest-Neighbors (KNN) (Batista; Monard, 2002) (K = 25) e indicados por uma variável binária adicional.

– Sexo dos pacientes: codificado com One-Hot Encoding (OHE) (Seger, 2018) (Masculino, Feminino ou Valor Ausente).

- Plano de saúde: codificado com OHE; planos de saúde pouco frequentes (< 0,5%) foram agrupados.
- Distância do paciente até a clínica: estimada por geolocalização usando a API do Bing Maps em Python; valores ausentes foram imputados por KNN (K = 25) e indicados por uma variável binária adicional.
- Tempo de espera: calculado em dias entre data de agendamento e data agendada para o procedimento;
- Aspectos temporais: Dia da semana, mês do ano e hora do dia foram codificados com OHE. Adicionalmente, foram criadas variáveis binárias indicando se houve feriado nacional no dia anterior ou no dia posterior à data agendada para o procedimento.
- Histórico completo de faltas: calculado como a taxa histórica completa de nãocomparecimento por paciente; valores ausentes foram imputados por KNN (K = 25) e indicados por uma variável binária adicional.
- Estatísticas temporais: para cada procedimento agendado, foram calculadas as taxas históricas de não-comparecimento considerando a taxa geral da clínica (todas as modalidades de serviço) e a taxa específica da modalidade de consultas. Essas taxas foram computadas diariamente e, para cada instância, foram extraídas a média, desvio padrão, quartis (1º e 3º), mediana e valores mínimo e máximo, resultando num total de 30 valores. Assim, utilizou-se Análise de Componente Principal (PCA) (Bro; Smilde, 2014) para reduzir a dimensionalidade desse conjunto de medidas, mantendose apenas os componentes responsáveis por explicar 95% da variância total.
- Tipo de consulta: foi realizada uma clusterização automática com base na similaridade textual dos nomes dos tipos de consulta. Cada nome foi dividido em n-gramas (1 a 3 caracteres) para capturar diferentes níveis de similaridade textual, com as distâncias entre as strings sendo calculadas usando a similaridade de Dice. Os tipos de consulta foram então

agrupados por meio de Clusterização Hierárquica (Murtagh; Contreras, 2012) (método average) com limiar de agrupamento definido como 70% da distância máxima observada. Posteriormente, aplicou-se OHE aos clusters resultantes, agrupando-se categorias com frequência inferior a 0,5% das instâncias em uma única categoria adicional. Essa abordagem reduziu significativamente a dimensionalidade da variável original, de 49 para 18.

2.3 Modelo de classificação: Balanced Random Forest

Neste estudo será utilizado o modelo de classificação Balanced Random Forest (BRF) (More; Rana, 2017), uma variação do algoritmo tradicional Random Forest (RF) (Breiman, 2001) especialmente desenvolvida para lidar com problemas de classificação com classes desbalanceadas. O RF tradicional constrói um conjunto (ensemble) de árvores de decisão a partir de amostras bootstrap do conjunto de treino, usando seleção aleatória de variáveis para gerar splits em cada árvore (Breiman, 2001). Apesar de sua robustez, o RF padrão tende a focar mais na classe majoritária, o que pode resultar em baixa performance na identificação da classe minoritária, especialmente em contextos altamente desbalanceados.

O BRF aborda diretamente essa limitação através da aplicação sistemática de subamostragem aleatória da classe majoritária para garantir que cada árvore individual seja treinada em um conjunto balanceado de dados (Lee, 2014; More; Rana, 2017). Em outras palavras, para cada árvore gerada no ensemble, é amostrado o mesmo número de instâncias de ambas as classes, obtidas por meio de amostragem com reposição. Esse procedimento garante que a classe minoritária esteja sempre adequadamente representada, resultando em uma melhor capacidade preditiva para esta classe.

Outra vantagem importante do BRF é sua relativa independência em relação à tunagem extensiva de hiperparâmetros. Devido à natureza da

sua construção (ensemble de árvores treinadas com dados balanceados), o BRF costuma apresentar boa performance preditiva mesmo utilizando valores padrão para seus hiperparâmetros principais. Isso é particularmente vantajoso no contexto deste estudo, cujo objetivo é avaliar o efeito do retreino periódico e da janela temporal de treino sobre as previsões futuras, demandando múltiplos treinamentos sucessivos e tornando inviável realizar a tunagem intensiva dos hiperparâmetros.

