



UNIVERSIDADE FEDERAL DE GOIÁS (UFG)
ESCOLA DE ENGENHARIA ELÉTRICA, MECÂNICA E DE COMPUTAÇÃO
PROGRAMA DE PÓS GRADUAÇÃO EM ENGENHARIA ELÉTRICA E DE
COMPUTAÇÃO (EMC)

CARLOS EDUARDO GONÇALVES DE OLIVEIRA

**Predição do não-comparecimento de pacientes em uma clínica
de diagnóstico por imagem usando aprendizado de máquina**

GOIÂNIA

04/07/2025



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE FÍSICA

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a [Lei 9.610/98](#), o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

Dissertação Tese Outro*: _____

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

Carlos Eduardo Gonçalves de Oliveira

3. Título do trabalho

“Predição do não-comparecimento de pacientes em uma clínica de diagnóstico por imagem usando aprendizado de máquina”

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento SIM NÃO¹

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:

a) consulta ao(à) autor(a) e ao(à) orientador(a);

b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação. O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.



Documento assinado eletronicamente por **Emerson Nobuyuki Itikawa, Professor do Magistério Superior**, em 14/08/2025, às 16:50, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Carlos Eduardo Gonçalves De Oliveira, Discente**, em 14/08/2025, às 17:02, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5574771** e o código CRC **4B53F1E7**.

CARLOS EDUARDO GONÇALVES DE OLIVEIRA

Predição do não-comparecimento de pacientes em uma clínica de diagnóstico por imagem usando aprendizado de máquina

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e de Computação, da Escola de Engenharia Elétrica, Mecânica e de Computação, da Universidade Federal de Goiás (UFG), como requisito para obtenção do título de Mestre em Engenharia Elétrica e de Computação.

Área de concentração: Sistemas e Sinais Biomédicos

Orientador: Professor Doutor Emerson Nobuyuki Itikawa

GOIÂNIA

04/07/2025

Ficha de identificação da obra elaborada pelo autor, através do
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Oliveira, Carlos Eduardo Gonçalves de
Predição do não-comparecimento de pacientes em uma clínica de
diagnóstico por imagem usando aprendizado de máquina [manuscrito]
/ Carlos Eduardo Gonçalves de Oliveira. - 2025.
ccxxi, 221 f.

Orientador: Prof. Dr. Emerson Nobuyuki Itikawa.
Dissertação (Mestrado) - Universidade Federal de Goiás, Escola
de Engenharia Elétrica, Mecânica e de Computação (EMC), Programa
de Pós-Graduação em Engenharia Elétrica e de Computação, Goiânia,
2025.

Bibliografia. Anexos.
Inclui gráfico, tabelas, lista de figuras, lista de tabelas.

1. Aprendizado de máquina. 2. Clínica. 3. Não-comparecimento. 4.
Otimização. 5. SHAP. I. Itikawa, Emerson Nobuyuki, orient. II. Título.

CDU 621.3



UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE FÍSICA

ATA DE DEFESA DE DISSERTAÇÃO

Ata nº 18 da sessão de Defesa de Dissertação de CARLOS EDUARDO GONÇALVES DE OLIVEIRA, que confere o título de Mestre em **Engenharia Elétrica e de Computação**, na área de concentração em **Engenharia de Computação**.

Aos **quatro dias do mês de julho de dois mil e vinte e cinco**, a partir das **14:00h**, na **sala Caryocar brasiliense da Escola de Engenharia Elétrica, Mecânica e de Computação**, realizou-se a sessão pública de Defesa de Dissertação intitulada “**Predição do não-comparecimento de pacientes em uma clínica de diagnóstico por imagem usando aprendizado de máquina**”. Os trabalhos foram instalados pelo Orientador, Professor Doutor **Emerson Nobuyuki Itikawa - (IF-UFG)**, com as participações dos demais membros da Banca Examinadora: Professor Doutor **Alisson Assis Cardoso - (EMC-UFG)**, Membro Titular Interno, Professor Doutor **Aldo André Diaz-Salazar - (IF-UFG)** Membro Titular Externo; **cuja participações ocorreram de forma híbrida**. Durante a arguição os membros da banca **não fizeram** sugestão de alteração do título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Dissertação, tendo sido o candidato **aprovado** pelos seus membros. Proclamados os resultados pelo Professor Doutor **Emerson Nobuyuki Itikawa**, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, aos **quatro dias do mês de julho de dois mil e vinte e cinco**.

TÍTULO SUGERIDO PELA BANCA



Documento assinado eletronicamente por **Emerson Nobuyuki Itikawa, Professor do Magistério Superior**, em 14/08/2025, às 16:54, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Alisson Assis Cardoso, Professor do Magistério Superior**, em 15/08/2025, às 09:18, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Aldo Andre Diaz Salazar, Professor do Magistério Superior**, em 18/08/2025, às 17:24, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5574794** e o código CRC **8E303BBA**.

Este trabalho é dedicado aos meus pais, cujo apoio incondicional e, por vezes, irracional, permitiu-me o prosseguimento de minha jornada acadêmica.

Agradecimentos

Agradeço aos meus pais que, com amor, me apoiaram em minha jornada acadêmica desde o início. A companhia e a confiança que depositaram em mim foram essenciais no desenvolvimento deste trabalho e em todas as conquistas que obtive até o momento. Devo tudo a vocês, meus queridos pais!

Agradeço ao professor Emerson pela orientação e pela amizade que, desde a graduação, me ajudaram a visualizar as oportunidades no meu caminho e a construir o pesquisador capaz e independente que venho me tornando.

Agradeço à doutora Ana Beatriz por confiar em minha competência para conduzir e desenvolver este trabalho, fornecendo-me todo o apoio e recursos necessários. Foi graças a ela que a parceria com a Clínica Radiológica de Anápolis foi possível.

Agradeço à instituição Clínica Radiológica de Anápolis como um todo pela parceria, confiando-me o manejo dos dados que me foram fornecidos. Em especial, agradeço ao doutor Paulo por oficializar formalmente a parceria e ao Mateus Simão pelo auxílio com a extração dos dados e com o esclarecimento de dúvidas.

Por fim, agradeço a todos os meus amigos por acreditarem no meu potencial e por terem me dado o devido apoio nos momentos em que mais precisei. Agradeço também à minha namorada pela companhia, paciência e apoio emocional que me foi dado durante o desenvolvimento deste trabalho.

“[...] mas para que tu, indivíduo, estás aí? – isso te pergunto, e, se ninguém te pode dizê-lo, tenta apenas uma vez legitimar o sentido de tua existência como que a posteriori, propondo tu a ti mesmo um fim, um alvo, um ‘para quê’, um alto e nobre ‘para quê’. Morre por ele – não conheço nenhuma finalidade melhor para a vida do que morrer pelo grandioso e pelo impossível.

Friedrich Nietzsche“

Resumo

O objetivo deste trabalho consiste em aplicar e analisar a performance de modelos de *Machine Learning* para a predição do não-comparecimento de pacientes em uma clínica de diagnóstico por imagem, utilizando dados de 2015 a 2023 provenientes de duas unidades da Clínica Radiológica de Anápolis (CRA), em Anápolis, Goiás, Brasil. A relevância deste estudo fundamenta-se na possibilidade de construção de um aplicativo final e na recorrência e no impacto negativo de não-comparecimentos de pacientes em centros de saúde, sendo necessário métodos para otimizar o uso dos recursos clínicos e reduzir perdas financeiras e de eficiência. As modalidades de procedimentos consideradas neste trabalho foram Ressonância Magnética, Tomografia Computadorizada, consultas e Ultrassom.

Os dados coletados foram idade do paciente, sexo do paciente, histórico de não-comparecimento do paciente, detalhes do agendamento (data e horário), tipo de procedimento, distância do endereço cadastrado do paciente até a clínica e entre outras. Os modelos testados, Regressão Logística, Perceptron Multicamadas, XGBoost, LightGBM e CatBoost, passaram por tunagem de hiperparâmetros e ajuste de limiar de probabilidade com base na área sobre a curva *Precision-Recall* e em uma métrica personalizada de "Custo". O *framework SHAP* foi utilizado para a interpretação das predições.

Comparações com a literatura indicaram a concordância dos resultados obtidos e o potencial dos métodos deste trabalho para servirem como solução de predição de não-comparecimentos para a otimização de tarefas como o *overbooking*. A análise por meio do *framework SHAP*, por sua vez, foi capaz de, para as diferentes modalidades de procedimentos, destacar as variáveis mais influentes na probabilidade de não-comparecimento, reforçando a utilidade desse método para a identificação de variáveis acionáveis.

Palavras-chave: SHAP (SHapley Additive exPlanations). Aprendizado de máquina. Clínica. não-comparecimento. Otimização.

Abstract

The objective of this work is to apply and analyze the performance of *Machine Learning* models for predicting patient no-shows at a diagnostic imaging clinic, using data from 2015 to 2023 from two units of Clínica Radiológica de Anápolis (CRA), in Anápolis, Goiás, Brazil. The relevance of this study is based on the possibility of building a final application and on the recurrence and negative impact of patient no-shows in health centers, requiring methods to optimize the use of clinical resources and reduce financial and efficiency losses. The procedure modalities considered in this work were Magnetic Resonance Imaging, Computed Tomography, consultations, and Ultrasound.

The collected data included patient age, patient gender, patient no-show history, scheduling details (date and time), procedure type, distance from the patient's registered address to the clinic, among others. The tested models, Logistic Regression, Multilayer Perceptron, XGBoost, LightGBM, and CatBoost, underwent hyperparameter tuning and probability threshold adjustment based on the *Precision-Recall* curve area and a customized "Cost" metric. The *SHAP framework* was used for interpreting the predictions.

Comparisons with the literature indicated the agreement of the obtained results and the potential of the methods in this work to serve as a no-show prediction solution for optimizing tasks such as overbooking. The analysis using the *SHAP framework*, in turn, was able to highlight the most influential variables in the probability of no-show for different procedure modalities, reinforcing the utility of this method for identifying actionable variables.

Keywords: SHAP (SHapley Additive exPlanations). Machine Learning. Clinic. no-show. Optimization.

Lista de ilustrações

Figura 1 – Boxplot das taxas de não-comparecimento agrupadas por década. Fonte: adaptada de (DANTAS et al., 2018).	29
Figura 2 – Boxplot das taxas de não-comparecimento agrupadas por modalidade médica. Fonte: adaptada de (DANTAS et al., 2018).	30
Figura 3 – a) Número de artigos publicados por ano. b) Número médio de apontamentos por ano usado nos bancos de dados numa escala logarítmica. Fonte: adaptada de (CARRERAS-GARCÍA et al., 2020).	35
Figura 4 – Diagrama de um neurônio artificial. Fonte: adaptada de (COSTA et al., 2023).	45
Figura 5 – Comportamento das diferentes funções de ativação citadas no texto. Fonte: próprio autor.	48
Figura 6 – Exemplo de arquitetura de MLP para um problema de classificação binária de entrada d-dimensional. Note que, para as camadas ocultas, qualquer outra função de ativação seria possível no lugar da sigmoide. Fonte: adaptada de (BOUZGOU, 2012).	49
Figura 7 – Exemplo de árvore de decisão, particionando o espaço de entrada contendo x_1 e x_2 nas regiões A, B, C, D e E. Fonte: (POTTS; SAMMUT, 2004).	53
Figura 8 – Ilustração de uma floresta aleatória de árvores de decisão (<i>Random Forest</i>). Fonte: próprio autor.	56
Figura 9 – Ilustração dos processos envolvidos na construção do <i>ensemble</i> ABT para k árvores de decisão. Fonte: adaptada de (GHOJOGH; CROWLEY, 2019).	58
Figura 10 – Ilustração dos processos envolvidos na construção do <i>ensemble</i> GBT para K árvores de decisão. Fonte: adaptada de (ALADSANI; BURTON; SE, 2021).	60
Figura 11 – Ilustração das técnicas de reamostragem clássicas: sobreamostragem e subamostragem. Fonte: adaptada de https://www.kaggle.com/code/rafjaa/resampling-strategies-for-imbalanced-datasets#t1 (Acessado em: 01/11/2024).	62
Figura 12 – Comportamento das métricas acurácia e AUC-ROC para diferentes valores de assimetria dos dados e diferentes taxas de erro do modelo. Fonte: adaptada de (JENI; COHN; TORRE, 2013).	64
Figura 13 – Comportamento das métricas F_1 -score e AUC-PR para diferentes valores de assimetria dos dados e diferentes taxas de erro do modelo. Fonte: adaptada de (JENI; COHN; TORRE, 2013).	65
Figura 14 – Breve apresentação do conjunto de dados utilizado. Fonte: próprio autor.	69

Figura 15 – Total de instâncias para cada modalidade de procedimentos na unidade do St. Central. Fonte: próprio autor.	70
Figura 16 – Total de instâncias para cada modalidade de procedimentos na unidade de Jundiaí. Fonte: próprio autor.	70
Figura 17 – Total de instâncias por ano para cada modalidade de procedimentos na unidade do St. Central. Fonte: próprio autor.	71
Figura 18 – Total de instâncias por ano para cada modalidade de procedimentos na unidade de Jundiaí. Fonte: próprio autor.	71
Figura 19 – Total de instâncias agendadas no mesmo dia ou não para cada modalidade de procedimentos na unidade do St. Central. Fonte: próprio autor.	72
Figura 20 – Total de instâncias agendadas no mesmo dia ou não para cada modalidade de procedimentos na unidade de Jundiaí. Fonte: próprio autor.	73
Figura 21 – Total de instâncias agendadas que foram canceladas ou não para cada modalidade de procedimentos na unidade do St. Central. Fonte: próprio autor.	74
Figura 22 – Total de instâncias agendadas que foram canceladas ou não para cada modalidade de procedimentos na unidade de Jundiaí. Fonte: próprio autor.	75
Figura 23 – Ilustração das etapas envolvidas ao gerar as variáveis referentes às estatísticas em janelas de 7 e 30 dias a partir da série temporal de taxa de não-comparecimentos. São mostradas as taxas de não-comparecimento da clínica e da modalidade de MR. Fonte: próprio autor.	77
Figura 24 – Ilustração do uso de Validação Cruzada com Divisão Temporal para a tunagem de hiperparâmetros. Note que, para um <i>fold</i> de teste, apenas os <i>folds</i> do passado são usados para treinamento. Ademais, cada <i>fold</i> de teste apresenta um peso correspondente à quantidade de <i>folds</i> usados para treino no cálculo da média ponderada do AUC-PR. Fonte: próprio autor.	80
Figura 25 – Sumário da metodologia geral deste trabalho. Fonte: próprio autor.	83
Figura 26 – Distribuição da proporção de não-comparecimentos por ano e mês para a modalidade de MR da unidade do St. Central. São também mostradas as proporções totais de não-comparecimentos separadamente para cada mês e cada ano (distribuições marginais). O "*" indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos, bem como o destaque em vermelho em torno das células mês-ano. Fonte: próprio autor.	86
Figura 27 – Distribuição do total de procedimentos por ano e mês para a modalidade de MR da unidade do St. Central. São também mostrados o total de procedimentos separadamente para cada mês e cada ano (distribuições marginais). Fonte: próprio autor.	87

Figura 28 – Distribuição da proporção de não-comparecimentos por dia da semana e horário para a modalidade de MR da unidade do St. Central. São também mostradas as proporções totais de não-comparecimentos separadamente para cada dia da semana e cada horário (distribuições marginais). O "*"indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos, bem como o destaque em vermelho em torno das células dia-horário. Fonte: próprio autor.	88
Figura 29 – Distribuição do total de procedimentos por dia da semana e horário para a modalidade de MR da unidade do St. Central. São também mostrados o total de procedimentos separadamente para cada dia da semana e cada horário (distribuições marginais). Fonte: próprio autor.	88
Figura 30 – Distribuição da proporção de não-comparecimentos por sexo dos pacientes para a modalidade de MR da unidade do St. Central. São também mostrados o total de procedimentos separadamente para sexo. O "*"indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos. Fonte: próprio autor.	90
Figura 31 – Distribuição da proporção de não-comparecimentos por idade dos pacientes para a modalidade de MR da unidade do St. Central. São também mostrados o total de procedimentos separadamente para cada <i>bin</i> de idade. O "*"indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos. Fonte: próprio autor.	90
Figura 32 – Distribuição da proporção de não-comparecimentos por distância até o CRA a partir do endereço cadastrado dos pacientes para a modalidade de MR da unidade do St. Central. São também mostrados o total de procedimentos separadamente para cada <i>bin</i> de distância. O "*"indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos. Fonte: próprio autor.	91
Figura 33 – Distribuição da proporção de não-comparecimentos por tempo de espera dos pacientes para a modalidade de MR da unidade do St. Central. São também mostrados o total de procedimentos separadamente para cada <i>bin</i> de tempo de espera. O "*"indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos. Fonte: próprio autor.	91
Figura 34 – Distribuição da proporção de não-comparecimentos por histórico prévio dos pacientes para a modalidade de MR da unidade do St. Central. São também mostrados o total de procedimentos separadamente para cada <i>bin</i> de histórico prévio. O "*"indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos. Fonte: próprio autor.	92

Figura 35 – Distribuição do total de procedimentos por plano de saúde dos pacientes para a modalidade de MR da unidade do St. Central. Fonte: próprio autor.	93
Figura 36 – Distribuição da proporção de não-comparecimentos por plano de saúde dos pacientes para a modalidade de MR da unidade do St. Central. O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos. Fonte: próprio autor.	94
Figura 37 – Performance geral dos modelos de classificação testados para a modalidade de MR da unidade do St. Central. O modelo com a melhor combinação de hiperparâmetros em termos de média ponderada de AUC-PR foi selecionado. Fonte: próprio autor.	96
Figura 38 – Performance para diferentes combinações de hiperparâmetros para o modelo de classificação selecionado. Fonte: próprio autor.	97
Figura 39 – Ajuste do limiar de probabilidade com base na média ponderada da métrica personalizada Custo para o modelo de classificação selecionado. Fonte: próprio autor.	97
Figura 40 – Matrizes de confusão resultantes da aplicação do modelo selecionado sobre os dados de treino e teste. Fonte: próprio autor.	98
Figura 41 – Curva ROC resultante da aplicação do modelo selecionado sobre os dados de treino e teste. Fonte: próprio autor.	99
Figura 42 – Curva PR resultante da aplicação do modelo selecionado sobre os dados de treino e teste. Fonte: próprio autor.	99
Figura 43 – Média dos valores SHAP absolutos resultantes da aplicação do modelo selecionado sobre os dados de teste. Fonte: próprio autor.	101
Figura 44 – Exemplo de previsões individuais feitas pelo modelo selecionado em instâncias do conjunto de teste, com seus respectivos valores de contribuição das variáveis. O eixo horizontal representa as contribuições em log-odds. As setas pretas indicam contribuições positivas para a probabilidade de não-comparecimento, enquanto as setas cinzas indicam contribuições negativas. O valor base $E[f(x)]$ corresponde à saída média do modelo e serve como ponto de referência. A saída final do modelo $f(x)$ reflete a contribuição cumulativa de todos os valores de atributos para a instância específica. Fonte: próprio autor.	102

Figura 45 – Distribuição da proporção de não-comparecimentos por ano e mês para a modalidade de CT da unidade do St. Central. São também mostradas as proporções totais de não-comparecimentos separadamente para cada mês e cada ano (distribuições marginais). O "*"indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos, bem como o destaque em vermelho em torno das células mês-ano. Fonte: próprio autor.	104
Figura 46 – Distribuição do total de procedimentos por ano e mês para a modalidade de CT da unidade do St. Central. São também mostrados o total de procedimentos separadamente para cada mês e cada ano (distribuições marginais). Fonte: próprio autor.	104
Figura 47 – Distribuição da proporção de não-comparecimentos por dia da semana e horário para a modalidade de CT da unidade do St. Central. São também mostradas as proporções totais de não-comparecimentos separadamente para cada dia da semana e cada horário (distribuições marginais). O "*"indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos, bem como o destaque em vermelho em torno das células dia-horário. Fonte: próprio autor.	105
Figura 48 – Distribuição do total de procedimentos por dia da semana e horário para a modalidade de CT da unidade do St. Central. São também mostrados o total de procedimentos separadamente para cada dia da semana e cada horário (distribuições marginais). Fonte: próprio autor.	106
Figura 49 – Distribuição da proporção de não-comparecimentos por sexo dos pacientes para a modalidade de CT da unidade do St. Central. São também mostrados o total de procedimentos separadamente para sexo. O "*"indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos. Fonte: próprio autor.	107
Figura 50 – Distribuição da proporção de não-comparecimentos por idade dos pacientes para a modalidade de CT da unidade do St. Central. São também mostrados o total de procedimentos separadamente para cada <i>bin</i> de idade. O "*"indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos. Fonte: próprio autor.	108

Figura 51 – Distribuição da proporção de não-comparecimentos por distância até o CRA a partir do endereço cadastrado dos pacientes para a modalidade de CT da unidade do St. Central. São também mostrados o total de procedimentos separadamente para cada <i>bin</i> de distância. O "*" indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos. Fonte: próprio autor.	108
Figura 52 – Distribuição da proporção de não-comparecimentos por tempo de espera dos pacientes para a modalidade de CT da unidade do St. Central. São também mostrados o total de procedimentos separadamente para cada <i>bin</i> de tempo de espera. O "*" indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos. Fonte: próprio autor.	109
Figura 53 – Distribuição da proporção de não-comparecimentos por histórico prévio dos pacientes para a modalidade de CT da unidade do St. Central. São também mostrados o total de procedimentos separadamente para cada <i>bin</i> de histórico prévio. O "*" indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos. Fonte: próprio autor.	109
Figura 54 – Distribuição do total de procedimentos por plano de saúde dos pacientes para a modalidade de CT da unidade do St. Central. Fonte: próprio autor.	110
Figura 55 – Distribuição da proporção de não-comparecimentos por plano de saúde dos pacientes para a modalidade de CT da unidade do St. Central. O "*" indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos. Fonte: próprio autor.	111
Figura 56 – Performance geral dos modelos de classificação testados para a modalidade de CT da unidade do St. Central. O modelo com a melhor combinação de hiperparâmetros em termos de média ponderada de AUC-PR foi selecionado. Fonte: próprio autor.	113
Figura 57 – Performance para diferentes combinações de hiperparâmetros para o modelo de classificação selecionado. Fonte: próprio autor.	114
Figura 58 – Ajuste do limiar de probabilidade com base na média ponderada da métrica personalizada Custo para o modelo de classificação selecionado. Fonte: próprio autor.	114
Figura 59 – Matrizes de confusão resultantes da aplicação do modelo selecionado sobre os dados de treino e teste. Fonte: próprio autor.	115
Figura 60 – Curva ROC resultante da aplicação do modelo selecionado sobre os dados de treino e teste. Fonte: próprio autor.	116
Figura 61 – Curva PR resultante da aplicação do modelo selecionado sobre os dados de treino e teste. Fonte: próprio autor.	116

Figura 62 – Média dos valores SHAP absolutos resultantes da aplicação do modelo selecionado sobre os dados de teste. Fonte: próprio autor.	118
Figura 63 – Exemplo de predições individuais feitas pelo modelo selecionado em instâncias do conjunto de teste, com seus respectivos valores de contribuição das variáveis. O eixo horizontal representa as contribuições em log-odds. As setas pretas indicam contribuições positivas para a probabilidade de não-comparecimento, enquanto as setas cinzas indicam contribuições negativas. O valor base $E[f(x)]$ corresponde à saída média do modelo e serve como ponto de referência. A saída final do modelo $f(x)$ reflete a contribuição cumulativa de todos os valores de atributos para a instância específica. Fonte: próprio autor.	119
Figura 64 – Distribuição da proporção de não-comparecimentos por ano e mês para a modalidade de consulta da unidade de Jundiaí. São também mostradas as proporções totais de não-comparecimentos separadamente para cada mês e cada ano (distribuições marginais). O "*" indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos, bem como o destaque em vermelho em torno das células mês-ano. Fonte: próprio autor.	121
Figura 65 – Distribuição do total de procedimentos por ano e mês para a modalidade de consulta da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para cada mês e cada ano (distribuições marginais). Fonte: próprio autor.	121
Figura 66 – Distribuição da proporção de não-comparecimentos por dia da semana e horário para a modalidade de consulta da unidade de Jundiaí. São também mostradas as proporções totais de não-comparecimentos separadamente para cada dia da semana e cada horário (distribuições marginais). O "*" indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos, bem como o destaque em vermelho em torno das células dia-horário. Fonte: próprio autor.	122
Figura 67 – Distribuição do total de procedimentos por dia da semana e horário para a modalidade de consulta da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para cada dia da semana e cada horário (distribuições marginais). Fonte: próprio autor.	123

Figura 68 – Distribuição da proporção de não-comparecimentos por sexo dos pacientes para a modalidade de consulta da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para sexo. O "*"indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos. Fonte: próprio autor.	124
Figura 69 – Distribuição da proporção de não-comparecimentos por idade dos pacientes para a modalidade de consulta da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para cada <i>bin</i> de idade. O "*"indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos. Fonte: próprio autor.	125
Figura 70 – Distribuição da proporção de não-comparecimentos por distância até o CRA a partir do endereço cadastrado dos pacientes para a modalidade de consulta da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para cada <i>bin</i> de distância. O "*"indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos. Fonte: próprio autor.	125
Figura 71 – Distribuição da proporção de não-comparecimentos por tempo de espera dos pacientes para a modalidade de consulta da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para cada <i>bin</i> de tempo de espera. O "*"indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos. Fonte: próprio autor.	126
Figura 72 – Distribuição da proporção de não-comparecimentos por histórico prévio dos pacientes para a modalidade de consulta da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para cada <i>bin</i> de histórico prévio. O "*"indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos. Fonte: próprio autor.	126
Figura 73 – Distribuição do total de procedimentos por plano de saúde dos pacientes para a modalidade de consulta da unidade de Jundiaí. Fonte: próprio autor.	127
Figura 74 – Distribuição da proporção de não-comparecimentos por plano de saúde dos pacientes para a modalidade de consulta da unidade de Jundiaí. O "*"indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos. Fonte: próprio autor.	128
Figura 75 – Performance geral dos modelos de classificação testados para a modalidade de consulta da unidade de Jundiaí. O modelo com a melhor combinação de hiperparâmetros em termos de média ponderada de AUC-PR foi selecionado. Fonte: próprio autor.	130

Figura 76 – Performance para diferentes combinações de hiperparâmetros para o modelo de classificação selecionado. Fonte: próprio autor.	131
Figura 77 – Ajuste do limiar de probabilidade com base na média ponderada da métrica personalizada Custo para o modelo de classificação selecionado. Fonte: próprio autor.	131
Figura 78 – Matrizes de confusão resultantes da aplicação do modelo selecionado sobre os dados de treino e teste. Fonte: próprio autor.	132
Figura 79 – Curva ROC resultante da aplicação do modelo selecionado sobre os dados de treino e teste. Fonte: próprio autor.	133
Figura 80 – Curva PR resultante da aplicação do modelo selecionado sobre os dados de treino e teste. Fonte: próprio autor.	133
Figura 81 – Média dos valores SHAP absolutos resultantes da aplicação do modelo selecionado sobre os dados de teste. Fonte: próprio autor.	135
Figura 82 – Exemplo de previsões individuais feitas pelo modelo selecionado em instâncias do conjunto de teste, com seus respectivos valores de contribuição das variáveis. O eixo horizontal representa as contribuições em log-odds. As setas pretas indicam contribuições positivas para a probabilidade de não-comparecimento, enquanto as setas cinzas indicam contribuições negativas. O valor base $E[f(x)]$ corresponde à saída média do modelo e serve como ponto de referência. A saída final do modelo $f(x)$ reflete a contribuição cumulativa de todos os valores de atributos para a instância específica. Fonte: próprio autor.	136
Figura 83 – Distribuição da proporção de não-comparecimentos por ano e mês para a modalidade de US da unidade de Jundiaí. São também mostradas as proporções totais de não-comparecimentos separadamente para cada mês e cada ano (distribuições marginais). O "*" indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos, bem como o destaque em vermelho em torno das células mês-ano. Fonte: próprio autor.	138
Figura 84 – Distribuição do total de procedimentos por ano e mês para a modalidade de US da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para cada mês e cada ano (distribuições marginais). Fonte: próprio autor.	138

Figura 85 – Distribuição da proporção de não-comparecimentos por dia da semana e horário para a modalidade de US da unidade de Jundiaí. São também mostradas as proporções totais de não-comparecimentos separadamente para cada dia da semana e cada horário (distribuições marginais). O "*"indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos, bem como o destaque em vermelho em torno das células dia-horário. Fonte: próprio autor.	139
Figura 86 – Distribuição do total de procedimentos por dia da semana e horário para a modalidade de US da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para cada dia da semana e cada horário (distribuições marginais). Fonte: próprio autor.	140
Figura 87 – Distribuição da proporção de não-comparecimentos por sexo dos pacientes para a modalidade de US da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para sexo. O "*"indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos. Fonte: próprio autor.	141
Figura 88 – Distribuição da proporção de não-comparecimentos por idade dos pacientes para a modalidade de US da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para cada <i>bin</i> de idade. O "*"indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos. Fonte: próprio autor.	142
Figura 89 – Distribuição da proporção de não-comparecimentos por distância até o CRA a partir do endereço cadastrado dos pacientes para a modalidade de US da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para cada <i>bin</i> de distância. O "*"indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos. Fonte: próprio autor.	142
Figura 90 – Distribuição da proporção de não-comparecimentos por tempo de espera dos pacientes para a modalidade de US da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para cada <i>bin</i> de tempo de espera. O "*"indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos. Fonte: próprio autor.	143
Figura 91 – Distribuição da proporção de não-comparecimentos por histórico prévio dos pacientes para a modalidade de US da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para cada <i>bin</i> de histórico prévio. O "*"indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos. Fonte: próprio autor.	143

Figura 92 – Distribuição do total de procedimentos por plano de saúde dos pacientes para a modalidade de US da unidade de Jundiaí. Fonte: próprio autor.	144
Figura 93 – Distribuição da proporção de não-comparecimentos por plano de saúde dos pacientes para a modalidade de US da unidade de Jundiaí. O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos. Fonte: próprio autor.	145
Figura 94 – Performance geral dos modelos de classificação testados para a modalidade de US da unidade de Jundiaí. O modelo com a melhor combinação de hiperparâmetros em termos de média ponderada de AUC-PR foi selecionado. Fonte: próprio autor.	147
Figura 95 – Performance para diferentes combinações de hiperparâmetros para o modelo de classificação selecionado. Fonte: próprio autor.	148
Figura 96 – Ajuste do limiar de probabilidade com base na média ponderada da métrica personalizada Custo para o modelo de classificação selecionado. Fonte: próprio autor.	148
Figura 97 – Matrizes de confusão resultantes da aplicação do modelo selecionado sobre os dados de treino e teste. Fonte: próprio autor.	149
Figura 98 – Curva ROC resultante da aplicação do modelo selecionado sobre os dados de treino e teste. Fonte: próprio autor.	150
Figura 99 – Curva PR resultante da aplicação do modelo selecionado sobre os dados de treino e teste. Fonte: próprio autor.	150
Figura 100 – Média dos valores SHAP absolutos resultantes da aplicação do modelo selecionado sobre os dados de teste. Fonte: próprio autor.	152
Figura 101 – Exemplo de previsões individuais feitas pelo modelo selecionado em instâncias do conjunto de teste, com seus respectivos valores de contribuição das variáveis. O eixo horizontal representa as contribuições em log-odds. As setas pretas indicam contribuições positivas para a probabilidade de não-comparecimento, enquanto as setas cinzas indicam contribuições negativas. O valor base $E[f(x)]$ corresponde à saída média do modelo e serve como ponto de referência. A saída final do modelo $f(x)$ reflete a contribuição cumulativa de todos os valores de atributos para a instância específica. Fonte: próprio autor.	153

Lista de tabelas

Tabela 1	– Breve descrição e espaço de valores dos hiperparâmetros para os modelos testados.	79
Tabela 2	– Métricas de performance utilizadas neste trabalho. VP : Verdadeiros Positivos, VN : Verdadeiros Negativos, FP : Falsos Positivos, FN : Falsos Negativos, $N = VP + VN + FP + FN$. Neste trabalho, $P_I = 0.333$ e $P_{II} = 1$	81
Tabela 3	– Informações adicionais sobre a performance geral dos modelos de classificação testados para a modalidade de MR da unidade do St. Central. O modelo com maior média ponderada de AUC-PR foi selecionado (coluna "Máx." da tabela).	96
Tabela 4	– Métricas de performance do modelo selecionado conforme calculadas a partir das matrizes de confusão.	98
Tabela 5	– Informações adicionais sobre a performance geral dos modelos de classificação testados para a modalidade de CT da unidade do St. Central. O modelo com maior média ponderada de AUC-PR foi selecionado (coluna "Máx." da tabela).	113
Tabela 6	– Métricas de performance do modelo selecionado conforme calculadas a partir das matrizes de confusão.	115
Tabela 7	– Informações adicionais sobre a performance geral dos modelos de classificação testados para a modalidade de consultas da unidade de Jundiaí. O modelo com maior média ponderada de AUC-PR foi selecionado (coluna "Máx." da tabela).	130
Tabela 8	– Métricas de performance do modelo selecionado conforme calculadas a partir das matrizes de confusão.	132
Tabela 9	– Informações adicionais sobre a performance geral dos modelos de classificação testados para a modalidade de US da unidade de Jundiaí. O modelo com maior média ponderada de AUC-PR foi selecionado (coluna "Máx." da tabela).	147
Tabela 10	– Métricas de performance do modelo selecionado conforme calculadas a partir das matrizes de confusão.	149
Tabela 11	– Métricas de performance do melhor modelo para cada especialidade médica conforme obtido em (DUNSTAN et al., 2023). As métricas Precisão (P), Sensibilidade (S) e $F1$ -score (F1) foram separadas para as classes de não-comparecimento (1) e comparecimento (0).	155

Tabela 13 – Métricas de performance do melhor modelo para cada modalidade de procedimentos conforme obtido neste projeto de mestrado. As métricas Precisão (P), Sensibilidade (S) e <i>F1-score</i> (F1) foram separadas para as classes de não-comparecimento (1) e comparecimento (0).	155
Tabela 15 – Métricas de performance das diferentes combinações entre técnicas de amostragem e modelos de classificação testados para o conjunto de dados 1 conforme obtido em (DEINA et al., 2024). As métricas Precisão (P), Sensibilidade (S) e <i>F1-score</i> (F1) foram calculadas somente para a classe de não-comparecimento. Também são apresentados a área sobre a curva ROC (AUC-ROC) e a acurácia total (A).	158
Tabela 17 – Métricas de performance das diferentes combinações entre técnicas de amostragem e modelos de classificação testados para o conjunto de dados 2 conforme obtido em (DEINA et al., 2024). As métricas Precisão (P), Sensibilidade (S) e <i>F1-score</i> (F1) foram calculadas somente para a classe de não-comparecimento. Também são apresentados a área sobre a curva ROC (AUC-ROC) e a acurácia total (A).	158
Tabela 19 – Métricas de performance dos melhores modelos aplicados a diferentes modalidades de procedimentos no CRA, conforme obtido para este projeto de mestrado. As métricas Precisão (P), Sensibilidade (S) e <i>F1-score</i> (F1) foram calculadas somente para a classe de não-comparecimento. Também são apresentados a área sobre a curva ROC (AUC-ROC) e a acurácia total (A). .	159

Trabalhos Submetidos e Publicados

Trabalhos aprovados e/ou publicados:

- Quintino, C., Malheiros-Lima, M. R., Ghazale, P. P., Braga, P. P. P., Maia, O. A., **de Oliveira, C. E. G.**, ... & Pansani, A. P. (2024). *The latency to awake from induced-obstructive sleep apnea is reduced in rats with chronic epilepsy. Epilepsy & Behavior*, 157, 109848.
- **de Oliveira, C. E. G.**, Vieira, S. L., Paranaíba, C. F. B., & Itikawa, E. N. (2025). *Breast tumor segmentation in ultrasound images: comparing U-net and U-net++*. *Research on Biomedical Engineering*, 41(1), 1-11.
- **de Oliveira, C. E. G.**, de Lima Júnior, C. Q., Colugnati, D. B., Schoorlemmer, G. H. M., da Matta, D. H., & Pansani, A. P. (2025). *Machine learning for sleep stage classification in control and pilocarpine-induced epileptic rats. Research on Biomedical Engineering*, 41(2), 1-14.
- **de Oliveira, C. E. G.**, de Jesus Teixeira, A. B. M., Oliveira, M. S., Salazar, A. A. D., & Itikawa, E. N. (2025). *AVALIAÇÃO DO IMPACTO DO RETREINO PERIÓDICO NA PREDIÇÃO DE NÃO-COMPARECIMENTOS DE PACIENTES EM CONSULTAS MÉDICAS UTILIZANDO BALANCED RANDOM FOREST. Revista FT*, 29(146).

Sumário

1	Introdução	25
1.1	Contextualização e problemática	25
1.2	Objetivos	27
2	Fundamentação Teórica	28
2.1	Não-comparecimento: impactos negativos e prevalência nas instituições de saúde	28
2.2	Fatores relacionados ao não-comparecimento de pacientes	30
2.3	Redução dos efeitos negativos do não-comparecimento de pacientes	33
2.4	Predição do não-comparecimento de pacientes	35
2.5	Análise de Componente Principal (PCA)	39
2.5.1	Definição qualitativa	39
2.5.2	Definição quantitativa	40
2.6	Regressão logística	41
2.6.1	Descrição matemática	41
2.6.2	Função de custo e regularização	43
2.7	Perceptron Multicamadas	45
2.7.1	Neurônio artificial	45
2.7.2	Perceptron Multicamadas: uma arquitetura de rede neural artificial	48
2.7.3	Treinamento	50
2.7.4	Algoritmo de retropropagação	51
2.8	Árvores de decisão	52
2.8.1	Estrutura das árvores de decisão	53
2.8.2	Formulação matemática das divisões em cada nó	53
2.8.3	Poda da árvore de decisão	54
2.9	Floresta Aleatória	55
2.10	Árvores de <i>Boosting</i> Adaptativo	56
2.11	Árvores de <i>Gradient Boosting</i>	58
2.12	Desbalanceamento de classes	60
2.12.1	Técnicas para lidar com desbalanceamento de classes	61
2.12.2	Métricas apropriadas	63
2.13	Interpretando as previsões de modelos: <i>framework SHAP</i>	65
2.13.1	Definição	65
2.13.2	Propriedades dos valores SHAP	66
3	Materiais e Métodos	68

3.1	Conjunto de dados	68
3.1.1	Caracterização inicial	68
3.1.2	Coleta dos dados	68
3.1.3	Exclusão de instâncias	69
3.1.3.1	Modalidades de procedimentos com poucas instâncias	69
3.1.3.2	Anos com poucas instâncias	70
3.1.3.3	Procedimentos agendados no mesmo dia	71
3.1.3.4	Cancelamentos e função de determinação das classes	73
3.1.4	Processamento dos dados: do sistema de registros a variáveis preditoras	75
3.1.5	Dados de treino, dados de teste, tunagem de hiperparâmetros e seleção dos modelos	78
3.1.6	Métricas de performance e limiar de probabilidade	81
3.1.7	Interpretando as previsões do modelo: <i>framework SHAP</i>	82
3.1.8	Sumário final	82
4	Resultados e Discussão	84
4.1	Introdução	84
4.2	Modalidade de Ressonância Magnética (unidade do St. Central)	85
4.2.1	Análise exploratória	85
4.2.1.1	Distribuição temporal da proporção de não-comparecimentos	85
4.2.1.2	Características dos pacientes e proporção de não-comparecimentos	89
4.2.2	Tunagem de hiperparâmetros e seleção do modelo de classificação	94
4.2.3	Resultados sobre os dados de teste	97
4.2.4	Importância das variáveis	100
4.3	Modalidade de Tomografia Computadorizada (unidade do St. Central)	102
4.3.1	Análise exploratória	103
4.3.1.1	Distribuição temporal da proporção de não-comparecimentos	103
4.3.1.2	Características dos pacientes e proporção de não-comparecimentos	106
4.3.2	Tunagem de hiperparâmetros e seleção do modelo de classificação	111
4.3.3	Resultados sobre os dados de teste	114
4.3.4	Importância das variáveis	117
4.4	Modalidade de consultas (unidade de Jundiaí)	119
4.4.1	Análise exploratória	120
4.4.1.1	Distribuição temporal da proporção de não-comparecimentos	120
4.4.1.2	Características dos pacientes e proporção de não-comparecimentos	123
4.4.2	Tunagem de hiperparâmetros e seleção do modelo de classificação	128
4.4.3	Resultados sobre os dados de teste	131
4.4.4	Importância das variáveis	134

4.5	Modalidade de Ultrassom (unidade de Jundiá)	136
4.5.1	Análise exploratória	137
4.5.1.1	Distribuição temporal da proporção de não-comparecimentos	137
4.5.1.2	Características dos pacientes e proporção de não-comparecimentos	140
4.5.2	Tunagem de hiperparâmetros e seleção do modelo de classificação	145
4.5.3	Resultados sobre os dados de teste	148
4.5.4	Importância das variáveis	151
4.6	Comparações com a literatura	153
4.6.1	Comparação de performance com um trabalho realizado em um hospital pediátrico no Chile envolvendo múltiplas especialidades médicas (2023)	154
4.6.2	Comparação de performance com um trabalho realizado com dois conjuntos de dados brasileiros (2024)	156
5	Conclusão	160
	Referências	162
	Anexos	168
ANEXO A	Artigo publicado	169
ANEXO B	Carta de anuência	192
ANEXO C	Termo de compromisso dos pesquisadores	194
ANEXO D	Parecer consubstanciado do Comitê de Ética em Pesquisa da UFG	196
ANEXO E	Certificado: Cientista de Dados	202
ANEXO F	Certificado: Cientista de Machine Learning	204
ANEXO G	Certificado: The Science of Uncertainty and Data	206
ANEXO H	Certificado: Machine Learning with Python - From Linear Models to Deep Learning	208
ANEXO I	Certificado: Participação no 1º Datathon, VI ENGOPE, IME/UFG	210
ANEXO J	Prêmio: Campeão do 1º Datathon, VI ENGOPE, IME/UFG	212
ANEXO K	Pôster apresentado, VI ENGOPE, IME/UFG	214
ANEXO L	Indicação, prêmio FAPEG N° 27/2024: pós-graduando destaque	216

1 Introdução

1.1 Contextualização e problemática

Com o aumento da demanda e da pressão por redução de custos, é essencial para as instituições de saúde melhorar a eficácia e a eficiência de seus serviços. Apesar dos esforços nesse sentido, há eventos inevitáveis que levam ao aumento de custos e à subutilização de recursos (MARBOUH et al., 2020). O não-comparecimento de pacientes é um desses eventos, resultando em um descompasso entre a oferta e a demanda esperada, o que afeta a qualidade do serviço e o desempenho da clínica (MARBOUH et al., 2020; GLOWACKA; HENRY; MAY, 2009). De fato, o impacto pode ser ainda maior em departamentos de radiologia, por exemplo, onde máquinas de Tomografia Computadorizada (CT) e Ressonância Magnética (MR) são caras e não é apropriado que sejam subutilizadas. Ademais, a saúde do paciente também fica em risco, uma vez que o tratamento ou o diagnóstico é adiado, consequência essa que se estende aos pacientes que poderiam ser agendados no lugar do paciente que não-compareceu (MARBOUH et al., 2020).

O não-comparecimento de pacientes ocorre quando um paciente perde uma consulta ou procedimento agendado sem aviso prévio (MARBOUH et al., 2020). Em uma revisão sistemática sobre o tema, relatou-se que a prevalência de não-comparecimentos variou por continente, com taxas mais altas na África (43,0%), seguida da América do Sul (27,8%), Ásia (25,1%), América do Norte (23,5%), Europa (19,3%) e Oceania (13,2%) (DANTAS et al., 2018). As taxas também variaram entre especialidades médicas, sendo mais altas em fisioterapia (57,3%), endocrinologia (36,0%) e cardiologia (30,0%) (DANTAS et al., 2018). Reduzir essas taxas teria um impacto financeiro significativo, especialmente para instituições públicas de saúde (ELLIS; JENKINS, 2012). Por exemplo, no Serviço Nacional de Saúde do Reino Unido, reduzir a taxa de faltas de 12% para 10,8% diminuiria os gastos públicos em 10% (ELLIS; JENKINS, 2012), o que é uma redução relevante.

Para enfrentar os efeitos negativos dos não-comparecimentos de pacientes, centros de saúde têm implementado várias estratégias, incluindo sanções, lembretes e excesso de agendamentos (do inglês, *overbooking*) (SATIANI; MILLER; PATEL, 2009; CHARIATTE et al., 2008; BERG et al., 2013). No entanto, essas estratégias têm custos: os lembretes, por exemplo, podem gerar um custo estimado de 0.41 euros por lembrete (HASVOLD; WOOTTON, 2011). Focar em pacientes com alta probabilidade de faltas e utilizar modelos de aprendizado de máquina (do inglês *Machine Learning*, doravante referido como ML) pode ajudar a mitigar esses custos. Dependendo do desempenho do modelo, ele pode inclusive ser aplicado para melhorar o

agendamento de pacientes (CARRERAS-GARCÍA et al., 2020). Aliás, o aumento de pesquisas nessa área se deve à maior disponibilidade de dados e aos avanços nos sistemas de registro (CARRERAS-GARCÍA et al., 2020).

Devido à complexidade do tema na literatura científica, não há consenso sobre as características ou modelos a serem usados para a predição de não-comparecimento (CARRERAS-GARCÍA et al., 2020). Por exemplo, mais de 40 possíveis características foram identificadas, incluindo idade, histórico de faltas, mês do ano, dia da semana, horário do dia, tempo de espera do paciente e outras (DANTAS et al., 2018). Modelos comumente empregados incluem Regressão Logística (LIU; CHEN; YE, 2009), Árvores de Decisão (STEINBERG; COLLA, 1995), Redes Neurais Artificiais (HAYKIN, 2008), modelos Bayesianos (SCHOOT et al., 2021) e modelos de ensemble/stacking (GHOJOGH; CROWLEY, 2019; BREIMAN, 1996). Um desafio adicional é que o desempenho dos modelos testados é medido de maneiras significativamente diferentes entre os estudos. As métricas mais frequentemente usadas são a área sob a curva ROC (AUC-ROC) (NAHM, 2022) e a acurácia geral (CARRERAS-GARCÍA et al., 2020). No entanto, para a predição de faltas de pacientes, onde há um desequilíbrio acentuado de classes (CARRERAS-GARCÍA et al., 2020), confiar apenas na acurácia geral e AUC-ROC pode não ser adequado. Essas métricas podem falhar em revelar diferenças de desempenho entre os modelos (SAITO; REHMSMEIER, 2015), por vezes indicando um bom desempenho quando esse não é o caso (OZENNE; SUBTIL; MAUCORT-BOULCH, 2015).