Portanto, a utilização do modelo BRF neste trabalho é justificada não somente pela sua capacidade intrínseca de tratar o desbalanceamento de classes, mas também pela eficiência computacional proporcionada ao eliminar a necessidade de ajustes exaustivos dos hiperparâmetros em cada ciclo de retreinamento. Neste trabalho, o modelo BRF foi aplicado em Python usando a classe `BalancedRandomForestClassifier`, disponível no módulo `imbalanced-learn` (Lemaître; Nogueira; Aridas, 2017). O único hiperparâmetro alterado em relação ao seu valor padrão foi `n_estimators`, definido para 250 em vez do valor padrão de 100. Esse aumento teve como objetivo garantir um maior número de árvores, proporcionando maior estabilidade nas previsões finais.

2.4 Avaliação do retreino periódico

Para avaliar o impacto do retreino periódico e das diferentes janelas temporais de treinamento sobre o desempenho preditivo do modelo, os dados coletados entre 2015 e 2023 foram divididos em trimestres. Essa divisão trimestral se justifica por representar um nível razoável de granularidade dos dados, especialmente considerando o desbalanceamento de classes. Divisões mais finas, como mensais, resultaram em um número pouco representativo de não-comparecimentos, afetando negativamente a estabilidade e a confiabilidade das métricas de performance utilizadas.

Nesse sentido, foram considerados dois cenários principais:

(1) Cenário sem retreino periódico: o modelo é treinado uma única vez utilizando os dados históricos correspondentes à janela temporal especificada (1, 2 ou 3 anos). A partir desse modelo inicial, são feitas previsões sequenciais para todos os trimestres subsequentes, sem nenhuma atualização do modelo.

(2) Cenário com retreino periódico: a cada novo trimestre, o modelo é retreinado utilizando apenas os dados históricos mais recentes, respeitando a janela temporal definida (1, 2 ou 3 anos). O modelo retreinado, é, então, utilizado exclusivamente para prever o trimestre seguinte.

Esses dois cenários foram aplicados separadamente para cada uma das três janelas temporais (1, 2 e 3 anos). A performance preditiva de cada cenário foi avaliada em cada trimestre utilizando a Acurácia total, Recall, Precisão, F1-score, Área sobre a curva ROC (AUC-ROC) e Área sobre a curva Precision-Recall (AUC-PR). As equações correspondentes às quatro primeiras métricas estão mostradas na Tabela 1.

Tabela 1 – Métricas utilizadas para avaliar o desempenho do modelo de classificação.

Métrica	Equação
Acurácia (A)	$A = \frac{TP+TN}{TP+TN+FP+FN}$
F1-score (F1)	$F1 = 2 \frac{P \times R}{P + R}$
Precisão (P)	$P = \frac{TP}{TP + FP}$
Recall (R)	$R = \frac{TP}{TP + FN}$

A comparação entre os cenários e as diferentes janelas temporais foi realizada por meio da análise dessas métricas ao longo dos trimestres, com ênfase na avaliação da estabilidade e na identificação das estratégias que apresentaram as melhores performances.

3 RESULTADOS E DISCUSSÕES OU ANÁLISE DOS DADOS

Na Figura 1, é mostrado a evolução a cada trimestre das métricas de desempenho do modelo de classificação aplicado à predição de não-comparecimentos de pacientes, tanto no cenário sem retreino periódico quanto com retreino periódico, considerando janelas históricas para treino de 1, 2 e 3 anos. A Tabela 2, por sua vez, resume a Figura 1 através da média e do desvio padrão, servindo para indicar o desempenho do modelo ao longo de todos os trimestres e a estabilidade das métricas para um mesmo cenário e janela histórica de treinamento.

Figura 1 – Evolução trimestral das métricas de desempenho dos modelos.

À esquerda, o cenário sem retreino periódico; à direita, o cenário com retreino periódico. Os cenários foram divididos entre as janelas históricas usadas para treinamento (1, 2 e 3 anos).

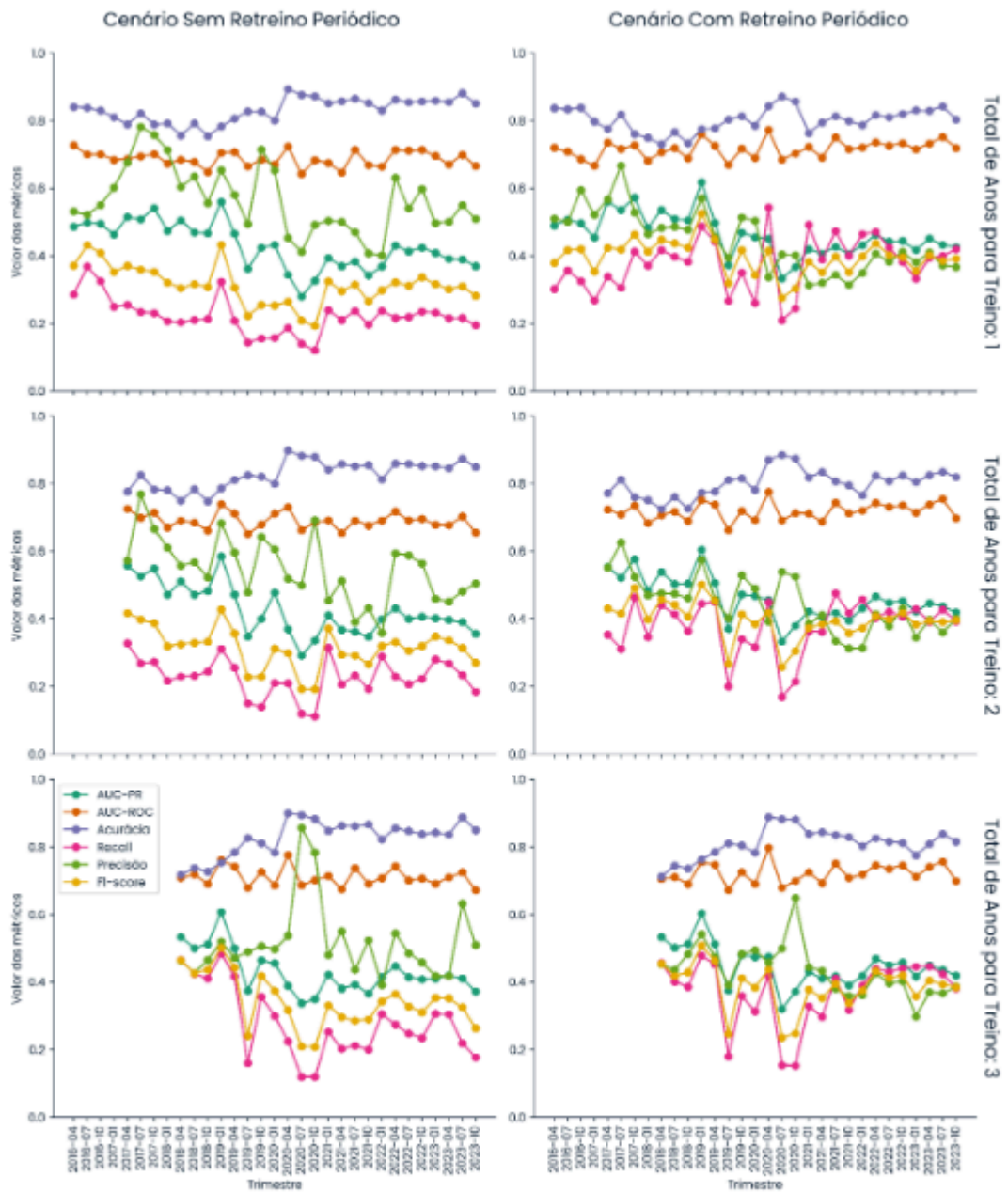


Tabela 2 – Média e desvio padrão (D.P.) das métricas de desempenho, considerando diferentes janelas históricas (1, 2, 3 anos) e os dois cenários avaliados (com e sem retreino periódico).

Métrica	Janela histórica	Cenário sem retreino periódico		Cenário com retreino periódico	
		Média	D.P.	Média	D.P.
Acurácia	1	0,8312	0,0363	0,8022	0,0352
	2	0,8286	0,0409	0,8026	0,0408
	3	0,8278	0,0533	0,8101	0,0452
AUC-PR	1	0,4287	0,0681	0,4634	0,0621
	2	0,4269	0,0751	0,4609	0,0638
	3	0,4298	0,0654	0,4492	0,0611
AUC-ROC	1	0,6869	0,0225	0,7156	0,0254
	2	0,6907	0,0237	0,7188	0,0256
	3	0,711	0,0274	0,7221	0,0302
F1-score	1	0,3128	0,057	0,3964	0,0497
	2	0,3154	0,06	0,3972	0,0555
	3	0,3428	0,0787	0,3862	0,0682
Precisão	1	0,564	0,1002	0,4432	0,0905
	2	0,5474	0,0965	0,4438	0,0826
	3	0,516	0,1102	0,4344	0,0741
Recall	1	0,221	0,0533	0,3778	0,0803
	2	0,2281	0,056	0,379	0,0804
	3	0,2792	0,1056	0,3699	0,0961