Neste contexto, utilizando dados de 2015 a 2023 de duas unidades (do St. Central e de Jundiaí) da Clínica Radiológica de Anápolis, localizadas em Anápolis, Goiás, Brasil, propõe-se o uso de modelos de ML para predizer o não-comparecimento de pacientes em procedimentos de diferentes modalidades da clínica. Mais especificamente, foram testados os modelos de Perceptron Multicamadas (do inglês, *Multilayer Perceptron*), Regressão Logística, *XGBoost*, *LightGBM* e *CatBoost*. No caso, os dois primeiros modelos tem sido utilizado com frequência em estudos anteriores (DUNSTAN et al., 2023; LIU et al., 2022; LENZI; BEN; STEIN, 2019), enquanto os três últimos apresentam a vantagem de lidar nativamente com variáveis categóricas. Neste trabalho, foram incluídas variáveis como histórico de faltas dos pacientes, estatísticas sobre taxas de não-comparecimento anteriores dentro de janelas de 7 e 30 dias, tipo de exame, distância do paciente até a clínica e plano de saúde. As contribuições metodológicas deste trabalho incluem o uso de Validação Cruzada com Divisão Temporal (do inglês, *Time Split Cross-Validation*) (SHERIDAN, 2013) durante o processo de tunagem de hiperparâmetros, que condiz com o modo de uso real do modelo final: prever o futuro usando dados do passado. Para lidar com o desbalanceamento de classes, usamos especificamente a AUC-PR para otimização de hiperparâmetros, conforme sugerido em (BISCHL et al., 2023), e uma métrica personalizada baseada em pesos distintos para erros de Tipo I e Tipo II para analisar o desempenho final do

modelo. Além disso, reconhecendo que existe também a necessidade de compreender as causas acionáveis dos não-comparecimentos, utilizamos o *framework* SHAP, que é baseado em Valores de Shapley (LUNDBERG, 2017), para interpretar as predições dos modelos de classificação. Esse *framework* estima a contribuição de cada variável utilizada durante a modelagem, proporcionando interpretabilidade individual (por predição) ou geral (para todas as predições).

1.2 Objetivos

O objetivo geral deste projeto de pesquisa consiste em estudar, aplicar e comparar os resultados de diferentes modelos de ML para o problema de predição de absenteísmo de pacientes em procedimentos de diferentes modalidades da Clínica Radiológica de Anápolis (CRA).

Em relação aos objetivos específicos, pode-se citar:

1. Efetuar a aquisição dos dados retrospectivos relevantes através do banco de dados do CRA;
2. Fazer o devido pré-processamento dos dados, com foco em torná-los utilizáveis por modelos de ML;
3. Explorar os dados adquiridos através de sumários estatísticos ou técnicas de visualização, buscando compreender a relação entre as diferentes variáveis e a proporção de não-comparecimentos de pacientes;
4. Aplicar as técnicas apropriadas para lidar com o desbalanceamento de classes do conjunto de dados (já que a proporção de não-comparecimento de pacientes tende a ser baixa comparado ao número total de procedimentos marcados);
5. Testar diferentes modelos de classificação aplicáveis aos dados em questão, selecionando o mais apropriado para cada modalidade de procedimentos;
6. Implementar a otimização de hiperparâmetros dos modelos de classificação selecionados usando o conjunto de dados de treino e a técnica de Validação Cruzada com Divisão Temporal;
7. Usar as melhores combinações de hiperparâmetros para os modelos de classificação e analisar os resultados sobre os dados de teste, buscando compreender as causas acionáveis dos não-comparecimentos;

Por fim, mas não menos importante, pode-se incluir como objetivo específico a revisão da literatura científica com relação à problemática em questão e aos métodos que serão empregados ao longo do trabalho.

2 Fundamentação Teórica

2.1 Não-comparecimento: impactos negativos e prevalência nas instituições de saúde

Com a crescente demanda e pressões em termos de custos, é imperativo para as instituições de saúde melhorar a eficácia e eficiência dos seus serviços (MARBOUH et al., 2020). Apesar de enormes esforços, existem eventos inevitáveis que levam a custos mais altos e subutilização (MARBOUH et al., 2020). Nesse sentido, o não-comparecimento de pacientes é um ótimo exemplo que desencadeia um descompasso significativo entre a oferta e a demanda esperada, gerando, portanto, um impacto negativo na qualidade do serviço e no desempenho do sistema (MARBOUH et al., 2020; GLOWACKA; HENRY; MAY, 2009).

Os casos de não-comparecimento (ou, em inglês, os casos de não-comparecimento) ocorrem quando um paciente não comparece a uma consulta ou a um procedimento agendado sem notificação prévia ao prestador de serviços de saúde (MARBOUH et al., 2020). Os não-comparecimentos representam horários e recursos subutilizados que podem afetar negativamente o uso ótimo do espaço e dos recursos humanos disponíveis (MARBOUH et al., 2020). Ademais, até mesmo a condição de saúde do paciente pode ser afetada, tendo em vista que o diagnóstico ou o tratamento são adiados (MARBOUH et al., 2020). Pacientes em potencial são impactados, haja vista que os não-comparecimentos acarretam a incapacidade desses pacientes em agendar seus procedimentos em tempo hábil (MARBOUH et al., 2020). Já em termos financeiros, a receita gerada pela instituição de saúde diminui devido à redução na eficiência operacional (MARBOUH et al., 2020). Essa redução pode ser ainda maior quando se considera, por exemplo, os departamentos de radiologia, onde recursos, como máquinas de raios-X e de MR, são muito caros, tornando inviável a subutilização dos mesmos (MARBOUH et al., 2020).

No que se refere mais especificamente aos impactos financeiros dos não-comparecimentos, no estudo (BERG et al., 2013), envolvendo uma clínica de endoscopia, foi estimado que uma proporção de não-comparecimentos igual a 18% resulta num prejuízo de 16.4%. Já uma pesquisa conduzida num laboratório vascular reportou que uma taxa de não-comparecimento de 12% pode custar ao laboratório uma perda bruta de \$89107 anualmente, sem contar as perdas relacionadas à ineficiência do sistema de agendamento (SATIANI; MILLER; PATEL, 2009). Adicionalmente foi demonstrado que uma redução na taxa de não-comparecimento de até 5% resultaria num aumento do lucro até \$51769,00 (SATIANI; MILLER; PATEL, 2009). Além

disso, tem sido descrito que um decréscimo na taxa de não-comparecimento pode resultar numa economia significativa especialmente nos sistemas de saúde pública (ELLIS; JENKINS, 2012). Por exemplo, no *National Health Service of the United Kingdom*, uma redução na prevalência do não-comparecimento de 12% até 10.8% iria diminuir os gastos públicos em 10% (ELLIS; JENKINS, 2012).

Numa revisão sistemática da literatura tratando sobre não-comparecimentos de pacientes nas instituições de saúde, foi reportado que a prevalência de não-comparecimentos variou com relação ao continente, ao ano e à modalidade médica nas quais a proporção foi estimada (DANTAS et al., 2018). A maior média de taxa de não-comparecimento foi reportada em estudos no continente africano (43.0%), seguido pela América do Sul (27.8%), Ásia (25.1%), América do Norte (23.5%), Europa (19.3%) e Oceania (13.2%) (DANTAS et al., 2018). A mediana da taxa de não-comparecimento, considerando todos os continentes citados previamente e todas as modalidades médicas consideradas nos estudos, têm caído com o tempo, conforme é mostrado na Figura 1, que mostra o boxplot das taxas de não-comparecimento para cada década (DANTAS et al., 2018).

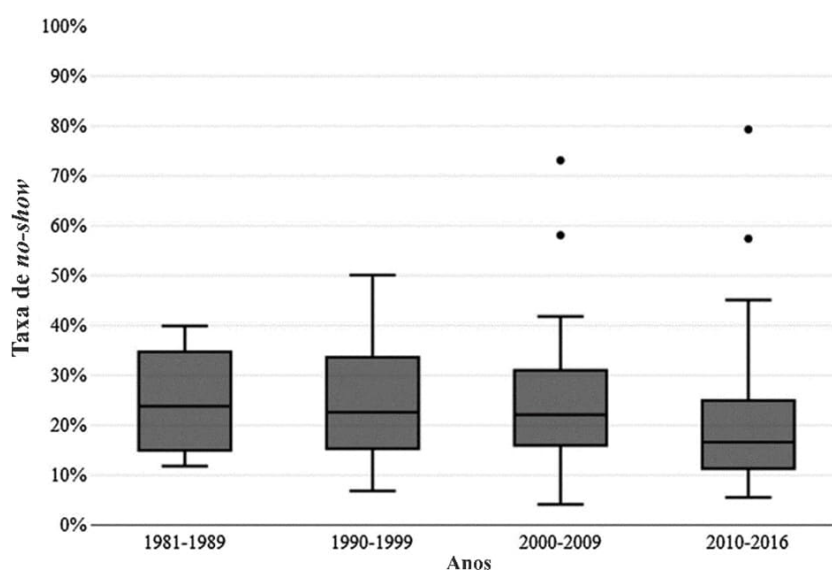


Figura 1 – Boxplot das taxas de não-comparecimento agrupadas por década. **Fonte:** adaptada de (DANTAS et al., 2018).

Já na Figura 2, são mostradas as taxas de não-comparecimento na forma de boxplot para as diferentes modalidades médicas consideradas. O termo "Outros" inclui as modalidades de tuberculose pulmonar, terapia intravenosa, reumatologia, cirurgia de mão, urologia, oftalmologia, obstétrica/ginecologia e oncologia, enquanto o termo "Várias especialidades" se refere a estudos que analisaram vários tipos diferentes de clínicas. As menores medianas de taxa de não-comparecimento estiveram associadas com outras especialidades (11.2%), seguida por

clínicas de exame médica (14.6%) e pediatria (17.0%) (DANTAS et al., 2018). Por outro lado, fisioterapia teve a maior mediana de taxa de não-comparecimento (57.3%), seguida por endocrinologia (36.0%) e cardiologia (30.0%) (DANTAS et al., 2018).

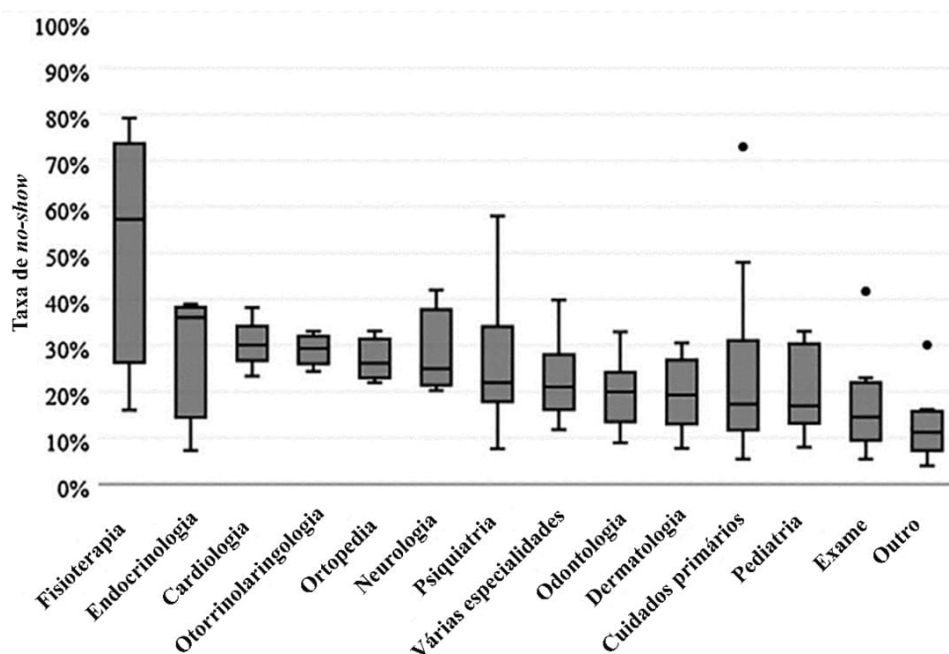


Figura 2 – Boxplot das taxas de não-comparecimento agrupadas por modalidade médica. **Fonte:** adaptada de (DANTAS et al., 2018).

2.2 Fatores relacionados ao não-comparecimento de pacientes

Os fatores associados ao não-comparecimento de pacientes incluem os relacionados à demografia do paciente (idade, sexo, condição socioeconômica etc.), às características de agendamento (dia da semana, horário do dia, mês do ano etc.) e à forma de pagamento da visita à instituição de saúde (isto é, se foi feita via plano de saúde ou se o pagamento foi particular) (DANTAS et al., 2018). Nesse contexto, as informações que serão expostas a seguir terão como referência a revisão sistemática de (DANTAS et al., 2018), que analisou uma grande quantidade de trabalhos distintos (pouco mais de 100 no total) que visaram determinar os fatores mais preditivos do não-comparecimento de pacientes. Eles foram determinados de acordo com técnicas estatísticas univariadas ou multivariadas, e por vezes através de modelos de ML. Para uma consulta pontual a algum dos trabalhos incluídos, sugere-se a leitura pormenorizada da revisão sistemática, com todas as suas referências.

Para a identificação dos fatores mais preditivos do não-comparecimento de pacientes,

a grande maioria dos estudos incluídos na revisão sistemática fizeram uso de alguma técnica estatística univariada, seguida na maioria das vezes por uma técnica multivariada. Das técnicas univariadas, foram utilizados o teste de chi-quadrado (para variáveis categóricas) e teste-t (para variáveis contínuas). Cerca de metade dos estudos considerados realizaram análise multivariada na forma de modelos de regressão logística (RL). Interessantemente, o número de estudos que conduziram esse tipo de análise cresceu com o tempo. De fato, apenas 11 dos 32 estudos que apareceram entre 1981 e 2004 produziram modelos de RL, mas o número salta para 45 de 73 entre 2005 e 2016. Essa tendência revela uma mudança das estatísticas descritivas para os modelos estatísticos inferenciais, e isso não surpreende dadas as limitações associadas com a primeira, como a inabilidade de estabelecer relações entre diferentes fatores.

Serão feitas considerações primeiramente aos fatores relacionados à demografia do paciente, isto é, idade, gênero, raça, status socioeconômico e nível de educação. A maioria dos estudos incluídos na revisão sistemática encontraram que a idade é inversamente proporcional à probabilidade de não-comparecimento dos pacientes: isto é, os jovens adultos são os mais propensos ao não-comparecimento. Por outro lado, na modalidade de pediatria, a probabilidade de não-comparecimento aumenta com a idade da criança. Grande parte dos estudos não identificaram o gênero como um preditor estatisticamente significativo de não-comparecimento, mas alguns estudos reportaram que os homens são mais propensos ao absenteísmo do que as mulheres. Grupos minoritários foram consistentemente associados com maior probabilidade de não-comparecimento, mas diferentes grupos em diferentes países foram considerados como minoria, não surpreendentemente. Menor condição socioeconômica também apresentou correlação com maiores probabilidades de não-comparecimento, e apesar do estado civil não ter sido um preditor significativo na maioria dos estudos, alguns poucos sugeriram que pacientes casados apresentam menor risco de não-comparecimento. A maioria dos estudos que analisaram o nível de educação dos pacientes não identificaram significância nesse fator. Entretanto, um dos estudos reportaram que menores níveis de escolaridade aumenta o risco de absenteísmo em 30% quando comparado com um paciente com qualquer nível de educação, e em 60% quando comparado a pacientes com um diploma universitário. No caso de consultas pediátricas, menor nível educacional parental foi associado com maior risco de não-comparecimento.

Em relação às características do agendamento, como tempo de espera (intervalo de tempo entre a data quando o agendamento foi registrado no sistema de agendamento e a data agendada), histórico prévio de não-comparecimento (existência de casos anteriores de absenteísmo), data e horário agendados, tipo de visita e o número de visitas previamente agendadas. Na maioria dos estudos, o tempo de espera foi identificado como o preditor mais importante de não-comparecimento: quanto maior a quantidade de dias entre a data de agendamento e a data agendada, maior o risco de não-comparecimento. Histórico prévio de absenteísmo também

foi um preditor forte: pacientes que se abstiveram de algum apontamento apresentaram maior probabilidade de não-comparecimento em agendamentos posteriores, mas não houve concórdância entre os estudos quanto ao número específico de apontamentos a serem considerados no passado. Na maioria dos estudos, dia da semana e mês agendado, assim como o horário, não foram preditores significativos de não-comparecimento. No entanto, entre os estudos que identificaram o fator dia da semana como preditor significativo, aproximadamente metade deles reportaram que há maiores taxas de absenteísmo às segundas-feiras, enquanto aproximadamente um terço reportou maiores taxas de absenteísmo às sextas-feiras. Procedimentos agendados através de outro provedor têm menor risco de não-comparecimento quando comparados com procedimentos agendados pelo próprio paciente ou através de outras fontes. Além disso, a maioria dos estudos encontraram que o tipo de visita é significativo, mas enquanto alguns reportam que novos pacientes (pacientes em sua primeira visita) têm risco maior de não-comparecimento, outros verificaram que visitas de retorno estavam mais propensas ao não-comparecimento. A maior parte dos estudos que analisou o número de visitas previamente agendadas encontraram que esse fator é significativo. Entretanto, quatro de sete estudos reportaram que a quantidade de visitas previamente agendadas é inversamente proporcional ao número de visitas com abstenção, enquanto o efeito oposto foi observado nos três estudos restantes.

A forma de pagamento para os serviços médicos foi identificada como um preditor significativo de não-comparecimento na maioria dos estudos. Pacientes segurados estiveram mais propensos a atender aos seus apontamentos do que os que precisavam pagar pelas despesas médicas. Vários estudos reportaram que pacientes com planos de saúde privados são menos propensos a faltar seus apontamentos do que pacientes com planos públicos.

Grande parte dos estudos que analisaram a distância até a instituição de saúde reportaram que esse fator é preditivo do comportamento de não-comparecimento, com um efeito diretamente proporcional sobre a probabilidade de absenteísmo. A distância entre a clínica e a casa dos pacientes foi estimada através do CEP (Código de Endereço Postal) da clínica e do fornecido pelo paciente. Todos os estudos que analisaram diferentes especialidades médicas reportaram que esse fator também é preditivo do comportamento de não-comparecimento. Ademais, pacientes com o número de contato registrados foram menos propensos a faltar seus apontamentos.

Fatores como o histórico médico do paciente, diagnóstico clínico e uso de medicação também foram identificados como fatores preditivos do comportamento de não-comparecimento. O histórico médico do paciente (isto é, se o paciente apresenta uma condição médica) foi reportado como tendo um efeito significativo na probabilidade de não-comparecimento na maioria dos estudos que consideraram esse fator. Seu efeito foi dependente do tipo de condição médica: por exemplo, pacientes com HIV eram menos propensos a faltar seus apontamentos, enquanto pacientes diagnosticados com depressão eram mais propensos ao não-comparecimento.

Diferentes estudos analisaram o efeito dos diferentes diagnósticos clínicos no comportamento de não-comparecimento. Pacientes diagnosticados com baixo nível glicêmico no sangue, distúrbios psiquiátricos ou doença arterial coronariana foram identificados como sendo mais propensos ao não-comparecimento. Ademais, o uso de medicações psiquiátricas e/ou antidepressivos, assim como abuso de substâncias (tabaco, drogas e/ou álcool) também foram associados com um maior risco de não-comparecimento.

Assim, como pode ser abstraído das informações previamente citadas a partir da revisão sistemática de (DANTAS et al., 2018), não existe um padrão único e universal de fatores preditivos do comportamento de não-comparecimento dos pacientes (DANTAS et al., 2018; LENZI; BEN; STEIN, 2019). Apesar disso, há o consenso de que o comportamento de não-comparecimento não é aleatório, e vários estudos reconhecem a necessidade de analisar estatisticamente os fatores que influenciam na probabilidade de não-comparecimento (SALAZAR et al., 2021). O padrão de fatores pode variar, por exemplo, entre populações diferentes e instituições de saúde (DANTAS et al., 2018; LENZI; BEN; STEIN, 2019). Nesse sentido, cabe a cada serviço a investigação local dos fatores preditivos de não-comparecimento para que seja possível tomar medidas cabíveis e razoáveis para a redução de seus efeitos negativos (LENZI; BEN; STEIN, 2019).

2.3 Redução dos efeitos negativos do não-comparecimento de pacientes

De modo a reduzir os efeitos negativos do não-comparecimento de pacientes, os centros de saúde têm implementado várias estratégias incluindo sanções e lembretes. Entretanto, apesar do fato de vários artigos questionarem sua eficiência (SATIANI; MILLER; PATEL, 2009; CHARIATTE et al., 2008), essas estratégias têm um custo significativo. Por exemplo, as sanções podem limitar o acesso de pacientes de renda inferior aos centros médicos (DAGGY et al., 2010). Já os lembretes possuem um impacto econômico que tem sido estimado em 0.41 euros por lembrete (HASVOLD; WOOTTON, 2011).

Nas últimas décadas, um número significativo de sistemas de agendamento tem sido desenvolvido para fornecer alternativas a essas estratégias (CARRERAS-GARCÍA et al., 2020). Esses sistemas têm como propósito uma melhor alocação dos apontamentos baseado na predição do não-comparecimento de pacientes (CARRERAS-GARCÍA et al., 2020). Um importante aspecto digno de nota é que a eficiência desses sistemas depende principalmente de dois elementos: a capacidade discriminatória dos preditores e a técnica de classificação usada para estimar as probabilidades de não-comparecimento (CARRERAS-GARCÍA et al., 2020).

Em relação ao primeiro desses dois elementos, conforme visto a partir da revisão sistemática citada na seção anterior (DANTAS et al., 2018), vários trabalhos têm agido no sentido de descobrir quais os fatores mais preditores que discriminam entre os pacientes que comparecem aos seus apontamentos e aqueles que não. Essas investigações têm levado à identificação de uma grande quantidade de preditores, como a porcentagem de não-comparecimentos prévios, tempo de espera, diagnóstico clínico ou idade. Conforme visto na seção anterior, (DANTAS et al., 2018) identificou mais de 40 potenciais preditores.

Em contraste às pesquisas que têm sido conduzidas para a identificação dos fatores preditores e para a construção de sistemas de agendamento acurados desde o ano de 1960, as pesquisas relacionadas ao desenvolvimento de modelos preditivos têm sido levadas à cabo somente na última década (CARRERAS-GARCÍA et al., 2020). Isso ocorre devido ao fato de que, até o surgimento recente dos sistemas eletrônicos de registro em saúde, não havia banco de dados suficiente para construir esses modelos acuradamente (CARRERAS-GARCÍA et al., 2020). É importante notar que para construir modelos acurados, é essencial que o sistema de agendamento funcione corretamente. No entanto, obter tais predições é ainda um problema não resolvido sobre o qual um número significativo de publicações tem surgido (CARRERAS-GARCÍA et al., 2020).

De modo mais específico, a distribuição de publicações por ano e pelo tamanho do banco de dados está mostrada na Figura 3. Na Figura 3-A, pode ser visto que a maioria dos artigos acerca de modelos preditores de não-comparecimento têm surgido na última década (CARRERAS-GARCÍA et al., 2020). Já na Figura 3-B, é possível perceber que o tamanho dos bancos de dados tem crescido quase em tendência exponencial, muito devido ao desenvolvimento da disponibilidade e desenvolvimento dos sistemas de registro na área de saúde (CARRERAS-GARCÍA et al., 2020).

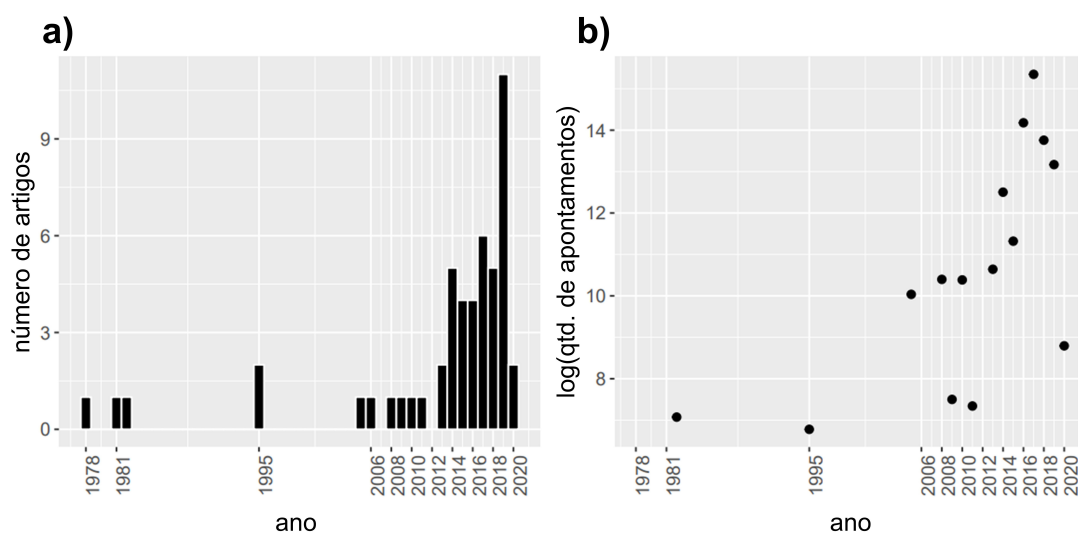


Figura 3 – a) Número de artigos publicados por ano. b) Número médio de apontamentos por ano usado nos bancos de dados numa escala logarítmica. **Fonte:** adaptada de (CARRERAS-GARCÍA et al., 2020).

2.4 Predição do não-comparecimento de pacientes

A relevância da aplicação de ML para predição do não-comparecimento de pacientes pode ser abstraída do fato de que 41 (82% do total) dos artigos analisados numa revisão sistemática foram publicados nos últimos 10 anos, e 32 (64% do total) nos últimos 5 anos (CARRERAS-GARCÍA et al., 2020). Essa mesma revisão identificou uma série de fatores que influenciam os resultados de cada um dos estudos analisados. Esses fatores incluem a escolha do modelo preditivo, as variáveis usadas por esses modelos, o método de seleção de variáveis, o *framework* de avaliação de performance, as técnicas usadas para lidar com o desbalanceamento de classes (visto que a classe positiva, isto é, o absenteísmo do paciente, é consideravelmente menos frequente do que a classe negativa, que corresponde ao comparecimento), o usufruto da dependência temporal intrapaciente e se os experimentos levaram em consideração os pacientes em primeira visita ou não (CARRERAS-GARCÍA et al., 2020).

A revisão encontrou que o algoritmo de classificação mais utilizado foi o de RL, que apareceu em 30 artigos, isto é, mais de 50% do total (CARRERAS-GARCÍA et al., 2020). Uma das razões para isso pode ser o fato de que os trabalhos iniciais foram focados em identificar os fatores mais impactantes no não-comparecimento de pacientes, papel que é desempenhado muito bem pela RL (CARRERAS-GARCÍA et al., 2020). O segundo modelo mais utilizado são as árvores de decisão (*decision trees*, abreviado como AD), que foram usadas como a técnica primária em 10 dos artigos analisados (20% do total) (CARRERAS-GARCÍA et al., 2020).

Entre os diferentes modelos, RL com regularização do tipo L2, proposto por (KURASAWA et al., 2016), se destacou, alcançando uma área sobre a curva de característica de operação do receptor (AUC-ROC) (*do inglês, receiver operating curve*) de 0.958. Outro trabalho que merece atenção especial é o de (SNOWDEN et al., 1995), que usou redes neurais (*neural networks*, abreviado como RN), alcançando uma acurácia de 91.11% num banco de dados cuja taxa de comparecimento foi de 80%. Com a atual explosão de RN profundas e o crescimento do tamanho dos bancos de dados, essa metodologia é promissora (CARRERAS-GARCÍA et al., 2020).

Em relação às variáveis utilizadas nos modelos, conforme indicado por (DANTAS et al., 2018), não há variáveis preditoras universais no banco de dados de não-comparecimento de pacientes. As variáveis mais apropriadas dependem, por exemplo, da população envolvida no estudo ou da especialidade médica. Entretanto, algumas variáveis mostram um poder discriminatório que é mais frequente na maioria dos estudos: como idade, gênero, plano de saúde, distância até a clínica, dia da semana, horário de visita, tempo de espera e histórico prévio de não-comparecimento do paciente. Essas variáveis apareceram em 50% dos estudos analisados pela revisão sistemática (CARRERAS-GARCÍA et al., 2020). Entre essas variáveis, o histórico prévio de não-comparecimentos (assim como a quantidade de apontamentos anteriores) tem sido reportado como a variável mais significativa. Isso mostra a importância da inclusão do histórico de apontamentos do paciente no estudo e reafirma a dependência intrapaciente nas observações. Apesar de em uma menor escala, variáveis como raça, estado civil e tipo de visita (se é a primeira vez do paciente na clínica ou um retorno) também têm sido frequentemente usados. Uma limitação dos estudos analisados pela revisão sistemática é que, em muitos casos, o uso de uma variável ou outra requer sua disponibilidade no sistema de registros do centro de saúde (CARRERAS-GARCÍA et al., 2020).

Um aspecto considerado importante por (CARRERAS-GARCÍA et al., 2020) é se os estudos analisados realizaram uma etapa de seleção de variáveis (em inglês, *feature selection*), já que isso pode impactar significativamente na performance dos modelos preditivos. Como uma regra geral, a adição de variáveis com baixo poder preditivo reduz o poder de generalização dos modelos. De acordo com (GUYON; ELISSEEFF, 2003), as técnicas de seleção de variáveis podem ser divididas em três principais grupos: (1) filtragem prévia, (2) testagem de combinações e (3) seleção automática. Os métodos de (1) filtragem prévia, que foram os mais frequentemente usados nos estudos analisados por (CARRERAS-GARCÍA et al., 2020), selecionam as variáveis preditoras antes de usá-las para treinamento dos modelos. A maioria dos trabalhos analisados empregaram modelos univariados para selecionar as variáveis significativas. Os métodos de (2) testagem de combinações testam múltiplos modelos que são treinados com diferentes combinações de variáveis. A técnica mais utilizada nesse sentido foi a seleção de variáveis por etapas (*do inglês, stepwise feature selection*). Outras técnicas são meta-heurísticas como algoritmo

genético (AHMADI et al., 2019) ou algoritmo de inteligência de coorte auto-adaptativo baseado em oposição (do inglês, *opposition-based self adaptive cohort intelligence*) (ALADEEMY et al., 2020). Finalmente, os (3) métodos de seleção automática envolvem a seleção de variáveis que é feita internamente em alguns modelos específicos. Nessa categoria, as técnicas mais utilizadas são AD e RL penalizada. De fato, estudos que aplicaram RL penalizada apresentaram as melhores medidas de ajuste.

Um aspecto muito importante é o *framework* de avaliação de performance, pois ele condiciona a generalização dos resultados. Na revisão sistemática em (CARRERAS-GARCÍA et al., 2020), em 13 estudos (26% do total), o desempenho dos modelos foi avaliado nos mesmos dados usados para treiná-los. É um fato bem conhecido que essa abordagem está sujeita ao sobreajuste, o que resulta em uma drástica diminuição no poder de generalização. 31 dos artigos (62% do total) conduziram uma única validação, na qual os dados foram divididos em treinamento e teste. Nesse caso, a desvantagem é que não há garantia de que as observações mais fáceis de classificar não estejam no conjunto de teste, o que pode levar a resultados excessivamente confiantes. Apenas 6 estudos (12% do total) realizaram uma validação repetida ou uma validação cruzada *k-fold*, que é o método que fornece resultados mais robustos. Esses números indicam que os resultados relatados na maioria dos estudos analisados na revisão sistemática podem não ser tão realistas quanto é possível desejar (CARRERAS-GARCÍA et al., 2020).

As medidas de desempenho são outro aspecto importante a ser destacado, pois a avaliação do desempenho na predição de não-comparecimento de pacientes varia muito entre os estudos analisados em (CARRERAS-GARCÍA et al., 2020). Em particular, 44 dos 50 artigos relatam pelo menos uma medida de desempenho. Entre essas, a métrica mais comumente usada foi a AUC, presente em 29 trabalhos. Desses 29 trabalhos, apenas um obteve um valor de AUC maior que 0,9, e apenas 6 artigos (cerca de 20%) relataram uma AUC superior a 0,85. A segunda medida de desempenho mais utilizada foi a acurácia, relatada em 21 artigos. Além disso, 17 estudos relataram especificidade e sensibilidade, 6 relataram valor preditivo positivo e valor preditivo negativo, e quatro utilizaram sensibilidade e precisão como métricas de desempenho. Por fim, 4 estudos relataram uma medida de erro (*mean absolute error*, ou erro absoluto médio: MAE, *mean squared error*, ou erro quadrático médio: MSE ou *root mean squared error*, ou raiz do erro quadrático médio: RMSE), e apenas um propôs o uso da métrica medida-F e outro o uso da medida-G. Essa heterogeneidade no uso de medidas de desempenho dificulta a comparação dos resultados entre os estudos.

O desbalanceamento de classes, por sua vez, é uma característica comum em todos os estudos analisados em (CARRERAS-GARCÍA et al., 2020). Um fato observado em vários desses estudos é que a acurácia obtida foi menor do que a taxa de comparecimento. Em particular, a acurácia supera a taxa de comparecimento em apenas 5 dos 15 estudos que relatam esses dois

valores. Essa baixa performance pode ser parcialmente atribuída ao desequilíbrio de classes, que enviesava os diferentes algoritmos para prever cada observação como um comparecimento. Adicionalmente, 26 trabalhos relatam uma taxa de não-comparecimento inferior a 20%, o que representa 68% dos 38 artigos que apresentaram esse índice. Várias abordagens foram propostas na literatura para lidar com o desequilíbrio de classes na classificação binária (KRAWCZYK, 2016). Elas podem ser categorizadas em três grupos: (1) aquelas baseadas em transformações de dados do conjunto de treinamento com o objetivo de reduzir o desequilíbrio entre as classes (por subamostragem ou superamostragem da classe majoritária ou minoritária, respectivamente), (2) aquelas baseadas no uso de algoritmos específicos que levam em consideração a distribuição desequilibrada das classes e (3) abordagens híbridas que combinam (1) e (2) (KRAWCZYK, 2016). Entre os artigos analisados, apenas o método sensível ao custo proposto por (HUANG; HANAUER, 2016) e os métodos de *ensemble/stacking* podem lidar com o desequilíbrio de classes. Eles se enquadram no segundo dos três grupos mencionados acima, ou seja, as abordagens em nível de algoritmo. Entre esses trabalhos, apenas (MOCHÓN et al., 2018) relaciona a escolha do classificador com o problema de desequilíbrio. Os autores deste estudo também apontaram que a acurácia não era uma medida de desempenho adequada e propuseram o uso da AUC. Alternativamente, (KURASAWA et al., 2016) propuseram o uso da métrica medida-F e (TOPUZ et al., 2018) propuseram a métrica medida-F.

Em relação à dependência temporal intrapaciente das observações, vários autores de trabalhos analisados por (CARRERAS-GARCÍA et al., 2020) evitaram lidar com esse problema através do uso de apenas o último apontamento do paciente (o mais recente) para treinar o modelo. No entanto, essa abordagem resulta em uma perda considerável de informação. Apenas 7 dos 50 artigos analisados incluíram a dependência temporal intrapaciente no modelo de predição. Essa dependência foi incorporada por meio de diferentes abordagens, incluindo cadeias de Markov, ponderação das observações com base em sua proximidade temporal, uso de uma soma exponencial para regressão, construção de vários modelos de RL com base no número de consultas anteriores ou uso de um modelo de RL de efeitos mistos (CARRERAS-GARCÍA et al., 2020). A última abordagem oferece uma aproximação promissora na resolução da dependência temporal, pois permite unificar o comportamento do paciente, as variáveis socioeconômicas e as variáveis ambientais (CARRERAS-GARCÍA et al., 2020).

No caso da primeira consulta do paciente, as informações disponíveis são muito limitadas. Apenas algumas variáveis ambientais estão disponíveis (por exemplo, mês, dia e horário da consulta) e talvez algumas variáveis socioeconômicas (idade e sexo do paciente). Essa limitação torna muito difícil prever o não-comparecimento na primeira consulta. Diferentes autores abordaram esse problema por meio de diferentes técnicas, como não incluir na análise pacientes que não têm um certo número de consultas anteriores, não incluir a primeira consulta no

estudo ou incluir uma variável que indica se a consulta corresponde à primeira visita, conforme (CARRERAS-GARCÍA et al., 2020).

A discussão acima mostrou que a identificação de pacientes que não comparecem às consultas é um problema desafiador e não resolvido. Como mencionado anteriormente, isso pode ser observado no fato de que apenas cinco artigos alcançaram uma acurácia maior do que a taxa de não comparecimento (CARRERAS-GARCÍA et al., 2020). Isso é consequência de vários obstáculos. Em primeiro lugar, os pesquisadores tiveram acesso a um número limitado de variáveis, que por vezes possuíam baixa capacidade de discriminação (CARRERAS-GARCÍA et al., 2020). Além disso, essas variáveis não foram as mesmas para todos os pesquisadores. Além disso, muitos estudos foram conduzidos com bancos de dados compostos por um pequeno número de pacientes, o que limitou as informações fornecidas aos classificadores (CARRERAS-GARCÍA et al., 2020). No entanto, a disponibilidade recente de bancos de dados mais informativos obtidos a partir de registros eletrônicos de saúde abre novas oportunidades de pesquisa. Esses bancos de dados atuais, contendo registros de centenas de milhares de consultas, permitem o uso de técnicas preditivas modernas, como RN profundas ou algoritmos de classificação binária inovadores para configurações de alta dimensão (CARRERAS-GARCÍA et al., 2020). Uma segunda linha de pesquisa consiste em desenvolver e incorporar estratégias que reduzam os efeitos negativos do desbalanceamento de classes (CARRERAS-GARCÍA et al., 2020). Por exemplo, o uso de técnicas de reamostragem, abordagens sensíveis ao custo ou modelos de *ensemble* pode melhorar o desempenho do classificador selecionado (CARRERAS-GARCÍA et al., 2020). Uma terceira possibilidade é a incorporação da dependência temporal intrapaciente, o que permitiria uma melhor caracterização do comportamento dos pacientes, unificando seus registros anteriores de comparecimento, suas características socioeconômicas e as variáveis ambientais (CARRERAS-GARCÍA et al., 2020). Essas estratégias podem levar a predições mais precisas que, quando incorporadas aos sistemas de agendamento, reduzirão as perdas econômicas sofridas pelos centros de saúde e o tempo de espera para acesso aos serviços médicos (CARRERAS-GARCÍA et al., 2020).

2.5 Análise de Componente Principal (PCA)

2.5.1 Definição qualitativa

A Análise de Componente Principal (PCA), introduzida por Karl Pearson no começo do século 20 e desenvolvido por Harold Hotteling em 1933, é uma técnica multivariada que busca explicar um conjunto de variáveis correlacionadas em termos de um número reduzido de variáveis não correlacionadas, de maior variância (SANTO, 2012; ABDI; WILLIAMS, 2010).

Resumidamente, os objetivos do PCA são (SANTO, 2012; ABDI; WILLIAMS, 2010):

1. Extrair somente o que é importante do conjunto de dados;
2. Comprimir o tamanho total do conjunto de dados;
3. Simplificar a estrutura do conjunto de dados;
4. Analisar a estrutura das observações e das variáveis;

De modo a cumprir tais objetivos, o PCA estima novas variáveis denominadas componentes principais, que são obtidos como combinações lineares das variáveis originais do conjunto de dados. O primeiro componente principal deve possuir a maior variância possível. Já o segundo componente é computado com a restrição de que deve ser ortogonal ao primeiro componente principal e possuir a maior variância possível. Os outros componentes principais, por sua vez, são computados da mesma forma. Os valores dessas novas variáveis para as observações são denominados escores, e eles podem ser interpretados geometricamente como sendo as projeções das observações sobre os componentes principais (SANTO, 2012; ABDI; WILLIAMS, 2010).

2.5.2 Definição quantitativa

A técnica de PCA será definida partindo-se de uma amostra de observações definidas por um conjunto definido de variáveis. Vamos considerar que haja p variáveis contínuas e n observações para essas variáveis de modo que pode-se resumir os dados numa matriz $\bar{X} = (x_{ij})_{n \times p}$. Os vetores-coluna pertencentes a essa matriz serão denotados como $\vec{X}_1, \vec{X}_2, \dots, \vec{X}_p$ (AGUILERA; ESCABIAS; VALDERRAMA, 2006; ABDI; WILLIAMS, 2010).

Seja $\bar{S} = (s_{jk})_{p \times p}$ a matriz de covariância amostral, cujos elementos s_{jk} são definidos por $s_{jk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \mu_j)(x_{ik} - \mu_k)$, com as médias amostrais μ_j calculadas por $\mu_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$, ($j = 1, \dots, p$). Como simplificação, será considerado, sem perda de generalidade, que as observações foram centralizadas para todas as variáveis, de modo que $\mu_1 = \mu_2 = \dots = \mu_p = 0$, o que faz com que a matriz de covariância seja expressa por $\bar{S} = \frac{1}{n-1} \bar{X}^T \bar{X}$ (AGUILERA; ESCABIAS; VALDERRAMA, 2006; ABDI; WILLIAMS, 2010).

Os componentes principais \vec{C}_j , por sua vez, são definidos como expansões lineares e ortogonais entre si com máxima variância calculadas das colunas da matriz \bar{X} . Matematicamente, tem-se: $\vec{C}_j = \bar{X} \vec{V}_j$, com $j = 1, \dots, p$. Assim, os vetores de coeficientes que definem os componentes principais, $\vec{V}_1, \dots, \vec{V}_p$, são os autovetores da matriz de covariância amostral \bar{S} associados aos seus autovalores $\lambda_1 \geq \dots \geq \lambda_p \geq 0$ correspondentes, que são as variâncias dos componentes principais. Se denotarmos por \bar{C} a matriz cujas colunas são os componentes

principais, ela pode ser expressa por $\bar{C} = \bar{X}\bar{V}$, com $\bar{V} = (v_{jk})_{p \times p}$ sendo a matriz que possui como colunas os autovetores da matriz de covariância amostral (AGUILERA; ESCABIAS; VALDERRAMA, 2006; ABDI; WILLIAMS, 2010).

Como a matriz de covariância amostral é decomposta como $\bar{S} = \bar{V}\Delta\bar{V}^T$, com \bar{V} sendo ortogonal e $\Delta = \text{diag}(\lambda_1, \dots, \lambda_p)$, então a matriz de observações \bar{X} pode ser escrita como $\bar{X} = \bar{C}\bar{V}^T$. Tal decomposição por componentes principais permite a obtenção de uma reconstrução aproximada de cada observação original em termos de um número reduzido de componentes principais (AGUILERA; ESCABIAS; VALDERRAMA, 2006; ABDI; WILLIAMS, 2010):

$$\vec{X}_j = \sum_{k=1}^s \vec{C}_k v_{jk}, \quad j = 1, \dots, p \quad (2.1)$$

que explica uma alta porcentagem da variabilidade total, calculada como (AGUILERA; ESCABIAS; VALDERRAMA, 2006; ABDI; WILLIAMS, 2010):

$$\frac{\sum_{k=1}^s \lambda_j}{\sum_{k=1}^p \lambda_j} \times 100, \quad s \leq p. \quad (2.2)$$

2.6 Regressão logística

2.6.1 Descrição matemática

Seja $\vec{X}_1, \vec{X}_2, \dots, \vec{X}_p$ um conjunto de variáveis contínuas observadas sem erro, com um total de n observações. Assim, os dados podem ser resumidos pela matriz $\bar{X} = (x_{ij})_{n \times p}$. Considera-se $Y = (y_1, y_2, \dots, y_n)^T$ uma amostra aleatória da variável resposta binária \tilde{Y} associada com as observações em \bar{X} , isto é, $y_i \in [0, 1]$, $i = 1, \dots, n$. Assim, o modelo de RL é dado por (AGUILERA; ESCABIAS; VALDERRAMA, 2006; LIU; CHEN; YE, 2009):

$$y_i = \pi_i + \epsilon_i, \quad i = 1, \dots, n, \quad (2.3)$$

onde π_i é a probabilidade de \tilde{Y} dado $(\vec{X}_1 = x_{i1}, \vec{X}_2 = x_{i2}, \dots, \vec{X}_p = x_{ip})$, calculada como (AGUILERA; ESCABIAS; VALDERRAMA, 2006; LIU; CHEN; YE, 2009):

$$\pi_i = P[\vec{X}_1 = x_{i1}, \vec{X}_2 = x_{i2}, \dots, \vec{X}_p = x_{ip}] = \frac{\exp\left\{\left(\beta_0 + \sum_{j=1}^p x_{ij}\beta_j\right)\right\}}{1 + \exp\left\{\left(\beta_0 + \sum_{j=1}^p x_{ij}\beta_j\right)\right\}}, \quad (2.4)$$

onde $\beta_0, \beta_1, \dots, \beta_p$ são os parâmetros do modelo e ϵ_i são erros independentes com média 0 e variâncias dadas por $\text{Var}[\epsilon_i] = \pi_i(1 - \pi_i)$, com $i = 1, \dots, n$ (AGUILERA; ESCABIAS; VALDERRAMA, 2006; LIU; CHEN; YE, 2009).

Define-se a transformação logit como $l_i = \log(\pi_i/(1 - \pi_i))$, com $i = 1, \dots, n$, onde $\pi_i/(1 - \pi_i)$ representa os *odds* da resposta $\tilde{Y} = 1$ para o valor observado $x_i = (x_{i1}, \dots, x_{ip})$. Assim, o modelo de RL pode ser entendido como um modelo linear generalizado com a transformação *logit* como função de ligação (AGUILERA; ESCABIAS; VALDERRAMA, 2006; LIU; CHEN; YE, 2009).

Antes de estimar o modelo de RL, deve-se lembrar que a relação entre a variável resposta e cada variável preditora pode ser interpretada em termos de razões de *odds* a partir dos parâmetros do modelo. Da Equação (2.4), tem-se que a exponencial do j -ésimo parâmetro ($j = 1, \dots, p$) é a razão de chances de sucesso ($\tilde{Y} = 1$) quando o j -ésimo preditor é acrescido de uma unidade e os outros preditores são controlados (fixados como constantes) (AGUILERA; ESCABIAS; VALDERRAMA, 2006; LIU; CHEN; YE, 2009). Matematicamente:

$$\theta(\Delta \vec{X}_j | \vec{X}_k = \vec{X}_{0k}, \forall k \neq j) = \frac{\pi(x_1, \dots, x_j+1, \dots, x_p)}{1 - \pi(x_1, \dots, x_j, \dots, x_p)} = \exp\{(\beta_j)\} \quad (2.5)$$

com x_1, x_2, \dots, x_p sendo uma única observação das variáveis predictoras. Assim, a exponencial do j -ésimo parâmetro do modelo de RL dá a mudança multiplicativa nos *odds* de sucesso se o parâmetro for positivo. Essa "razão de *odds*" ajuda a medir, em um exemplo mais intuitivo, a mudança relativa na probabilidade de recuperação de um paciente quando aumentamos a dose de um determinado medicamento. Essa interpretação dos parâmetros necessita, portanto, de uma estimativa acurada dos parâmetros do modelo de RL (AGUILERA; ESCABIAS; VALDERRAMA, 2006; LIU; CHEN; YE, 2009).