De modo geral, considerando a Tabela 2 e a Figura 1, observa-se que a acurácia foi ligeiramente superior no cenário sem retreino periódico, com destaque para a janela histórica de 1 ano, que apresentou a maior média ao longo dos trimestres. Entretanto, deve-se considerar que a acurácia tende a ser inflada em conjuntos de dados desbalanceados ao privilegiar a performance do modelo na classe majoritária (neste caso, os comparecimentos) (He; Garcia, 2009). Assim, essa métrica não necessariamente reflete uma boa capacidade de identificar os não-comparecimentos, que são justamente o foco do presente estudo (He; Garcia, 2009).

Ao considerarmos métricas mais sensíveis à classe minoritária, como AUC-PR e F1score (Jeni; Cohn; De La Torre, 2013), os resultados da Tabela 2 favorecem claramente o uso do retreino periódico. A AUC-PR foi superior em todas as janelas históricas no cenário com retreino, com a maior média ocorrendo na janela de 1 ano. Além disso, apresentou menor desvio

padrão, sugerindo um comportamento mais estável ao longo do tempo. De forma semelhante, o F1-score, que representa o equilíbrio entre Precisão e Recall, foi consideravelmente maior com o uso do retreino periódico, especialmente para a janela de 2 anos.

A métrica AUC-ROC também foi superior com o retreino periódico, atingindo seu melhor desempenho médio com a janela de 3 anos, conforme pode ser visto na Tabela 2. No entanto, sua estabilidade foi ligeiramente melhor no cenário sem retreino, como indicado pelos valores de desvio padrão na Tabela 2.

A Precisão foi maior no cenário sem retreino periódico, mas isso ocorreu às custas de um menor Recall, evidenciando um trade-off. Interpretando esse resultado de forma mais intuitiva, modelos sem retreino periódico são mais conservadores ao prever nãocomparecimentos (errando menos entre os classificados como tal, conforme sugere os valores de Precisão), mas identificam menos casos efetivos de ausência (como consequência de um Recall menor). Em contrapartida, o retreino periódico leva a modelos mais sensíveis, capazes de capturar mais casos de não-comparecimentos, ainda que com maior risco de falsos positivos. Note que isso pode ser vantajoso dependendo da estratégia adotada pela instituição de saúde, por exemplo envio de lembretes ou confirmações.

Para verificar se essas diferenças relatadas são estatisticamente significativas, aplicouse o teste de Wilcoxon pareado entre os valores trimestrais correspondentes de cada métrica. Os resultados do teste, apresentados na Tabela 3, revelam que o cenário com retreino periódico apresentou diferenças estatisticamente significativas em relação ao cenário sem retreino para métricas como AUC-PR, AUC-ROC, Recall e F1-score ($p < 0.05$), principalmente nas janelas de 1 e 2 anos. Já para acurácia e precisão, não houve diferença significativa entre os cenários, conforme a análise superficial por média e desvio padrão sugeriram.

Tabela 3 – Resultados do teste de Wilcoxon pareado entre os cenários com e sem retreino periódico, considerando diferentes janelas históricas (1, 2 e 3 anos). O teste foi aplicado com hipótese unilateral, avaliando se o desempenho do modelo com retreino periódico foi estatisticamente superior ao desempenho sem retreino, ao longo dos trimestres.