Um dos métodos utilizados para estimar o modelo de RL é o de máxima verossimilhança (*maximum likelihood*) (AGUILERA; ESCABIAS; VALDERRAMA, 2006; LIU; CHEN; YE, 2009). Seja $L(Y; \beta)$ a verossimilhança dada por

$$L(Y; \beta) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{(1-y_i)}. \quad (2.6)$$

onde a forma matricial das equações de verossimilhança é $\bar{X}^T(Y - \Pi) = 0$, com $\Pi = (\pi_1, \dots, \pi_n)^T$. Essas equações não são lineares com respeito aos parâmetros, de modo que elas precisam ser maximizadas utilizando um método numérico tal como o de Newton-Raphson (AGUILERA; ESCABIAS; VALDERRAMA, 2006; LIU; CHEN; YE, 2009).

Uma medida de validade do modelo é a taxa de classificações corretas, definida como a porcentagem de acertos no conjunto de dados. De modo a classificar uma observação de acordo com o modelo estimado, é comum fixar-se um limiar entre 0 e 1, de modo a diferenciar a resposta binária. Assim, uma instância é corretamente classificada quando os valores observados e preditos são concordantes. O limiar mais comumente utilizado é 0.5 mas é possível utilizar outros valores (AGUILERA; ESCABIAS; VALDERRAMA, 2006; LIU; CHEN; YE, 2009).

Outra tarefa importante atribuída à RL é a seleção das melhores covariáveis que explicam a variável resposta. Muitos procedimentos de seleção de modelo existem, mas nenhum é "o melhor". Os cuidados em regressão linear também se aplicam à RL. Por exemplo, um modelo com excesso de variáveis preditoras corre o risco de multicolinearidade de modo que uma das variáveis pode ter um efeito pequeno simplesmente porque ela age conjuntamente com outras variáveis do modelo (AGUILERA; ESCABIAS; VALDERRAMA, 2006; LIU; CHEN; YE, 2009).

2.6.2 Função de custo e regularização

Conforme explicado anteriormente, no modelo de RL a variável resposta é modelada como uma distribuição binomial. Para estimar todos os parâmetros do modelo, é preciso maximizar a verossimilhança, definida matematicamente na Equação (2.6). Pode-se aplicar uma transformação logarítmica sobre a verossimilhança, de modo que (SALEHI; ABBASI; HASSIBI, 2019)

$$\log(L) = \sum_{i=1}^n \left[\log(1 - \pi_i) + y_i \log\left(\frac{\pi_i}{1 - \pi_i}\right) \right]. \quad (2.7)$$

Em termos computacionais, a Equação (2.7) é mais simples do que a Equação (2.6), e por isso a transformação logarítmica é conveniente (SALEHI; ABBASI; HASSIBI, 2019).

Em linguagens de programação amplamente utilizadas para modelagem estatística e *Machine Learning*, como *R* e *Python*, as funcionalidades são intrinsecamente projetadas para otimizar os processos computacionais envolvidos na minimização de funções matemáticas. Consequentemente, para a Regressão Logística (RL), torna-se essencial a especificação de uma função de custo, ou função de perda, que possa ser efetivamente minimizada durante o processo de estimação dos parâmetros do modelo. A estratégia convencionalmente adotada para este fim

é a minimização do negativo da função de verossimilhança logarítmica (SALEHI; ABBASI; HASSIBI, 2019).:

$$F_{\text{custo}} = -\log(L) = -\sum_{i=1}^n \left[\log(1 - \pi_i) + y_i \log\left(\frac{\pi_i}{1 - \pi_i}\right) \right] \quad (2.8)$$

Definida a função de custo do modelo de RL, é possível limitar a flexibilidade do modelo de modo a tentar aumentar sua generalidade para dados futuros ou não vistos durante o treinamento. Em geral, qualquer modificação no método de aprendizagem de um modelo com o objetivo de melhorar sua performance é denominado "regularização". Neste caso em específico, a regularização é introduzida com a adição de um termo à função de custo, de modo a limitar os valores dos parâmetros estimados (SALEHI; ABBASI; HASSIBI, 2019).

Um dos métodos de regularização é o do tipo L2. Ele é aplicado adicionando um termo de penalização $\lambda \sum_{j=1}^p \beta_j^2$ à função de custo, que penalizará o modelo para valores altos dos coeficientes, limitando, portanto, seus valores para os mais próximos de 0. O parâmetro λ controla a ênfase que é dada ao termo de penalização: quanto maior λ , mais os coeficientes tenderão a 0. Apesar de os coeficientes tenderem a 0, poucos coeficientes chegam realmente a possuir valor nulo. Isso não é desejável, por exemplo, se o objetivo for selecionar as variáveis preditoras mais importantes. Apesar disso, a regularização do tipo L2 é extremamente útil na melhora da generalização dos modelos de RL, controlando o sobreajuste (SALEHI; ABBASI; HASSIBI, 2019).

De modo a selecionar as variáveis com maior poder preditivo, há o método de regularização do tipo L1. O termo que é adicionado à função de custo é parecido com o da regularização do tipo L2: $\lambda \sum_{j=1}^p |\beta_j|$. Com esse termo de penalização, as variáveis preditoras menos importantes são forçadas a apresentar valor nulo (SALEHI; ABBASI; HASSIBI, 2019).

Existem vantagens e desvantagens (FRIEDMAN; HASTIE; TIBSHIRANI, 2010) em cada um dos métodos de regularização. Por exemplo, quando as variáveis estão correlacionadas, a regularização do tipo L1 tende a selecionar aleatoriamente as variáveis preditoras. Na regressão de Ridge, por sua vez, as variáveis correlacionadas têm suas contribuições diminuídas conjuntamente, mas mantendo a contribuição de todas elas. Mais recentemente, um método chamado *Elastic Net* foi proposto para incluir o melhor das duas técnicas de regularização, introduzindo um termo de penalização do tipo $\lambda \sum_{j=1}^p (\alpha \beta_j^2 + (1 - \alpha) |\beta_j|)$ (FRIEDMAN; HASTIE; TIBSHIRANI, 2010).

2.7 Perceptron Multicamadas

2.7.1 Neurônio artificial

O neurônio artificial é inspirado pelo neurônio biológico. Em sistemas biológicos, impulsos nervosos são recebidos pelos dendritos e processados no corpo da célula. Dependendo do resultado da integração dos sinais recebidos, se eles são maiores que um dado limiar, o neurônio pode ou não produzir um novo impulso (um potencial de ação), que será, por sua vez, transmitido para outros neurônios conectados ao seu terminal axônico (CASTRO, 2006). A união dos axônios com dendritos é denominada sinapse, que funciona como uma válvula que controla o fluxo de informações entre os neurônios.

Na Figura 4 é mostrado um neurônio artificial usado em redes neurais artificiais. Esse dispositivo recebe um conjunto $x = [x_1, x_2, \dots, x_I]$ como entrada, que pode vir de outros neurônios ou diretamente das variáveis de entrada do problema a ser resolvido. Quando elas correspondem às variáveis de entrada do problema, $I = d$ é a dimensionalidade do problema. Esses dados (o conjunto x) são processados, e o neurônio responde com um sinal \hat{y}_j , onde j é o neurônio em questão. Essa resposta, que é uma transformação dos valores de entrada recebidos, ou é propagada para outros neurônios, ou dada como saída da rede neural artificial.

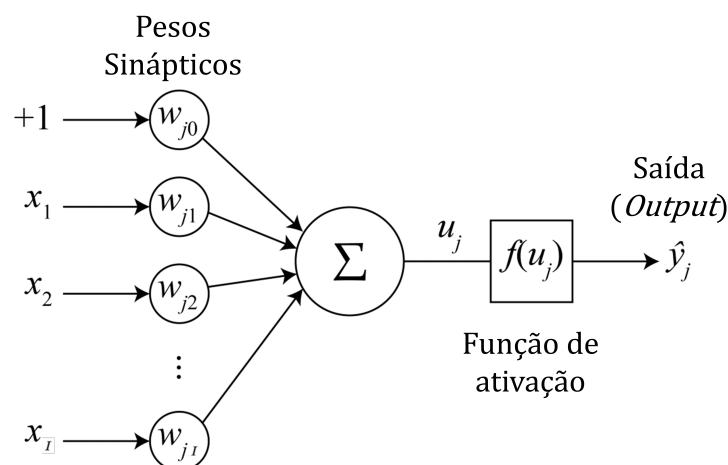


Figura 4 – Diagrama de um neurônio artificial. **Fonte:** adaptada de (COSTA et al., 2023).

O processamento realizado pelo neurônio j consiste em atribuir pesos w_{ji} (análogos a "coeficientes angulares") para os sinais de entrada x_i ($i = 1, 2, \dots, I$) e um valor fixo de viés w_{j0} (análogo a um "coeficiente linear"). A soma u_j dos valores de entrada multiplicados pelos seus respectivos pesos é calculada e passa por uma função de ativação $f(\cdot)$, gerando a saída \hat{y}_j . A representação matemática disso é dada na Equação 2.9:

$$\hat{y}_j = f(u_j) = f\left(\sum_{i=0}^I w_{ji}x_i\right) \quad (2.9)$$

Uma função de ativação $f(\cdot)$ define a entrada de um dado neurônio, dado o valor de ativação u . Várias funções têm sido usadas ou desenvolvidas nesse sentido, com a maioria delas consistindo em funções não-lineares, o que confere maior flexibilidade aos valores de saída das redes neurais artificiais. Alguns exemplos serão citados logo a seguir.

A função de ativação linear é uma função que pode ser descrita pela Equação 2.10:

$$f(u) = \alpha u \quad (2.10)$$

Outra função padrão, mas não-linear, é a função "sinal", também conhecida como *Heaviside* (HAYKIN, 2008), que resulta em saídas binárias $\{0, 1\}$ ou bipolares $\{-1, 1\}$. A função com o primeiro modo de saída é mostrada na Equação 2.11:

$$f(u) = \begin{cases} 1, & \text{se } u \geq 0 \\ 0, & \text{se } u < 0 \end{cases} \quad (2.11)$$

A função de ativação sigmoide, por sua vez, apresenta o formato peculiar de "S" (com valores entre 0 e 1) e representa um balanço entre os comportamentos linear e não-linear (HAYKIN, 2008; CASTRO, 2006), que pode ser obtido pelo uso da função logística (JEFFREY; DAI, 2008), conforme Equação 2.12:

$$f(u) = \frac{1}{1 + e^{-u}} \quad (2.12)$$

De forma geral, o uso de funções como a função de ativação sigmoide é motivado devido a algumas características muito úteis (MENON et al., 1996):

- É uma função contínua e diferenciável em todos os pontos;
- Existe saturação para os valores de saída quando os valores de ativação u são muito baixos ou muito altos;
- É possível utilizá-la para criar diferentes "mapeamentos", haja vista que essa função tem comportamento praticamente linear quando u é próximo de 0, e comportamento fortemente não-linear quando u é muito diferente de 0.

Uma função com um comportamento parecido com o da função sigmoide (mas com valores entre -1 e 1) é a função de ativação tangente hiperbólica (HAYKIN, 2008), cuja expressão está mostrada na Equação 2.13:

$$f(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}} = \tanh u \quad (2.13)$$

Mais recentemente, com o avanço dos estudos envolvendo redes neurais artificiais ou, mais especificamente, aprendizado profundo (*deep learning*), novas funções de ativação tem sido desenvolvidas para resolver o problema de gradiente evanescente (FAHLMAN et al., 1988), que pode surgir quando há o encadeamento de muitas camadas de neurônios utilizando funções de ativação tais como a sigmoide ou tangente hiperbólica.

Uma função de ativação que previne esse problema é a Unidade Linear Retificada (do inglês, *Rectified Linear Unit*, ou ReLU) (MAAS et al., 2013), que é similar à função de ativação linear para valores de u maiores do que 0, conforme é mostrado na Equação 2.14:

$$f(u) = \begin{cases} u, & \text{se } u > 0 \\ 0, & \text{se } u \leq 0 \end{cases} \quad (2.14)$$

Outra função de ativação que surgiu recentemente foi a Unidade Linear Exponencial (do inglês, *Exponential Linear Unit*, ou ELU) (HOCHREITER; UNTERTHINER; CLEVERT, 2016). Ela é definida conforme a Equação 2.15 e, diferentemente da função de ativação ReLU, a ELU permite saídas negativas.

$$f(u) = \begin{cases} u, & \text{se } u > 0 \\ \alpha(e^u - 1), & \text{se } u \leq 0 \end{cases} \quad (2.15)$$

Na Figura 5 é mostrado o comportamento de cada uma das funções de ativação citadas nesta seção.

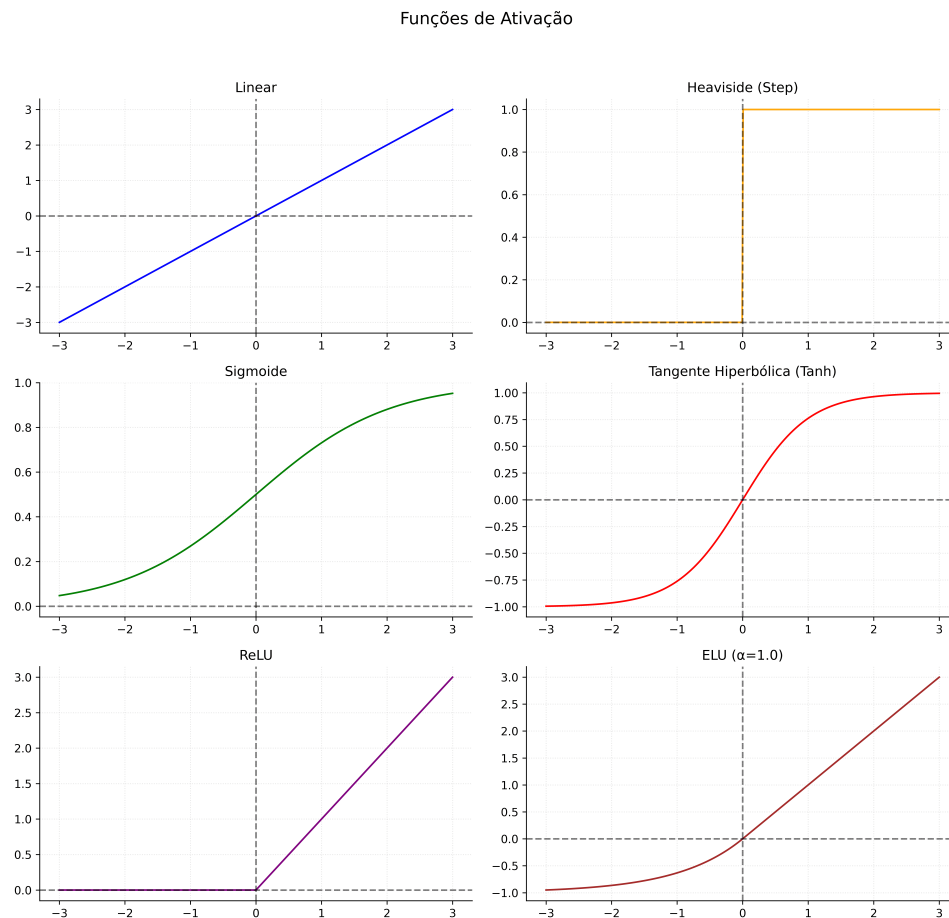


Figura 5 – Comportamento das diferentes funções de ativação citadas no texto. **Fonte:** próprio autor.

2.7.2 Perceptron Multicamadas: uma arquitetura de rede neural artificial

Uma das arquiteturas de redes neurais artificiais mais bem conhecidas é o Perceptron Multicamadas (do inglês *Multilayer Perceptron*, doravante referido como MLP), que generaliza estruturalmente o neurônio artificial denominado perceptron Rosenblatt, que faz uso da função de ativação de *Heaviside*. Conforme demonstrado por (CYBENKO, 1989), o MLP tem capacidade de aproximação universal: pode aproximar qualquer função contínua, limitada, diferenciável e não-linear de valores de entrada definidos num espaço compacto de precisão arbitrária. Isso é possível graças à composição aditiva das funções de base, que, para o MLP, são funções de Ridge. Entretanto, esse teorema não especifica a quantidade de neurônios artificiais requeridos por cada problema nem algum método que sirva para ajustar os pesos do MLP de modo que a configuração ótima da rede neural seja garantida.

Os MLPs organizam os neurônios em várias camadas encadeadas entre si, onde cada

camada é o arranjo de neurônios paralelos. Neurônios de uma mesma camada não se comunicam entre si e somente enviam suas saídas para "frente" (isto é, para camadas posteriores), o que é conhecido como uma estrutura de propagação direta (*feedforward*). Os MLPs possuem uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. A Figura 6 mostra um exemplo de possível estrutura de MLP para um problema de classificação binária, com variável de entrada d-dimensional. Como os neurônios são organizados em camadas, usa-se números sobrescritos para a identificação da mesma. A primeira camada ($l = 0$) é a camada de entrada, onde os neurônios recebem diretamente as variáveis de entrada do problema em questão. Os neurônios das camadas ocultas ($0 < l < L$) e da camada de saída ($l = L$) são neurônios artificiais padrões que recebem como entrada os valores de saída dos neurônios da camada anterior. A última camada é a camada de saída, onde os neurônios produzem os valores de saída para o problema em questão.

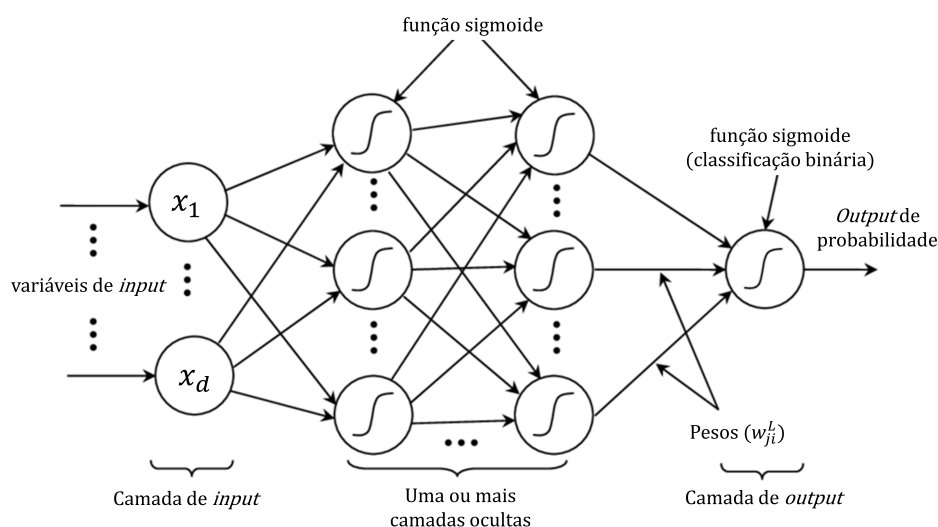


Figura 6 – Exemplo de arquitetura de MLP para um problema de classificação binária de entrada d-dimensional. Note que, para as camadas ocultas, qualquer outra função de ativação seria possível no lugar da sigmoide. **Fonte:** adaptada de (BOUZGOU, 2012).

Uma ou mais camadas ocultas para o MLP podem ser utilizadas. Elas são responsáveis por mapear o sinal de entrada de modo não-linear para outro espaço, de acordo com a demanda do problema. Como a combinação de funções lineares é também uma função linear, as funções de ativação lineares (conforme citado anteriormente) quase nunca são utilizadas.

Ao lidar com problemas de classificação binária ou de regressão, um único nó de saída é utilizado. Ao lidar com problemas multiclasse, um nó de saída é utilizado para cada classe. Após o treinamento do MLP (que será discutido adiante), predições são feitas através do mapeamento dos valores numéricos dados pelos nós de saída em classes ao lidar com problemas de classificação. Em particular, funções de ativação sigmoide são tipicamente utilizadas no nó de

saída para problemas de classificação binária. Se o valor de saída for maior que 0.5, assume-se que a predição é a classe positiva. Se a saída for menor ou igual a 0.5, a classe negativa é assumida.

2.7.3 Treinamento

Algoritmos de treinamento têm em vista o ajuste dos pesos de um MLP. Por essa razão, os pesos do MLP são também entendidos como "parâmetros" que devem ser aprendidos no processo de treinamento. Haverá o foco aqui no método de gradiente descendente juntamente com o famoso algoritmo de retropropagação (do inglês, *backpropagation*) (RUMELHART; HINTON; WILLIAMS, 1986).

O processo de treinamento de um MLP é tipicamente supervisionado, sendo necessário um conjunto de dados de treino devidamente rotulado. A partir do algoritmo de retropropagação, o treinamento consiste em duas etapas: *forward* e *backward*. Na etapa de *forward*, os pesos da rede neural artificial não mudam; eles são fixados, e os dados de entrada (provenientes do conjunto de treino) são propagados da primeira até a última camada para a obtenção da saída da rede neural. Em seguida, o erro em relação ao valor de saída real é calculado, assim como o vetor gradiente da função de erro em relação aos parâmetros da rede neural. Baseado nisso, a etapa de *backward* aplica o método de otimização gradiente descendente da última até a primeira camada do MLP, de modo que os pesos dos neurônios são ajustados na direção oposta do gradiente (na direção oposta ao crescimento da função de erro) (HAYKIN, 2008). As propagações *forward* e *backward* de todo o conjunto de treino são aplicados iterativamente até que algum critério de parada seja atendido (CASTRO, 2006). Cada iteração através de todo o conjunto de dados de treino é chamada de "época".

Vale notar que esse procedimento nada mais é do que um algoritmo de otimização (gradiente descendente) sendo aplicado a um problema de otimização não-linear sem restrições. Nesse sentido, qualquer outro algoritmo de otimização aplicável para este problema pode ser utilizado.

Existem diferentes formas de aplicar o método de gradiente descendente, haja vista que a quantidade de dados utilizada no processo interfere diretamente na velocidade de ajuste dos pesos da rede neural:

- Aprendizado por lote de dados (do inglês, *batch*): os novos pesos são calculados após todos os exemplos disponíveis para treino terem sido apresentados para a rede neural, ao final de cada época durante o processo de treinamento. Em outras palavras, o gradiente da

função de erro é calculado como a média entre os gradientes calculados para os erros em cada exemplo de treino;

- Aprendizado *online* (gradiente descendente estocástico): os novos pesos são calculados a medida que cada exemplo de treino é apresentado. Em outras palavras, eles são calculados através do gradiente de um único exemplo de treino que é apresentado, escolhido aleatoriamente (por isso é estocástico e é um processo bastante "ruidoso");
- Aprendizado por mini-lotes de dados (do inglês, *mini-batch*): esse é um método intermediário entre os dois métodos citados anteriormente. A média do gradiente da função de erro é calculada através de lotes menores dos dados de treino e posteriormente é utilizada para o cálculo dos novos pesos da rede neural, até o final da época.

2.7.4 Algoritmo de retropropagação

Aqui serão seguidas as informações contidas em (HAYKIN, 2008) para a formalização do algoritmo de retropropagação. Mais especificamente, a versão "online"(ou de gradiente descendente estocástico) será apresentada.

Antes de iniciar o processo de aprendizado, os pesos do MLP são tipicamente inicializados de forma aleatória no intervalo $[-1, 1]$. É possível que os pesos também sejam inicialmente fixados de acordo com algum conhecimento prévio, mas esse não é o cenário mais comum.

O primeiro passo na aplicação do algoritmo de retropropagação é a propagação forward do sinal de entrada, de modo a obter um valor de saída. Considerando que um exemplo de treino seja da forma $(\mathbf{x}(n), \mathbf{y}(n))$ para a iteração n , onde $\mathbf{x}(n)$ representa as variáveis de entrada e $y(n)$ a saída desejada. Nesse sentido, o valor de ativação u_j^l para o neurônio j da camada l está expresso na Equação 2.16:

$$u_j^l(n) = \sum_i \omega_{ji}^l(n) y_i^{(l-1)}(n) \quad (2.16)$$

Desse modo, a saída final do neurônio j da camada l será:

$$\hat{y}_j^l(n) = f_j(u_j^l(n)) \quad (2.17)$$

onde f_j é a função de ativação do neurônio j . Note que, com base na Equação 2.16, se $l = 1$, tem-se que $\hat{y}_j^0(n) = x_j(n)$, onde $x_j(n)$ é o j -ésimo elemento do vetor de entrada $\mathbf{x}(n)$.

O segundo passo do algoritmo de *backpropagation* consiste em retropropagar o valor resultante da função de erro (isto é, propagar o valor de erro da camada de saída até a camada

inicial do MLP). Vamos supor que o valor da função de erro na última camada do MLP seja conforme a Equação 2.18:

$$e_j(n) = F_{erro}(\hat{y}_j^L(n), y_j^L(n)) \quad (2.18)$$

Nesse sentido, o gradiente local em relação aos valores de ativação para a camada de saída é:

$$\delta_j^L(n) = f'_j(u_j^L(n))e_j(n) \quad (2.19)$$

Propagando o erro para as camadas anteriores, tem-se que, para os neurônios das camadas ocultas:

$$\delta_j^l(n) = f'_j(u_j^l(n)) \sum_k \delta_k^{l+1}(n) \omega_{kj}^{l+1}(n) \quad (2.20)$$

Assim, para ajustar os pesos com base no erro $e_j(n)$, faz-se:

$$\omega_{ji}^l(n+1) = \omega_{ji}^l(n) + \eta \delta_j^l(n) \hat{y}_i^{(l-1)}(n) \quad (2.21)$$

Note que, da Equação 2.21, a variável η indica a taxa de aprendizado (do inglês, *learning rate*), que é um valor que controla a escala de ajuste dos pesos do MLP. Quanto maior η , maior é a escala do ajuste, porém também maior é a instabilidade do processo. Essas atualizações dos pesos vão sendo repetidas até que seja atingido um número específico de iterações (HAYKIN, 2008).

2.8 Árvores de decisão

Um dos principais algoritmos utilizados para gerar árvores de decisão é o CART (*Classification And Regression Trees*), originalmente proposto em (BREIMAN, 2017). Nesse sentido, as árvores de decisão são modelos não-paramétricos muito utilizados em problemas de classificação e regressão devido à sua flexibilidade e robustez a *outliers* (STEINBERG; COLLA, 1995). O método CART não possui nenhum pressuposto de distribuição das variáveis envolvidas no problema, tornando-o utilizável em uma grande variedade de aplicações.

As árvores de decisão são populares devido à sua interpretabilidade e habilidade em lidar com dados numéricos e categórico, assim como por sua robustez a *outliers*. Entretanto, trata-se de um modelo sujeito ao sobreajuste, especialmente em problemas com conjuntos de

dados pequenos ou quando os dados contêm muitas variáveis preditoras. Técnicas como poda (*pruning*), *bagging* e *boosting* são comumente empregadas envolvendo árvores de decisão a fim de mitigar tais limitações (SUTTON, 2005).

2.8.1 Estrutura das árvores de decisão

O método CART constrói árvores de decisão para problemas de classificação a partir do particionamento recursivo do espaço de entrada \mathbf{X} em regiões disjuntas, conforme o exemplo mostrado na Figura 7. Isso é feito com o objetivo de prever a classe da nova observação associando-a a uma das regiões particionadas. Cada divisão (*split*) da árvore de decisão corresponde a uma regra de decisão (nó) baseada em uma ou mais variáveis preditoras. O processo de crescimento da árvore continua até que certo critério de parada seja satisfeito, resultando numa estrutura hierárquica (de nós até as "folhas"/"decisões") que pode ser utilizada para a tarefa de classificação.

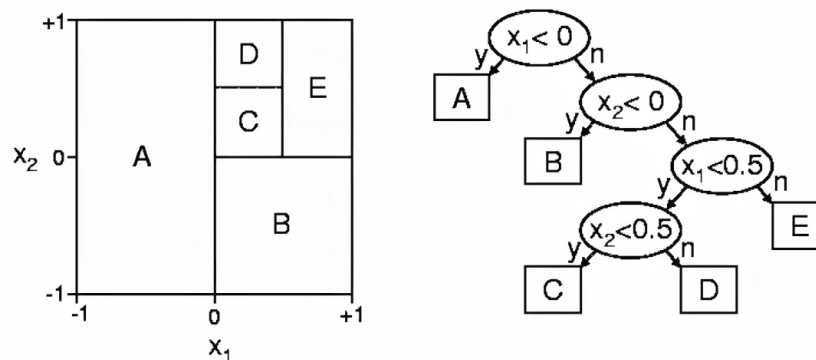


Figura 7 – Exemplo de árvore de decisão, particionando o espaço de entrada contendo x_1 e x_2 nas regiões A, B, C, D e E. Fonte: (POTTS; SAMMUT, 2004).

2.8.2 Formulação matemática das divisões em cada nó

O algoritmo CART emprega um particionamento recursivo binário para, em cada nó, dividir os dados. A divisão ótima é escolhida baseado no critério que minimiza a função de impureza escolhida. Duas medidas de impureza comumente utilizadas são o índice de Gini e entropia (ganho de informação/*information gain*) (SUTTON, 2005).

O índice de Gini mede o grau de diferença na distribuição de classes em um dado nó e é definido matematicamente conforme a Equação 2.22 (SUTTON, 2005), onde p_j representa a proporção de casos no nó que corresponde à classe j :

$$g(p_1, \dots, p_j) = 1 - \sum_{j=1}^J p_j^2 \quad (2.22)$$

A função de entropia (ganho de informação), por sua vez, é usada para quantificar a incerteza na distribuição das classes, e é dada conforme a Equação 2.23 (SUTTON, 2005):

$$h(p_1, \dots, p_j) = - \sum_{j=1}^J p_j \log(p_j) \quad (2.23)$$

De forma geral, o objetivo em cada divisão é maximizar a redução na impureza, que é a diferença na impureza do nó "pai" e a soma ponderada das impurezas dos nós "filhos". Matematicamente, para cada nó n que se divide em dois nós filhos n_{esquerda} e n_{direita} , pode-se calcular a redução na impureza devido ao split conforme a Equação 2.24. Note que $I(n)$ corresponde à impureza associada à distribuição de classes no nó pai e $I(n_{\text{filhos}}) = \frac{N_{\text{esquerda}}}{N} I(n_{\text{esquerda}}) + \frac{N_{\text{direita}}}{N} I(n_{\text{direita}})$ corresponde à soma ponderada das impurezas dos nós filhos. Ademais, a função de impureza I pode corresponder tanto ao índice Gini quanto à entropia.

$$\Delta I = I(n) - I(n_{\text{filhos}}) = I(n) - \left(\frac{N_{\text{esquerda}}}{N} I(n_{\text{esquerda}}) + \frac{N_{\text{direita}}}{N} I(n_{\text{direita}}) \right) \quad (2.24)$$

Nesse sentido, a árvore de decisão cresce recursivamente, fazendo o split de nós pai em nós filhos. Esse processo ocorre até que algum critério de parada seja satisfeito, como por exemplo número mínimo de casos em um nó ou uma profundidade máxima de árvore. Os critérios de parada podem ser ajustados e funcionam como hiperparâmetros úteis no controle da complexidade do modelo e do sobreajuste.

2.8.3 Poda da árvore de decisão

Outra medida contra o sobreajuste inclui a técnica de "*pruning*" (em português: poda) da árvore de decisão. Isso ocorre com base numa medida de custo-complexidade $R_\alpha(T)$, que é um balanço entre os erros do modelo (taxa de erro $R(T)$) e a complexidade da árvore (medida como o número total de nós na árvore: $|T|$), conforme é dado na Equação 2.25:

$$R_\alpha(T) = R(T) + \alpha|T| \quad (2.25)$$

Perceba que o parâmetro α controla a ênfase que é dada à complexidade do modelo em detrimento da taxa de erro, e seus valores são dados *a priori* pelo usuário. Para aplicar a técnica de poda, primeiro é necessário deixar a árvore de decisão crescer ao máximo (com base

no algoritmo CART). Posteriormente, a métrica de custo-complexidade é aplicada a árvore de decisão completa avaliando-se suas "sub-árvores", de modo a buscar alguma sub-árvore com o valor mínimo de custo-complexidade.

2.9 Floresta Aleatória

O modelo de Floresta Aleatória (do inglês *Random Forest*, doravante referido como RF) se baseia no princípio de *bagging* para limitar o sobreajuste das árvores de decisão (SUTTON, 2005). Múltiplas árvores de decisão são treinadas com porções aleatórias do dataset de treino, de modo que, na predição final, é considerado o voto da maioria das árvores de decisão. Esse método, no geral, é mais robusto e produz melhores resultados do que quando consideramos uma única árvore de decisão treinada sobre todo o conjunto de dados de treino (BREIMAN, 2001).

Na definição original de RF por (BREIMAN, 2001), o modelo consiste em um conjunto (*ensemble*) de K árvores de decisão $h(x, \Theta_k)$, onde Θ_k representa um vetor aleatório independente e identicamente distribuído para cada árvore. Cada árvore vota na classe mais popular dada uma entrada x .

$$h(x) = \text{Moda}\{h_1(x, \Theta_1), h_2(x, \Theta_2), \dots, h_K(x, \Theta_K)\} \quad (2.26)$$

Para que haja variedade entre as árvores, um subconjunto dos dados de treino é escolhido aleatoriamente com reposição (método conhecido como *bootstrap*). Para forçar ainda mais uma diferenciação entre as árvores, é feita uma seleção aleatória de m variáveis preditoras (correspondente ao vetor Θ_k) para que sejam utilizadas nas divisões e na construção das regras de decisão dos modelos (BREIMAN, 2001). Note que o número m é geralmente menor do que o número total de variáveis M . Dado o subconjunto de variáveis, a métrica de impureza (como o índice Gini ou a entropia) é utilizada para efetuar a divisão nos nós de cada árvore de decisão. Uma ilustração do modelo RF é mostrada na Figura 8.

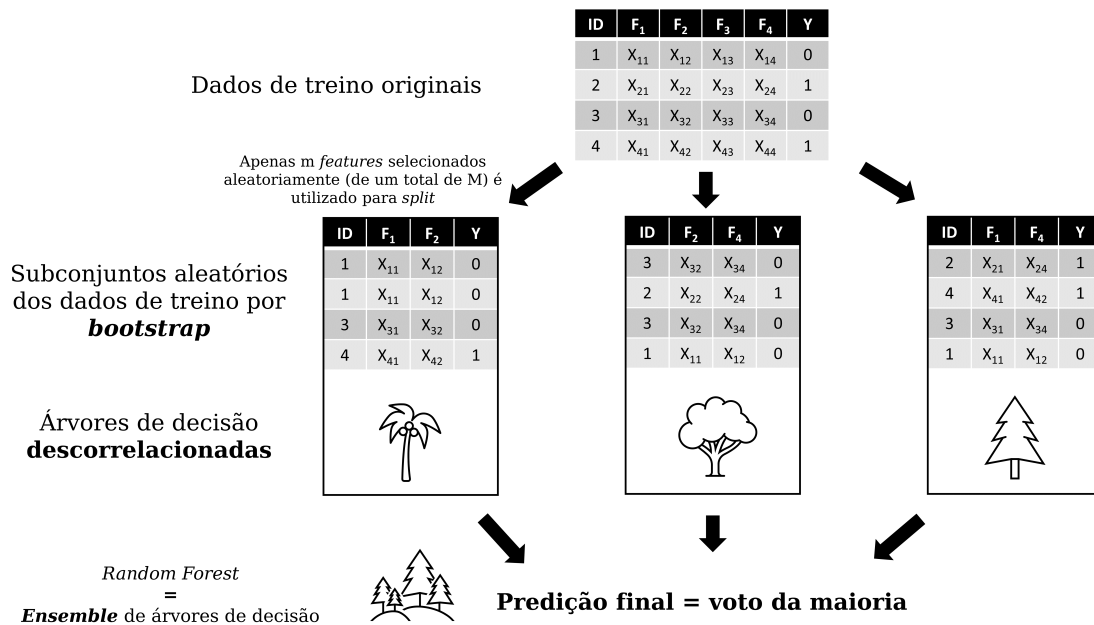


Figura 8 – Ilustração de uma floresta aleatória de árvores de decisão (*Random Forest*). Fonte: próprio autor.

É relevante citar que é relevante ter modelos distintos em um *ensemble* (isto é, modelos descorrelacionados, o que é o caso para o RF, haja vista os métodos de amostragem citados anteriormente), pois a variância da predição, que está intimamente relacionada ao sobreajuste, melhora significativamente (GHOJOGH; CROWLEY, 2019). Isso torna o RF preferível em relação a uma árvore de decisão isolada. Ademais, o *bagging* nunca é destrutivo: ou ele simplesmente não é efetivo em melhorar as predições ou as melhora significativamente, principalmente em termos de variância (GHOJOGH; CROWLEY, 2019; BREIMAN, 1996).

2.10 Árvores de *Boosting* Adaptativo

No mesmo sentido da solução proposta pelo modelo de RF, o modelo de Árvores de *Boosting* Adaptativo (do inglês, *Adaptive Boosting Trees*, doravante referido como ABT) faz uso de um *ensemble* de árvores de decisão para superar as limitações devido ao uso de uma única árvore, que envolvem sobretudo o sobreajuste. Entretanto, diferentemente de fazer uso de um *ensemble* de árvores em paralelo (como ocorre para o RF), o ABT, proposto por (FREUND; SCHAPIRE et al., 1996), treina árvores de decisão em sequência (técnica conhecida como *boosting*), onde os dados de treino de cada árvore possui um peso distinto para as instâncias de treino com base na performance da árvore de decisão anterior, dando mais ênfase a exemplos que foram incorretamente classificados. Ademais, também são atribuídos pesos para cada árvore

de decisão a depender de sua performance, fazendo com que a predição final seja uma "média ponderada" das predições de todas as árvores do *ensemble* (SUTTON, 2005).

O algoritmo começa atribuindo pesos iguais a todos as instâncias dos dados de treino (x_i, y_i) , onde $i = 1, 2, \dots, N$ e y_i é a saída verdadeira da instância x_i . Assim, inicialmente os pesos são dados por:

$$w_i^{(1)} = \frac{1}{N}, \quad \forall i = 1, 2, \dots, N \quad (2.27)$$

Como o ABT é um processo iterativo que constrói o modelo final através de uma série de iterações $t = 1, 2, \dots, T$, onde T é o número total de classificadores fracos (árvores de decisão sem muita profundidade) a serem combinados, então, primeiramente, uma única árvore de decisão $h_t(x)$ é treinada usando o conjunto de dados inicial com seus respectivos pesos. O erro ponderado ϵ_t dessa árvore de decisão é calculado como:

$$\epsilon_t = \sum_{i=1}^N w_i^{(t)} I(y_i \neq h_t(x_i)) \quad (2.28)$$

onde $I(\cdot)$ é a função indicadora que retorna 1 se $y_i \neq h_t(x_i)$ e 0 caso contrário. Tendo-se calculado o erro ponderado da árvore de decisão, é possível calcular seu peso α_t como:

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \epsilon_t}{\epsilon_t} \right) \quad (2.29)$$

Intuitivamente, dado o erro ponderado da árvore de decisão, esse peso indica a importância relativa da mesma na composição final do *ensemble*. Quanto menor o erro ϵ_t , maior será o valor de α_t . Para o treino da próxima árvore de decisão, os pesos das instâncias de treinamento são atualizadas para que os exemplos incorretamente classificados recebam mais ênfase nas iterações seguintes:

$$w_i^{(t+1)} = w_i^{(t)} \cdot \exp(\alpha_t \cdot I(y_i \neq h_t(x_i))) \quad (2.30)$$

Os pesos são então normalizados para que sua soma total seja igual a 1:

$$w_i^{(t+1)} = \frac{w_i^{(t+1)}}{\sum_{i=1}^N w_i^{(t+1)}} \quad (2.31)$$

Após T iterações, o *ensemble* final $H(x)$ é uma combinação ponderada de todas as árvores de decisão treinadas sequencialmente:

$$H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right) \quad (2.32)$$

Aqui, $\text{sign}(\cdot)$ é a função sinal que retorna 1 se o valor é positivo e -1 se negativo (para o caso de um problema de classificação binária). Na Figura 9 é ilustrado todo o processo que acabou de ser descrito.

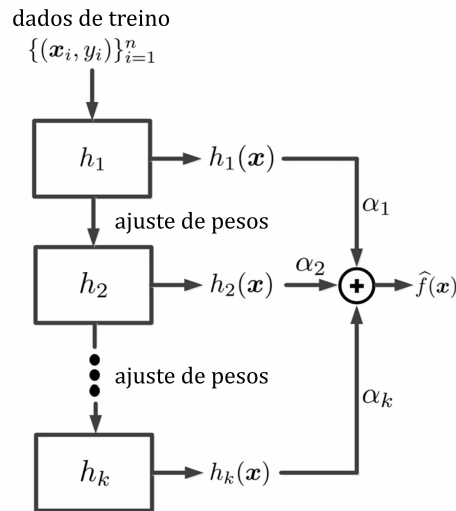


Figura 9 – Ilustração dos processos envolvidos na construção do *ensemble* ABT para k árvores de decisão. **Fonte:** adaptada de (GHOJOGH; CROWLEY, 2019).

Perceba que a combinação final da previsão das árvores de decisão é capaz de reduzir significativamente o erro de generalização em comparação com qualquer árvore de decisão isolada (SUTTON, 2005). Ademais, assim como ocorre para as técnicas de *bagging* (modelo RF incluso), as técnicas de *boosting* (como ABT) melhoram significativamente a variância das previsões (GHOJOGH; CROWLEY, 2019).

2.11 Árvores de *Gradient Boosting*

O modelo de Árvores de *Gradient Boosting* (do inglês *Gradient Boosting Trees*, doravante referido como GBT) foi primeiramente apresentado por Friedman (FRIEDMAN, 2001) e é uma técnica de aprendizado de máquina amplamente utilizada para problemas de classificação e regressão. Na mesma linha dos dois últimos modelos apresentados, este método combina a simplicidade das árvores de decisão com o poder preditivo dos métodos de *boosting*, formando um modelo robusto e eficiente. Apesar de ser parecido com o modelo ABT, o GBT faz uso de funções de erro, e as árvores de decisão são treinadas sequencialmente sobre os resíduos (erros)

das árvores anteriores, de modo que o *ensemble* final visa minimizar a função de erro total, assim como uma rede neural artificial.

O GBT constrói um modelo forte a partir de uma sequência de árvores de decisão de baixa profundidade. Cada nova árvore é construída para "corrigir os erros" do modelo anterior. O modelo inicial é definido como a média dos valores da variável dependente y . Para problemas de regressão, temos:

$$F_0(x) = \arg \min_c \sum_{i=1}^N L(y_i, c), \quad (2.33)$$

onde $L(y, F(x))$ é uma função de erro, geralmente o erro quadrático médio (MSE) para regressão ou a perda logarítmica para classificação. No caso do MSE, a inicialização é a média dos valores y .

A cada iteração m , o modelo atual $F_{m-1}(x)$ é atualizado adicionando uma nova árvore $h_m(x)$ que minimiza a função de erro. Esta nova árvore é ajustada aos resíduos, que são definidos como o gradiente negativo da função de perda em relação ao modelo atual:

$$r_{im} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}. \quad (2.34)$$

O novo modelo é dado por:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x), \quad (2.35)$$

onde γ_m é o peso (ou taxa de aprendizado) que controla a contribuição da nova árvore. Esse peso é determinado minimizando a função de perda:

$$\gamma_m = \arg \min_{\gamma} \sum_{i=1}^N L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)). \quad (2.36)$$

Note que cada árvore $h_m(x)$ é ajustada aos resíduos r_{im} . Isso significa que as árvores são construídas para prever as diferenças entre os valores de saída anteriores e os valores reais do problema. Essas árvores são tipicamente restritas em profundidade ou número de nós, para evitar sobreajuste e garantir que cada árvore seja um modelo fraco.

O processo é repetido por um número predeterminado de iterações ou até que a melhora na função de erro seja desprezível. O resultado final é um *ensemble* que combina várias árvores de decisão, cada uma "corrigindo" os erros das anteriores. A predição final após M iterações é obtida considerando todas as contribuições das árvores ao valor inicial $F_0(x)$:

$$F_M(x) = F_0(x) + \sum_{m=1}^M \gamma_m h_m(x) \quad (2.37)$$

Todo esse processo pode ser ilustrado resumidamente pela Figura 10.

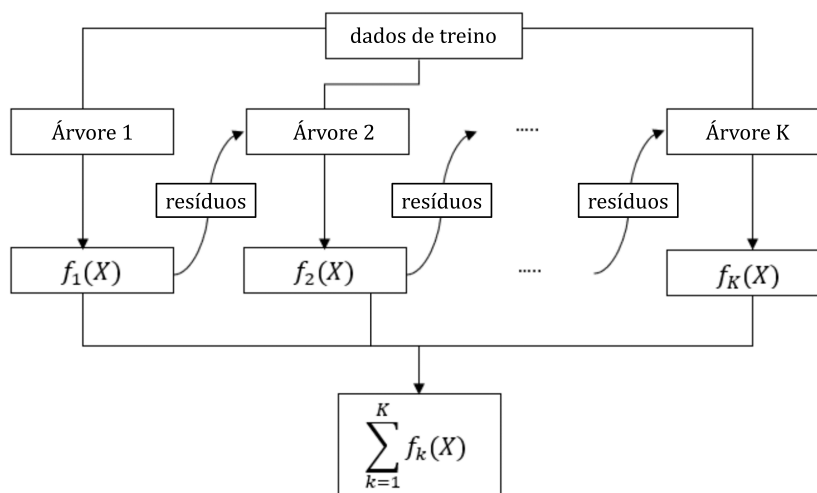


Figura 10 – Ilustração dos processos envolvidos na construção do *ensemble* GBT para K árvores de decisão. **Fonte:** adaptada de (ALADSANI; BURTON; SE, 2021).

O GBT é altamente eficaz, mas também sensível a hiperparâmetros, como a taxa de aprendizado e a profundidade das árvores. A escolha inadequada desses parâmetros pode levar a problemas de sobreajuste ou subajuste (HASTIE et al., 2009). Além disso, o tempo de treinamento pode ser elevado, especialmente em conjuntos de dados grandes. Recentemente, variações como o *XGBoost* (CHEN; GUESTRIN, 2016) e o *LightGBM* (GUOLINKE et al., 2017) têm se destacado por suas melhorias em eficiência computacional e precisão.