Métrica	Janela Histórica (anos)	Estatística	p-valor
Acurácia	1	5	1,0000
	2	1	1,0000
	3	22	0,9999
AUC-PR	1	490	0,0000
	2	372	0,0000
	3	257	0,0000
AUC-ROC	1	481	0,0000
	2	370	0,0000
	3	228	0,0024
F1-score	1	492	0,0000
	2	378	0,0000
	3	254	0,0001
Precisão	1	4	1,0000
	2	2	1,0000
	3	16	1,0000
Recall	1	464	0,0000
	2	378	0,0000
	3	258	0,0000

Ademais, foram realizados testes de Wilcoxon para comparar as diferentes janelas históricas dentro do cenário com retreino periódico, restringindo-se aos trimestres comuns entre todas elas. Os resultados, apresentados na Tabela 4, indicam que não houve diferenças significativas entre as janelas de 1, 2 e 3 anos em nenhuma das métricas analisadas. Apesar de isso ser negativo em termos de seleção da melhor janela histórica para treinamento, serve para reforçar que a janela ideal pode considerar não apenas a performance pontual, mas também aspectos como viabilidade computacional, haja vista que menores janelas demandam menos tempo de treinamento para o modelo de classificação.

Tabela 4 – Resultados do teste de Wilcoxon para comparação entre janelas históricas de 1, 2 e 3 anos dentro do cenário com retreino periódico. O teste foi aplicado com hipótese unilateral, avaliando se a janela B teve desempenho estatisticamente superior à janela A. Não foram observadas diferenças estatisticamente significativas.

Métrica	Janela A (anos)	Janela B (anos)	Estatística	p-valor
Acurácia	1	2	155	0,3013
	1	3	96	0,1056
	2	3	119	0,2902
AUC-PR	1	2	164	0,2810
	1	3	106	0,1723
	2	3	113	0,2317
AUC-ROC	1	2	164	0,2810
	1	3	97	0,1113
	2	3	92	0,0848
F1-score	1	2	183	0,4483
	1	3	156	0,7098
	2	3	159	0,7399
Precisão	1	2	168	0,3138
	1	3	98	0,1172
	2	3	118	0,2800
Recall	1	2	223	0,7934
	1	3	184	0,9199
	2	3	168	0,8198

Considerando todo o conjunto de resultados, a janela de 2 anos com retreino periódico parece oferecer o melhor equilíbrio entre desempenho preditivo (especialmente F1-score e Recall), estabilidade das métricas e custo computacional. Deve-se salientar, no entanto, que, embora estes achados sejam específicos e particularmente úteis para a CRA, a metodologia pode ser facilmente replicada e adaptada para outros conjuntos de dados, servindo para revelar a melhor abordagem para utilizar o modelo de classificação de forma contínua e operacional.

4 CONCLUSÃO

Os resultados indicaram que o cenário com retreino periódico apresentou desempenho superior em métricas críticas para contextos com classes desbalanceadas, como o F1-score e a AUC-PR, com destaque para as janelas temporais de 1 e 2 anos. Embora a acurácia tenha se apresentado com uma média ligeiramente maior no cenário sem retreino periódico, é importante destacar que essa métrica não é a ideal para avaliar o desempenho em relação à predição da classe minoritária (não-comparecimento).

Já a análise comparativa das janelas temporais para treinamento revelou que não houve diferenças significativas entre as janelas 1, 2 e 3 anos em termos de performance para todas as métricas testadas. Apesar disso, para este conjunto de dados, sugeriu-se o uso da janela de 2 anos, por representar um bom balanço entre desempenho preditivo e viabilidade computacional.

Estudos futuros podem explorar e comparar os resultados atuais com técnicas de aprendizado incremental, com potencial para anular a necessidade de retreino periódico dos modelos. Adicionalmente, um fato a ser levado em consideração, é que os resultados mostrados aqui poderiam ser melhores se houvesse uma otimização intensiva de hiperparâmetros do modelo BRF. No entanto, o objetivo deste estudo foi exclusivamente comparar os cenários com e sem retreino periódico sob condições padronizadas.

AGRADECIMENTOS

Os autores são gratos pela Clínica Radiológica de Anápolis pela parceria e suporte no fornecimento dos dados e dos recursos necessários para este estudo.

DECLARAÇÃO DE INTERESSES

Os autores declaram que não há conflitos de interesse.

DECLARAÇÃO DE FINANCIAMENTO

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

DECLARAÇÃO DE DISPONIBILIDADE DOS DADOS

Os dados que suportam os achados deste estudo estão disponíveis através dos autores, mediante solicitação razoável.

REFERÊNCIAS

ALKHAWALDEH, Ibraheem M; ALBALKHI, Ibrahem; NASWHAN, Abdulqadir Jeprel. Challenges and limitations of synthetic minority oversampling techniques in machine learning. **World Journal of Methodology**, [s. l.], v. 13, n. 5, p. 373–378, 2023.