2.12 Desbalanceamento de classes

Em vários campos da ciência, eventos raros são difíceis de detectar devido à sua infrequência e casualidade. No entanto, o interesse na predição correta desses eventos está associado aos altos custos comumente associados à classificação incorreta ou à mera ocorrência desses eventos, para os quais é vantajoso uma postura de antecipação premeditada. Isso se aplica ao diagnóstico de doenças raras e à detecção de fraudes financeiras ou transações inválidas, por exemplo. Nesse último contexto, esses eventos podem surgir isoladamente em meio a centenas de milhares de registros de transações, mas note que a falha em identificar uma transação fraudulenta grave pode ser a causa de perdas enormes (HAIXIANG et al., 2017). Em problemas de classificação, a raridade dos eventos de interesse é definida como desbalanceamento de classes.

Matematicamente, o desbalanceamento de classes e sua intensidade podem ser caracterizados pela grandeza assimetria (*skew*), definida conforme a Equação 2.38 (JENI; COHN; TORRE, 2013). Perceba que, se a assimetria é bem menor que 1, isso significa que a classe negativa é minoria em relação ao total de classes positivas. Por outro lado, se a assimetria é muito maior que 1, a classe negativa tende a ser maioria em relação ao total de classes positivas. No caso do problema envolvendo a predição de não-comparecimento de pacientes, a assimetria dos dados, em geral, é sempre maior do que 1, pois a proporção de pacientes que comparecem tende a ser maior do que a proporção de pacientes que não comparece (CARRERAS-GARCÍA et al., 2020).

$$\text{assimetria} = \frac{\text{total de classes negativas}}{\text{total de classes positivas}} \quad (2.38)$$

Até o momento, o aprendizado de modelos de ML a partir de dados desbalanceados é um desafio em aberto e ainda há muitas questões a serem investigadas, como por exemplo as que envolvem o desenvolvimento de técnicas para lidar com o desbalanceamento de classes ou em relação ao desenvolvimento de métricas mais adaptadas para mensurar a performance dos modelos.

2.12.1 Técnicas para lidar com desbalanceamento de classes

É possível lidar com o desbalanceamento de classes já na etapa de pré-processamento dos dados, que é frequentemente realizado antes de alimentar o modelo de ML com os dados de entrada. Enquadradas nessa categoria, as técnicas de reamostragem são utilizadas para rebalancear o espaço amostral de um conjunto de dados desbalanceado, a fim de amenizar o efeito da distribuição desigual das classes no processo de treinamento dos modelos. Métodos de reamostragem são mais versáteis porque são independentes do modelo de classificação selecionado. As técnicas clássicas de reamostragem podem ser de dois tipos (HAIXIANG et al., 2017):

- Sobreamostragem (Figura 11): lidam com os efeitos do desbalanceamento de classes a partir da criação de novas amostras para a classe minoritária. Dois métodos amplamente usados para criar amostras sintéticas da classe minoritária são a duplicação aleatória das amostras minoritárias (técnica conhecida como Sobreamostragem Aleatória/*Random Oversampling*) e a Sobreamostragem Sintética da Classe Minoritária/*Synthetic Minority Over-sampling Technique* (SMOTE);
- Subamostragem (Figura 11): lidam com os efeitos do desbalanceamento de classes a partir do descarte de amostras da classe majoritária. O método mais simples e eficaz é a

Subamostragem Aleatória (*Random UnderSampling* - RUS), que envolve a eliminação aleatória de exemplos da classe majoritária;

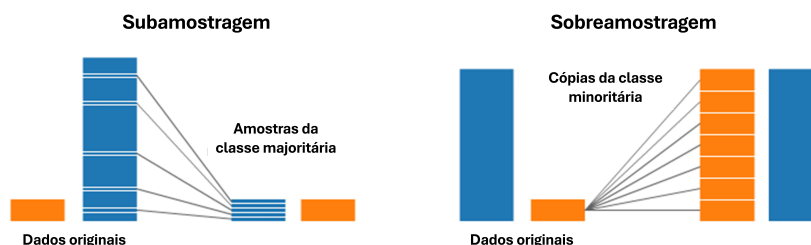


Figura 11 – Ilustração das técnicas de reamostragem clássicas: sobreamostragem e subamostragem. **Fonte:** adaptada de <https://www.kaggle.com/code/rafjaa/resampling-strategies-for-imbalanced-datasets#1> (Acessado em: 01/11/2024).

Outra forma de lidar com o desbalanceamento de classes é atribuindo pesos mais altos para a classificação incorreta de amostras da classe minoritária em relação às amostras da classe majoritária. Modelos de ML como Regressão Logística e Redes Neurais têm a implementação nativa de atribuição de diferentes pesos para as classes. Esses pesos normalmente são aplicados durante o cálculo dos valores das funções de erro e podem ser extremamente eficazes para lidar com o desbalanceamento de classes. Tratando-se de um contexto específico e bem conhecido, os pesos podem ser determinadas pela opinião de especialistas ou, em cenários de fluxo de dados, podem variar para cada registro, de acordo com a distribuição de dados. Em comparação com os métodos de reamostragem, o aprendizado sensível a custos é mais eficiente em termos computacionais, o que pode torná-lo mais adequado para grandes fluxos de dados (HAI XIANG et al., 2017). Outra vantagem é que não há o uso de dados sintéticos, os quais podem ser problemáticos em contextos como o de diagnóstico de doenças.

É importante ressaltar que os métodos de reamostragem, assim como os de atribuição de pesos distintos para as classes, já possuem implementações disponíveis em pacotes no Python, por exemplo. Mais especificamente, o pacote *imbalanced-learn* (LEMAÎTRE; NOGUEIRA; ARIDAS, 2017) une a vantagem dos algoritmos de *bagging* e *boosting* das árvores de decisão com o uso de técnicas de subamostragem para cada árvore de decisão. Isso é extremamente vantajoso do ponto de vista do uso de dados, pois não são criados dados sintéticos e, ao mesmo tempo, são utilizados todos os dados disponíveis, o que não seria possível numa aplicação simplificada da técnica de subamostragem.

2.12.2 Métricas apropriadas

As métricas de performance mais amplamente utilizadas em problemas de classificação são a acurácia total e a área sobre a Curva Característica de Operação do Receptor (do inglês *Receiver Operating Characteristic*, doravante referido como AUC-ROC), que é uma curva que mostra a taxa de verdadeiros positivos e a taxa de falsos positivos para a classe positiva considerando diferentes limiares de probabilidade para um mesmo modelo de classificação. Ambas as métricas, no entanto, são problemáticas na presença de desbalanceamento de classes. No caso da métrica acurácia, que corresponde à porcentagem de resultados corretos entre o número total de instâncias avaliadas em um conjunto de teste, um classificador que a maximiza isoladamente pode ser completamente ineficaz para dados extremamente desbalanceados. Por exemplo, no caso em que 99% de todos os dados pertencem à classe negativa e os 1% restantes pertencem à classe positiva, o classificador pode produzir uma acurácia de 99% mesmo que simplesmente esteja predizendo todos os dados como pertencentes à classe majoritária. Outro problema da métrica acurácia é que ela implicitamente assume que o classificador operará em dados provenientes da mesma distribuição dos dados de treinamento. No entanto, é claro que o desbalanceamento entre classes presente no conjunto de treinamento nem sempre é o que será encontrado no conjunto de dados de teste (FU; YI; PAN, 2019).

A AUC-ROC, por sua vez, é frequentemente considerada a técnica padrão para avaliar o desempenho de um classificador em muitos campos da ciência. Suas propriedades e índices associados tem sido amplamente estudados, caracterizados e aplicados. No entanto, a AUC-ROC tende a fornecer uma visão excessivamente otimista do desempenho de um algoritmo para dados com classes desbalanceadas, principalmente em cenários com excesso de classes negativas (alta assimetria) (FU; YI; PAN, 2019). Outro fator digno de nota é a insensibilidade dessa métrica em relação a diferentes valores de assimetria (JENI; COHN; TORRE, 2013). Por exemplo, se a proporção de instâncias positivas para negativas mudar no conjunto de dados de teste em relação ao conjunto de treino, a AUC-ROC tenderá a não mudar, e essa é uma característica indesejada em problemas com desbalanceamento de classes.

No estudo (JENI; COHN; TORRE, 2013), são aplicadas e comparadas diferentes métricas de performance para diferentes valores de assimetria dos dados. Na Figura 12 é mostrado o que foi obtido nesse estudo para o comportamento das métricas acurácia e AUC-ROC para diferentes taxas de erro do modelo e diferentes valores de assimetria para os dados. É importante ressaltar que essas taxas de erro se aplicam tanto para a classe positiva quanto para a negativa, isto é, a taxa de erro é homogênea entre as classes (JENI; COHN; TORRE, 2013). Levando em consideração que, em problemas onde há o interesse de classificação automática, geralmente o valor de assimetria está associado ao tamanho do impacto do erro para a classe positiva (ou seja,

se existe interesse em prever um evento raríssimo, é porque o impacto do mesmo é altamente relevante), o comportamento das curvas mostrado na Figura 12 é completamente indesejado, pois não há diferenciação na performance para valores diferentes de assimetria.

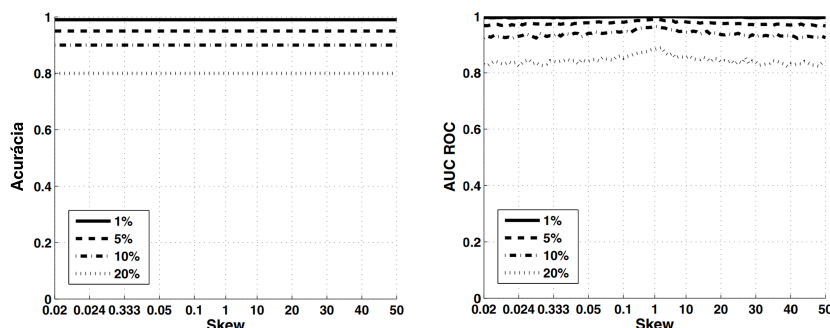


Figura 12 – Comportamento das métricas acurácia e AUC-ROC para diferentes valores de assimetria dos dados e diferentes taxas de erro do modelo. **Fonte:** adaptada de (JENI; COHN; TORRE, 2013).

As métricas F_1 -score (média harmônica entre Sensibilidade e Precisão, vide definição na Tabela 2) e área sobre a curva Precisão-Recall (do inglês *Precision-Recall*, doravante referido como AUC-PR), que é uma curva que mostra a Precisão e a Sensibilidade do modelo de classificação para diferentes limiares de probabilidade, são mais apropriadas para penalizar a má performance na classe positiva em problemas com dados de alta assimetria, que é o caso no problema de predição de não-comparecimentos. A AUC-PR, inclusive, é mais recomendada na tunagem de hiperparâmetros em problemas com desbalanceamento de classes (FU; YI; PAN, 2019), pois num problema com excesso de classes negativas, a AUC-PR leva em consideração a Precisão e a Sensibilidade na classe positiva para limiares de probabilidade variados, sendo mais apropriada para comparar a performance geral de diferentes modelos em dados com classes desbalanceadas. Na Figura 13, é mostrado como os valores do F_1 -score e da AUC-PR se comportam para diferentes taxas de erro e diferentes valores de assimetria do conjunto de dados, evidenciando o que foi discutido neste parágrafo. Note que, para essas métricas, quanto maior o valor de assimetria, maior é a penalização dos erros do modelo para as classes positivas, o que é desejável.

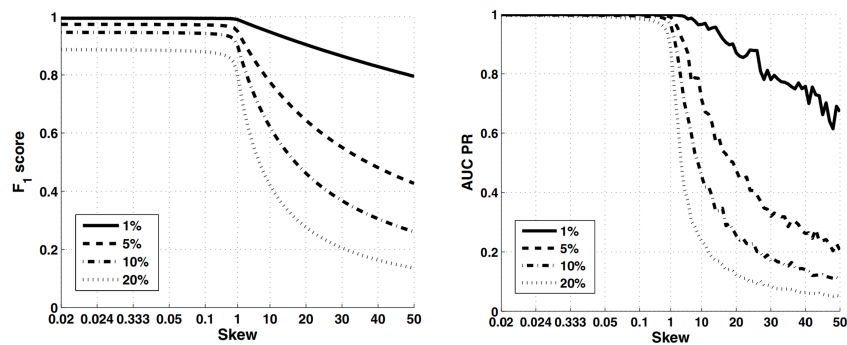


Figura 13 – Comportamento das métricas F_1 -score e AUC-PR para diferentes valores de assimetria dos dados e diferentes taxas de erro do modelo. **Fonte:** adaptada de (JENI; COHN; TORRE, 2013).

2.13 Interpretando as previsões de modelos: *framework SHAP*

A complexidade de modelos de ML, especialmente modelos de aprendizado profundo (*deep learning*) e *ensemble*, muitas vezes dificulta a interpretação dos resultados. Nesse sentido, os valores SHAP surgem como uma técnica baseada na teoria dos jogos para atribuir importância a cada variável de entrada com base em uma previsão específica do modelo (LUNDBERG, 2017). Esta abordagem permite a interpretação de modelos complexos ao calcular a contribuição marginal de cada variável, e uma propriedade importante é que há garantia de que a soma das contribuições reproduza a saída do modelo.

2.13.1 Definição

Os valores SHAP são baseados na teoria dos valores de Shapley, originalmente desenvolvida para jogos cooperativos. A ideia central é distribuir a contribuição de cada variável de forma justa, considerando todas as combinações possíveis de variáveis (SHAPLEY, 1953). Assim, dado um modelo $f(x)$, que gera uma previsão com base em um vetor de variáveis de entrada x , os valores SHAP atribuem a cada variável i um valor ϕ_i que reflete a sua importância ou contribuição para a previsão. A equação geral para o cálculo dos valores de Shapley é dada em 2.39.

$$\begin{aligned}
\phi_i &= \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)] \\
&= \frac{1}{|N|} \sum_{S \subseteq N \setminus \{i\}} \binom{|N| - 1}{|S|}^{-1} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)]
\end{aligned} \tag{2.39}$$

onde N representa o conjunto total de variáveis, S é um subconjunto de variáveis sem i , e $f_S(S)$ é a predição do modelo considerando apenas as variáveis em S . O termo $f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)$ representa a contribuição marginal da variável i em relação a S . Note que a Equação 2.39 considera todos os tamanhos de subconjuntos S e combinações possíveis de serem realizados em N , sendo ϕ_i a média ponderada da contribuição marginal da variável i .

2.13.2 Propriedades dos valores SHAP

Os valores SHAP possuem propriedades que garantem sua consistência e interpretação (LUNDBERG, 2017).

1. A propriedade de **acurácia local** (ou *local accuracy*) assegura que a soma dos valores SHAP de todas as variáveis deve ser igual à previsão do modelo para uma entrada específica x . Formalmente, para um modelo $f(x)$ e um ponto de dados x , temos:

$$f(x) = \phi_0 + \sum_{i=1}^M \phi_i, \tag{2.40}$$

onde ϕ_0 representa o valor base do modelo, ou seja, a expectativa da saída quando nenhuma informação das variáveis é conhecida, ou $E[f(z)]$. Este valor serve como uma "linha de base", e ϕ_i é o valor SHAP da variável i para a entrada x . Assim, a soma dos valores SHAP reflete a saída do modelo para a entrada específica x .

2. A propriedade de **ausência** afirma que se uma variável não contribui para a predição do modelo, seu valor SHAP deve ser zero. Em outras palavras, se $f(x)$ não depende de uma variável i em qualquer subconjunto S , então:

$$x_i = 0 \Rightarrow \phi_i = 0. \tag{2.41}$$

Essa propriedade garante que variáveis irrelevantes, ou seja, aquelas que não afetam a saída do modelo, não recebam qualquer valor de importância.

3. A propriedade de **consistência** assegura que, se a contribuição de uma variável aumenta ou permanece a mesma entre dois modelos f e f' , então o valor SHAP dessa variável também não deve diminuir. Formalmente, para dois modelos f e f' , se:

$$f'(S \cup \{i\}) - f'(S) \geq f(S \cup \{i\}) - f(S) \quad (2.42)$$

para todos os subconjuntos S que não contêm i , então os valores SHAP devem satisfazer:

$$\phi_i(f', x) \geq \phi_i(f, x). \quad (2.43)$$

Essa propriedade garante que se a inclusão de uma variável aumenta sua contribuição à previsão do modelo em qualquer contexto, o valor SHAP dessa variável não deve diminuir. Isso mantém a consistência dos valores SHAP ao longo de diferentes versões ou variações do modelo.

Essas três propriedades (acurácia local, ausência e consistência) definem uma solução única para o valor de importância de cada variável com base na teoria dos jogos. Segundo (YOUNG, 1985), o valor de Shapley é a única solução que satisfaz essas propriedades em jogos cooperativos. Portanto, os SHAP values, que se baseiam no valor de Shapley, são a única forma de atribuição de importância que garante essas propriedades ao mesmo tempo.

3 Materiais e Métodos

3.1 Conjunto de dados

3.1.1 Caracterização inicial

Os dados foram coletados diretamente do sistema de registros (*SmartRis*) da Clínica Radiológica de Anápolis (CRA), a partir das unidades do St. Central e de Jundiaí, localizadas na cidade de Anápolis, Goiás, Brasil. A distância entre as unidades é relativamente pequena (1.8 km, ou cerca de 8 minutos), mas uma diferença relevante é que a unidade do St. Central oferece serviços relacionados à área de medicina nuclear enquanto na unidade de Jundiaí há serviços relacionados a consultas, de modo que não necessariamente atendem aos mesmos grupos de pacientes.

Sobre a clínica, ela é uma instituição de destaque em diagnóstico por imagem, oferecendo uma ampla gama de serviços, incluindo Angiografia por Tomografia Computadorizada de Coronárias, Tomografia Computadorizada, Ressonância Magnética, Densitometria Óssea, Medicina Nuclear, Ultrassonografia e consultas médicas. Devido à amplitude dos serviços oferecidos, a clínica contribui para reduzir a necessidade de os moradores de Anápolis (398869 habitantes, segundo o Censo Demográfico de 2022 (([IBGE](#)), 2022)) se deslocarem até Goiânia – capital de Goiás, localizada a aproximadamente 55 km de distância – em busca de serviços avançados de diagnóstico. Moradores de cidades menores da região também se beneficiam do acesso aos serviços da clínica.

3.1.2 Coleta dos dados

O acesso ao sistema de registros para a coleta de dados foi feito através da equipe de tecnologia da informação do CRA, com a devida anuência da clínica (vide carta de anuência junto aos apêndices) e com a aprovação do Comitê de Ética em Pesquisa da Universidade Federal de Goiás (n° CAAE: 81394924.0.0000.5083, vide documento de aprovação junto aos apêndices).

Os dados coletados dos pacientes incluíram: ID no sistema de registros, data de nascimento, sexo, CEP, bairro, cidade e estado de residência, plano de saúde, modalidade e tipo de exame agendados, status do agendamento (se foi realizado com sucesso, se foi cancelado ou se continua pendente no sistema), observações referentes ao agendamento (que pode incluir informações relevantes como se o paciente compareceu ou não), data/horário em que o agen-

damento foi feito, data/horário agendados para o procedimento e data/horário de realização do procedimento. O ID original dos pacientes no sistema de registros foi remapeado para valores aleatórios de modo a dificultar o rastreo dos mesmos através dos dados. A coleta das informações supracitadas foi feita separadamente para as duas unidades do CRA, resultando em 525030 instâncias para a unidade do St. Central e 487398 instâncias para a unidade de Jundiá, com ambos os conjuntos de dados envolvendo os anos de 2014 a 2023. Uma breve apresentação dos dados pode ser encontrada na Figura 14.

patient_id	nascimento	sex	cep	city	uf	bairro	plano	us_procedime	modality	exame	data_realizacao_agendamento	procedimento_agendado_pari	final
111149488335...	1959-09-05	M	14051188	ANAPOLIS	GO	VILA FORMOSA	PARTICULAR	done	US	US DE COTOVELO	2014-08-27 16:33:27	2014-08-27 13:00:00	2017-
111149488335...	1959-09-05	M	14051188	ANAPOLIS	GO	VILA FORMOSA	PARTICULAR	cancelado	DX	RX DA ARTICULACAO ACROMIO-CLAVICULAR	2014-09-18 14:18:29	2014-09-18 12:50:00	2014-
111149488335...	1959-09-05	M	14051188	ANAPOLIS	GO	VILA FORMOSA	PARTICULAR	cancelado	MR	ARM AORTA E ILIACAS	2014-09-18 16:52:47	2014-09-18 17:18:00	NaT
111149488335...	1959-09-05	M	14051188	ANAPOLIS	GO	VILA FORMOSA	PARTICULAR	cancelado	ECG	ELETRICARDIOGRAMA - ECG - ATE 12 DERIVACES	2014-09-24 17:22:10	2014-09-24 13:00:00	NaT
111149488335...	1959-09-05	M	14051188	ANAPOLIS	GO	VILA FORMOSA	PARTICULAR	cancelado	DX	RX DE COLUNA PARA ESCOLIOSE DINAMICA	2014-09-26 18:47:28	2014-09-26 09:20:00	NaT
111149488335...	1959-09-05	M	14051188	ANAPOLIS	GO	VILA FORMOSA	PARTICULAR	cancelado	CT	TC DE ABOOME SUPERIOR	2014-09-26 18:56:05	2014-09-26 17:15:00	2014-
111149488335...	1959-09-05	M	14051188	ANAPOLIS	GO	VILA FORMOSA	PARTICULAR	cancelado	MR	RM DE COLUNA CERVICAL	2014-09-26 19:01:04	2014-09-26 18:18:00	2014-
111149488335...	1959-09-05	M	14051188	ANAPOLIS	GO	VILA FORMOSA	PARTICULAR	cancelado	CT	ATC DAS CAROTIDAS E VERTEBRAIS	2014-09-26 21:07:43	2014-09-30 11:15:00	2014-
111149488335...	1959-09-05	M	14051188	ANAPOLIS	GO	VILA FORMOSA	PARTICULAR	cancelado	CONSULTA	OBSTETRICIA	2014-09-28 20:33:14	2014-09-28 20:39:00	NaT
111149488335...	1959-09-05	M	14051188	ANAPOLIS	GO	VILA FORMOSA	PARTICULAR	cancelado	DX	RX DA ARTICULACAO ACROMIO-CLAVICULAR	2014-09-28 20:33:14	2014-09-28 20:39:00	NaT
111149488335...	1959-09-05	M	14051188	ANAPOLIS	GO	VILA FORMOSA	PARTICULAR	cancelado	CONSULTA	OBSTETRICIA	2014-09-28 20:33:14	2014-09-28 20:39:00	NaT
111149488335...	1959-09-05	M	14051188	ANAPOLIS	GO	VILA FORMOSA	PARTICULAR	cancelado	CONSULTA	OBSTETRICIA	2014-09-28 20:33:14	2014-09-28 20:39:00	2014-
111149488335...	1959-09-05	M	14051188	ANAPOLIS	GO	VILA FORMOSA	PARTICULAR	cancelado	ECG	ELETRICARDIOGRAMA - ECG - ATE 12 DERIVACES	2014-09-29 15:35:06	2014-09-29 13:15:00	NaT
111149488335...	1959-09-05	M	14051188	ANAPOLIS	GO	VILA FORMOSA	PARTICULAR	cancelado	DX	RX DA ARTICULACAO ACROMIO-CLAVICULAR	2014-09-30 00:50:55	2014-09-30 11:30:00	NaT
111149488335...	1959-09-05	M	14051188	ANAPOLIS	GO	VILA FORMOSA	PARTICULAR	done	CT	TC DE COTOVELO - UNILATERAL	2014-09-30 01:19:54	2014-09-30 10:30:00	2014-
111149488335...	1959-09-05	M	14051188	ANAPOLIS	GO	VILA FORMOSA	PARTICULAR	cancelado	ECG	ELETRICARDIOGRAMA - ECG - ATE 12 DERIVACES	2014-09-30 10:28:49	2014-09-30 08:15:00	2014-
111149488335...	1959-09-05	M	14051188	ANAPOLIS	GO	VILA FORMOSA	PARTICULAR	cancelado	CT	TC DE ABOOME SUPERIOR	2014-09-30 17:46:26	2014-09-30 17:30:00	NaT

Figura 14 – Breve apresentação do conjunto de dados utilizado. **Fonte:** próprio autor.

3.1.3 Exclusão de instâncias

3.1.3.1 Modalidades de procedimentos com poucas instâncias

Para que fossem treinados modelos para cada uma das modalidades de procedimentos, foi preciso que houvesse um total de instâncias que permitisse o treinamento dos mesmos. Como critério de exclusão para as modalidades, foi decidido que modalidades com total de instâncias menor que 500 seriam excluídas do estudo. Assim, para a unidade do St. Central, foram excluídas as modalidades "TILT", "TX", e "ANEST"(vide Figura 15); enquanto que para a unidade de Jundiá foi excluída a modalidade de "ANEST"(vide Figura 16).

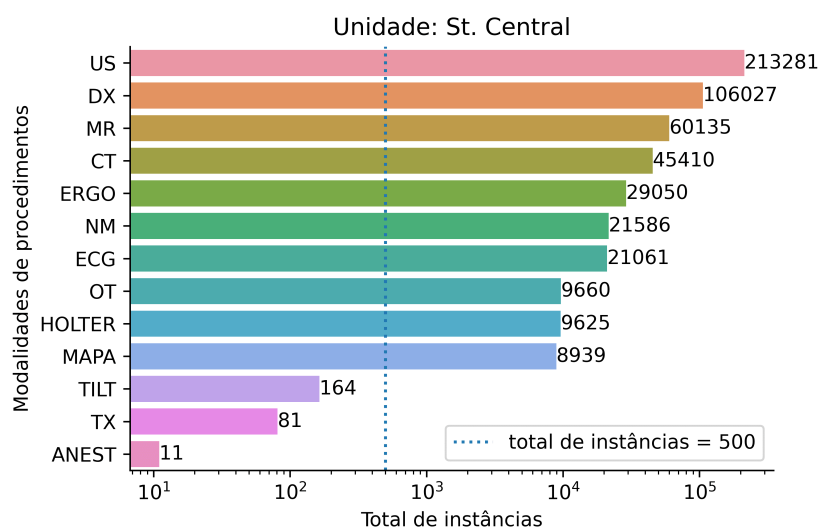


Figura 15 – Total de instâncias para cada modalidade de procedimentos na unidade do St. Central.

Fonte: próprio autor.

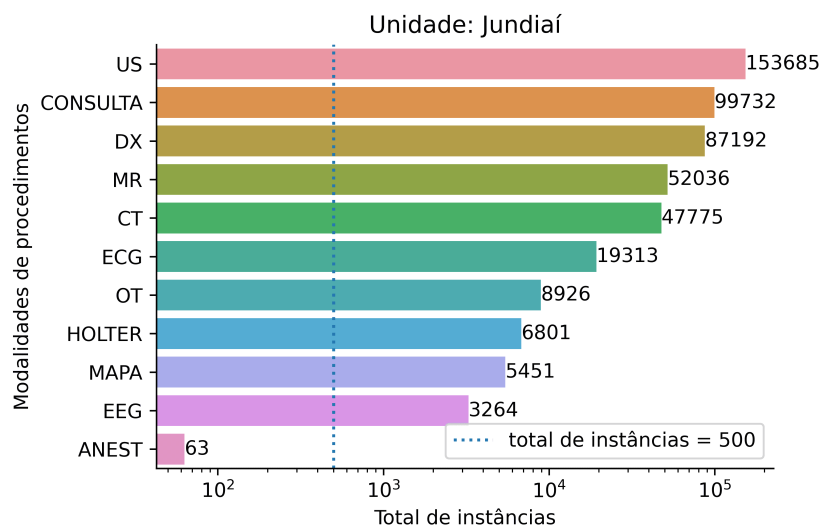


Figura 16 – Total de instâncias para cada modalidade de procedimentos na unidade de Jundiaí.

Fonte: próprio autor.

3.1.3.2 Anos com poucas instâncias

Como o sistema de registros *SmartRis* foi implantado em 2014 e há diversas instâncias de "teste" do sistema, todo o ano de 2014 foi descartado do estudo para que não houvesse nenhum viés. Note que o total de instâncias no ano de 2014 em cada umas das unidades do CRA é menor do que o ano de 2015, de modo que o descarte do ano de 2014 não leva a uma perda de dados significativa. Além disso, pelo motivo de poucas instâncias, o ano de 2024 também foi excluído

do estudo. Vide as Figura 17 e 18 para entender o total de instâncias por modalidade em cada ano disponível.

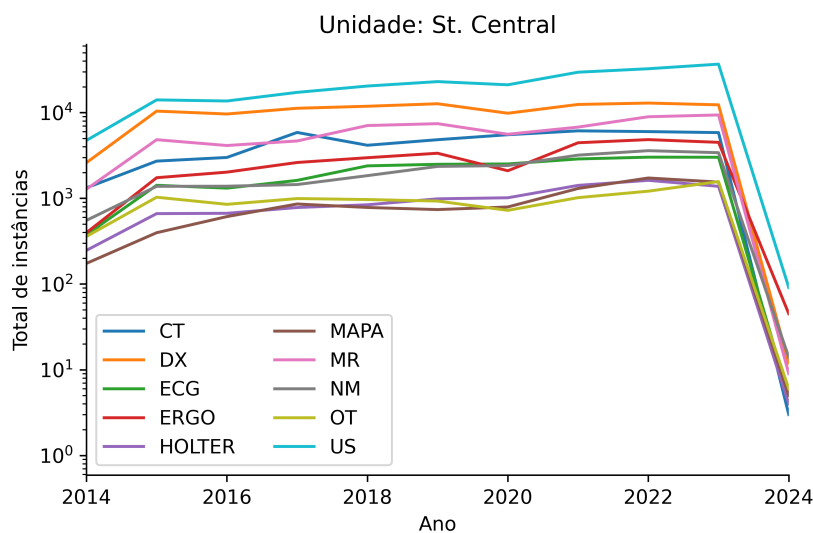


Figura 17 – Total de instâncias por ano para cada modalidade de procedimentos na unidade do St. Central. **Fonte:** próprio autor.

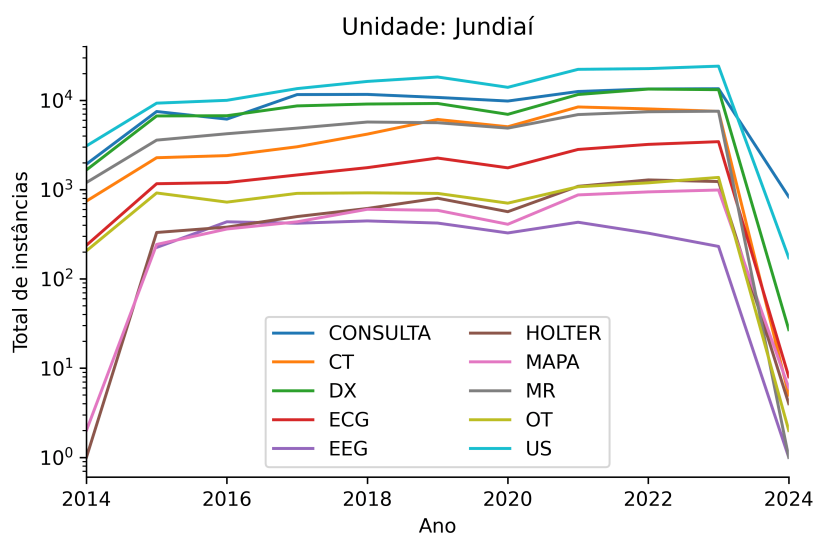


Figura 18 – Total de instâncias por ano para cada modalidade de procedimentos na unidade de Jundiá. **Fonte:** próprio autor.

3.1.3.3 Procedimentos agendados no mesmo dia

Como o propósito deste trabalho é treinar um modelo para prever o não-comparecimento de pacientes com pelo menos um dia de antecedência, não faz sentido considerar as instâncias em

que os procedimentos foram agendados no mesmo dia. Assim, essas instâncias foram excluídas. Veja pelas Figuras 19 e 20 a quantidade de procedimentos agendados no mesmo dia para cada modalidade de procedimentos e em cada unidade do CRA. Note que para algumas modalidades a quantidade de procedimentos agendados no mesmo dia é maior do que a quantidade de procedimentos agendados com um dia ou mais de antecedência. Neste caso, portanto, a redução na quantidade total de dados para cada unidade do CRA foi significativa, sendo de 39.07% para a unidade do St. Central (total de instâncias caiu de 512530 para 312278) e 33.77% para a unidade de Jundiaí (total de instâncias caiu de 474055 para 313958).

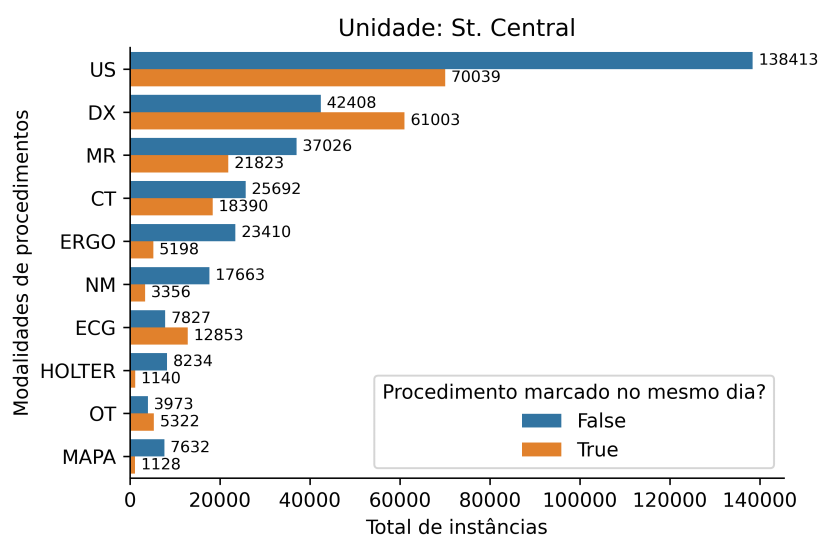


Figura 19 – Total de instâncias agendadas no mesmo dia ou não para cada modalidade de procedimentos na unidade do St. Central. **Fonte:** próprio autor.

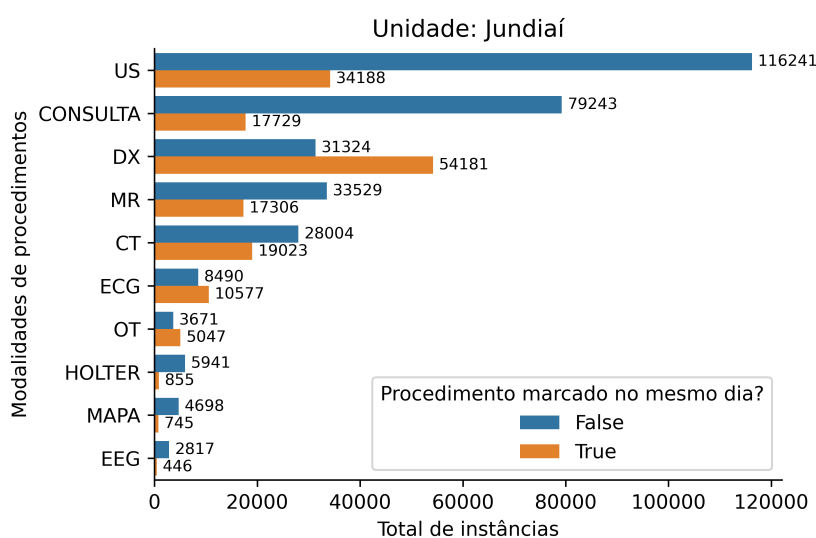


Figura 20 – Total de instâncias agendadas no mesmo dia ou não para cada modalidade de procedimentos na unidade de Jundiaí. **Fonte:** próprio autor.

3.1.3.4 Cancelamentos e função de determinação das classes

Tão fundamental quanto excluir as instâncias agendadas para o mesmo dia é excluir as instâncias que foram canceladas. Tendo em vista que o trabalho tem como objetivo prevenir a clínica de não-comparecimentos sem aviso prévio, os cancelamentos não se enquadram nesse contexto, pois são como não-comparecimentos com o devido aviso prévio.

Para identificar se houve cancelamento por parte do paciente para uma determinada instância do conjunto de dados, basta averiguar a coluna referente ao status do procedimento. No caso, instâncias canceladas aparecem com o valor "*cancelado*", de modo que a filtragem do conjunto de dados foi simples. Entretanto, é preciso levar em consideração que em algumas instâncias na coluna de status do procedimento consta como "*cancelado*", mas na coluna de observações referentes ao agendamento consta "não compareceu". É preciso, portanto, fazer a devida distinção entre essas possibilidades. As instâncias em que aparece "*confirmed*" também indicam absenteísmo por parte do paciente, visto que após o término de todas as etapas do procedimento o valor do status muda para "done", único valor que garante a informação de que o paciente compareceu.

Nas Figuras 21 e 22, são mostradas as quantidades de instâncias canceladas e não-canceladas para as diferentes modalidades de procedimentos nas unidades do St. Central e de Jundiaí para o CRA. Assim como as instâncias canceladas, a classe "indeterminada" foi excluída do conjunto de dados. Note que a redução na quantidade total de dados para cada unidade do CRA foi novamente significativa, sendo de 17.66% para a unidade do St. Central (total de

instâncias caiu de 312278 para 257129) e 20.21% para a unidade de Jundiá (total de instâncias caiu de 313958 para 250492).

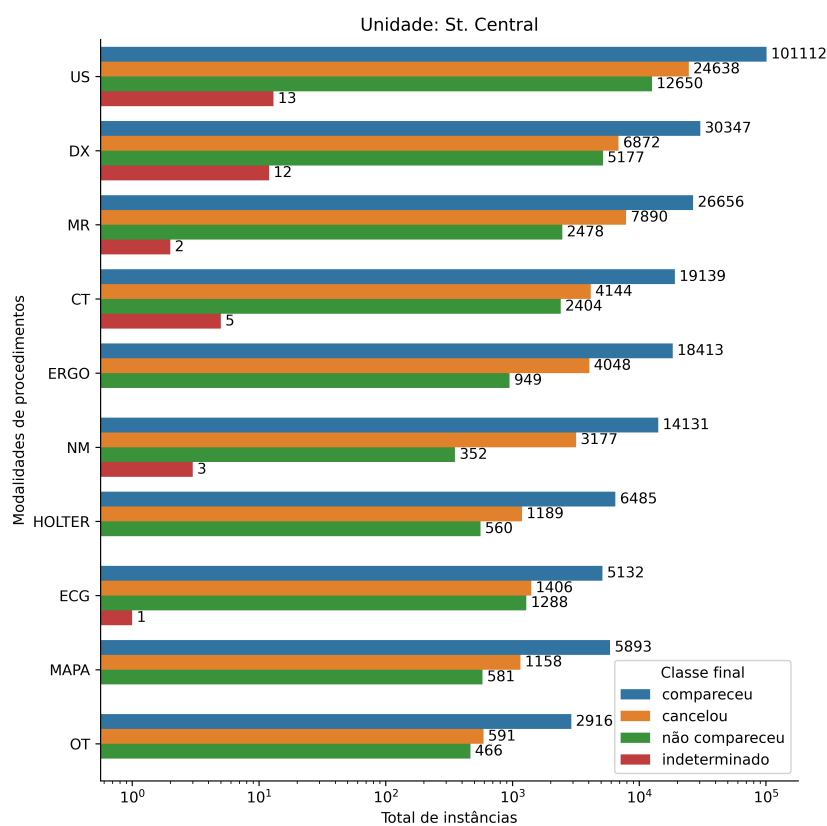


Figura 21 – Total de instâncias agendadas que foram canceladas ou não para cada modalidade de procedimentos na unidade do St. Central. **Fonte:** próprio autor.

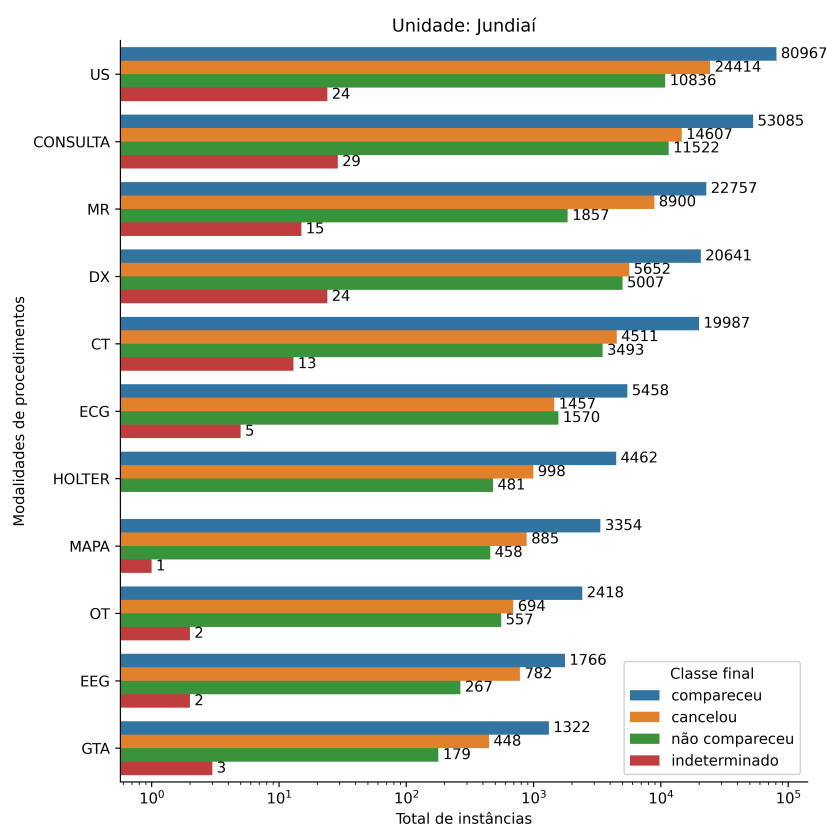


Figura 22 – Total de instâncias agendadas que foram canceladas ou não para cada modalidade de procedimentos na unidade de Jundiaí. **Fonte:** próprio autor.

3.1.4 Processamento dos dados: do sistema de registros a variáveis preditoras

Os dados foram processados de tal modo que as informações extraídas do sistema de registros fossem úteis como variáveis de entrada para o modelo de classificação. Como serão treinados modelos separadamente para cada modalidade, as etapas de processamento também foram feitas por modalidade de procedimento. Assim:

- **Idade:** foi calculada a idade dos pacientes na data agendada do procedimento através da diferença entre a data agendada do procedimento e a data de nascimento do paciente. A mediana foi utilizada para imputação de valores faltantes;
- **Sexo:** para o sexo dos pacientes (masculino, M, ou feminino, F), há valores faltantes (NA) no sistema de registros. Os valores faltantes foram imputados utilizando a moda, e em seguida houve a etapa de *One Hot Encoding* para que os dados pudessem servir como entrada numérico para os modelos de classificação;

- **Plano de saúde:** foram criadas por *One Hot Encoding* variáveis indicando o plano de saúde dos pacientes;
- **Tipo de procedimento:** foram criadas por *One Hot Encoding* variáveis indicando o tipo de procedimento agendado pelos pacientes;
- **Distância até a clínica:** foi estimada a distância até a clínica a partir do endereço cadastrado do paciente através do *Bing Maps API*. Os valores faltantes foram imputados utilizando a mediana;
- **Tempo de espera:** o tempo de espera do paciente (definido como a quantidade de dias de antecedência com a qual o agendamento foi realizado) foi calculado, em número de dias, a partir da subtração da data agendada do procedimento e a data com que o agendamento foi realizado;
- **Especificações temporais do apontamento:** usando-se a data agendada do procedimento, foram criadas variáveis indicando: o dia da semana, o mês do ano e o horário do dia. Todas essas variáveis passaram por uma etapa de *One Hot Encoding* para que pudessem servir como entrada numérica para os modelos de classificação. Adicionalmente, usando o pacote *holiday* em Python, foram criadas variáveis binárias indicando se no dia anterior ou no dia posterior à data agendada do procedimento houve algum feriado nacional brasileiro;
- **Histórico prévio de não-comparecimento:** para cada paciente e em cada instância, foi calculado o histórico prévio de não-comparecimentos na clínica como a fração de não-comparecimentos do total de procedimentos agendados (independentemente da modalidade). A mediana foi utilizada para imputação de valores faltantes. Adicionalmente, os valores faltantes foram indicados através de uma variável binária, visto que podem apresentar poder preditivo. Aliás, neste caso em específico, os valores faltantes trazem uma informação muito importante, que é a de se é a primeira vez do paciente na clínica;
- **Estatísticas baseadas nas séries temporais de taxas de não-comparecimento e de demanda em janelas de 7 e 30 dias:** para cada instância, foi calculada a taxa de não-comparecimentos e a demanda de pacientes total (quantidade de agendamentos) no dia anterior de funcionamento da clínica. Isso foi feito tanto para todas as modalidades (para indicar tendências mais gerais) quanto para a modalidade em questão. Adicionalmente, a partir desses valores e numa janela móvel, foram calculadas as estatísticas de média, desvio-padrão, 1º quartil, mediana, 3º quartil, máximo e mínimo. O tamanho utilizado das janelas para o cálculo das estatísticas foi de 7 e 30 dias anteriores. Contabilizando o total de estatísticas de acordo com as duas janelas, foram geradas 28 variáveis para a demanda e para a taxa de não-comparecimentos separadamente. Para evitar sobreajuste por excesso de

variáveis preditoras, as variáveis passaram por PCA (BRO; SMILDE, 2014), mantendo-se somente os componentes que somam, juntos, 95% da variância total. Além da redução de dimensionalidade, outra vantagem do PCA é que ele retorna variáveis descorrelacionadas (BRO; SMILDE, 2014). A sequência de etapas descrita aqui está ilustrada na Figura 23;

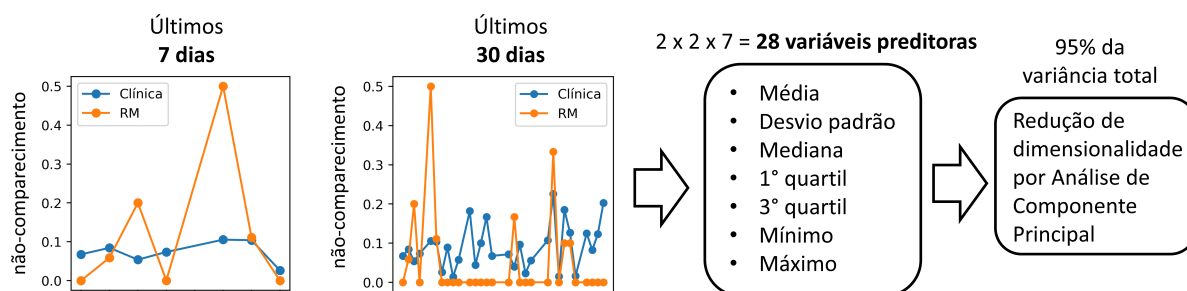


Figura 23 – Ilustração das etapas envolvidas ao gerar as variáveis referentes às estatísticas em janelas de 7 e 30 dias a partir da série temporal de taxa de não-comparecimentos. São mostradas as taxas de não-comparecimento da clínica e da modalidade de MR. **Fonte:** próprio autor.