<https://doi.org/10.5662/wjm.v13.i5.373>.

BATISTA, Gustavo E. A. P. A.; MONARD, Maria C. A study of k-nearest neighbour as an imputation method. **Frontiers in Artificial Intelligence and Applications**, [s. l.], v. 87, p. 251–260, 2002. .

BERG, Bjorn P.; MURR, Michael; CHERMAK, David; WOODALL, Jonathan; PIGNONE, Michael; SANDLER, Robert S.; DENTON, Brian T. Estimating the cost of no-shows and evaluating the effects of mitigation strategies.

Medical Decision Making, [s. l.], v. 33, n. 8, p. 976–985, 2013.

<https://doi.org/10.1177/0272989X13478194>.

BREIMAN, Leo. Random Forests. **Machine learning**, [s. l.], v. 45, p. 5–32, 2001. .

BRO, Rasmus; SMILDE, Age K. Principal component analysis. **Analytical Methods**, [s. l.], v. 6, n. 9, p. 2812–2831, 2014.

<https://doi.org/10.1039/c3ay41907j>.

CARRERAS-GARCÍA, Danae; DELGADO-GÓMEZ, David; LLORENTE-FERNÁNDEZ, Fernando; ARRIBAS-GIL, Ana. Patient no-show prediction: A systematic literature review. **Entropy**, [s. l.], v. 22, n. 6, p. 1–19, 2020. <https://doi.org/10.3390/E22060675>.

CHEN, Tianqi; GUESTRIN, Carlos. XGBoost: A scalable tree boosting system. **Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**, [s. l.], v. 13-17-Augu, p. 785–794, 2016. <https://doi.org/10.1145/2939672.2939785>.

DANTAS, Leila F.; FLECK, Julia L.; CYRINO OLIVEIRA, Fernando L.; HAMACHER, Silvio. No-shows in appointment scheduling – a systematic literature review. **Health Policy**, [s. l.], v. 122, n. 4, p. 412–421, 2018. DOI 10.1016/j.healthpol.2018.02.002. Available at: <https://doi.org/10.1016/j.healthpol.2018.02.002>.

GARCÍA-LAENCINA, Pedro J.; SANCHO-GÓMEZ, José Luis; FIGUEIRAS-VIDAL, Aníbal R. Pattern classification with missing data: A review. **Neural Computing and Applications**, [s. l.], v. 19, n. 2, p. 263–282, 2010. <https://doi.org/10.1007/s00521-0090295-6>.

HE, Haibo; GARCIA, Eduardo A. Learning from imbalanced data. **IEEE Transactions on Knowledge and Data Engineering**, [s. l.], v. 21, n. 9, p. 1263–1284, 2009. <https://doi.org/10.1109/TKDE.2008.239>.

JENI, László A; COHN, Jeffrey F; DE LA TORRE, Fernando. Facing Imbalanced Data Recommendations for the Use of Performance Metrics. **International Conference on Affective Computing and Intelligent Interaction and workshops: [proceedings]. ACII (Conference)**, United States, v. 2013, p. 245–251, 2013. <https://doi.org/10.1109/ACII.2013.47>.

LAVALLEY, Michael P. Logistic regression. **Circulation**, [s. l.], v. 117, n. 18, p. 2395–2399, 2008. <https://doi.org/10.1161/CIRCULATIONAHA.106.682658>.

LEE, Paul H. Resampling methods improve the predictive power of modeling in classimbalanced datasets. **International Journal of Environmental Research and Public Health**, [s. l.], v. 11, n. 9, p. 9776–9789, 2014. <https://doi.org/10.3390/ijerph110909776>.

LEMAÎTRE, Guillaume; NOGUEIRA, Fernando; ARIDAS, Christos K. Imbalancedlearn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. **Journal of machine learning research**, [s. l.], v. 18, n. 17, p. 1–5, 2017. .

LU, Jie; LIU, Anjin; DONG, Fan; GU, Feng; GAMA, Joao; ZHANG, Guangquan. Learning under Concept Drift: A Review. **IEEE Transactions on Knowledge and Data Engineering**, [s. l.], v. 31, n. 12, p. 2346–2363, 2019. <https://doi.org/10.1109/TKDE.2018.2876857>.