É crucial notar que, embora indicadores de valores faltantes sejam frequentemente utilizados para agregar informações da falta de dados, no contexto das variáveis idade, sexo e distância da clínica, optou-se deliberadamente por não criá-los. Essa escolha foi motivada pela forte correlação entre a ausência de valores para essas variáveis e o não-comparecimento dos pacientes. Essa correlação se deve a um fenômeno sutil de vazamento de dados. Em particular, quando se trata da primeira consulta agendada de um paciente, a equipe de atendimento da clínica frequentemente deixa alguns detalhes de registro para serem preenchidos na chegada do paciente. Consequentemente, se o paciente não comparece a essa primeira consulta, esses campos permanecem em branco no sistema, e os próprios valores ausentes se tornam um indicativo da ausência do paciente.

Das variáveis geradas, apenas as descritas por último não têm sido usados com frequência em outros trabalhos tratando sobre a predição do não-comparecimento de pacientes (CARRERAS-GARCÍA et al., 2020). Apesar disso, essas variáveis são essenciais em problemas de classificação envolvendo séries temporais, conforme descrito e aplicado em (GUO; LU; CHEN, 2020). Já no contexto do tema deste trabalho, a utilidade de tais variáveis consiste principalmente em serem capazes de indicar tendências que dependem exclusivamente do comportamento temporal da taxa de não-comparecimentos na clínica, que não podem ser capturadas meramente pelas informações contidas no sistema de registros.

Ademais, vale ressaltar que todos as variáveis não-binárias foram padronizados subtraindo-se a média e redimensionando para uma variância igual a 1. Isso foi feito para que o modelo não

sofra vieses devido às variáveis de maior escala numérica, como comumente ocorre nos modelos que fazem uso de métodos de regularização (WUJEK; HALL; GÜNES, 2016).

3.1.5 Dados de treino, dados de teste, tunagem de hiperparâmetros e seleção dos modelos

Para que fosse possível avaliar a performance dos modelos de classificação aplicados sobre os dados de cada modalidade de procedimentos, as instâncias do ano de 2023 foram separadas para teste enquanto as entre 2015 e 2022 foram separadas para treinamento. A depender da modalidade de procedimentos, há uma maior ou menor quantidade de dados disponível para treino e teste, conforme pode-se deduzir das Figuras 17 e 18.

Com o propósito de garantir uma melhor performance dos modelos, os dados de treinamento também foram utilizados para a tunagem de hiperparâmetros. Essa etapa consiste em treinar e testar os modelos para diferentes combinações de valores para os hiperparâmetros, de modo a selecionar a melhor combinação dado um critério predefinido, como uma métrica específica de performance (YANG; SHAMI, 2020). O método de busca da melhor combinação de hiperparâmetros utilizado foi o de otimização bayesiana (YANG; SHAMI, 2020), com um total máximo de 300 iterações. Os modelos de classificação testados e seus respectivos espaços de valores dos hiperparâmetros estão mostrados na Tabela 1.

Tabela 1 – Breve descrição e espaço de valores dos hiperparâmetros para os modelos testados.

Modelo e breve descrição	Espaço de valores dos hiperparâmetros
<p>Logistic Regression (LR) (LAVALLEY, 2008): modela a probabilidade de um resultado binário usando a função logística (sigmoide) aplicada a uma combinação linear das características de entrada. Estima o logaritmo das chances (log-odds) da probabilidade de um evento. O modelo foi implementado usando o pacote <i>scikit-learn</i> em Python.</p>	$C \sim \text{log-uniform}[10^{-4}, 10^6]$ $\text{penalty} \in \{11, 12\}$ peso da classe positiva $\in [1.0, 50.0]$
<p>Multilayer Perceptron (MLP) (COSTA et al., 2023): consiste em uma camada de entrada, uma ou mais camadas ocultas com funções de ativação não-lineares e uma camada de saída. Treinado usando o algoritmo de retropropagação (<i>backpropagation</i>) para minimizar o erro de predição. O modelo foi implementado usando os pacotes <i>pytorch</i> e <i>skorch</i> em Python.</p>	camadas ocultas $\in [1, 4]$ neurônios por camada $\in [64, 256]$ taxa de aprendizado $\sim \text{log-uniform}[10^{-5}, 10^{-1}]$ decaimento de peso $\sim \text{log-uniform}[10^{-5}, 10]$ taxa de dropout $\in [0.0, 0.5]$ peso da classe positiva $\in [1.0, 50.0]$
<p>XGBoost (XGB) (CHEN; GUESTRIN, 2016): aprimora as Árvores de Decisão com Gradient Boosting tradicionais com otimizações de sistema como processamento paralelo, poda de árvores e regularização (L1 e L2). Emprega uma estratégia de crescimento de árvore por nível (level-wise) e suporta o manuseio de dados esparsos. O modelo foi implementado usando o pacote <i>xgboost</i> em Python.</p>	$n_estimators \in [50, 3000]$ $\text{max_depth} \in [3, 100]$ $\text{colsample_bytree} \in [0.25, 1.0]$ $\text{scale_pos_weight} \in [1.0, 50.0]$ taxa de aprendizado $\sim \text{log-uniform}[10^{-6}, 1.5]$ $\text{reg_alpha} \sim \text{log-uniform}[10^{-5}, 10^3]$ $\text{reg_lambda} \sim \text{log-uniform}[10^{-5}, 10^3]$
<p>LightGBM (LGBM) (KE et al., 2017): otimizado para velocidade e eficiência de memória. Utiliza uma estratégia de crescimento de árvore por folha (leaf-wise), Amostragem Unilateral Baseada em Gradiente (GOSS) e Agrupamento Exclusivo de Características (EFB) para lidar com dados em grande escala com características de alta dimensionalidade. O modelo foi implementado usando o pacote <i>lightgbm</i> em Python.</p>	$n_estimators \in [50, 3000]$ $\text{max_depth} \in [3, 100]$ $\text{colsample_bytree} \in [0.25, 1.0]$ $\text{scale_pos_weight} \in [1.0, 50.0]$ taxa de aprendizado $\sim \text{log-uniform}[10^{-6}, 1.5]$ $\text{reg_alpha} \sim \text{log-uniform}[10^{-5}, 10^3]$ $\text{reg_lambda} \sim \text{log-uniform}[10^{-5}, 10^3]$
<p>CatBoost (CB) (PROKHORENKOVA et al., 2018): projetado para lidar com características categóricas eficientemente sem pré-processamento extensivo. Implementa boosting ordenado para reduzir o sobreajuste (overfitting) e usa árvores simétricas para predição mais rápida. O modelo foi implementado usando o pacote <i>catboost</i> em Python.</p>	iterações $\in [10, 300]$ profundidade $\in [1, 16]$ taxa de aprendizado $\sim \text{log-uniform}[10^{-5}, 1]$ $\text{l2_leaf_reg} \sim \text{log-uniform}[10^{-5}, 10]$ $\text{bagging_temperature} \in [0.0, 2.0]$ $\text{rsm} \in [0.1, 1.0]$ $\text{border_count} \in [4, 64]$ $\text{random_strength} \in [0.0, 1.0]$ $\text{scale_pos_weight} \in [1.0, 50.0]$

O motivo pelo qual esses modelos em específico foram testados é que, no caso da Regressão Logística e MLP, eles têm sido empregados com frequência no contexto de predição de não-comparecimentos (CARRERAS-GARCÍA et al., 2020). No entanto, neste trabalho, foi decidido incluir modelos avançados de *boosting* de gradiente, que apresentam histórico de vencer campeonatos de ciência de dados e que são capazes de lidar nativamente com variáveis categóricas. Esta última característica talvez seja a mais importante, pois de fato o conjunto de dados deste trabalho apresenta variáveis categóricas de alta cardinalidade, o que pode ser uma

dificuldade para os algoritmos de otimização da Regressão Logística ou MLP.

O critério de seleção da melhor combinação de valores para os hiperparâmetros foi a AUC-PR, visto que essa métrica é mais apropriada para a avaliação da performance de modelos em problemas com desbalanceamento de classes acentuado (SAITO; REHMSMEIER, 2015; OZENNE; SUBTIL; MAUCORT-BOULCH, 2015) e tem sido recomendada nesse mesmo contexto para a tunagem de hiperparâmetros (FU; YI; PAN, 2019). Outra vantagem dessa métrica é que ela leva em consideração diferentes limiares de probabilidade.

Ademais, para uma validação mais realista e um uso maximizado dos dados de treino, foi utilizado o método de Validação Cruzada com Divisão Temporal (do inglês *Time Split Cross Validation*) para a obtenção dos resultados. Nesse sentido, os dados foram ordenados temporalmente e em seguida divididos em 7 *folds* de teste, de modo que apenas os dados do passado fossem utilizados para a predição (SHERIDAN, 2013). É mostrado na Figura 24 uma ilustração da divisão de *folds* decorrente da Validação Cruzada com Divisão Temporal para a tunagem de hiperparâmetros. Como para cada *fold* de teste existe uma quantidade diferente de dados disponíveis para treino, uma média simples para a avaliação da melhor combinação de valores para os hiperparâmetros não é apropriada. Nesse sentido, optou-se por fazer uso da média ponderada, onde o peso de cada resultado corresponde à quantidade de *folds* utilizados para treino. A vantagem dessa estratégia é que o resultado final não será tão penalizado pela indisponibilidade de dados para os primeiros *folds* de teste. Os pesos também estão indicados na Figura 24.

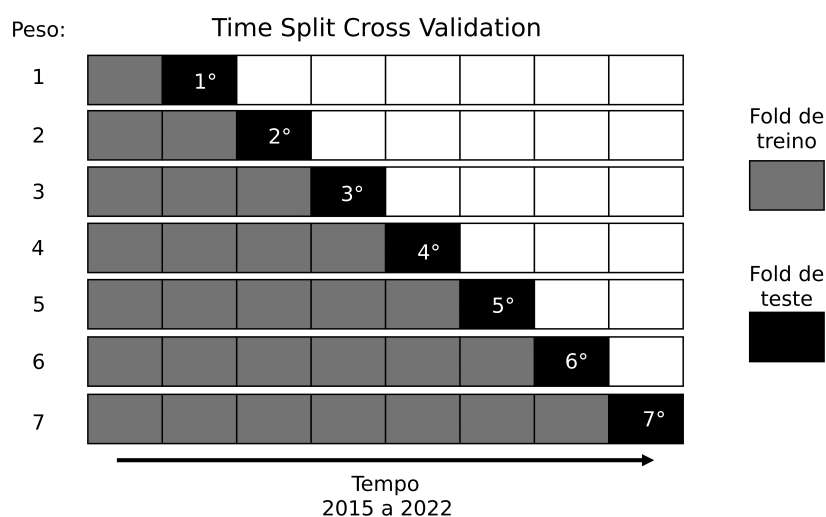


Figura 24 – Ilustração do uso de Validação Cruzada com Divisão Temporal para a tunagem de hiperparâmetros. Note que, para um *fold* de teste, apenas os *folds* do passado são usados para treinamento. Ademais, cada *fold* de teste apresenta um peso correspondente à quantidade de *folds* usados para treino no cálculo da média ponderada do AUC-PR. **Fonte:** próprio autor.

Tabela 2 – Métricas de performance utilizadas neste trabalho. VP : Verdadeiros Positivos, VN : Verdadeiros Negativos, FP : Falsos Positivos, FN : Falsos Negativos, $N = VP + VN + FP + FN$. Neste trabalho, $P_I = 0.333$ e $P_{II} = 1$.

Métrica	Definição matemática
Acurácia total (A)	$A = \frac{VN+VP}{N}$
Sensibilidade (S)	$S = \frac{VP}{VP+FN}$
Precisão (P)	$P = \frac{VP}{VP+FP}$
F1-score (F_1)	$F_1 = 2\frac{P \times S}{P+S}$
Custo (C)	$C = P_I \times \frac{FP}{N} + P_{II} \times \frac{FN}{N}$

Por fim, tendo-se determinado a melhor combinação de valores para cada modelo, o modelo (incluindo sua respectiva combinação de hiperparâmetros) que apresentou a maior média ponderada de AUC-PR foi selecionado, sendo treinado sobre todo o conjunto de dados de treino para posterior aplicação sobre os dados de teste. Note que esse processo ocorre separadamente para cada modalidade de procedimentos, de modo que cada modalidade apresenta um respectivo "melhor modelo".

3.1.6 Métricas de performance e limiar de probabilidade

As métricas de performance usadas neste trabalho incluem as comumente utilizadas em problemas de classificação, isto é, acurácia total, sensibilidade, precisão, F1-score e AUC-ROC, assim como o AUC-PR e a métrica Custo. Todas essas métricas estão definidas na Tabela 2.

No caso, a métrica Custo é uma métrica personalizada, ajustada ao impacto dos erros do modelo. Assim, quanto menor o Custo, menor o impacto dos erros do modelo e, portanto, melhor a performance. Métricas parecidas foram propostas em (BERG et al., 2013; DUNSTAN et al., 2023; HUANG; HANAUER, 2014). A decisão do impacto dos erros do tipo I (Falso Positivo, FP) e tipo II (Falso Negativo, FN) deve ser feita junto aos gestores responsáveis, que têm o conhecimento a respeito da rotina dos procedimentos da modalidade em questão e, portanto, sabem avaliar o significado de um FN ou de um FP do modelo na prática. Nesse sentido, após discutir com os gestores da Clínica Radiológica de Anápolis, decidiu-se que o peso para o FP é 3 vezes menor que o peso para o FN , ou seja, $P_I = 0.333$ e $P_{II} = 1$. A justificativa para essa decisão foi que, para o FN , o paciente que não comparece é predito como se fosse comparecer. Assim, há a perda de uma série de recursos que foram ajustados para uma demanda que, na

verdade, não existe. Por outro lado, no caso do *FP* (quando o paciente é predito como se não fosse comparecer, mas comparece), existe a demanda, mas os recursos ainda não foram ajustados. Apesar de isso ainda ser indesejável por introduzir imprevistos na rotina, continua preferível comparado ao desperdício de recursos.

Tendo em vista que o Custo reflete mais diretamente o impacto na rotina e que as taxas de *FN* e *FP* variam de acordo com o limiar de probabilidade utilizado, foi feito o ajuste do mesmo. Ele foi feito analogamente à etapa de tunagem de hiperparâmetros, sobre os mesmos *folds* de treino e teste decorrentes do uso de Validação Cruzada com Divisão Temporal. Já que o ajuste de limiar ocorre posteriormente à tunagem de hiperparâmetros, a melhor combinação de valores para os hiperparâmetros foi utilizada. Variou-se de 0.01 a 0.99, em intervalos de 0.01, o limiar de probabilidade, armazenando-se a média ponderada da métrica Custo para os *folds* de teste. O valor de limiar que resultou no menor valor para o Custo foi tomado para o classificador em questão.

3.1.7 Interpretando as predições do modelo: *framework SHAP*

Com o propósito de conferir interpretabilidade para as predições dos modelos de classificação, foi utilizado o *framework SHAP* (LUNDBERG, 2017). Ele é baseado nos *Shapley Values*, proveniente da teoria dos jogos, e estima valores que se relacionam com a contribuição de cada variável para uma dada predição do modelo. Devido ao fato de a estimativa exata da contribuição das variáveis ser desafiadora, existem métodos de aproximação que são específicos para determinados modelos. Neste estudo, os métodos mais apropriados aos modelos de classificação selecionados foram utilizados.

3.1.8 Sumário final

Sumarizando todas as etapas que foram descritas até aqui para a metodologia do trabalho, tem-se:

1. Para uma determinada modalidade de procedimentos, as instâncias são separadas em conjunto de treino (anos de 2015 a 2022) e em conjunto de teste (ano de 2023);
2. Usando Validação Cruzada com Divisão Temporal, os modelos LR, MLP, XGB, CB e LGBM são testados para diferentes combinações de valores para seus respectivos hiperparâmetros. O modelo (juntamente com sua respectiva combinação de hiperparâmetros) que apresentar maior média ponderada para a métrica AUC-PR é selecionado;

3. O limiar de probabilidade é otimizado utilizando os mesmos folds provenientes da Validação Cruzada com Divisão Temporal de modo a minimizar a métrica personalizada de Custo;
4. Tendo ajustado o limiar de probabilidade, o modelo é treinado sobre todo o conjunto de dados de treino;
5. O modelo treinado é aplicado sobre os dados de teste, a partir dos quais são calculadas as métricas de performance e estimadas a contribuição das variáveis através do *framework SHAP*;

O sumário descrito acima está ilustrado na Figura 25.

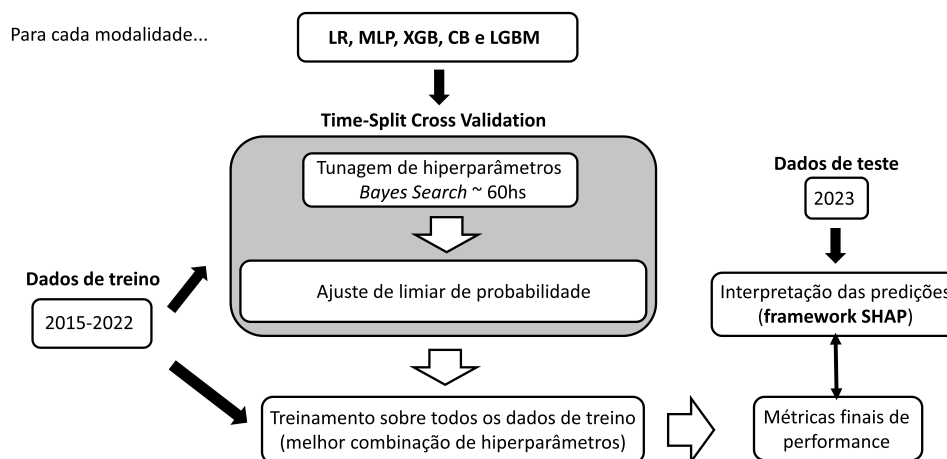


Figura 25 – Sumário da metodologia geral deste trabalho. **Fonte:** próprio autor.

4 Resultados e Discussão

4.1 Introdução

A seguir, serão mostrados os resultados pertinentes ao projeto, envolvendo as modalidades de MR e CT para a unidade do St. Central e as modalidades de consulta e Ultrassom (US) para a unidade de Jundiaí. A justificativa para a escolha dessas modalidades é que as máquinas de MR e CT são muito caras, de modo que a urgência em otimizar o uso das mesmas é maior. Para enriquecer a discussão e ter modalidades distintas, selecionamos a modalidade de consultas e de US para a unidade de Jundiaí. Note que essas modalidades são relevantes devido ao fato de apresentarem as maiores quantidades de instâncias, conforme pode ser visto pela Figura 22. Decidiu-se por incluir somente essas modalidades para evitar uma sobrecarga de informações.

De forma geral, os resultados envolvem a análise exploratória, onde foi feita a análise da relação entre as diferentes variáveis extraídas do sistema de registros e a proporção de não-comparecimentos de pacientes, e os resultados de classificação e de análise de importância de variáveis provenientes dos modelos de ML treinados e selecionados. Note que, para os resultados de análise exploratória, foram empregados testes Z de proporção (AKINKUNMI, 2019) para avaliar a significância da diferença entre a proporção estimada para grupos específicos da variável em questão e a proporção geral (considerando uma mesma modalidade de procedimentos).

Chama-se a atenção, no entanto, para os resultados de "perfil superficial" dos pacientes, que indicam uma maior proporção de não-comparecimentos quando há valores faltantes para determinadas variáveis. Conforme explicado na seção de Materiais e Métodos, para o treinamento dos modelos de predição não foram utilizadas variáveis indicadoras de valores faltantes, mesmo com os resultados de análise exploratória sugerindo alto poder preditivo das mesmas. A razão para isso é que a alta correlação entre valores faltantes e a proporção de não-comparecimentos se deve a um fenômeno sutil de vazamento de dados. Em particular, quando se trata da primeira consulta agendada de um paciente, a equipe de atendimento da clínica frequentemente deixa alguns detalhes de registro para serem preenchidos na chegada do paciente. Conseqüentemente, se o paciente não comparece a essa primeira consulta, esses campos permanecem em branco no sistema, e os próprios valores ausentes se tornam um indicativo da ausência do paciente. Portanto, antes de servir como dados de entrada para os modelos, foi feita a imputação dos valores faltantes para garantir que os modelos aprendam padrões generalizáveis, e não padrões de artefatos no sistema de registros.

Para maior organização, os resultados serão apresentados separadamente por modalidade

de procedimentos.

4.2 Modalidade de Ressonância Magnética (unidade do St. Central)

Para a modalidade de MR na unidade do St. Central do CRA, houve um total de 28991 instâncias devidamente pré-processadas, datando entre 2015 e 2023, dentre as quais 2337 houve não-comparecimento de pacientes, o que resulta em uma proporção total de não-comparecimentos próxima de 8.1%. Logo abaixo, é feita uma análise de como essa proporção varia de acordo com diferentes agrupamentos das variáveis extraídas do sistema de registros.

4.2.1 Análise exploratória

4.2.1.1 Distribuição temporal da proporção de não-comparecimentos

Nas Figuras 26 e 27 são mostrados, respectivamente, o comportamento da proporção de não-comparecimentos e o total de instâncias considerando os diferentes anos e meses do ano. Mais especificamente, na Figura 26 vemos que existe uma proporção de não-comparecimentos menor do que a proporção geral para os anos de 2015, 2016 e 2019. Em relação aos anos de 2015 e 2016, as menores proporções de não-comparecimento estão concentradas nos primeiros meses do ano, o que pode ser indicativo de que houve uma abordagem mais robusta em relação aos agendamentos na primeira metade desses anos. Por outro lado, houve maior proporção de não-comparecimentos nos anos de 2018, 2019 e 2022. Em relação ao ano de 2022, vemos que as altas proporções de não-comparecimento estão bem espalhadas ao longo dos meses do ano. Um fator a ser levado em consideração é que o ano de 2022 representa o retorno (mesmo que gradual e "controlado") das pessoas às suas rotinas normais após a pandemia da *covid-19* (MOURA et al., 2022). Esse retorno veio somado a uma maior busca a procedimentos de MR - houve um aumento no total de agendamentos 43.13% entre os anos de 2021 e 2022 (vide Figura 27). Interessantemente, levando em consideração que o maior e mais perigoso pico da pandemia da *covid-19* ocorreu entre Março e Abril de 2021 (MOURA et al., 2022), nesse mesmo ano houve maior proporção de não-comparecimentos nos meses de Junho e Agosto, podendo ter relação com a maior insegurança que os pacientes sentiram. Posteriormente, houve uma diminuição na proporção de não-comparecimentos nos meses de Setembro e Novembro. Analisando, agora, somente os meses, tem-se que os meses do ano com menores proporções de não-comparecimento foram Fevereiro, Abril e Dezembro. O único mês que apresentou uma maior proporção de

não-comparecimento foi Junho, e isso pode ter relação com o fato de Junho/Julho ser o período de férias escolares.

Levando em consideração, agora, somente a Figura 27, é possível perceber que, ao longo dos anos, há uma tendência de aumento no total de procedimentos agendados para a modalidade de MR. Isso pode estar relacionado a uma maior oferta da clínica através de, por exemplo, mais máquinas de MR, ou, simplesmente, a uma maior busca por exames nessa modalidade. De forma geral, não parece existir uma tendência importante envolvendo o total de agendamentos de exames de MR entre os meses do ano. Essas quantidades parecem estar bem distribuídas entre os meses, com variações apenas entre os anos. Um fato curioso é que, no mês de Agosto do ano de 2020, houve apenas 10 procedimentos agendados. Isso pode ter relação com a indisponibilidade da máquina, por exemplo.

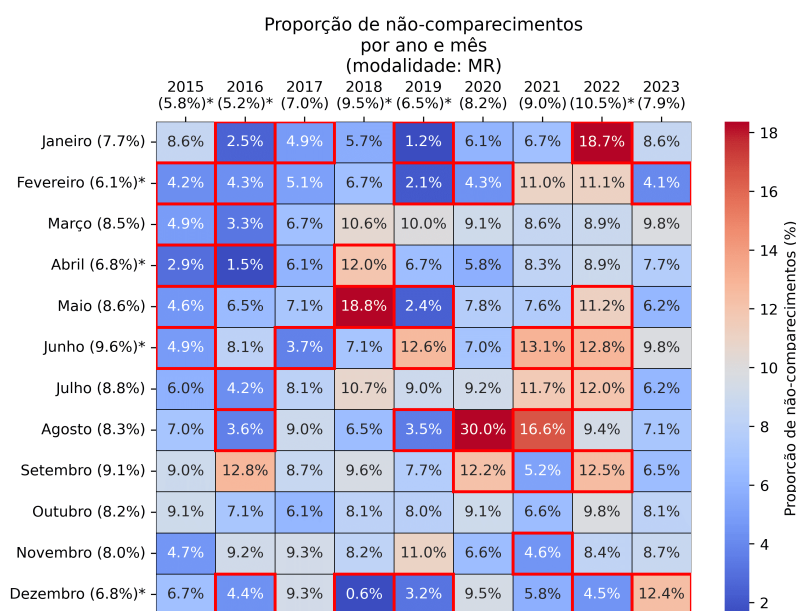


Figura 26 – Distribuição da proporção de não-comparecimentos por ano e mês para a modalidade de MR da unidade do St. Central. São também mostradas as proporções totais de não-comparecimentos separadamente para cada mês e cada ano (distribuições marginais). O "*" indica uma diferença significativa ($p\text{-valor} < 0.05$) em relação à proporção geral de não-comparecimentos, bem como o destaque em vermelho em torno das células mês-ano. **Fonte:** próprio autor.

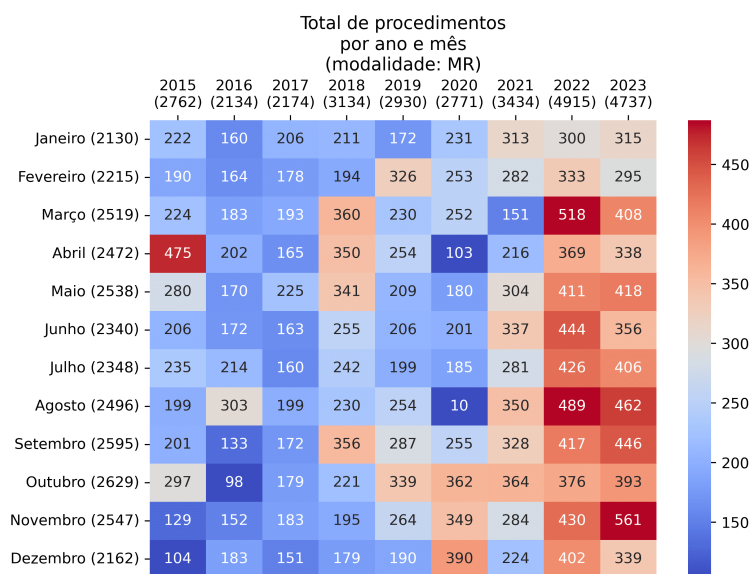


Figura 27 – Distribuição do total de procedimentos por ano e mês para a modalidade de MR da unidade do St. Central. São também mostrados o total de procedimentos separadamente para cada mês e cada ano (distribuições marginais). **Fonte:** próprio autor.

Analisando a proporção de não-comparecimentos e de total de procedimentos agendados para os dias da semana e horários do dia, tem-se as Figuras 28 e 29. Para a modalidade de MR, nota-se da Figura 29 que os horários preferidos para o agendamento de exames é das 7:00hs até as 17:00hs. A partir das 18:00hs, o volume de agendamentos cai de forma relevante. Para o dia de sábado, os agendamentos se concentram entre os horários de 7:00hs e 10:00hs. Analisando-se, agora, a Figura 28, vemos que nos dias úteis há maiores proporções de não-comparecimento somente para horários a partir das 18:00hs: para terça-feira, são os horários 20:00hs e 21:00hs; para quarta-feira, são os horários das 19:00hs até 21:00hs; na quinta-feira, são os horários das 19:00hs e 20:00hs; na sexta-feira, são os horários das 18:00hs até as 22:00hs. Esses resultados fazem sentido, visto que da 29 percebe-se que há menor busca no agendamento de procedimentos nesses horários por parte dos pacientes. Interessantemente, no dia de quarta-feira há menores proporções de não-comparecimento nos horários de 8:00hs e a partir das 10:00hs até 13:00hs. Na quinta-feira, há menores proporções de não-comparecimento nos horários de 9:00hs e 10:00hs; enquanto que, na sexta-feira, isso ocorre somente no horário de 11:00hs.

De forma isolada, nenhum dia da semana se destacou por possuir menores ou maiores proporções de comparecimento estatisticamente significativas. Já em termos dos horários do dia, os horários de 9:00hs, 10:00hs, 11:00hs e 13:00hs apresentaram menores proporções de não-comparecimento; enquanto que, nos horários a partir das 18:00hs, todos apresentaram maiores proporções de não-comparecimento.

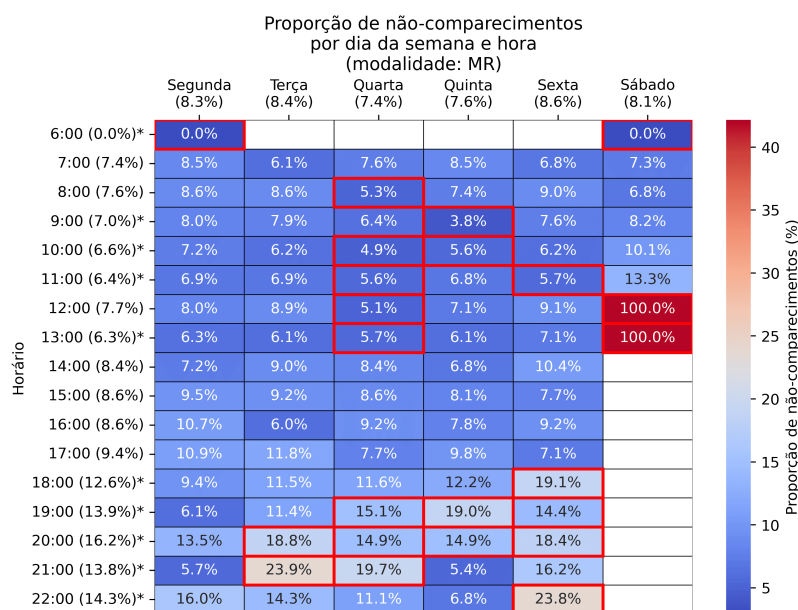


Figura 28 – Distribuição da proporção de não-comparecimentos por dia da semana e horário para a modalidade de MR da unidade do St. Central. São também mostradas as proporções totais de não-comparecimentos separadamente para cada dia da semana e cada horário (distribuições marginais). O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos, bem como o destaque em vermelho em torno das células dia-horário. **Fonte:** próprio autor.

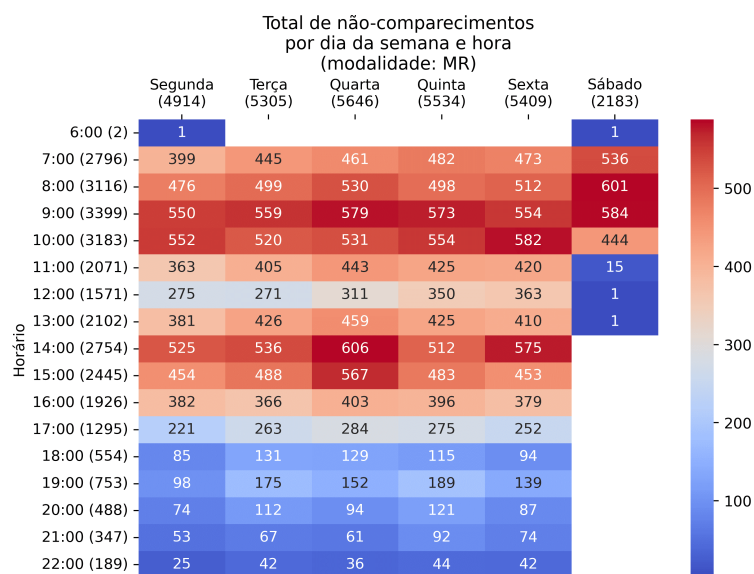


Figura 29 – Distribuição do total de procedimentos por dia da semana e horário para a modalidade de MR da unidade do St. Central. São também mostrados o total de procedimentos separadamente para cada dia da semana e cada horário (distribuições marginais). **Fonte:** próprio autor.

4.2.1.2 Características dos pacientes e proporção de não-comparecimentos

Apesar de as Figuras 30, 31, 32, 33, 34, 35 e 36 representarem efeitos isolados (isto é, sem levar em conta interações) de diferentes características dos pacientes sobre a proporção de não-comparecimento na modalidade de MR, é possível esboçar, mesmo que de forma inicial, o "perfil" do paciente que tende a não comparecer.

Analisando as figuras desta subseção, tem-se que o "perfil" superficial do paciente que tende a não comparecer a procedimentos agendados em MR é:

- Não apresenta a informação "sexo" preenchida no sistema de registros ou é do sexo feminino (note que a proporção de não-comparecimentos para sexo = "F" continua sendo menor que a proporção geral, mas é maior numericamente que a proporção para sexo = "M") - Figura 30;
- Não apresenta a informação "idade" preenchida no sistema de registros ou não apresenta idade entre 40 e 90 anos (nesse range de idade há menores proporções de não-comparecimento) - 31;
- Não apresenta a informação de "endereço" no sistema de registros - Figura 32;
- Agendou o procedimento pelo menos 6 dias de antecedência - Figura 33;
- Ou agendou pela primeira vez algum procedimento no CRA, ou apresenta histórico prévio de não-comparecimento de no mínimo 25% - Figura 34;
- Agendaram o procedimento via SUS ou via PARTICULAR (levando-se em consideração somente os planos de saúde mais comuns no conjunto de dados) - Figuras 35 e 36;

Todas essas características, além de serem úteis para a construção superficial do "perfil" do paciente que não comparece a procedimentos de MR, também servem como referência ao analisar a importância das variáveis via *framework SHAP*, que será feita mais adiante.

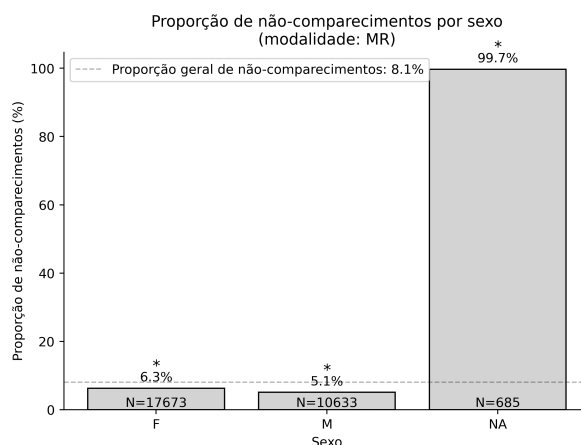


Figura 30 – Distribuição da proporção de não-comparecimentos por sexo dos pacientes para a modalidade de MR da unidade do St. Central. São também mostrados o total de procedimentos separadamente para sexo. O "*" indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos. **Fonte:** próprio autor.

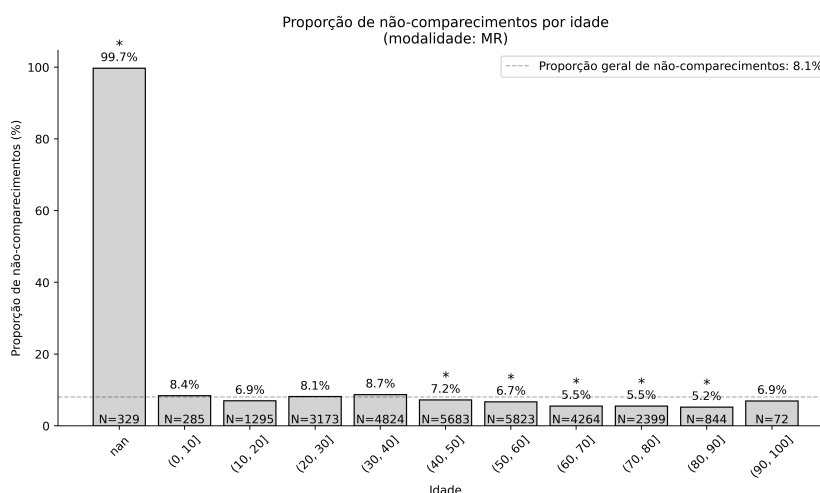


Figura 31 – Distribuição da proporção de não-comparecimentos por idade dos pacientes para a modalidade de MR da unidade do St. Central. São também mostrados o total de procedimentos separadamente para cada bin de idade. O "*" indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos. **Fonte:** próprio autor.

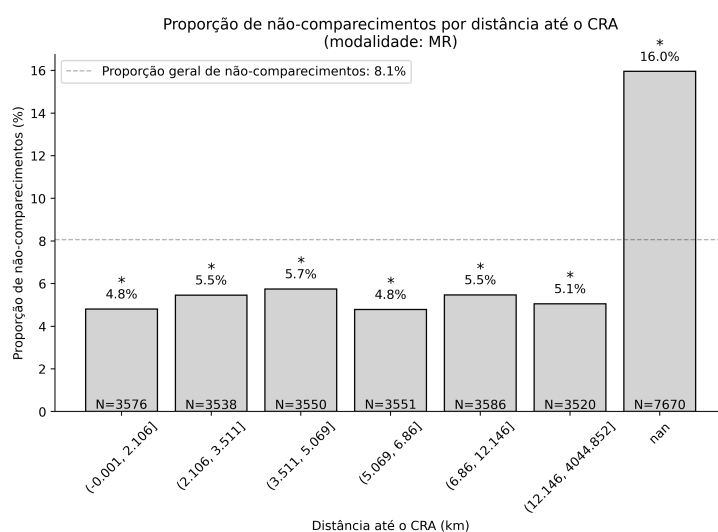


Figura 32 – Distribuição da proporção de não-comparecimentos por distância até o CRA a partir do endereço cadastrado dos pacientes para a modalidade de MR da unidade do St. Central. São também mostrados o total de procedimentos separadamente para cada bin de distância. O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos. **Fonte:** próprio autor.

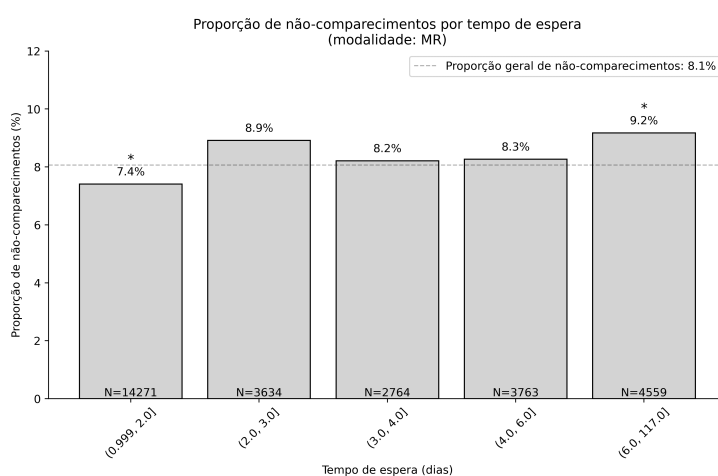


Figura 33 – Distribuição da proporção de não-comparecimentos por tempo de espera dos pacientes para a modalidade de MR da unidade do St. Central. São também mostrados o total de procedimentos separadamente para cada bin de tempo de espera. O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos. **Fonte:** próprio autor.

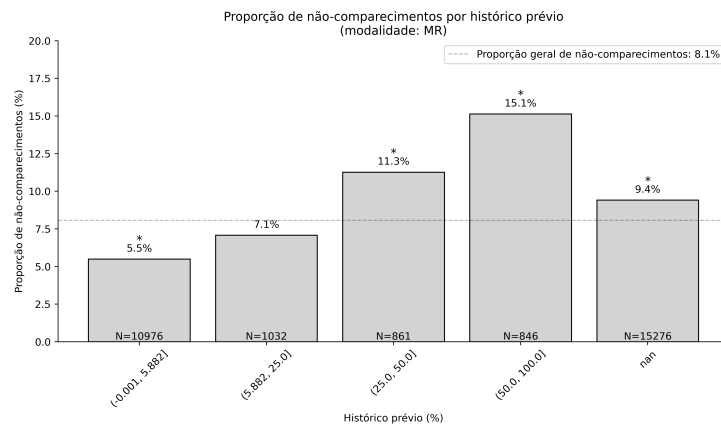


Figura 34 – Distribuição da proporção de não-comparecimentos por histórico prévio dos pacientes para a modalidade de MR da unidade do St. Central. São também mostrados o total de procedimentos separadamente para cada *bin* de histórico prévio. O "*" indica uma diferença significativa ($p\text{-valor} < 0.05$) em relação à proporção geral de não-comparecimentos. **Fonte:** próprio autor.

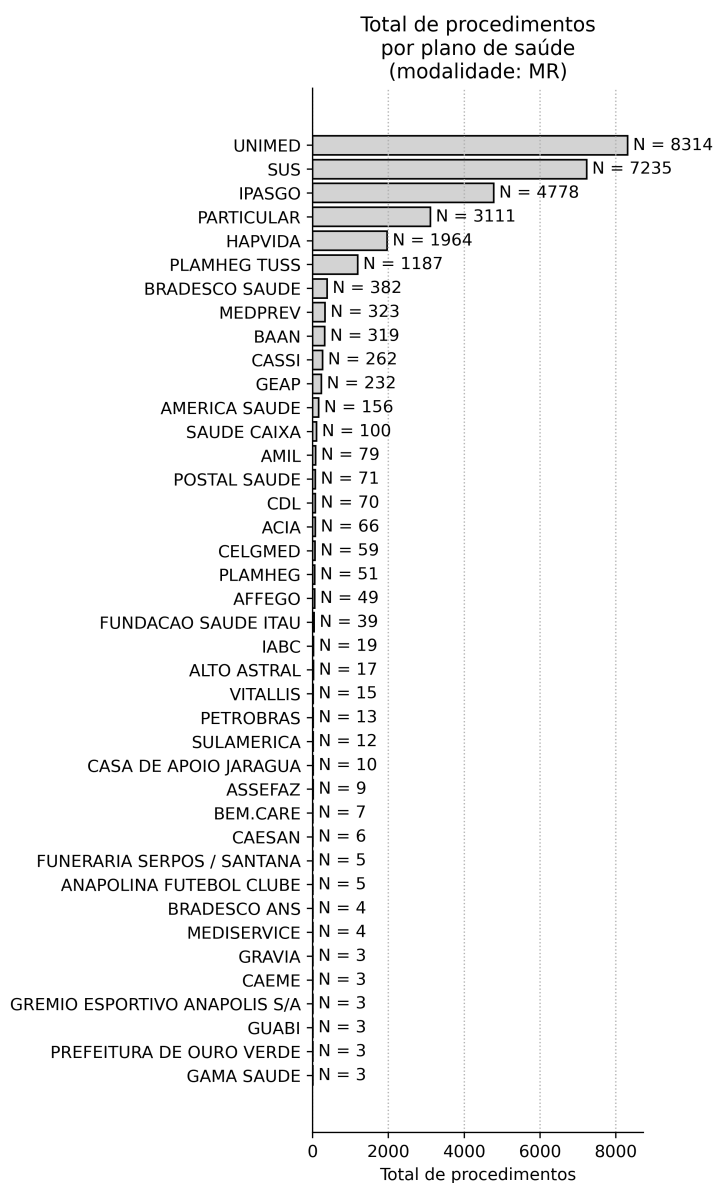


Figura 35 – Distribuição do total de procedimentos por plano de saúde dos pacientes para a modalidade de MR da unidade do St. Central. **Fonte:** próprio autor.

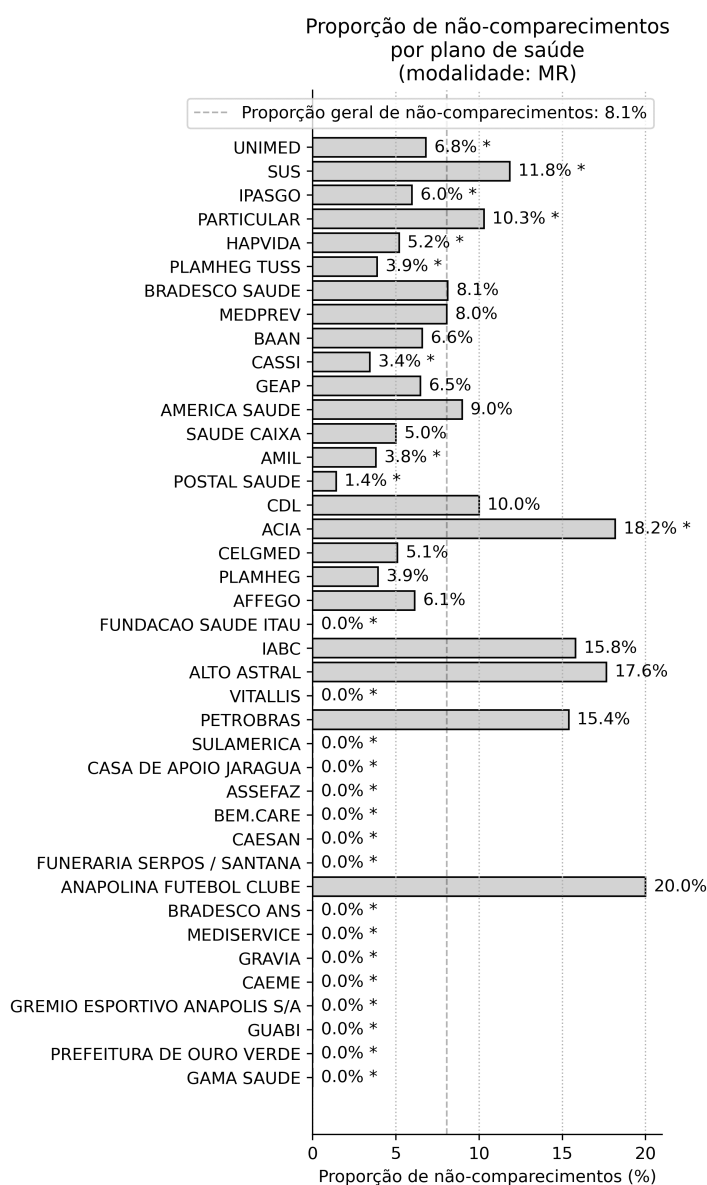


Figura 36 – Distribuição da proporção de não-comparecimentos por plano de saúde dos pacientes para a modalidade de MR da unidade do St. Central. O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos. **Fonte:** próprio autor.

4.2.2 Tunagem de hiperparâmetros e seleção do modelo de classificação

Utilizando os dados de 2015 a 2022, foi realizada a etapa de tunagem de hiperparâmetros e o treinamento dos modelos LR, MLP, XGB, CB e LGBM. Para selecionar o melhor modelo (com a melhor combinação de hiperparâmetros) foi utilizado como critério a média

ponderada de AUC-PR obtido para cada *fold* de teste proveniente da Validação Cruzada com Divisão Temporal. A Figura 37 mostra a performance geral dos 5 modelos de classificação testados, para diferentes combinações de hiperparâmetros. As distribuições de valores mostradas na figura por meio de *boxplots* estão melhor detalhadas na Tabela 3. Nota-se por meio da figura e da tabela que os modelos que tendem a apresentar melhor performance para diferentes combinações de hiperparâmetros foram LGBM, CB e XGB. Já os modelos com maior variabilidade para diferentes combinações de hiperparâmetros foram LGBM e CB. A Tabela 3 indica que o melhor modelo e combinação de hiperparâmetros foi o LGBM, com média ponderada máxima de AUC-PR igual a 0.3363. A melhor combinação de hiperparâmetros correspondente foi: $\{n_estimators: 3000, max_depth: 63, colsample_bytree: 0.7886372708925217, scale_pos_weight: 7.472410910017989, learning_rate: 0.0007589929361363828, reg_alpha: 0.0011885229743602958, reg_lambda: 0.7041794245156696\}$.

Para a melhor combinação de hiperparâmetros do modelo LGBM, os resultados em cada *fold* de teste proveniente da Validação Cruzada com Divisão Temporal estão detalhados na Figura 38, inclusive para outras combinações de hiperparâmetros. É esperado que exista uma tendência de melhora na performance de acordo com os *folds* de teste, pois o modelo faz uso de uma maior quantidade de dados de treino. Apesar disso, há um decréscimo de performance entre os folds 4 e 5, de modo que pode ter existido mudanças no comportamento dos pacientes para o fold de teste 5. Para os folds posteriores, a performance continua a tendência de crescimento.

Usando o modelo LGBM com a melhor combinação de hiperparâmetros, foi feito o ajuste de limiar de probabilidade do modelo utilizando os mesmos folds de teste obtidos por Validação Cruzada com Divisão Temporal. O critério para seleção do limiar foi de melhor média ponderada para a métrica Custo, conforme definida na Tabela 2. Os resultados estão mostrados na Figura 39, e o limiar que minimizou a média ponderada da métrica custo foi 0.55.

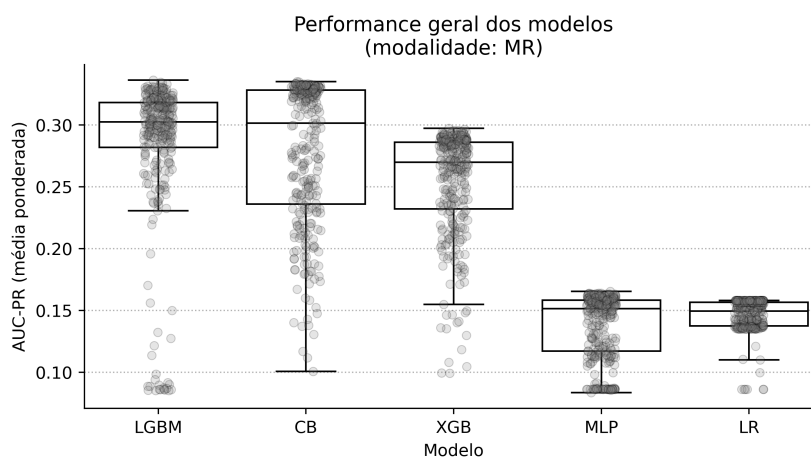


Figura 37 – Performance geral dos modelos de classificação testados para a modalidade de MR da unidade do St. Central. O modelo com a melhor combinação de hiperparâmetros em termos de média ponderada de AUC-PR foi selecionado. **Fonte:** próprio autor.

Tabela 3 – Informações adicionais sobre a performance geral dos modelos de classificação testados para a modalidade de MR da unidade do St. Central. O modelo com maior média ponderada de AUC-PR foi selecionado (coluna "Máx." da tabela).

Modelo	Mín.	Mediana	Média	Máx.	Desv. Pad.
LGBM	0.0852	0.3024	0.2841	0.3363	0.0592
CB	0.1007	0.3015	0.2783	0.3351	0.0575
XGB	0.0992	0.2700	0.2536	0.2973	0.0427
LR	0.0863	0.1495	0.1463	0.1580	0.0130
MLP	0.0835	0.1514	0.1378	0.1656	0.0260

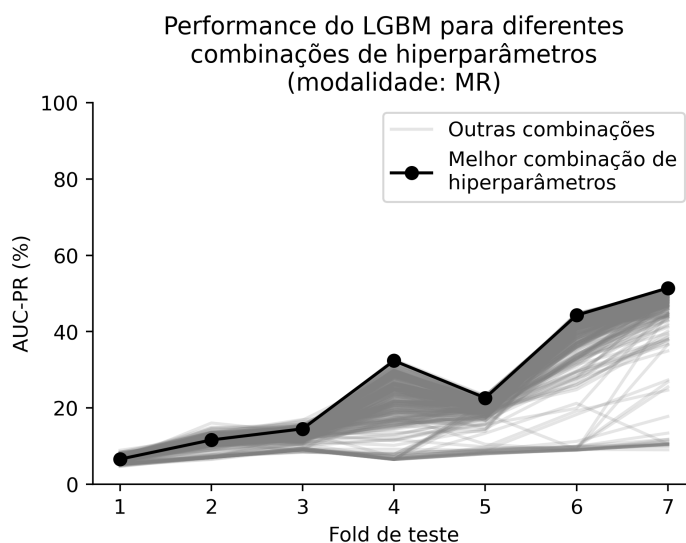


Figura 38 – Performance para diferentes combinações de hiperparâmetros para o modelo de classificação selecionado. **Fonte:** próprio autor.

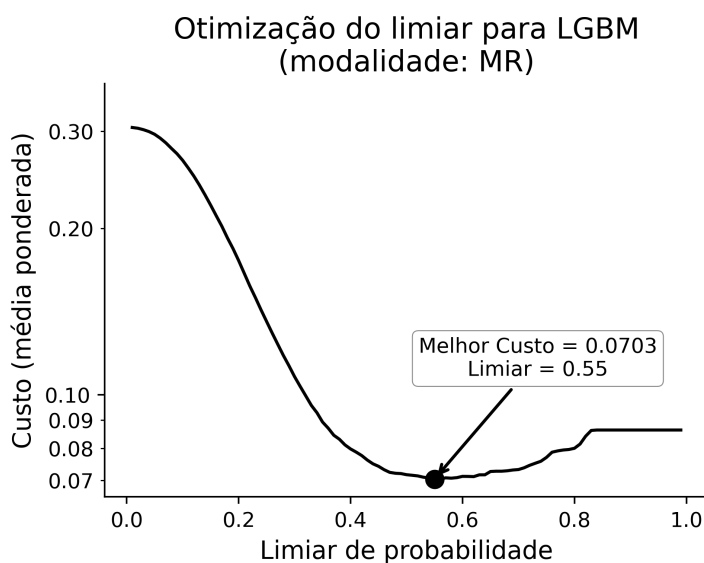


Figura 39 – Ajuste do limiar de probabilidade com base na média ponderada da métrica personalizada Custo para o modelo de classificação selecionado. **Fonte:** próprio autor.

4.2.3 Resultados sobre os dados de teste

Com o treino do modelo selecionado (ajustado com a melhor combinação de hiperparâmetros) sobre os dados entre os anos de 2015 e 2022, ele foi aplicado sobre os dados de teste (ano de 2023) para avaliação de performance. Na Figura 40, é mostrada a matriz de confusão do modelo sobre os dados de teste, a partir da qual, para a classe positiva (paciente não compa-

rece), foram calculadas as métricas de Precisão = 32.99%, Sensibilidade = 17.02% e F1-score = 22.46%. Tendo em vista que a proporção de comparecimento dos pacientes é de 92.06% nos dados de teste e a Acurácia do modelo foi de 90.67%, isso implica que o modelo está acertando menos do que um modelo que prediz somente a classe positiva, o que é um indicativo ruim de performance, conforme sugerido em (CARRERAS-GARCÍA et al., 2020). Apesar disso, o modelo de ML conseguiu reduzir a métrica Custo de 0.07937 (assumindo um cenário onde é esperado que todos os pacientes compareçam) para 0.07501 (cenário com o uso de ML para prever não-comparecimentos), o que representa uma redução na métrica Custo de 5.5%, o que não é muito significativo. Em relação às curvas de performance, a curva ROC (AUC = 0.6559) e Precision-Recall (AUC = 0.2203) estão mostradas nas Figuras 41 e 42, respectivamente. Ambas as curvas indicam que o modelo de ML é melhor do que um classificador aleatório.

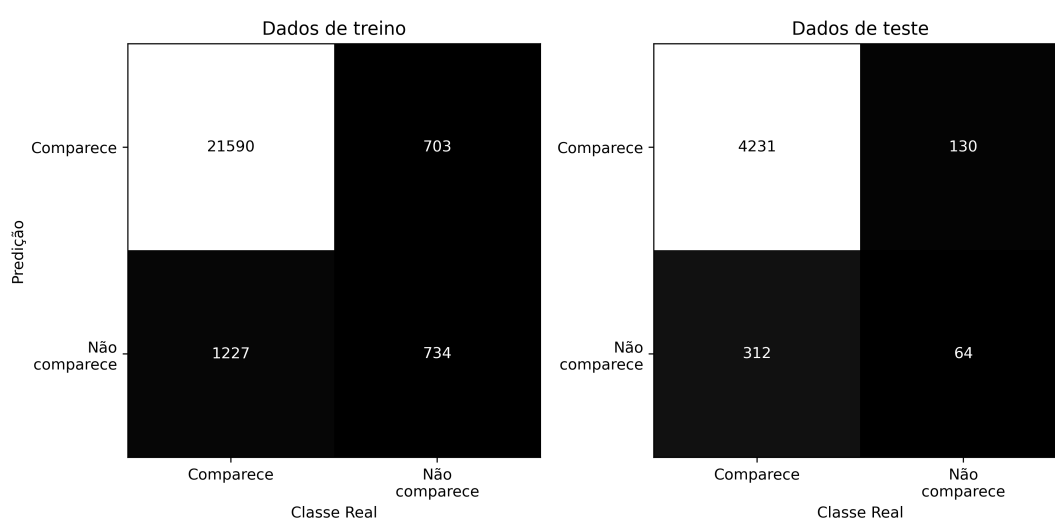


Figura 40 – Matrizes de confusão resultantes da aplicação do modelo selecionado sobre os dados de treino e teste. **Fonte:** próprio autor.

Tabela 4 – Métricas de performance do modelo selecionado conforme calculadas a partir das matrizes de confusão.

Dados	Classe	Precisão (%)	Sensibilidade (%)	F1-score (%)	N
Treino	Comparece	94.62	96.85	95.72	22293
	Não comparece	51.08	37.43	43.20	1961
	Acurácia (Treino)	92.04% (total de 24254)			
Teste	Comparece	93.13	97.02	95.04	4361
	Não comparece	32.99	17.02	22.46	376
	Acurácia (Teste)	90.67% (total de 4737)			

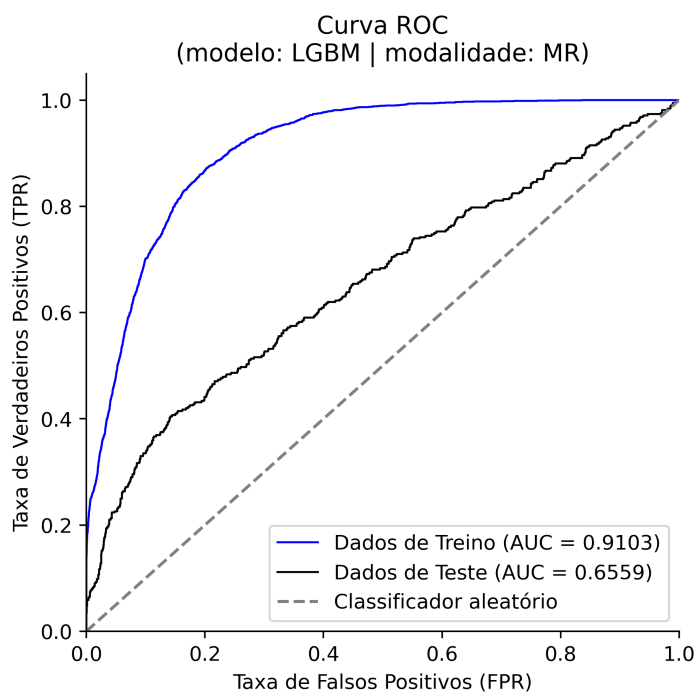


Figura 41 – Curva ROC resultante da aplicação do modelo selecionado sobre os dados de treino e teste. **Fonte:** próprio autor.

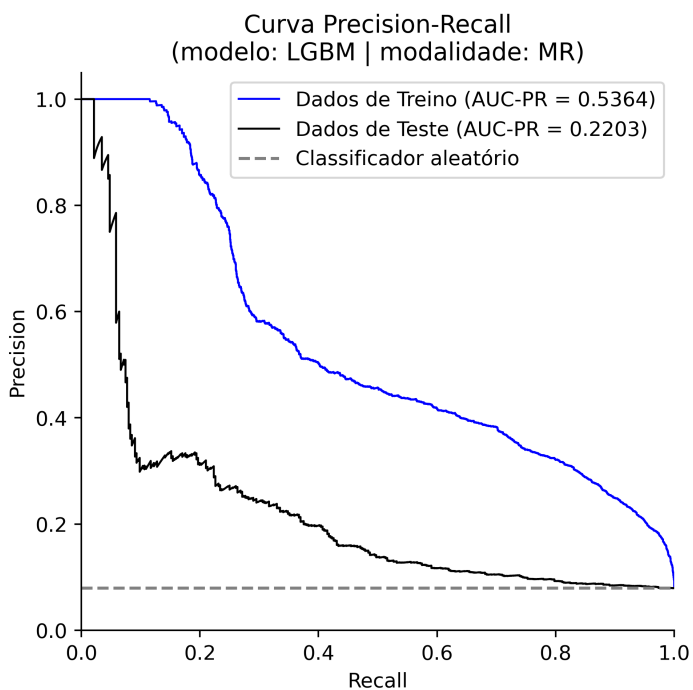


Figura 42 – Curva PR resultante da aplicação do modelo selecionado sobre os dados de treino e teste. **Fonte:** próprio autor.

4.2.4 Importância das variáveis

Aplicando o *framework SHAP* sobre as predições do modelo selecionado nos dados de teste para a modalidade de MR, tem-se a Figura 43, que mostra as 20 variáveis preditoras que mais contribuem. Pela Figura, é possível perceber que as 10 variáveis mais importantes nesse contexto são distância até a clínica, o plano de saúde do paciente, o sexo do paciente, a idade do paciente, o tipo de procedimento agendado, o mês e o horário de agendamento, o tempo de espera do paciente, a variável indicadora de se é a primeira vez do paciente na clínica e o histórico de não-comparecimento do paciente. Todas essas variáveis tem sido destacadas em estudos anteriores como sendo relevantes para a predição de não-comparecimentos (DANTAS et al., 2018; CARRERAS-GARCÍA et al., 2020), e isso sugere que o modelo treinado apreendeu padrões importantes nesse contexto. Chama-se a atenção para as variáveis resultantes do PCA aparecendo no top 20, indicando que, mesmo originalmente servindo para séries temporais, apresentaram contribuição relevante no problema de predição de não-comparecimentos.

Note, entretanto, que a Figura 43 não indica a direcionalidade da contribuição das variáveis, de modo que é preciso abstrá-las parcialmente através da seção de exploração dos dados, que inclui análises univariadas. Ainda assim, devido às interações complexas que existem entre as variáveis, a direcionalidade de contribuição pode não se manter para todas as instâncias de predição, sendo necessário uma alternativa para compreender cada predição. Nesse sentido, a Figura 44 serve como uma ilustração de como o *framework SHAP* pode ser útil nesse sentido, podendo servir como uma ferramenta para interpretar as predições individualmente, para ambas as classes estudadas.

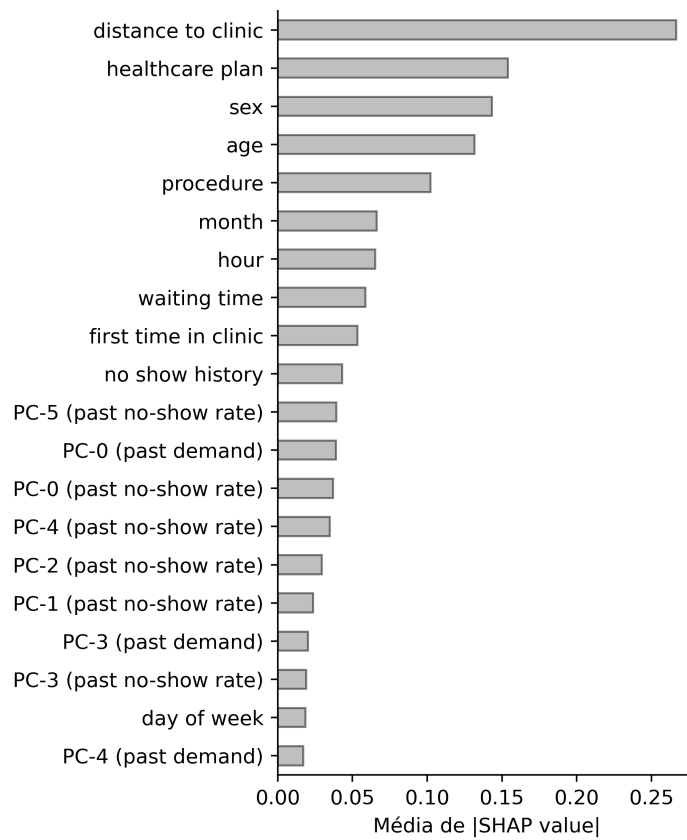


Figura 43 – Média dos valores SHAP absolutos resultantes da aplicação do modelo selecionado sobre os dados de teste. **Fonte:** próprio autor.

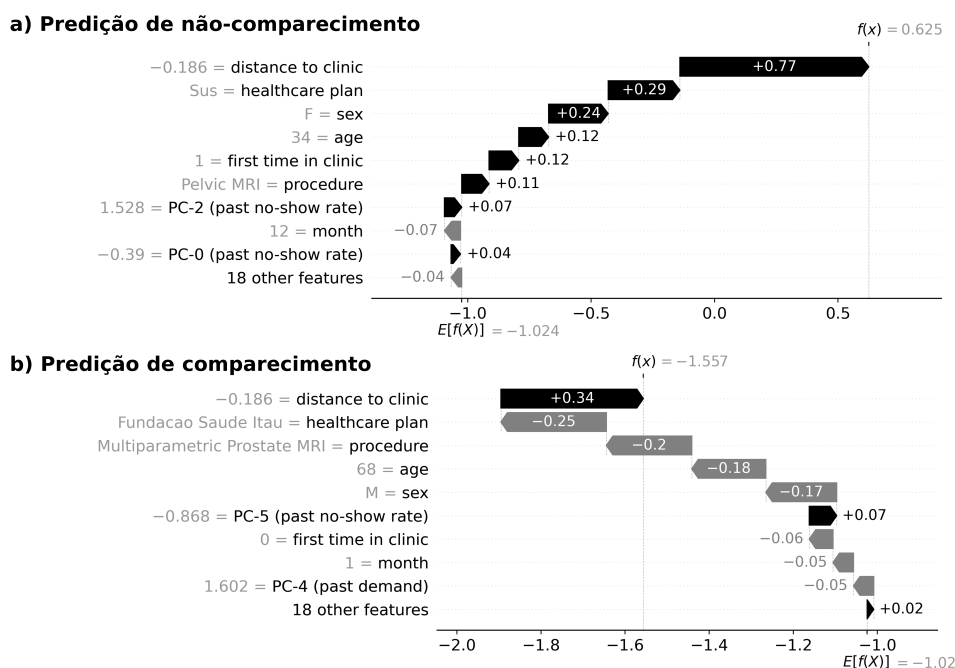


Figura 44 – Exemplo de predições individuais feitas pelo modelo selecionado em instâncias do conjunto de teste, com seus respectivos valores de contribuição das variáveis. O eixo horizontal representa as contribuições em log-odds. As setas pretas indicam contribuições positivas para a probabilidade de não-comparecimento, enquanto as setas cinzas indicam contribuições negativas. O valor base $E[f(x)]$ corresponde à saída média do modelo e serve como ponto de referência. A saída final do modelo $f(x)$ reflete a contribuição cumulativa de todos os valores de atributos para a instância específica. **Fonte:** próprio autor.

4.3 Modalidade de Tomografia Computadorizada (unidade do St. Central)

Para a modalidade de CT na unidade do St. Central do CRA, houve um total de 21509 instâncias devidamente pré-processadas, datando entre 2015 e 2023, dentre as quais 2376 houve não-comparecimento de pacientes, o que resulta em uma proporção total de não-comparecimentos próxima de 11.0%. Logo abaixo, é feita uma análise de como essa proporção varia de acordo com diferentes agrupamentos das variáveis extraídas do sistema de registros.

4.3.1 Análise exploratória

4.3.1.1 Distribuição temporal da proporção de não-comparecimentos

Nas Figuras 45 e 46 são mostrados, respectivamente, o comportamento da proporção de não-comparecimentos e o total de instâncias considerando os diferentes anos e meses do ano. Mais especificamente, na Figura 45 vemos que existe uma proporção de não-comparecimentos menor do que a proporção geral para os anos de 2015, 2016, 2018, 2019, 2020. Nesses anos, não parece haver um padrão perceptível de menores proporções de não-comparecimentos ao longo dos meses, estando bem distribuídas ao longo dos anos. Por outro lado, houve maior proporção de não-comparecimentos nos anos de 2021, 2022 e 2023, sendo perceptível, inclusive, uma tendência de crescimento entre esses anos. Por exemplo, no mês de 2021, há somente dois meses com maiores proporções de não-comparecimento (mês de Outubro apresenta uma proporção de 38.1%, mas isso pode ter relação com a baixa quantidade de procedimentos agendados nesse mês e nesse ano, vide Figura 46). Já nos meses de 2022 e 2023, a quantidade de meses com maiores proporções de não-comparecimento aumenta para 5 e 7, respectivamente, o que é um aumento relevante. Adicionalmente, não parece existir nenhum padrão associado aos meses, estando essas proporções maiores bem distribuídas ao longo desses mesmos anos. Muito provavelmente essa tendência de crescimento da proporção de não-comparecimentos está relacionada a uma pior gestão dos agendamentos ou, simplesmente, a uma diminuição na qualidade do serviço oferecido aos pacientes, pois somado ao aumento na proporção de não-comparecimentos, há uma diminuição no total de procedimentos agendados para essa modalidade (vide Figura 46). Analisando, agora, somente os meses, não há meses que se destacam significativamente por maiores ou menores proporções de não-comparecimentos.

Levando em consideração, agora, somente a Figura 46, é possível perceber que, ao longo dos anos, há uma tendência de aumento no total de procedimentos agendados para a modalidade de CT entre os anos de 2015 e 2021. Isso pode estar relacionado a uma maior oferta da clínica através de, por exemplo, mais máquinas de CT, ou, simplesmente, a uma maior busca por exames nessa modalidade. Entretanto, conforme foi discutido no parágrafo acima, entre os anos de 2021 e 2023, há uma diminuição no total de agendamentos. Isso pode ser justificado por uma piora na gestão dos agendamentos ou, simplesmente, uma diminuição na qualidade do serviço oferecido aos pacientes, motivando uma busca por outras clínicas. De forma geral, há uma maior quantidade de procedimentos agendados para os meses de Janeiro, Março, Maio, Junho e Agosto. Já um fato curioso é que, no mês de Outubro do ano de 2021, houve apenas 21 procedimentos agendados. Isso pode ter relação com a indisponibilidade da máquina, por exemplo.

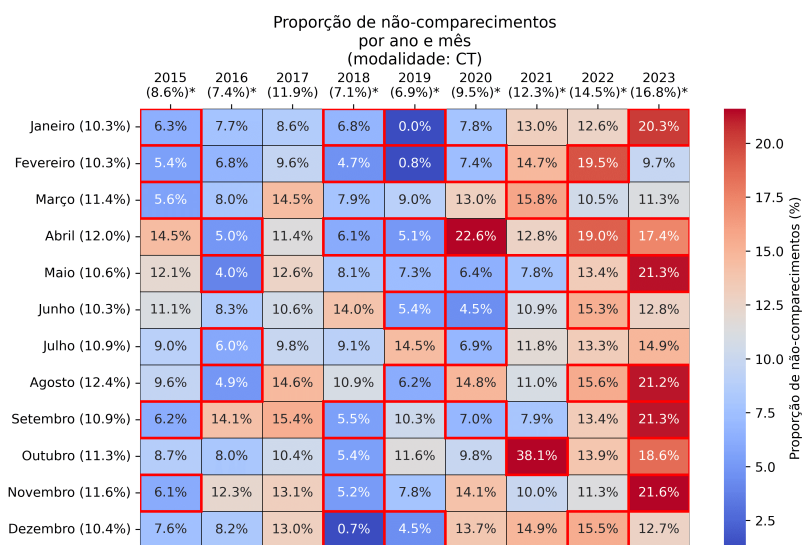


Figura 45 – Distribuição da proporção de não-comparecimentos por ano e mês para a modalidade de CT da unidade do St. Central. São também mostradas as proporções totais de não-comparecimentos separadamente para cada mês e cada ano (distribuições marginais). O "*" indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos, bem como o destaque em vermelho em torno das células mês-ano. **Fonte:** próprio autor.

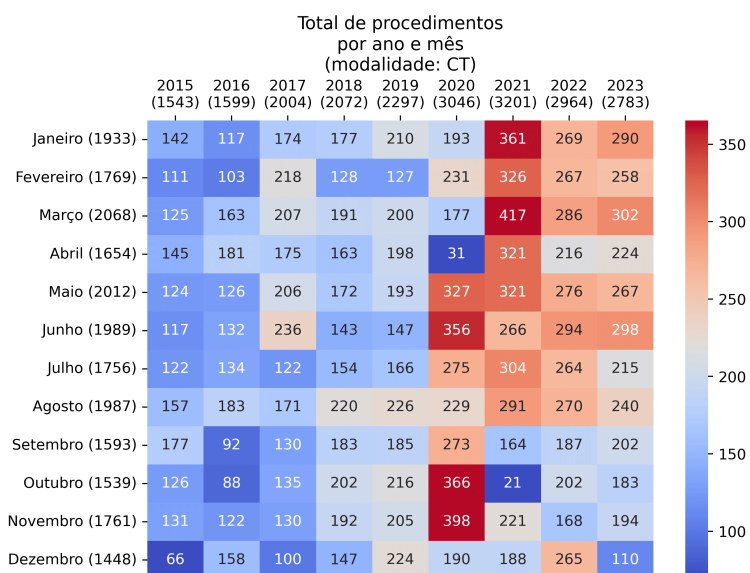


Figura 46 – Distribuição do total de procedimentos por ano e mês para a modalidade de CT da unidade do St. Central. São também mostrados o total de procedimentos separadamente para cada mês e cada ano (distribuições marginais). **Fonte:** próprio autor.

Analisando a proporção de não-comparecimentos e de total de procedimentos agendados

para os dias da semana e horários do dia, tem-se as Figuras 47 e 48. Para a modalidade de CT, nota-se da Figura 48 que os horários preferidos para o agendamento de exames é das 7:00hs até as 11:00hs, inclusive para o sábado. A partir das 12:00hs, o volume de agendamentos cai de forma relevante. Nota-se adicionalmente que, diferentemente da modalidade de MR, os exames de CT acontecem até no máximo 19:00hs. Analisando-se, agora, a Figura 47, vemos que há menores proporções de não-comparecimento para o horário das 7:00hs entre os dias de terça-feira e sexta-feira, assim como no horário de 11:00hs do dia de quinta-feira. Há maiores proporções de não-comparecimento nos horários de 12:00hs (horário de almoço) e 15:00hs para o dia de segunda-feira, assim como para o horário das 14:00hs no dia de sexta-feira. Perceba que as maiores proporções de não-comparecimento acontecem a partir das 12:00hs, que é o horário no qual a busca de agendamentos por parte dos pacientes cai drasticamente.

De forma isolada, o dia de segunda-feira se destacou por apresentar maiores proporções de não-comparecimento. O dia de quinta-feira, por sua vez, apresentou menores proporções. Já em termos dos horários do dia, o horário de 7:00hs apresentou menor proporção de não-comparecimento; enquanto que, nos horários de 14:00hs e 15:00hs, houve maiores proporções de não-comparecimento.

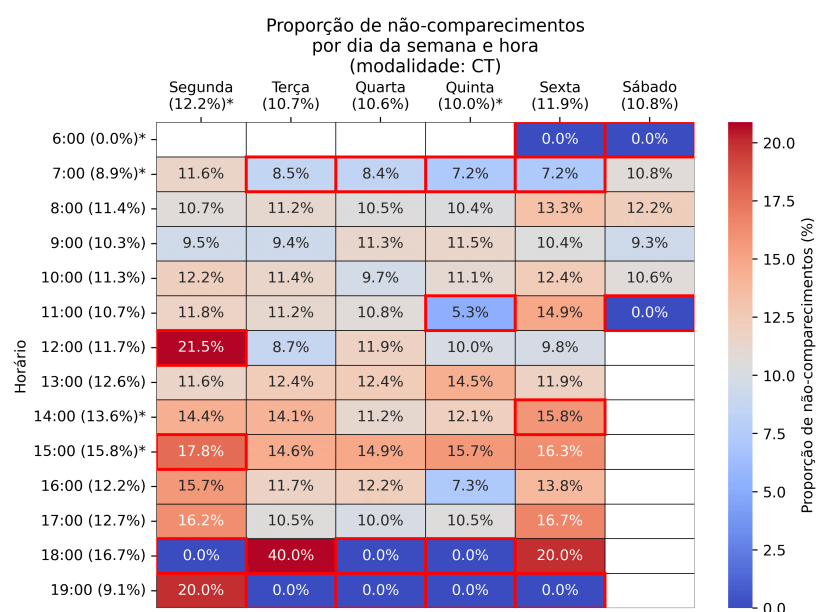


Figura 47 – Distribuição da proporção de não-comparecimentos por dia da semana e horário para a modalidade de CT da unidade do St. Central. São também mostradas as proporções totais de não-comparecimentos separadamente para cada dia da semana e cada horário (distribuições marginais). O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos, bem como o destaque em vermelho em torno das células dia-horário. **Fonte:** próprio autor.

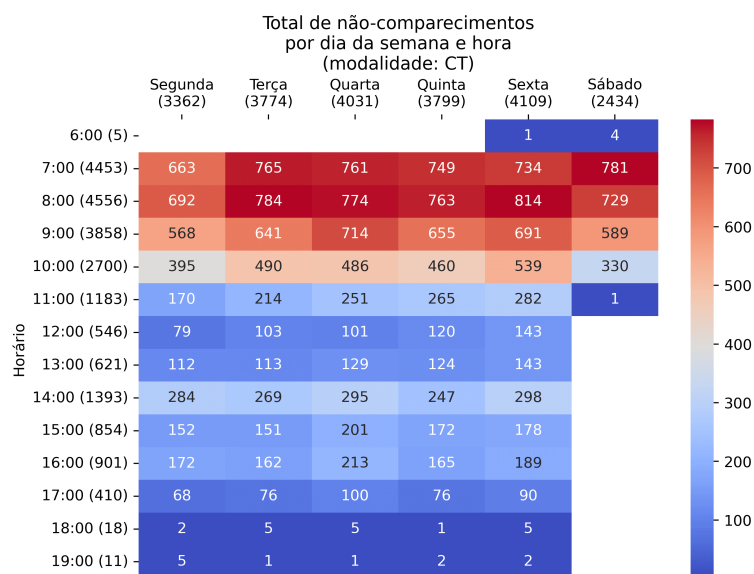


Figura 48 – Distribuição do total de procedimentos por dia da semana e horário para a modalidade de CT da unidade do St. Central. São também mostrados o total de procedimentos separadamente para cada dia da semana e cada horário (distribuições marginais).
Fonte: próprio autor.

4.3.1.2 Características dos pacientes e proporção de não-comparecimentos

Apesar de as Figuras 49, 50, 51, 52, 53, 54 e 55 representarem efeitos isolados (isto é, sem levar em conta interações) de diferentes características dos pacientes sobre a proporção de não-comparecimento na modalidade de CT, é possível esboçar, mesmo que de forma inicial, o "perfil" do paciente que tende a não comparecer.

Analisando as figuras desta subseção, tem-se que o "perfil" superficial do paciente que tende a não comparecer a procedimentos agendados em CT é:

- Não apresenta a informação "sexo" preenchida no sistema de registros ou é do sexo feminino (note que a proporção de não-comparecimentos para sexo = "F" continua sendo menor que a proporção geral, mas é maior numericamente que a proporção para sexo = "M") - Figura 49;
- Não apresenta a informação "idade" preenchida no sistema de registros ou apresenta idade nos bins (0, 10], (20, 30] ou (30, 40] - 50;
- Não apresenta a informação de "endereço" no sistema de registros - Figura 51;
- Agendou o procedimento com pelo menos 3 dias de antecedência - Figura 52;

- Ou agendou pela primeira vez algum procedimento no CRA, ou apresenta histórico prévio de não-comparecimento de no mínimo 6.67% - Figura 53;
- Agendaram o procedimento via SUS ou via PARTICULAR (levando-se em consideração somente os planos de saúde mais comuns no conjunto de dados) - Figuras 54 e 55;

Todas essas características, além de serem úteis para a construção superficial do "perfil" do paciente que não comparece a procedimentos de CT, também servem como um *sanity check* ao analisar a importância das variáveis via *framework SHAP*, que será feita mais adiante.

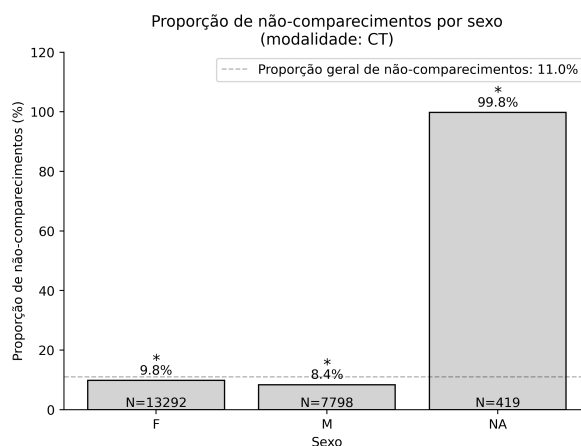


Figura 49 – Distribuição da proporção de não-comparecimentos por sexo dos pacientes para a modalidade de CT da unidade do St. Central. São também mostrados o total de procedimentos separadamente para sexo. O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos. **Fonte:** próprio autor.

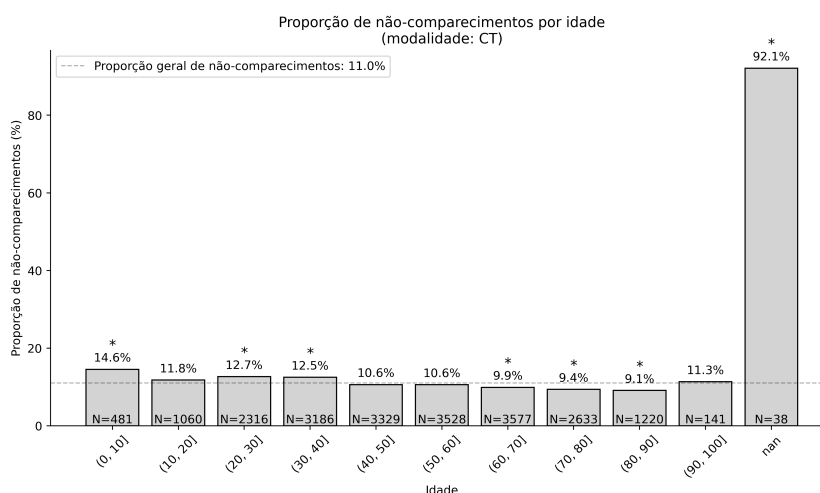


Figura 50 – Distribuição da proporção de não-comparecimentos por idade dos pacientes para a modalidade de CT da unidade do St. Central. São também mostrados o total de procedimentos separadamente para cada *bin* de idade. O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos. **Fonte:** próprio autor.

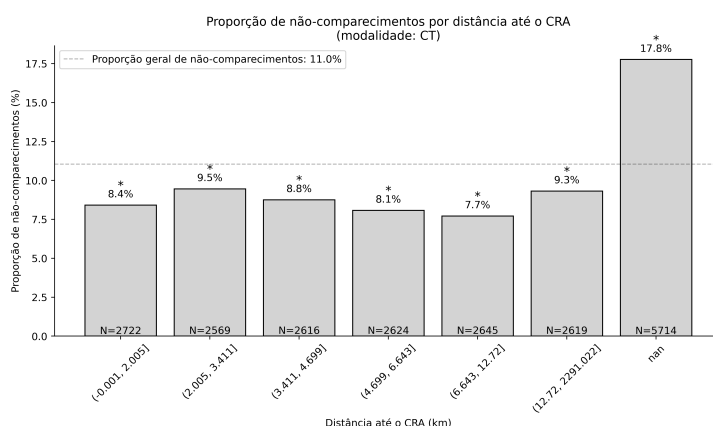


Figura 51 – Distribuição da proporção de não-comparecimentos por distância até o CRA a partir do endereço cadastrado dos pacientes para a modalidade de CT da unidade do St. Central. São também mostrados o total de procedimentos separadamente para cada *bin* de distância. O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos. **Fonte:** próprio autor.

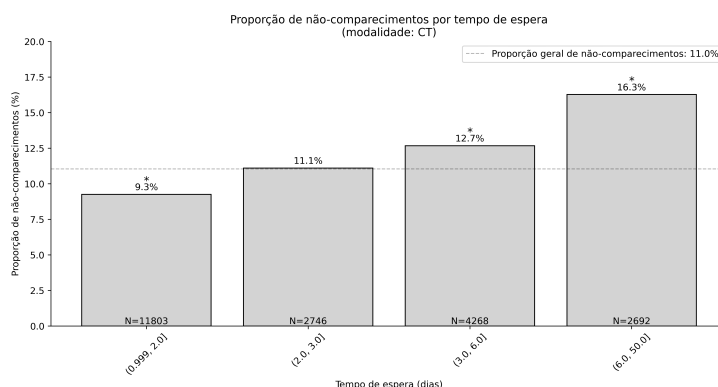


Figura 52 – Distribuição da proporção de não-comparecimentos por tempo de espera dos pacientes para a modalidade de CT da unidade do St. Central. São também mostrados o total de procedimentos separadamente para cada *bin* de tempo de espera. O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos. **Fonte:** próprio autor.

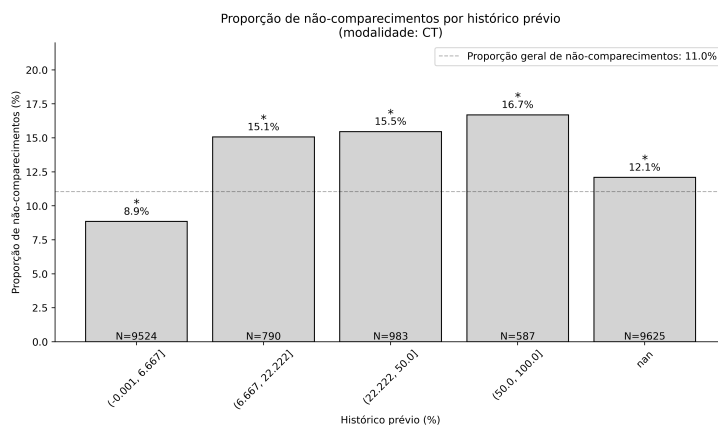


Figura 53 – Distribuição da proporção de não-comparecimentos por histórico prévio dos pacientes para a modalidade de CT da unidade do St. Central. São também mostrados o total de procedimentos separadamente para cada *bin* de histórico prévio. O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos. **Fonte:** próprio autor.

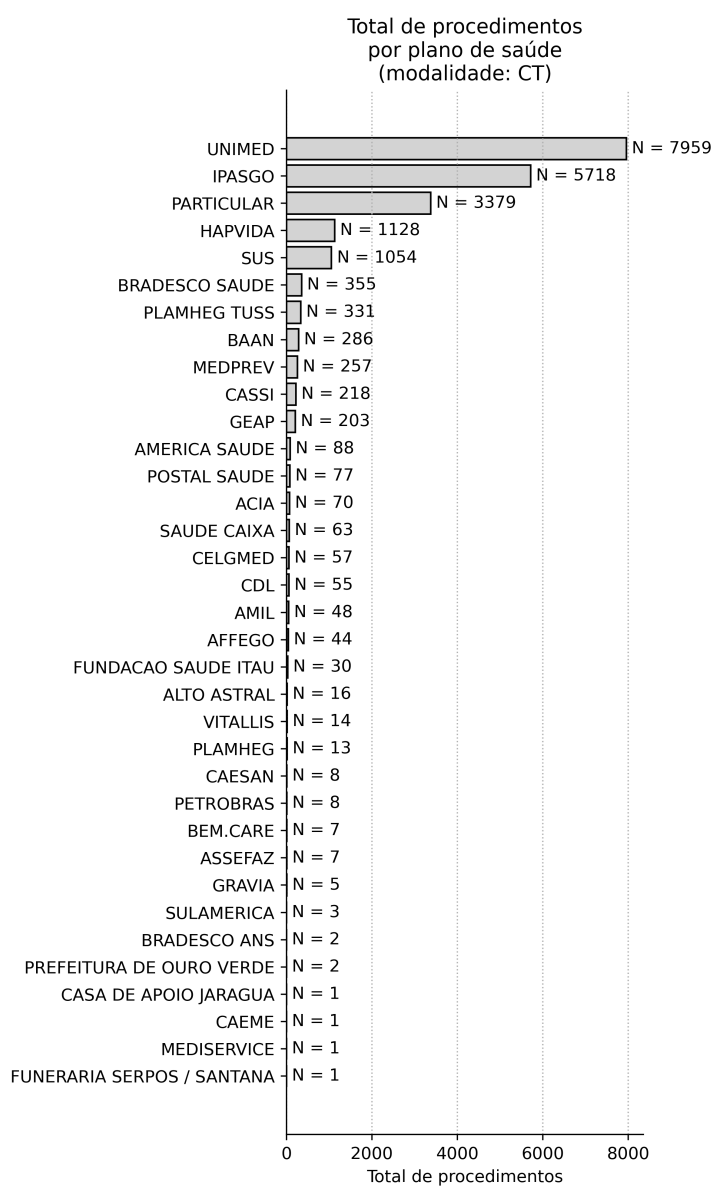


Figura 54 – Distribuição do total de procedimentos por plano de saúde dos pacientes para a modalidade de CT da unidade do St. Central. **Fonte:** próprio autor.

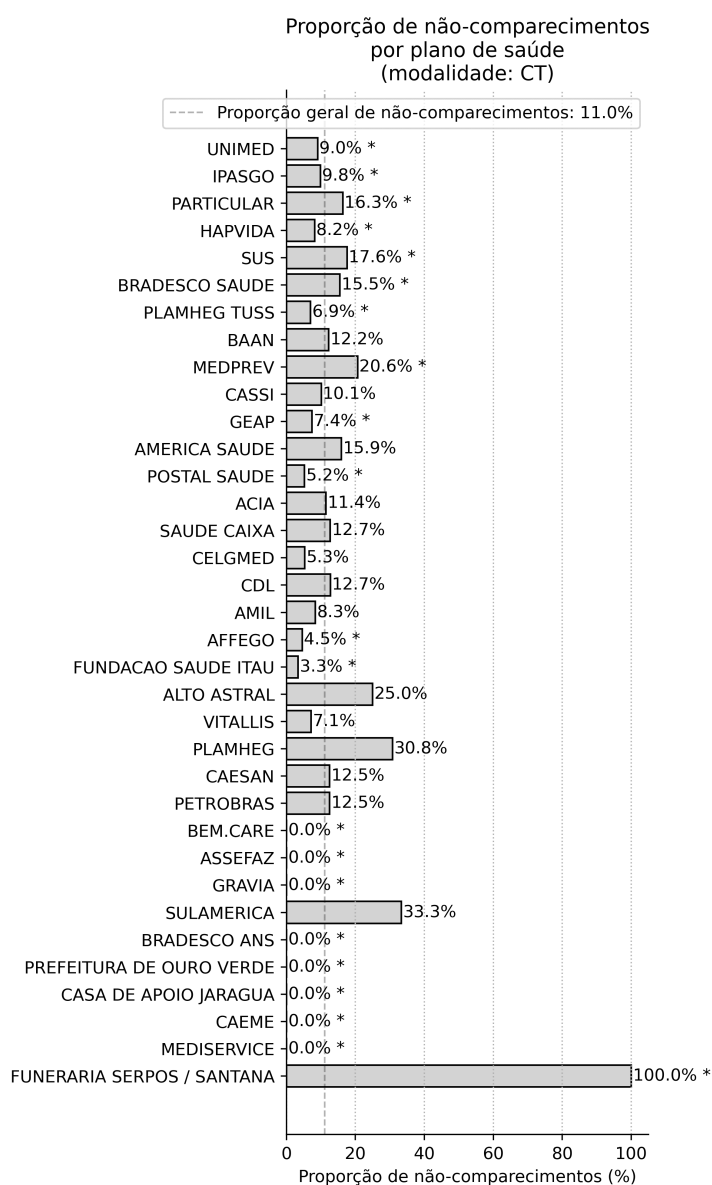


Figura 55 – Distribuição da proporção de não-comparecimentos por plano de saúde dos pacientes para a modalidade de CT da unidade do St. Central. O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos. **Fonte:** próprio autor.

4.3.2 Tunagem de hiperparâmetros e seleção do modelo de classificação

Utilizando os dados de 2015 a 2022, foi realizada a etapa de tunagem de hiperparâmetros e o treinamento dos modelos LR, MLP, XGB, CB e LGBM. Para selecionar o melhor modelo (com a melhor combinação de hiperparâmetros) foi utilizado como critério a média ponderada

de AUC-PR obtido para cada *fold* de teste proveniente da Validação Cruzada com Divisão Temporal. A Figura 56 mostra a performance geral dos 5 modelos de classificação testados, para diferentes combinações de hiperparâmetros. As distribuições de valores mostradas na figura por meio de *boxplots* estão melhor detalhadas na Tabela 5. Nota-se por meio da figura e da tabela que os modelos que tendem a apresentar melhor performance para diferentes combinações de hiperparâmetros foram CB e LGBM. Já os modelos com maior variabilidade para diferentes combinações de hiperparâmetros foram também CB e LGBM. A Tabela 5 indica que o melhor modelo e combinação de hiperparâmetros foi o CB, com média ponderada máxima de AUC-PR igual a 0.2444. A melhor combinação de hiperparâmetros correspondente foi: {*iterations*: 261, *depth*: 6, *learning_rate*: 0.012668500959553707, *l2_leaf_reg*: 10.0, *bagging_temperature*: 0.0, *rsm*: 1.0, *border_count*: 8, *random_strength*: 0.2304335349011509, *scale_pos_weight*: 1.0}.