MARBOUH, Dounia; KHALEEL, Iman; SHANQITI, Khawla Al; TAMIMI, Maryam Al; SIMSEKLER, Mecit Can Emre; ELLAHHAM, Samer; ALIBAZOGLU, Deniz;

ALIBAZOGLU, Haluk. Evaluating the impact of patient no-shows on service quality. **Risk Management and Healthcare Policy**, [s. l.], v. 13, p. 509–517, 2020. <https://doi.org/10.2147/RMHP.S232114>.

MORE, A; RANA, P. Review of Random Forest Classification Techniques to Resolve Data Imbalance. **1st International Conference on Intelligent Systems and Information Management: ICISIM-2017**, [s. l.], , p. 72–78, 2017. .

MURTAGH, Fionn; CONTRERAS, Pedro. Algorithms for hierarchical clustering: An overview. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, [s. l.], v. 2, n. 1, p. 86–97, 2012. <https://doi.org/10.1002/widm.53>.

SATIANI, Bhagwan; MILLER, Susan; PATEL, Darshan. No-Show Rates in the Vascular Laboratory: Analysis and Possible Solutions. **Journal of Vascular**

and Interventional Radiology, [s. l.], v. 20, n. 1, p. 87–91, 2009. DOI

10.1016/j.jvir.2008.09.027. Available at:

<http://dx.doi.org/10.1016/j.jvir.2008.09.027>.

SEGER, Cedric. An investigation of categorical variable encoding techniques in machine learning: binary versus one-hot and feature hashing. **Degree Project Technology**, [s. l.], , p. 41, 2018. .

TOFFAHA, Khaled M.; SIMSEKLER, Mecit Can Emre; OMAR, Mohammed Atif; ELKEBBI, Imad. Predicting patient no-shows using machine learning: A comprehensive review and future research agenda. **Intelligence-Based Medicine**, [s. l.], v. 11, n. October 2023, p. 100229, 2025. DOI

10.1016/j.ibmed.2025.100229. Available at:

<https://doi.org/10.1016/j.ibmed.2025.100229>.

WUJEK, Brett; HALL, Patrick; GÜNEŞ, Funda. Best Practices for Machine Learning Applications. **SAS Institute Inc**, [s. l.], , p. 1–23, 2016. .

¹Escola de Engenharia Elétrica, Mecânica e de Computação, Universidade Federal de Goiás, Goiânia, Goiás, Brasil
carlosedgonc@gmail.com

²Clínica Radiológica de Anápolis, Anápolis, Goiás, Brasil
teixeiramjanabeatriz@gmail.com

³Clínica Radiológica de Anápolis, Anápolis, Goiás, Brasil
mateussimao@live.com

⁴Instituto de Informática, Universidade Federal de Goiás, Goiânia, Goiás
aldo.diaz@ufg.br

⁵Escola de Engenharia Elétrica, Mecânica e de Computação, Universidade Federal de Goiás, Goiânia, Goiás, Brasil

Instituto de Física, Universidade Federal de Goiás, Goiânia, Goiás, Brasil

emersonitikawa@ufg.br

[← Post anterior](#)[Post seguinte →](#)

RevistaFT

A RevistaFT têm 30 anos. É uma **Revista Científica Eletrônica Interdisciplinar Indexada de Alto Impacto e Qualis/DOI**.

Periodicidade mensal e de acesso livre. Leia gratuitamente todos os artigos e publique o seu também [clikando aqui](#),



Contato

Queremos te ouvir.

WhatsApp: (21) 99451-7530

WhatsApp: (21) 99217-2623

WhatsApp SP: (11) 98597-3405

e-Mail: contato@revistaft.com.br

ISSN: 1678-0817

CNPJ: 48.728.404/0001-22

Fator de impacto FI= 6.684 (segundo maior do Brasil

Conselho Editorial

Editores

Fundadores:

Dr. Oston de Lacerda Mendes.
Dr. João Marcelo Gigliotti.

Editor

Científico:

Dr. Oston de Lacerda Mendes

Jornalista

Responsável:

Marcos Antônio Alves MTB 6036DRT-MG

Orientadoras:

Dra. Hevellyn Andrade Monteiro
Dra. Chimene Kuhn Nobre

Revisores:

Lista atualizada
periodicamente
em

revistaft.com.br/expresspediente

Venha

fazer parte de

nosso time de

revisores

também!

Copyright © Revista ft Ltda. 1996 -
2026

Rua José Linhares, 134 - Leblon | Rio
de Janeiro-RJ | Brasil