Para a melhor combinação de hiperparâmetros do modelo CB, os resultados em cada *fold* de teste proveniente da Validação Cruzada com Divisão Temporal estão detalhados na Figura 57, inclusive para outras combinações de hiperparâmetros. Sobre a Figura 57, é esperado que exista uma tendência de melhora na performance de acordo com os *folds* de teste, pois o modelo faz uso de uma maior quantidade de dados de treino. Apesar disso, esse padrão não acontece na figura, especialmente devido ao decréscimo de performance entre os *folds* de teste 2 e 3. Novamente, isso pode ter relação com mudanças abruptas no comportamento de não-comparecimento dos pacientes para o *fold* de teste 3. Perceba que para os *folds* seguintes, a tendência de crescimento se mantém.

Usando o modelo CB com a melhor combinação de hiperparâmetros, foi feito o ajuste de limiar de probabilidade do modelo utilizando os mesmos *folds* de teste obtidos por Validação Cruzada com Divisão Temporal. O critério para seleção do limiar foi de melhor média ponderada para a métrica Custo, conforme definida na Tabela 2. Os resultados estão mostrados na Figura 58, e o limiar que minimizou a média ponderada da métrica custo foi 0.17.

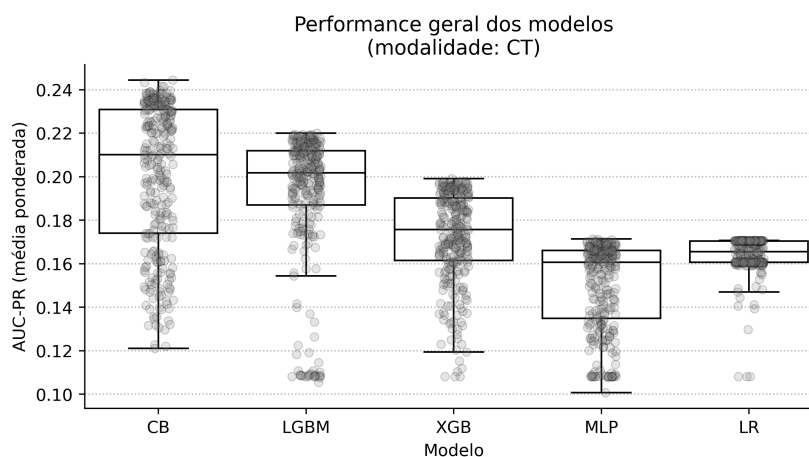


Figura 56 – Performance geral dos modelos de classificação testados para a modalidade de CT da unidade do St. Central. O modelo com a melhor combinação de hiperparâmetros em termos de média ponderada de AUC-PR foi selecionado. **Fonte:** próprio autor.

Tabela 5 – Informações adicionais sobre a performance geral dos modelos de classificação testados para a modalidade de CT da unidade do St. Central. O modelo com maior média ponderada de AUC-PR foi selecionado (coluna "Máx." da tabela).

Modelo	Mín.	Mediana	Média	Máx.	Desv. Pad.
CB	0.1211	0.2101	0.2008	0.2444	0.0335
LGBM	0.1055	0.2018	0.1921	0.2200	0.0301
XGB	0.1082	0.1758	0.1725	0.1991	0.0210
LR	0.1082	0.1655	0.1643	0.1708	0.0083
MLP	0.1007	0.1607	0.1503	0.1714	0.0200

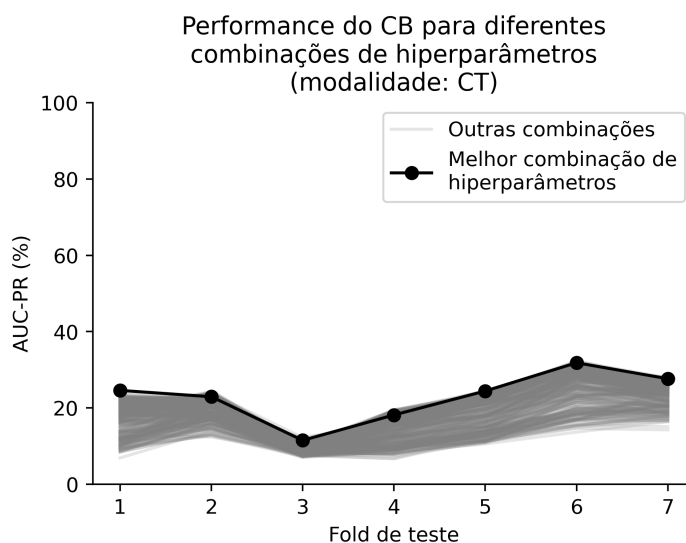


Figura 57 – Performance para diferentes combinações de hiperparâmetros para o modelo de classificação selecionado. **Fonte:** próprio autor.

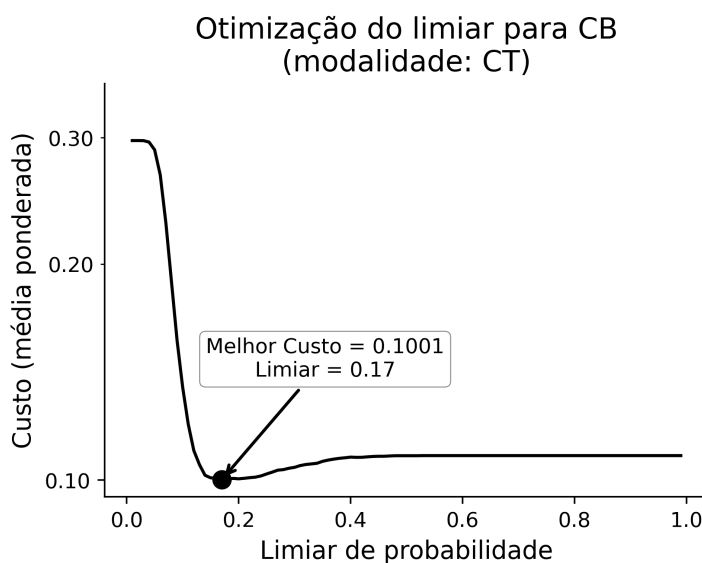


Figura 58 – Ajuste do limiar de probabilidade com base na média ponderada da métrica personalizada Custo para o modelo de classificação selecionado. **Fonte:** próprio autor.

4.3.3 Resultados sobre os dados de teste

Com o treino do modelo selecionado (ajustado com a melhor combinação de hiperparâmetros) sobre os dados entre os anos de 2015 e 2022, ele foi aplicado sobre os dados de teste (ano de 2023) para avaliação de performance. Na Figura 59, é mostrada a matriz de confusão do modelo sobre os dados de teste, a partir da qual, para a classe positiva (paciente não comparece),

foram calculadas as métricas de Precisão = 41.43%, Sensibilidade = 44.44% e F1-score = 42.89%. Tendo em vista que a proporção de comparecimento dos pacientes é de 83.18% nos dados de teste e a Acurácia do modelo foi de 80.09%, o que implica que o modelo está acertando menos do que tão somente a classe positiva, o que é um indicativo ruim de performance, conforme sugerido em (CARRERAS-GARCÍA et al., 2020). Apesar disso, o modelo de ML conseguiu reduzir a métrica Custo de 0.16816 (assumindo um cenário onde é esperado que todos os pacientes compareçam) para 0.12863 (cenário com o uso de ML para prever não-comparecimentos), o que representa uma redução na métrica Custo de 23.5%, que deixa claro a vantagem de se utilizar o modelo de ML. Em relação às curvas de performance, a curva ROC (AUC = 0.7043) e Precision-Recall (AUC = 0.3711) estão mostradas nas Figuras 60 e 61, respectivamente. Ambas as curvas indicam que o modelo de ML é melhor do que um classificador aleatório.

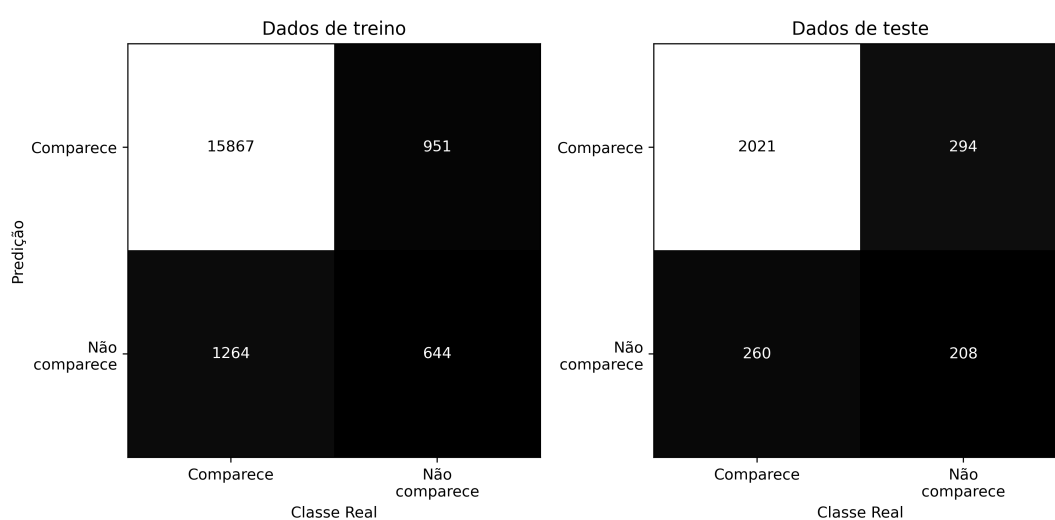


Figura 59 – Matrizes de confusão resultantes da aplicação do modelo selecionado sobre os dados de treino e teste. **Fonte:** próprio autor.

Tabela 6 – Métricas de performance do modelo selecionado conforme calculadas a partir das matrizes de confusão.

Dados	Classe	Precisão (%)	Sensibilidade (%)	F1-score (%)	N
Treino	Comparece	92.62	94.35	93.48	16818
	Não comparece	40.38	33.75	36.77	1908
	Acurácia (Treino)	88.17% (total de 18726)			
Teste	Comparece	88.60	87.30	87.95	2315
	Não comparece	41.43	44.44	42.89	468
	Acurácia (Teste)	80.09% (total de 2783)			

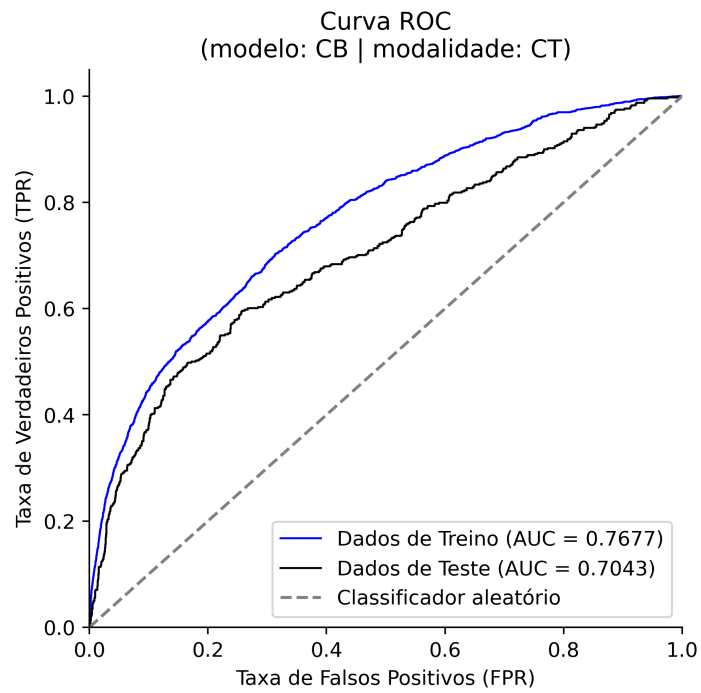


Figura 60 – Curva ROC resultante da aplicação do modelo selecionado sobre os dados de treino e teste. **Fonte:** próprio autor.

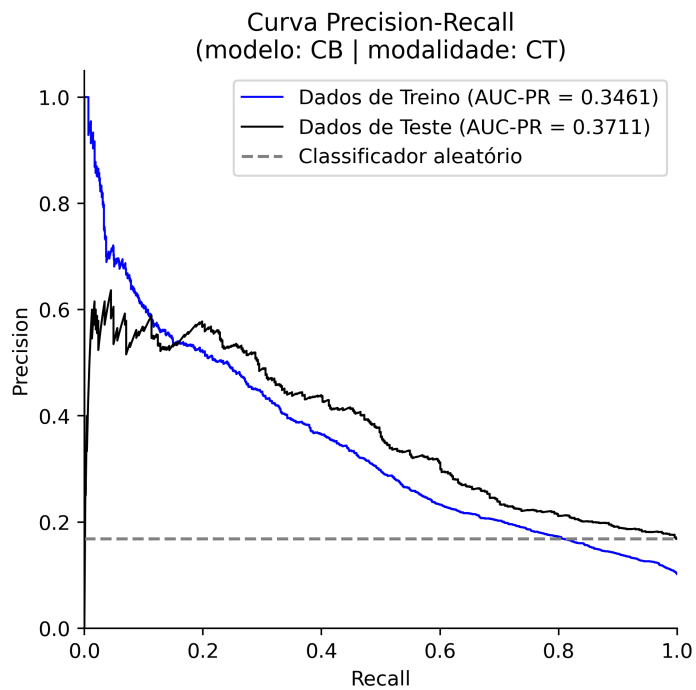


Figura 61 – Curva PR resultante da aplicação do modelo selecionado sobre os dados de treino e teste. **Fonte:** próprio autor.

4.3.4 Importância das variáveis

Aplicando o *framework SHAP* sobre as predições do modelo selecionado nos dados de teste para a modalidade de CT, tem-se a Figura 62, que mostra as 20 variáveis preditoras que mais contribuem. Pela Figura, é possível perceber que as variáveis mais importantes nesse contexto são distância até a clínica, tempo de espera do paciente, o sexo do paciente, o plano de saúde do paciente, a idade do paciente, a variável indicadora de se é a primeira vez do paciente na clínica, o componente principal 5 decorrente das taxas de não-comparecimento, o histórico de não-comparecimento do paciente, a hora agendada e o componente principal 0 decorrente das taxas de não-comparecimento. Todas essas variáveis (exceto as que envolveram PCA) tem sido destacadas em estudos anteriores como sendo relevantes para a predição de não-comparecimentos (DANTAS et al., 2018; CARRERAS-GARCÍA et al., 2020), e isso sugere que o modelo treinado apreendeu padrões importantes nesse contexto. Chama-se a atenção para as variáveis resultantes do PCA aparecendo no top 20 (e inclusive no top 10), indicando que, mesmo originalmente servindo para séries temporais, apresentaram contribuição relevante no problema de predição de não-comparecimentos.

Note, entretanto, que a Figura 62 não indica a direcionalidade da contribuição das variáveis, de modo que é preciso abstrai-las parcialmente através da seção de exploração dos dados, que inclui análises univariadas. Ainda assim, devido às interações complexas que existem entre as variáveis, a direcionalidade de contribuição pode não se manter para todas as instâncias de predição, sendo necessário uma alternativa para compreender cada predição. Nesse sentido, a Figura 63 serve como uma ilustração de como o *framework SHAP* pode ser útil nesse sentido, podendo servir como uma ferramenta para interpretar as predições individualmente, para ambas as classes estudadas.

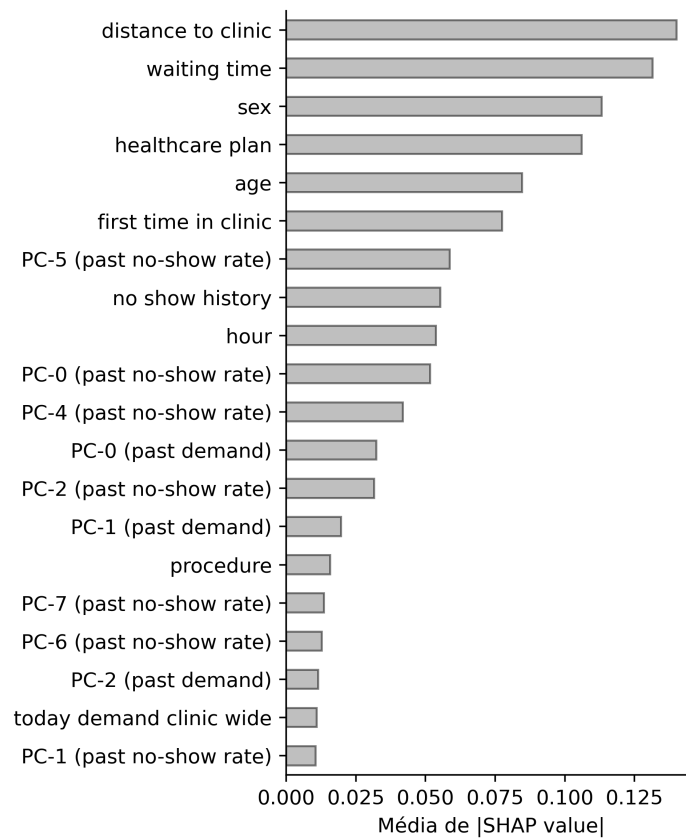


Figura 62 – Média dos valores SHAP absolutos resultantes da aplicação do modelo selecionado sobre os dados de teste. **Fonte:** próprio autor.

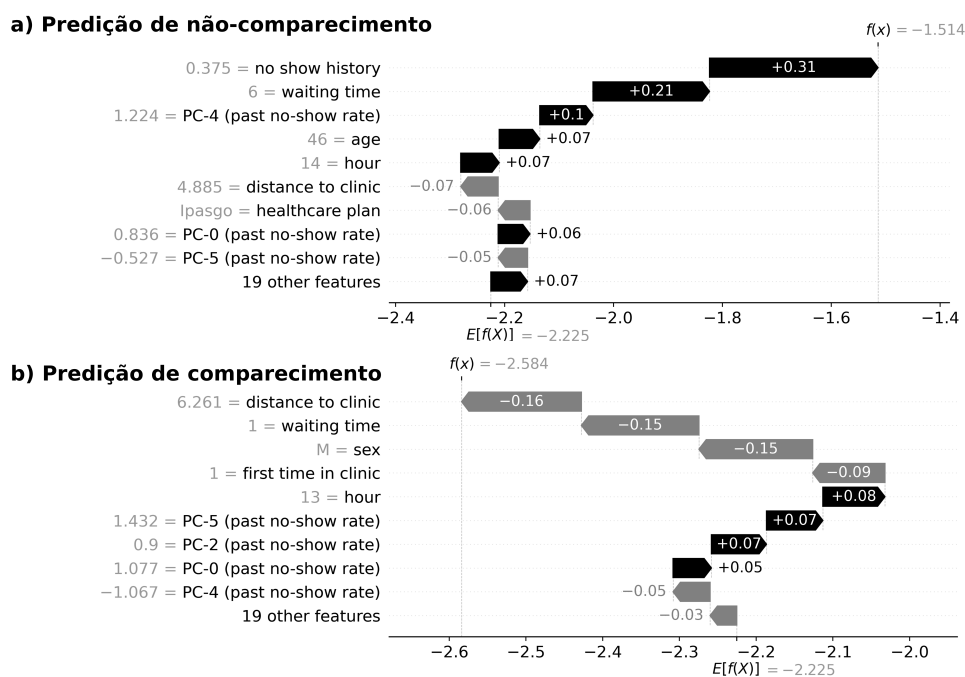


Figura 63 – Exemplo de predições individuais feitas pelo modelo selecionado em instâncias do conjunto de teste, com seus respectivos valores de contribuição das variáveis. O eixo horizontal representa as contribuições em log-odds. As setas pretas indicam contribuições positivas para a probabilidade de não-comparecimento, enquanto as setas cinzas indicam contribuições negativas. O valor base $E[f(x)]$ corresponde à saída média do modelo e serve como ponto de referência. A saída final do modelo $f(x)$ reflete a contribuição cumulativa de todos os valores de atributos para a instância específica. **Fonte:** próprio autor.

4.4 Modalidade de consultas (unidade de Jundiaí)

Para a modalidade de consultas na unidade de Jundiaí do CRA, houve um total de 64367 instâncias devidamente pré-processadas, datando entre 2015 e 2023, dentre as quais 11282 houve não-comparecimento de pacientes, o que resulta em uma proporção total de não-comparecimentos próxima de 17.5%. Logo abaixo, é feita uma análise de como essa proporção varia de acordo com diferentes agrupamentos das variáveis extraídas do sistema de registros.

4.4.1 Análise exploratória

4.4.1.1 Distribuição temporal da proporção de não-comparecimentos

Nas Figuras 64 e 65 são mostrados, respectivamente, o comportamento da proporção de não-comparecimentos e o total de instâncias considerando os diferentes anos e meses do ano. Mais especificamente, na Figura 64 vemos que existe uma proporção de não-comparecimentos menor do que a proporção geral para os anos de 2015, 2016, 2020, 2021, 2022 e 2023. Nesses anos, não parece haver um padrão perceptível de menores proporções de não-comparecimentos ao longo dos meses, estando bem distribuídas ao longo de cada ano. Por outro lado, houve maior proporção de não-comparecimentos nos anos de 2017, 2018 e 2019. Mais especificamente, nos anos de 2017 e 2018 não parece ter havido um padrão para as maiores proporções de não-comparecimento em termos dos meses, estando elas bem distribuídas ao longo de cada ano. Mas em 2019 (até mesmo em 2020) houve uma concentração das maiores proporções de não-comparecimento para os primeiros meses do ano (Janeiro até Abril). É interessante perceber que, a partir do ano de 2020, há uma melhora significativa e relevante na proporção de não-comparecimentos, que se manteve estável até 2023. O comportamento de crescimento da proporção de não-comparecimentos até 2018 (pico de 24.7%) e sua diminuição seguida de estabilidade até 2023 (proporção de 13.9%) sugere que existia um problema na gestão dos agendamentos que, a partir do ano de 2020 no mês de Maio, foi solucionado e culminou numa redução drástica e estável da proporção de não-comparecimentos. Analisando, agora, somente os meses do ano, temos que Janeiro, Março, Abril e Dezembro se destacam por possuírem proporções de não-comparecimentos maiores do que a proporção geral. Já nos meses de Junho, Julho e Setembro há menores proporções de não-comparecimentos.

Levando em consideração, agora, somente a Figura 65, é possível perceber que, ao longo dos anos, houve um aumento relevante no total de agendamentos entre os anos de 2016 e 2017 (quase 50%) que se manteve relativamente estável até o ano de 2021, onde houve outro aumento seguido de estabilidade. Isso pode estar relacionado a variações na oferta de consultas da clínica devido ao total de médicos disponíveis, por exemplo. Em relação aos meses do ano, os meses com maior quantidade de agendamentos foi Março, Maio, Agosto, Setembro e Outubro.

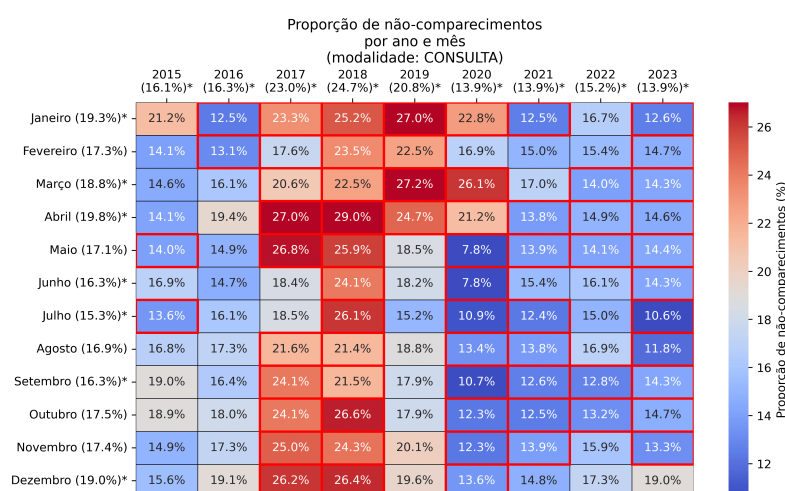


Figura 64 – Distribuição da proporção de não-comparecimentos por ano e mês para a modalidade de consulta da unidade de Jundiaí. São também mostradas as proporções totais de não-comparecimentos separadamente para cada mês e cada ano (distribuições marginais). O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos, bem como o destaque em vermelho em torno das células mês-ano. **Fonte:** próprio autor.

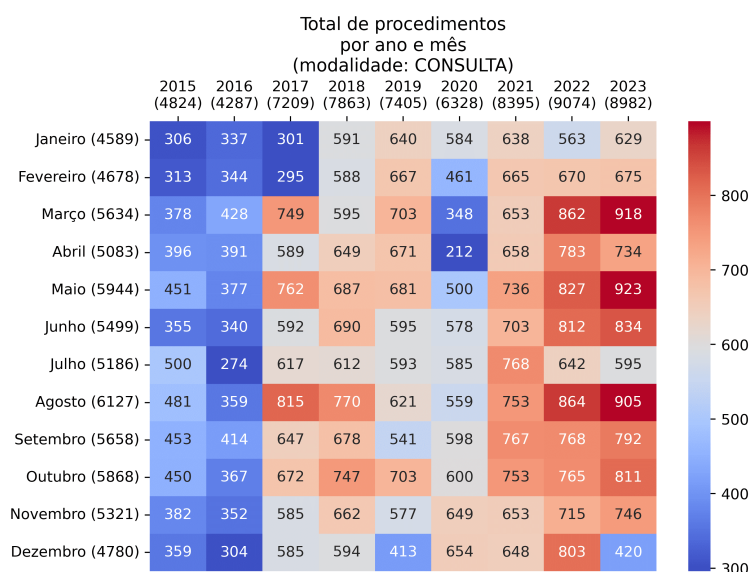


Figura 65 – Distribuição do total de procedimentos por ano e mês para a modalidade de consulta da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para cada mês e cada ano (distribuições marginais). **Fonte:** próprio autor.

Analisando a proporção de não-comparecimentos e de total de procedimentos agendados para os dias da semana e horários do dia, tem-se as Figuras 66 e 67. Para a modalidade de

consultas, nota-se da Figura 67 que os horários preferidos para o agendamento de consultas é das 7:00hs até as 11:00hs e das 13:00hs até as 16:00hs. Nos dias de segunda-feira, sexta-feira e sábado, há um volume consideravelmente menor de agendamentos. Existem, inclusive, consultas agendadas em horários estranhos, como 00:00hs, 3:00hs e 4:00hs, o que obviamente não faz sentido e se refere a erros no conjunto de dados. Não há risco de esses erros interferirem nos modelos de ML pois na etapa de pré-processamento todos grupos com total de instâncias significativamente menor foram agrupadas, de modo a não introduzirem nenhum viés. Na Figura 66, não há um padrão perceptível ou simples para as menores ou maiores proporções em termos da interação entre os dias da semana e os horários do dia, sendo útil somente a análise dos efeitos isolados dos dias da semana e dos horários do dia separadamente.

De forma isolada, os dias de segunda-feira e sábado se destacaram por apresentarem maiores proporções de não-comparecimento. Já nos dias de quinta-feira e sexta-feira houve menores proporções de não-comparecimento. Em termos de horários do dia, os horários das 7:00hs até as 9:00hs, 13:00hs e 14:00hs apresentaram maiores proporções de não-comparecimento; enquanto que, nos horários de 10:00hs até as 12:00hs e das 15:00hs até as 17:00hs, houve menores proporções de não-comparecimento.

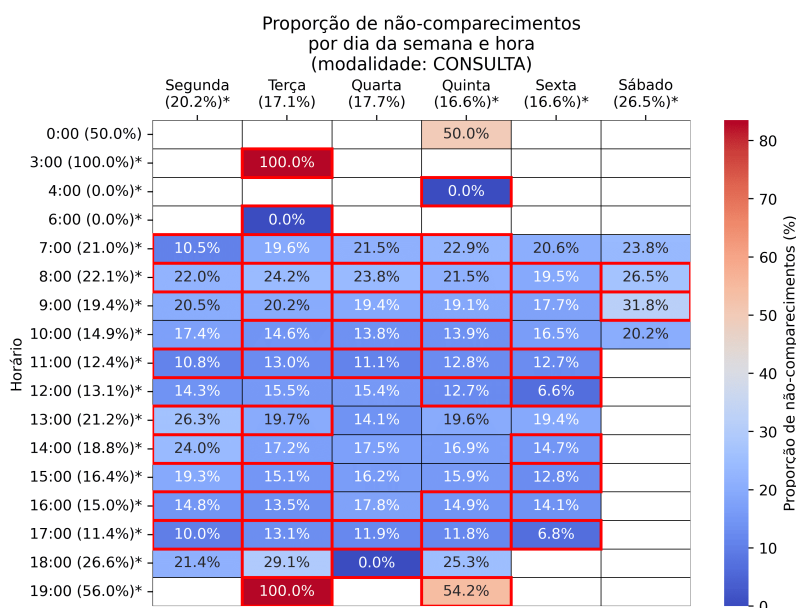


Figura 66 – Distribuição da proporção de não-comparecimentos por dia da semana e horário para a modalidade de consulta da unidade de Jundiaí. São também mostradas as proporções totais de não-comparecimentos separadamente para cada dia da semana e cada horário (distribuições marginais). O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos, bem como o destaque em vermelho em torno das células dia-horário. **Fonte:** próprio autor.

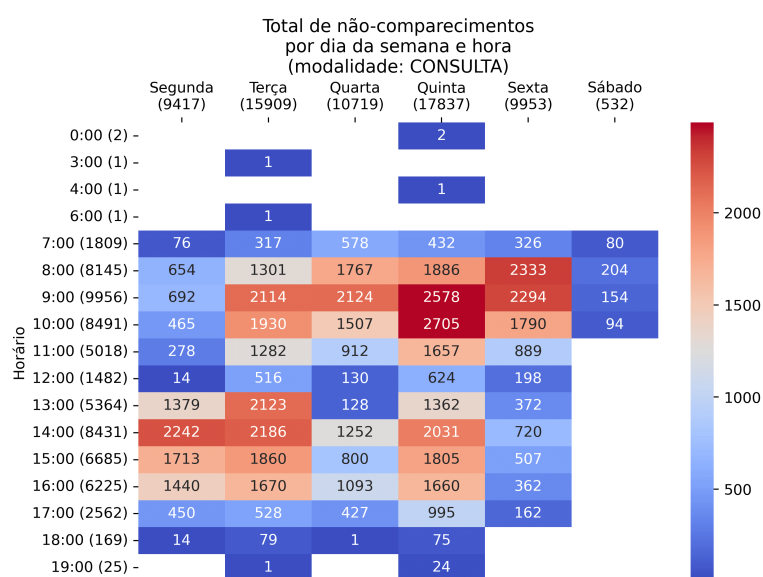


Figura 67 – Distribuição do total de procedimentos por dia da semana e horário para a modalidade de consulta da unidade de Jundiá. São também mostrados o total de procedimentos separadamente para cada dia da semana e cada horário (distribuições marginais).
Fonte: próprio autor.

4.4.1.2 Características dos pacientes e proporção de não-comparecimentos

Apesar de as Figuras 68, 69, 70, 71, 72, 73 e 74 representarem efeitos isolados (isto é, sem levar em conta interações) de diferentes características dos pacientes sobre a proporção de não-comparecimento na modalidade de consultas, é possível esboçar, mesmo que de forma inicial, o "perfil" do paciente que tende a não comparecer.

Analisando as figuras desta subseção, tem-se que o "perfil" superficial do paciente que tende a não comparecer a procedimentos agendados em consultas é:

- Não apresenta a informação "sexo" preenchida no sistema de registros ou é do sexo feminino (note que a proporção de não-comparecimentos para sexo = "F" continua sendo menor que a proporção geral, mas é maior numericamente que a proporção para sexo = "M") - Figura 68;
- Não apresenta a informação "idade" preenchida no sistema de registros ou apresenta idade nos bins (0, 10], (20, 30] ou (30, 40] - 69;
- Não apresenta a informação de "endereço" no sistema de registros - Figura 70;
- Agendou o procedimento com pelo menos 14 dias de antecedência - Figura 71;

- Ou agendou pela primeira vez algum procedimento no CRA, ou apresenta histórico prévio de não-comparecimento de no mínimo 25% - Figura 72;
- Agendou o procedimento via UNIMED (levando-se em consideração somente os planos de saúde mais comuns no conjunto de dados) - Figuras 73 e 74;

Todas essas características, além de serem úteis para a construção superficial do "perfil" do paciente que não comparece a procedimentos de consultas, também servem como um *sanity check* ao analisar a importância das variáveis via *framework SHAP*, que será feita mais adiante.

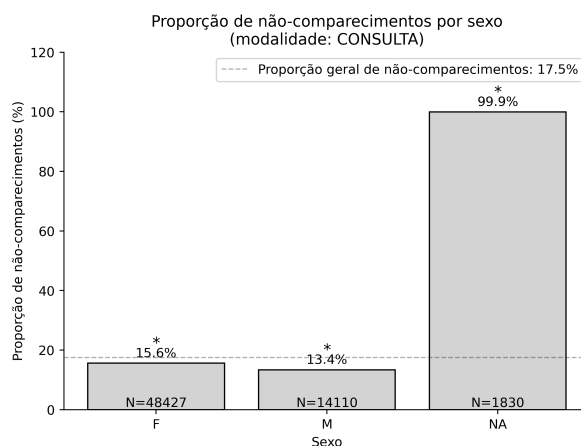


Figura 68 – Distribuição da proporção de não-comparecimentos por sexo dos pacientes para a modalidade de consulta da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para sexo. O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos. **Fonte:** próprio autor.

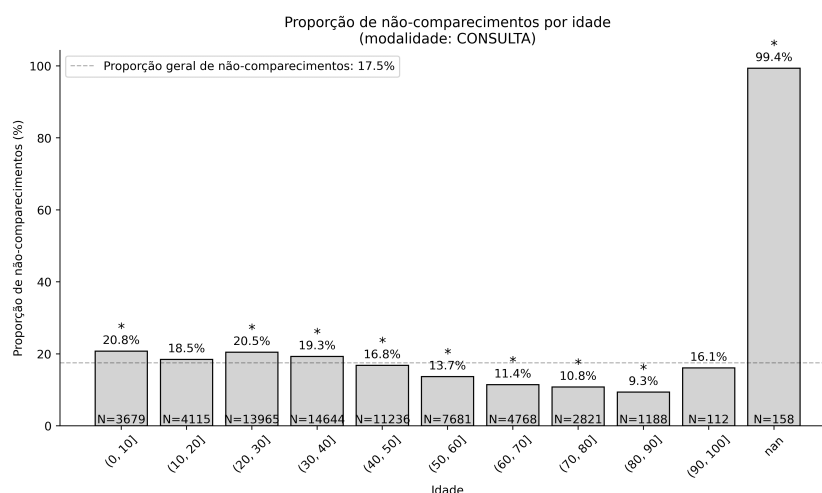


Figura 69 – Distribuição da proporção de não-comparecimentos por idade dos pacientes para a modalidade de consulta da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para cada *bin* de idade. O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos. **Fonte:** próprio autor.

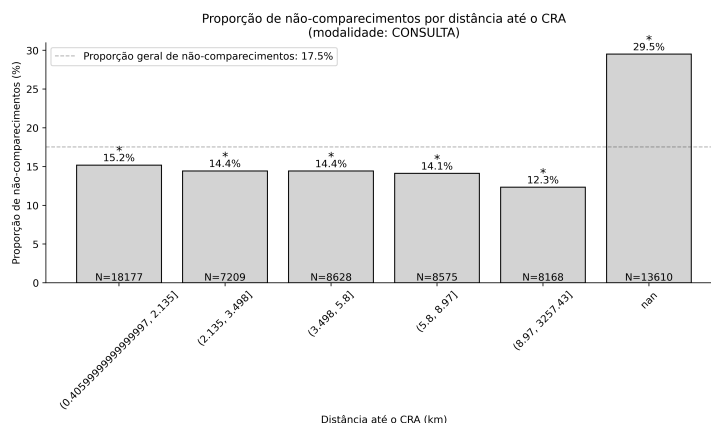


Figura 70 – Distribuição da proporção de não-comparecimentos por distância até o CRA a partir do endereço cadastrado dos pacientes para a modalidade de consulta da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para cada *bin* de distância. O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos. **Fonte:** próprio autor.

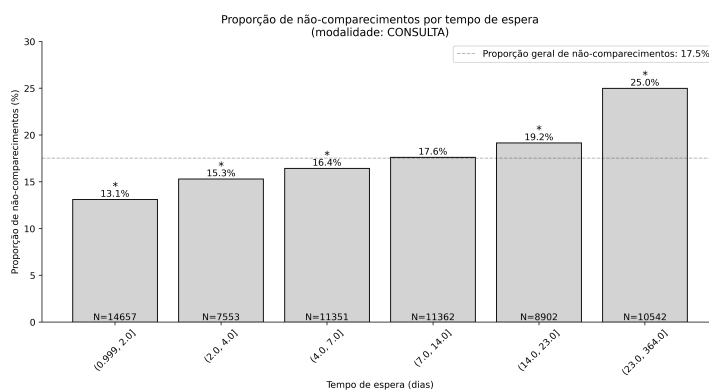


Figura 71 – Distribuição da proporção de não-comparecimentos por tempo de espera dos pacientes para a modalidade de consulta da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para cada bin de tempo de espera. O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos. **Fonte:** próprio autor.

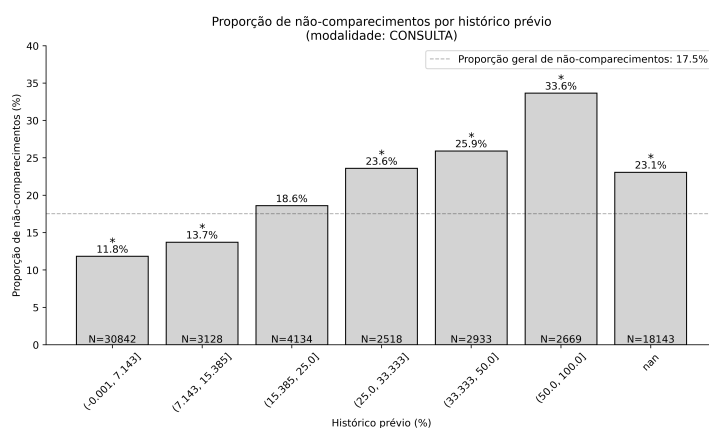


Figura 72 – Distribuição da proporção de não-comparecimentos por histórico prévio dos pacientes para a modalidade de consulta da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para cada bin de histórico prévio. O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos. **Fonte:** próprio autor.

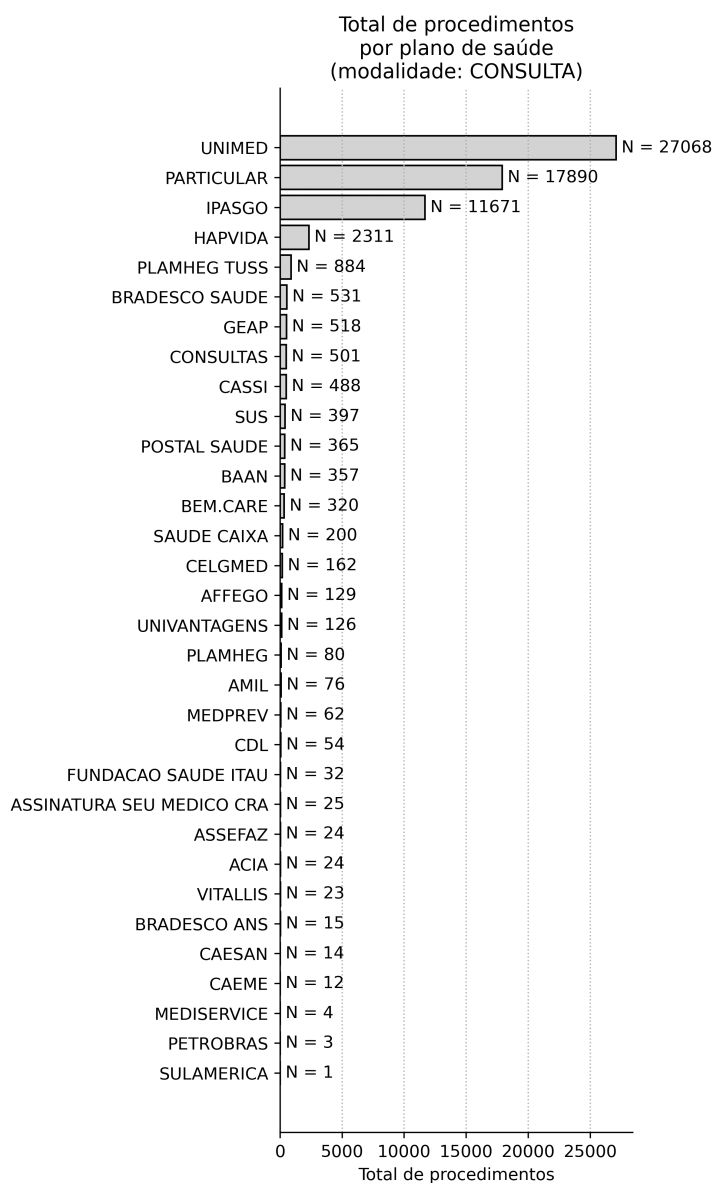


Figura 73 – Distribuição do total de procedimentos por plano de saúde dos pacientes para a modalidade de consulta da unidade de Jundiaí. **Fonte:** próprio autor.

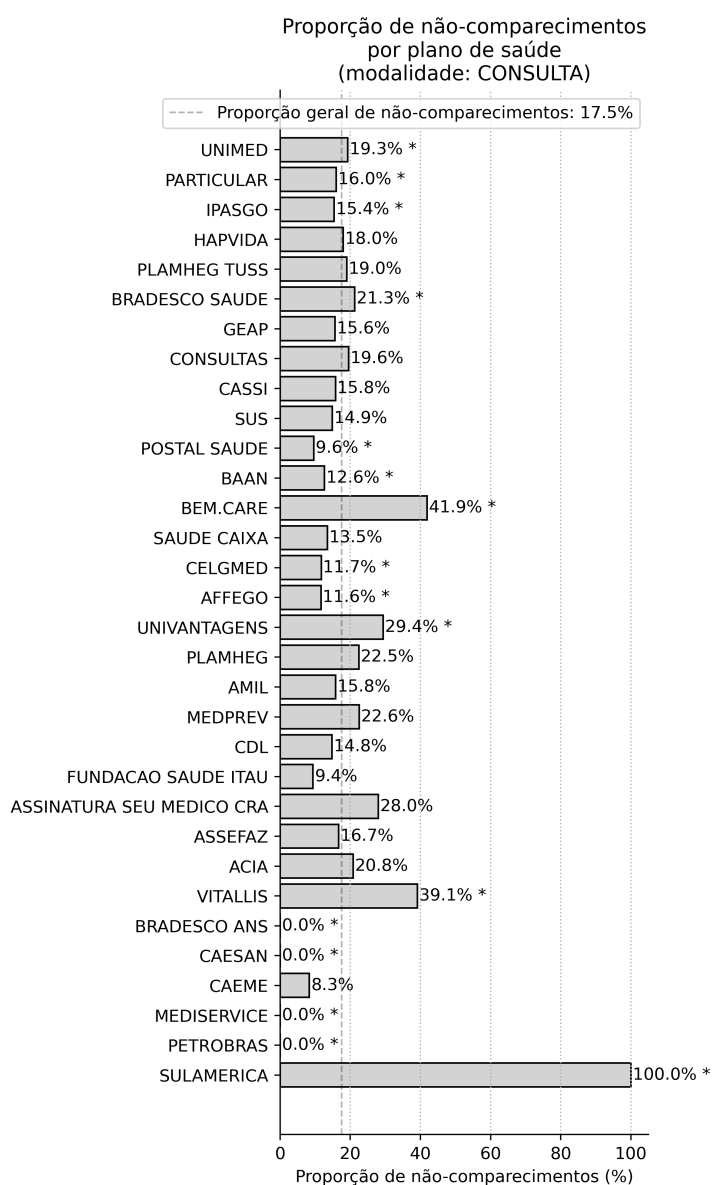


Figura 74 – Distribuição da proporção de não-comparecimentos por plano de saúde dos pacientes para a modalidade de consulta da unidade de Jundiaí. O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos. **Fonte:** próprio autor.

4.4.2 Tunagem de hiperparâmetros e seleção do modelo de classificação

Utilizando os dados de 2015 a 2022, foi realizada a etapa de tunagem de hiperparâmetros e o treinamento dos modelos LR, MLP, XGB, CB e LGBM. Para selecionar o melhor modelo (com a melhor combinação de hiperparâmetros) foi utilizado como critério a mé-

dia ponderada de AUC-PR obtido para cada *fold* de teste proveniente da Validação Cruzada com Divisão Temporal. A Figura 75 mostra a performance geral dos 5 modelos de classificação testados, para diferentes combinações de hiperparâmetros. As distribuições de valores mostradas na figura por meio de *boxplots* estão melhor detalhadas na Tabela 7. Nota-se por meio da figura e da tabela que os modelos que tendem a apresentar melhor performance para diferentes combinações de hiperparâmetros foram CB e LGBM. Já os modelos com maior variabilidade para diferentes combinações de hiperparâmetros foram LGBM e MLP. A Tabela 7 indica que o melhor modelo e combinação de hiperparâmetros foi o CB, com média ponderada máxima de AUC-PR igual a 0.3869. A melhor combinação de hiperparâmetros correspondente foi: {*iterations*: 259, *depth*: 6, *learning_rate*: 0.04855172569893545, *l2_leaf_reg*: 10.0, *bagging_temperature*: 1.81729137527627, *rsm*: 0.8105239589044421, *border_count*: 64, *random_strength*: 0.8116349602477629, *scale_pos_weight*: 1.0}.

Para a melhor combinação de hiperparâmetros do modelo CB, os resultados em cada *fold* de teste proveniente da Validação Cruzada com Divisão Temporal estão detalhados na Figura 76, inclusive para outras combinações de hiperparâmetros. Sobre a Figura 76, é esperado que exista uma tendência de melhora na performance de acordo com os *folds* de teste, pois o modelo faz uso de uma maior quantidade de dados de treino. Apesar disso, esse padrão não acontece na figura. A performance começa com valores altos, decresce entre os *folds* 4 e 6 e depois cresce novamente, mas de forma mais lenta. Nesse caso, pode ter existido uma diferença considerável no comportamento dos pacientes, que culminou em performances variadas do modelo de classificação.

Usando o modelo CB com a melhor combinação de hiperparâmetros, foi feito o ajuste de limiar de probabilidade do modelo utilizando os mesmos *folds* de teste obtidos por Validação Cruzada com Divisão Temporal. O critério para seleção do limiar foi de melhor média ponderada para a métrica Custo, conforme definida na Tabela 2. Os resultados estão mostrados na Figura 77, e o limiar que minimizou a média ponderada da métrica custo foi 0.26.

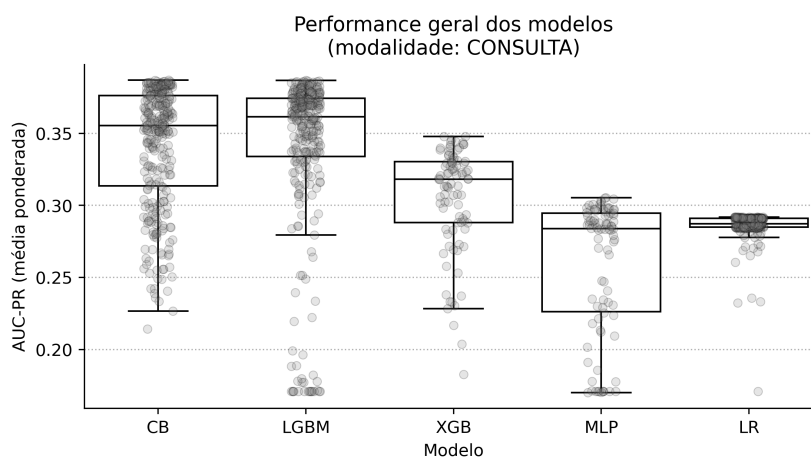


Figura 75 – Performance geral dos modelos de classificação testados para a modalidade de consulta da unidade de Jundiá. O modelo com a melhor combinação de hiperparâmetros em termos de média ponderada de AUC-PR foi selecionado. **Fonte:** próprio autor.

Tabela 7 – Informações adicionais sobre a performance geral dos modelos de classificação testados para a modalidade de consultas da unidade de Jundiá. O modelo com maior média ponderada de AUC-PR foi selecionado (coluna "Máx." da tabela).

Modelo	Mín.	Mediana	Média	Máx.	Desv. Pad.
CB	0.2143	0.3554	0.3415	0.3869	0.0418
LGBM	0.1708	0.3615	0.3395	0.3867	0.0575
XGB	0.1828	0.3181	0.3051	0.3479	0.0361
LR	0.1708	0.2871	0.2864	0.2919	0.0096
MLP	0.1700	0.2838	0.2589	0.3053	0.0463

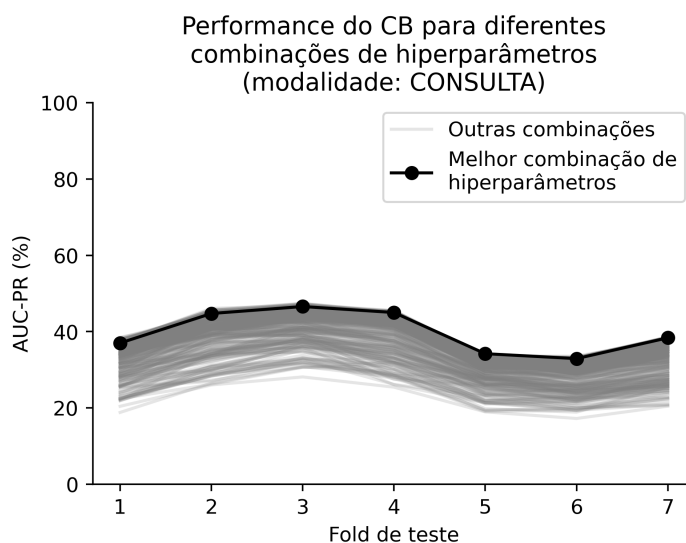


Figura 76 – Performance para diferentes combinações de hiperparâmetros para o modelo de classificação selecionado. **Fonte:** próprio autor.

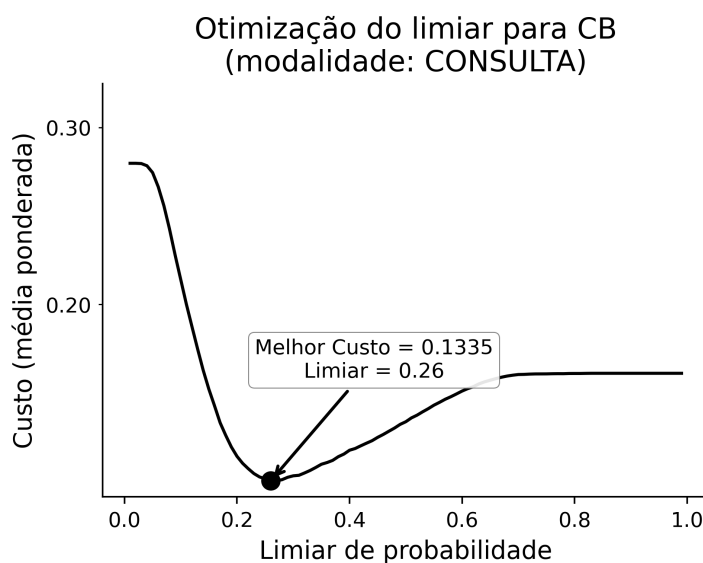


Figura 77 – Ajuste do limiar de probabilidade com base na média ponderada da métrica personalizada Custo para o modelo de classificação selecionado. **Fonte:** próprio autor.

4.4.3 Resultados sobre os dados de teste

Com o treino do modelo selecionado (ajustado com a melhor combinação de hiperparâmetros) sobre os dados entre os anos de 2015 e 2022, ele foi aplicado sobre os dados de teste (ano de 2023) para avaliação de performance. Na Figura 78, é mostrada a matriz de confusão do modelo sobre os dados de teste, a partir da qual, para a classe positiva (paciente não compa-

rece), foram calculadas as métricas de Precisão = 39.17%, Sensibilidade = 37.07% e F1-score = 38.09%. Tendo em vista que a proporção de comparecimento dos pacientes é de 86.09% nos dados de teste e a Acurácia do modelo foi de 83.24%, isso implica que o modelo está acertando menos do que um modelo que prediz somente a classe positiva, o que é um indicativo ruim de performance, conforme sugerido em (CARRERAS-GARCÍA et al., 2020). Apesar disso, o modelo de ML conseguiu reduzir a métrica Custo de 0.13905 (assumindo um cenário onde é esperado que todos os pacientes compareçam) para 0.11419 (cenário com o uso de ML para prever não-comparecimentos), o que representa uma redução na métrica Custo de 17.8%, o que deixa claro a vantagem de se utilizar o modelo de ML neste contexto. Em relação às curvas de performance, a curva ROC (AUC = 0.7282) e Precision-Recall (AUC = 0.3719) estão mostradas nas Figuras 79 e 80, respectivamente. Ambas as curvas indicam que o modelo de ML é melhor do que um classificador aleatório.

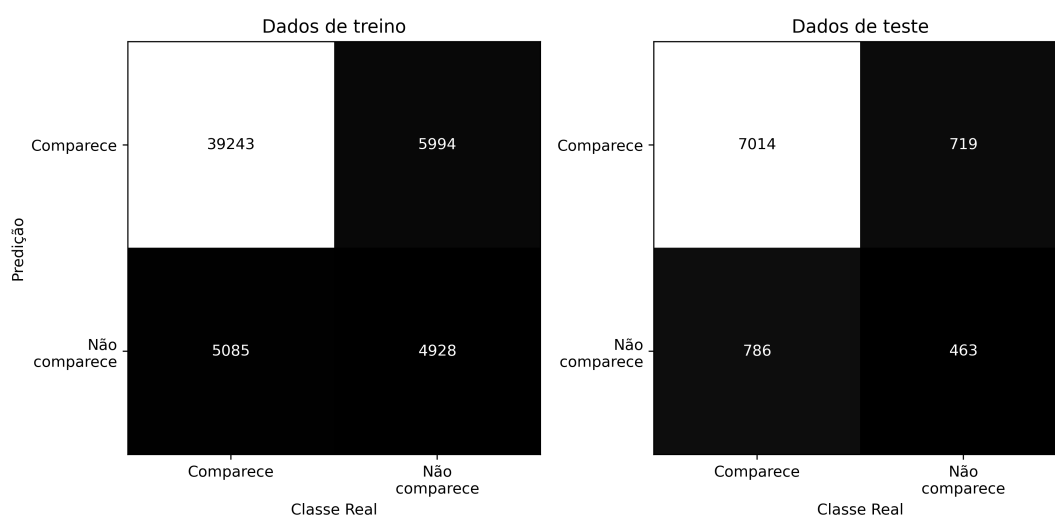


Figura 78 – Matrizes de confusão resultantes da aplicação do modelo selecionado sobre os dados de treino e teste. **Fonte:** próprio autor.

Tabela 8 – Métricas de performance do modelo selecionado conforme calculadas a partir das matrizes de confusão.

Dados	Classe	Precisão (%)	Sensibilidade (%)	F1-score (%)	N
Treino	Comparece	88.53	86.75	87.63	45237
	Não comparece	45.12	49.22	47.08	10013
	Acurácia (Treino)	79.95% (total de 55250)			
Teste	Comparece	89.92	90.70	90.31	7733
	Não comparece	39.17	37.07	38.09	1249
	Acurácia (Teste)	83.24% (total de 8982)			

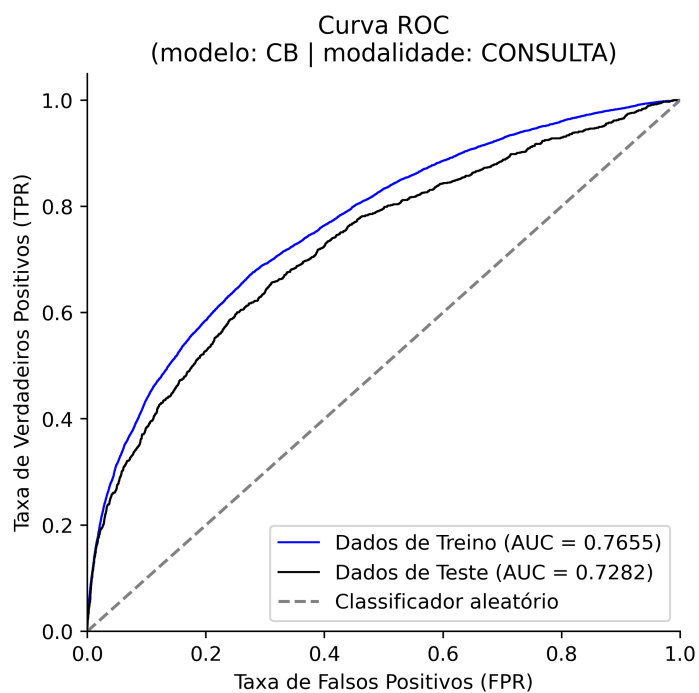


Figura 79 – Curva ROC resultante da aplicação do modelo selecionado sobre os dados de treino e teste. **Fonte:** próprio autor.

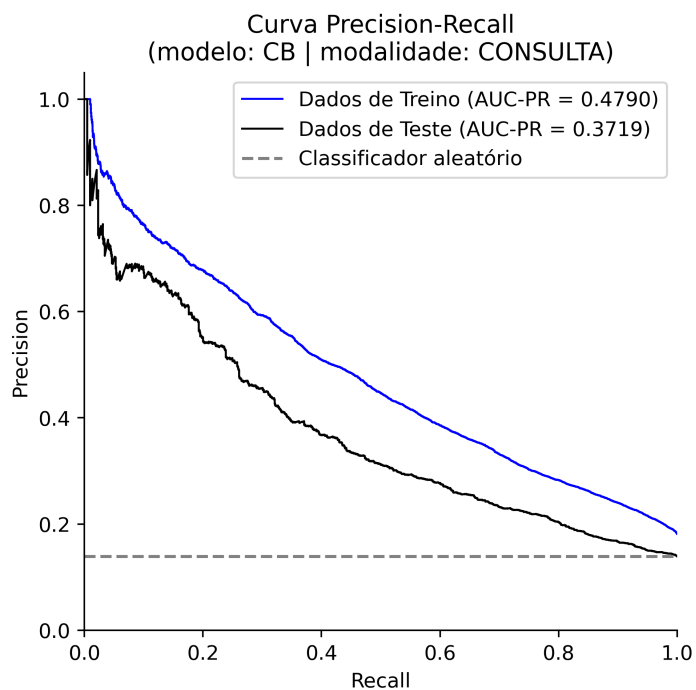


Figura 80 – Curva PR resultante da aplicação do modelo selecionado sobre os dados de treino e teste. **Fonte:** próprio autor.

4.4.4 Importância das variáveis

Aplicando o *framework SHAP* sobre as predições do modelo selecionado nos dados de teste para a modalidade de consultas, tem-se a Figura 81, que mostra as 20 variáveis preditoras que mais contribuem. Pela Figura, é possível perceber que as variáveis mais importantes nesse contexto são tipo de consulta, tempo de espera do paciente, distância até a clínica, histórico de não-comparecimento do paciente, idade do paciente, variável indicadora de se é a primeira vez do paciente na clínica, sexo do paciente e três variáveis decorrentes do PCA envolvendo taxas passadas de não-comparecimento na clínica. Todas essas variáveis (exceto as envolvendo PCA) tem sido destacadas em estudos anteriores como sendo relevantes para a predição de não-comparecimentos (DANTAS et al., 2018; CARRERAS-GARCÍA et al., 2020), e isso sugere que o modelo treinado apreendeu padrões importantes nesse contexto. Chama-se a atenção para as variáveis resultantes do PCA aparecendo no top 20 (e inclusive no top 10), indicando que, mesmo originalmente servindo para séries temporais, apresentaram contribuição relevante no problema de predição de não-comparecimentos.

Note, entretanto, que a Figura 81 não indica a direcionalidade da contribuição das variáveis, de modo que é preciso abstrai-las parcialmente através da seção de exploração dos dados, que inclui análises univariadas. Ainda assim, devido às interações complexas que existem entre as variáveis, a direcionalidade de contribuição pode não se manter para todas as instâncias de predição, sendo necessário uma alternativa para compreender cada predição. Nesse sentido, a Figura 82 serve como uma ilustração de como o *framework SHAP* pode ser útil nesse sentido, podendo servir como uma ferramenta para interpretar as predições individualmente, para ambas as classes estudadas.

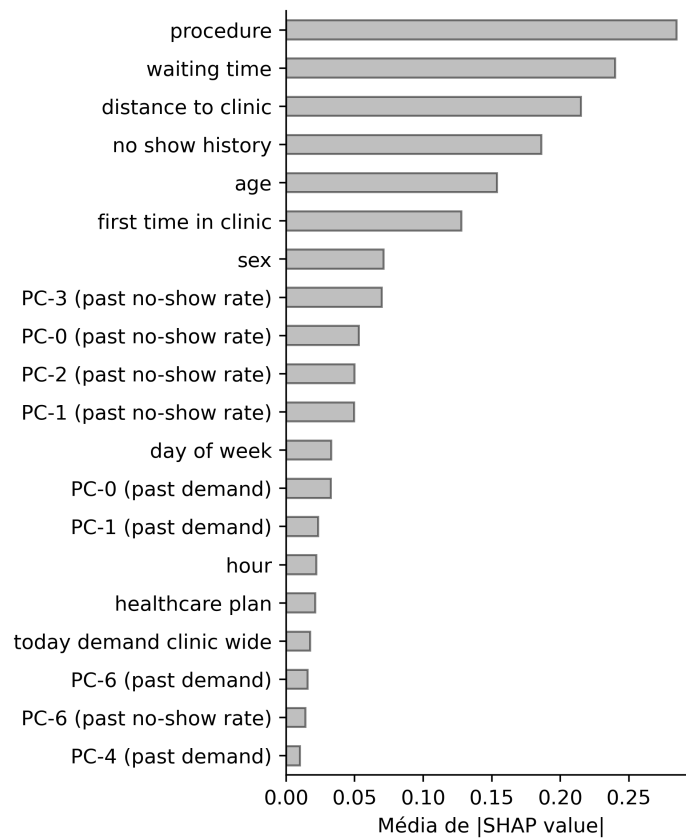


Figura 81 – Média dos valores SHAP absolutos resultantes da aplicação do modelo selecionado sobre os dados de teste. **Fonte:** próprio autor.

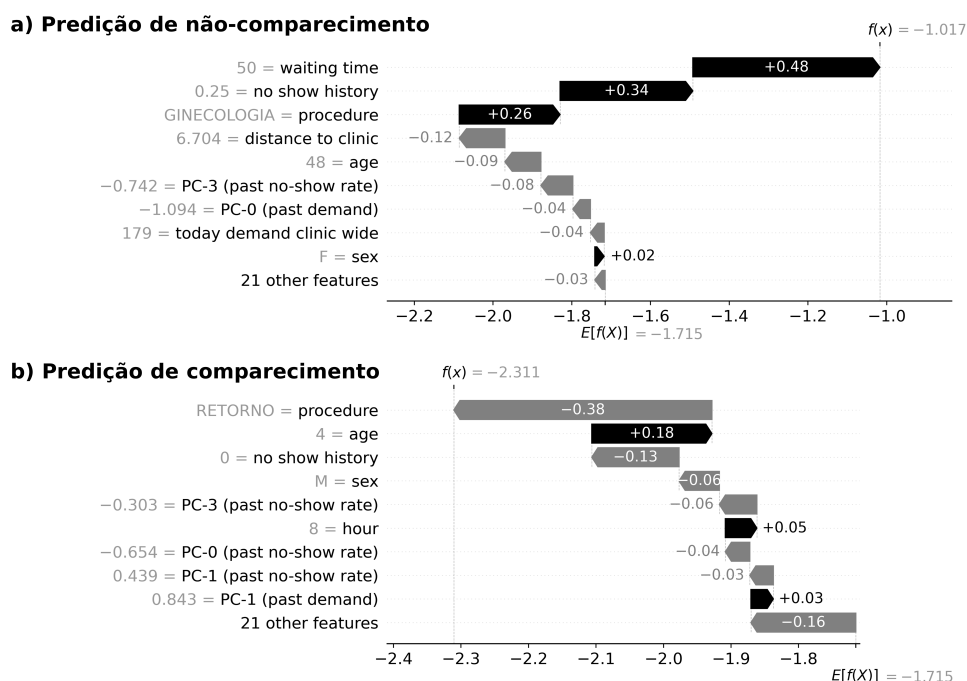


Figura 82 – Exemplo de predições individuais feitas pelo modelo selecionado em instâncias do conjunto de teste, com seus respectivos valores de contribuição das variáveis. O eixo horizontal representa as contribuições em log-odds. As setas pretas indicam contribuições positivas para a probabilidade de não-comparecimento, enquanto as setas cinzas indicam contribuições negativas. O valor base $E[f(x)]$ corresponde à saída média do modelo e serve como ponto de referência. A saída final do modelo $f(x)$ reflete a contribuição cumulativa de todos os valores de atributos para a instância específica. **Fonte:** próprio autor.

4.5 Modalidade de Ultrassom (unidade de Jundiaí)

Para a modalidade de US na unidade de Jundiaí do CRA, houve um total de 91438 instâncias devidamente pré-processadas, datando entre 2015 e 2023, dentre as quais 10498 houve não-comparecimento de pacientes, o que resulta em uma proporção total de não-comparecimentos próxima de 11.5%. Logo abaixo, é feita uma análise de como essa proporção varia de acordo com diferentes agrupamentos das variáveis extraídas do sistema de registros.

4.5.1 Análise exploratória

4.5.1.1 Distribuição temporal da proporção de não-comparecimentos

Nas Figuras 83 e 84 são mostrados, respectivamente, o comportamento da proporção de não-comparecimentos e o total de instâncias considerando os diferentes anos e meses do ano. Mais especificamente, na Figura 83 vemos que existe uma proporção de não-comparecimentos menor do que a proporção geral para os anos de 2016, 2018, 2019 e 2020. Em relação aos anos de 2018 a 2020 (que foram os anos com menores proporções de não-comparecimento, apresentando valores próximos a 9%), é interessante perceber que os meses de Agosto a Novembro apresentaram menores proporções de não-comparecimento. Fazendo a correspondência com a Figura 84, vemos que nesses meses e anos há também uma grande quantidade de agendamentos. Assim, é possível que tenha havido alguma abordagem em relação aos não-comparecimentos dos pacientes, passível de ser replicada no futuro. Entretanto, a partir de 2021, o padrão não se repetiu. Aliás, houve maior proporção de não-comparecimentos nos anos de 2015, 2021, 2022 e 2023. Mais especificamente, os valores de proporção de não-comparecimentos estabilizaram em torno de 13.5% a partir do ano de 2021, sugerindo margem para melhora. Desde 2021, os meses Março, Abril, Maio e Novembro têm apresentado proporções altas de não-comparecimento. Isso é algo que chama a atenção e é pertinente para análises futuras envolvendo mecanismos acionáveis para lidar com o absentéismo dos pacientes. Analisando, agora, somente os meses do ano, temos que Fevereiro e Setembro se destacam por possuírem proporções menores de não-comparecimentos. Já nos meses de Março e Novembro há maiores proporções de não-comparecimentos.

Levando em consideração, agora, somente a Figura 84, é possível perceber que, ao longo dos anos, houve um aumento relevante no total de agendamentos entre os anos de 2015 e 2019 (cerca de 50%). Mas no ano de 2020 houve uma diminuição do total de agendamentos, que posteriormente retornou para os patamares de 2019 já no ano de 2021, mantendo-se estável em valores em torno de 12000 agendamentos. A queda no total de agendamentos em 2020 pode ser devido a alguma limitação envolve a quantidade de equipamentos, por exemplo, o que limitou a oferta. Levando em consideração os meses do ano, os meses com maiores quantidades de agendamentos foram Março, Agosto, Setembro, Outubro e Novembro. Os meses de Abril, Julho e Dezembro se destacam por apresentarem menores quantidades de agendamento no total. Curiosamente, foram agendados somente 12 procedimentos no mês de Abril do ano de 2020, com proporção de não-comparecimento de 100%. Esses números podem indicar que, por algum motivo, exames de US estiveram indisponíveis nesse período.

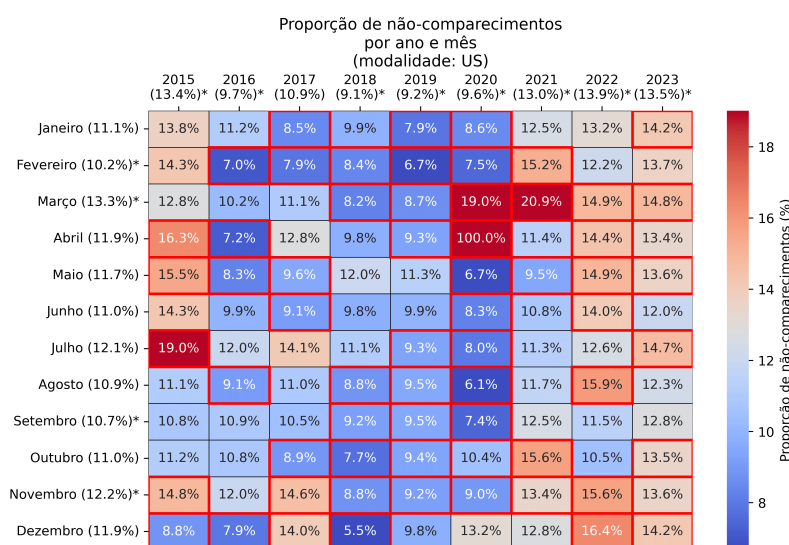


Figura 83 – Distribuição da proporção de não-comparecimentos por ano e mês para a modalidade de US da unidade de Jundiáí. São também mostradas as proporções totais de não-comparecimentos separadamente para cada mês e cada ano (distribuições marginais). O "*" indica uma diferença significativa ($p\text{-valor} < 0.05$) em relação à proporção geral de não-comparecimentos, bem como o destaque em vermelho em torno das células mês-ano. **Fonte:** próprio autor.

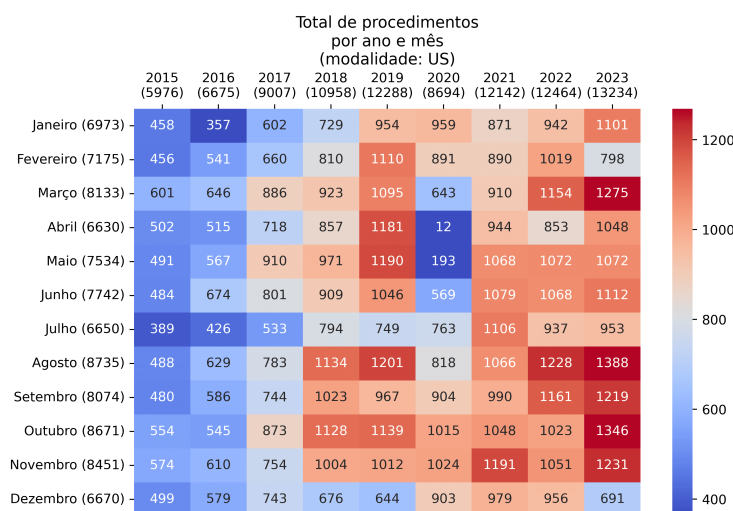


Figura 84 – Distribuição do total de procedimentos por ano e mês para a modalidade de US da unidade de Jundiáí. São também mostrados o total de procedimentos separadamente para cada mês e cada ano (distribuições marginais). **Fonte:** próprio autor.

Analisando a proporção de não-comparecimentos e de total de procedimentos agendados para os dias da semana e horários do dia, tem-se as Figuras 85 e 86. Para a modalidade de US, nota-se da Figura 86 que os horários preferidos para o agendamento de exames é das 7:00hs

até as 11:00hs e das 13:00hs até as 16:00hs, que é um padrão idêntico ao da modalidade de consultas, com um volume consideravelmente maior de agendamentos na terça-feira. Existem, inclusive, assim como na modalidade de consultas, exames agendados em horários estranhos, como 00:00hs, 01:00hs, 2:00hs e 5:00hs, o que obviamente não faz sentido e se refere a erros no conjunto de dados. Não há risco de esses erros interferirem nos modelos de ML pois na etapa de pré-processamento todos grupos com total de instâncias significativamente menor foram agrupadas, de modo a não introduzirem nenhum viés. Na Figura 85, percebe-se que há maior proporção de não-comparecimentos nos horários das 07:00hs até 11:00hs para o dia de segunda-feira, com o horário de 12:00hs (horário de almoço) apresentando menor proporção. Menores proporções também apareceram entre 13:00hs e 16:00hs para os dias entre terça-feira e quinta-feira, sugerindo um padrão. De forma avulsa, os horários de 10:00hs da quarta-feira e 08:00hs de sexta-feira também apresentaram menores proporções.

De forma isolada, os dias de segunda-feira e sábado se destacaram por apresentarem maiores proporções de não-comparecimento. Já nos dias de quarta-feira e quinta-feira houve menores proporções de não-comparecimento. Em termos de horários do dia, os horários das 7:00hs apresentou maior proporção de não-comparecimento; enquanto que, nos horários de 14:00hs até as 16:00hs, houve menores proporções de não-comparecimento.

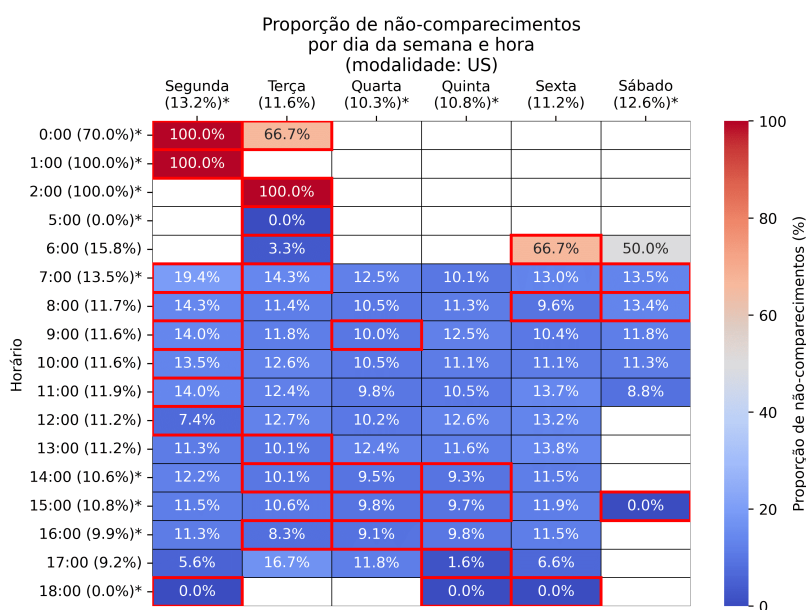


Figura 85 – Distribuição da proporção de não-comparecimentos por dia da semana e horário para a modalidade de US da unidade de Jundiá. São também mostradas as proporções totais de não-comparecimentos separadamente para cada dia da semana e cada horário (distribuições marginais). O "*" indica uma diferença significativa (p-valor < 0.05) em relação à proporção geral de não-comparecimentos, bem como o destaque em vermelho em torno das células dia-horário. **Fonte:** próprio autor.

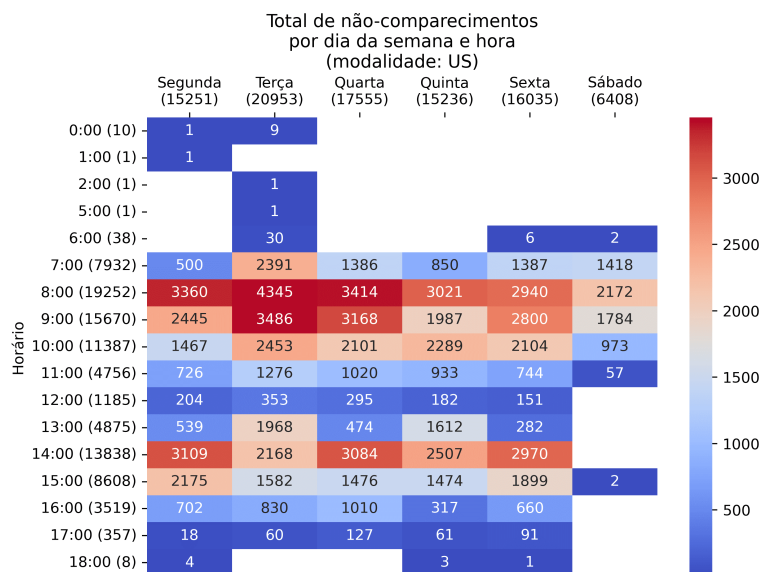


Figura 86 – Distribuição do total de procedimentos por dia da semana e horário para a modalidade de US da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para cada dia da semana e cada horário (distribuições marginais).
Fonte: próprio autor.

4.5.1.2 Características dos pacientes e proporção de não-comparecimentos

Apesar de as Figuras 87, 88, 89, 90, 91, 92 e 93 representarem efeitos isolados (isto é, sem levar em conta interações) de diferentes características dos pacientes sobre a proporção de não-comparecimento na modalidade de US, é possível esboçar, mesmo que de forma inicial, o "perfil" do paciente que tende a não comparecer.

Analisando as figuras desta subseção, tem-se que o "perfil" superficial do paciente que tende a não comparecer a procedimentos agendados em US é:

- Não apresenta a informação "sexo" preenchida no sistema de registros ou é do sexo feminino (note que a proporção de não-comparecimentos para sexo = "F" continua sendo menor que a proporção geral, mas é maior numericamente que a proporção para sexo = "M") - Figura 87;
- Não apresenta a informação "idade" preenchida no sistema de registros ou apresenta idade entre 0 e 40 anos - 88;
- Não apresenta a informação de "endereço" no sistema de registros - Figura 89;
- Agendou o procedimento com pelo menos 9 dias de antecedência - Figura 90;

- Ou agendou pela primeira vez algum procedimento no CRA, ou apresenta histórico prévio de não-comparecimento de no mínimo 22.2% - Figura 91;
- Agendou o procedimento via IPASGO ou PARTICULAR (levando-se em consideração somente os planos de saúde mais comuns no conjunto de dados) - Figuras 92 e 93;

Todas essas características, além de serem úteis para a construção superficial do "perfil" do paciente que não comparece a procedimentos de US, também servem como um *sanity check* ao analisar a importância das variáveis via *framework SHAP*, que será feita mais adiante.

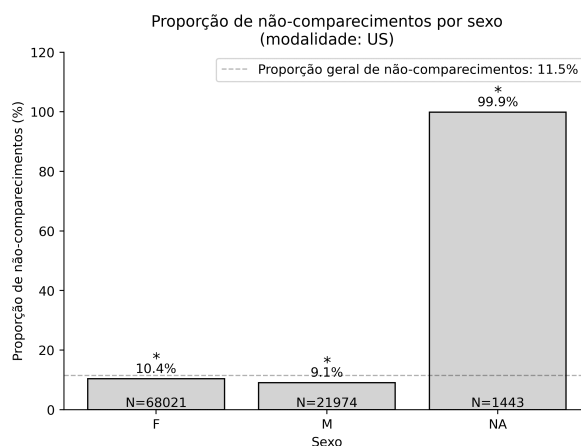


Figura 87 – Distribuição da proporção de não-comparecimentos por sexo dos pacientes para a modalidade de US da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para sexo. O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos. **Fonte:** próprio autor.

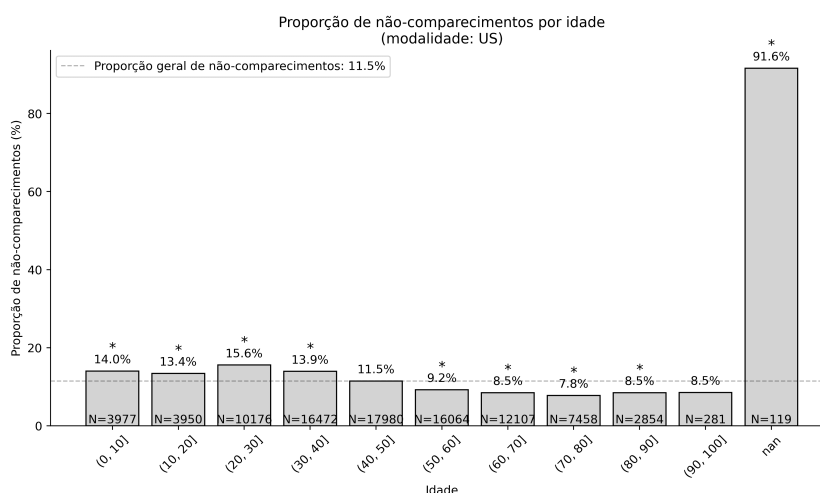


Figura 88 – Distribuição da proporção de não-comparecimentos por idade dos pacientes para a modalidade de US da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para cada *bin* de idade. O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos. **Fonte:** próprio autor.

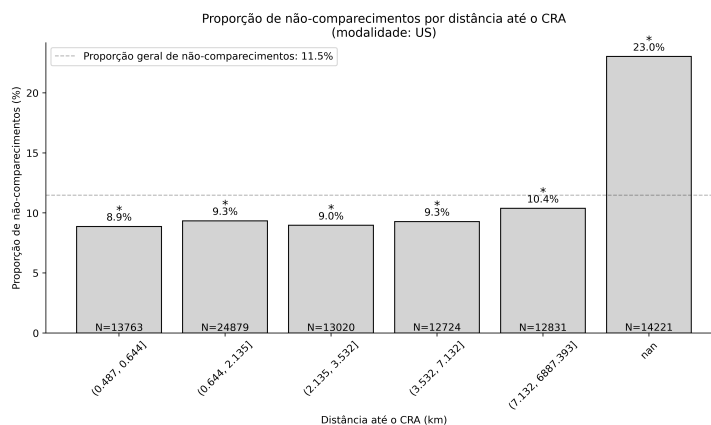


Figura 89 – Distribuição da proporção de não-comparecimentos por distância até o CRA a partir do endereço cadastrado dos pacientes para a modalidade de US da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para cada *bin* de distância. O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos. **Fonte:** próprio autor.

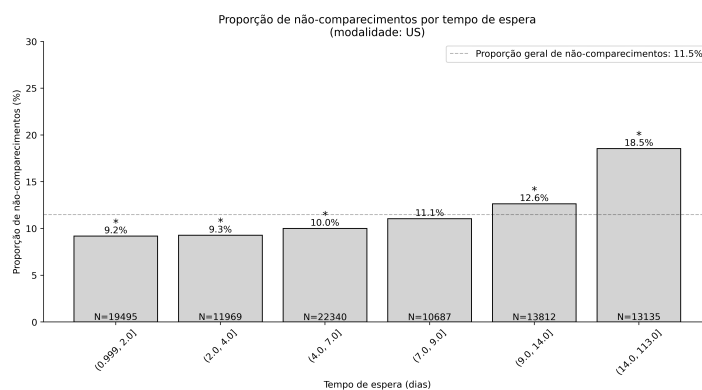


Figura 90 – Distribuição da proporção de não-comparecimentos por tempo de espera dos pacientes para a modalidade de US da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para cada bin de tempo de espera. O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos. **Fonte:** próprio autor.

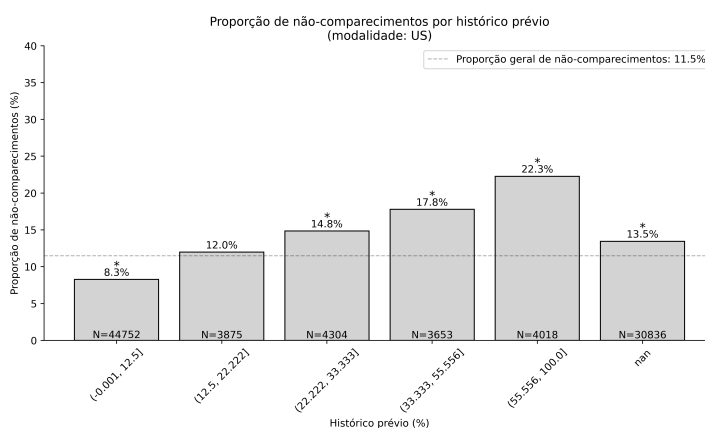


Figura 91 – Distribuição da proporção de não-comparecimentos por histórico prévio dos pacientes para a modalidade de US da unidade de Jundiaí. São também mostrados o total de procedimentos separadamente para cada bin de histórico prévio. O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos. **Fonte:** próprio autor.

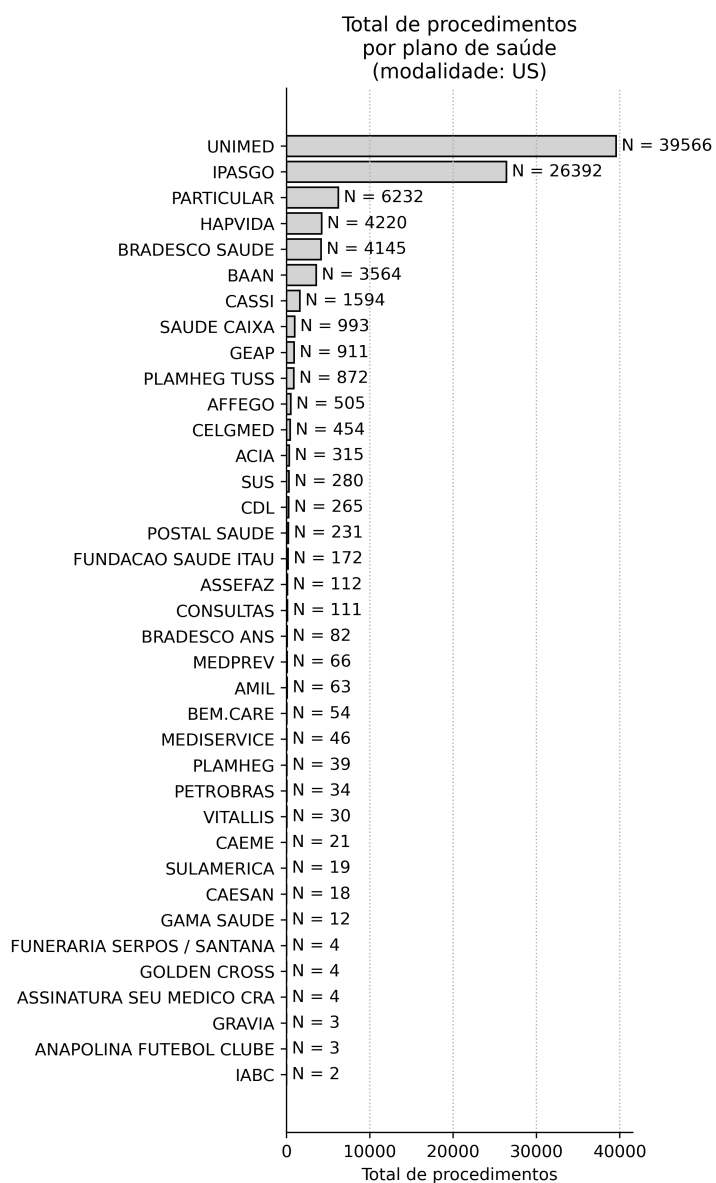


Figura 92 – Distribuição do total de procedimentos por plano de saúde dos pacientes para a modalidade de US da unidade de Jundiaí. **Fonte:** próprio autor.

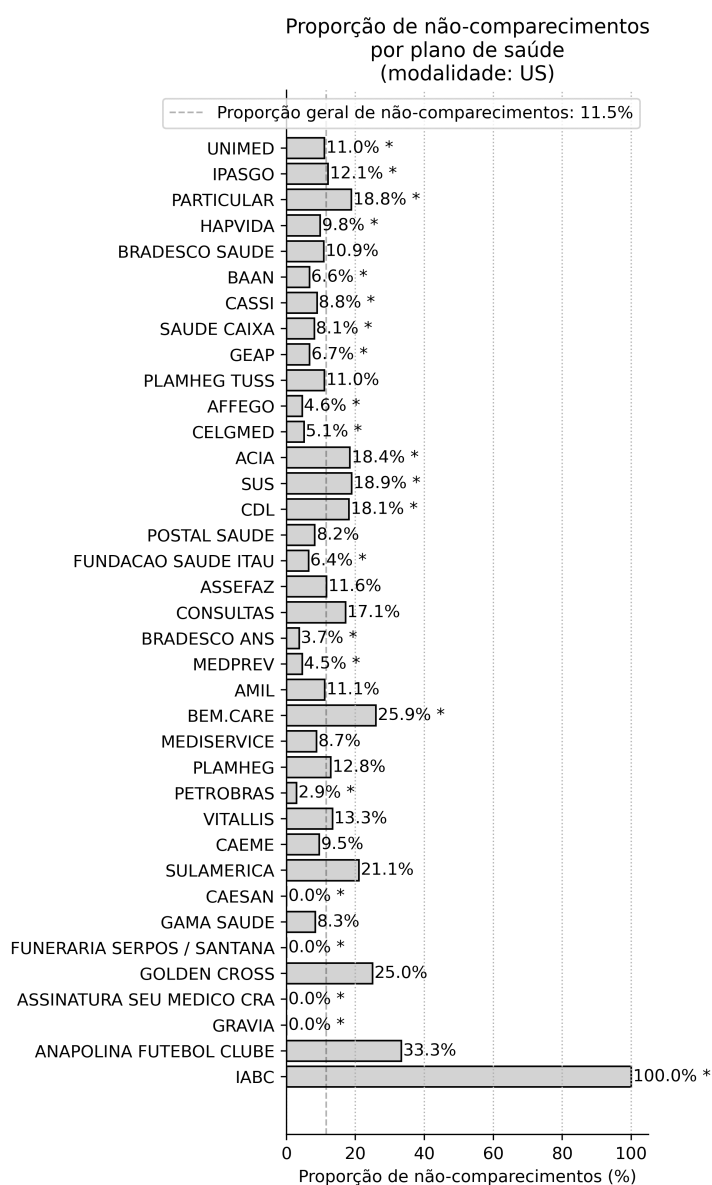


Figura 93 – Distribuição da proporção de não-comparecimentos por plano de saúde dos pacientes para a modalidade de US da unidade de Jundiaí. O "*" indica uma diferença significativa (p -valor < 0.05) em relação à proporção geral de não-comparecimentos. **Fonte:** próprio autor.

4.5.2 Tunagem de hiperparâmetros e seleção do modelo de classificação

Utilizando os dados de 2015 a 2022, foi realizada a etapa de tunagem de hiperparâmetros e o treinamento dos modelos LR, MLP, XGB, CB e LGBM. Para selecionar o melhor modelo (com a melhor combinação de hiperparâmetros) foi utilizado como critério a média ponderada

de AUC-PR obtido para cada *fold* de teste proveniente da Validação Cruzada com Divisão Temporal. A Figura 94 mostra a performance geral dos 5 modelos de classificação testados, para diferentes combinações de hiperparâmetros. As distribuições de valores mostradas na figura por meio de *boxplots* estão melhor detalhadas na Tabela 9. Nota-se por meio da figura e da tabela que os modelos que tendem a apresentar melhor performance para diferentes combinações de hiperparâmetros foram CB e LGBM. Já os modelos com maior variabilidade para diferentes combinações de hiperparâmetros foram MLP e LGBM. A Tabela 9, por sua vez, indica que o melhor modelo e combinação de hiperparâmetros foi o CB, com média ponderada máxima de AUC-PR igual a 0.2522. A melhor combinação de hiperparâmetros correspondente foi: {*iterations*: 300, *depth*: 7, *learning_rate*: 0.017035265463161607, *l2_leaf_reg*: 10.0, *bagging_temperature*: 1.8919330061496367, *rsm*: 1.0, *border_count*: 64, *random_strength*: 0.0, *scale_pos_weight*: 1.0}.

Para a melhor combinação de hiperparâmetros do modelo CB, os resultados em cada *fold* de teste proveniente da Validação Cruzada com Divisão Temporal estão detalhados na Figura 95, inclusive para outras combinações de hiperparâmetros. Sobre a Figura 95, é esperado que exista uma tendência de melhora na performance de acordo com os *folds* de teste, pois o modelo faz uso de uma maior quantidade de dados de treino. E isso ocorre na figura, diferentemente do que foi visto para os modelos das outras modalidades neste trabalho, sugerindo que o comportamento dos pacientes não mudou de forma abrupta ao longo do tempo.

Usando o modelo CB com a melhor combinação de hiperparâmetros, foi feito o ajuste de limiar de probabilidade do modelo utilizando os mesmos *folds* de teste obtidos por Validação Cruzada com Divisão Temporal. O critério para seleção do limiar foi de melhor média ponderada para a métrica Custo, conforme definida na Tabela 2. Os resultados estão mostrados na Figura 96, e o limiar que minimizou a média ponderada da métrica custo foi 0.20.

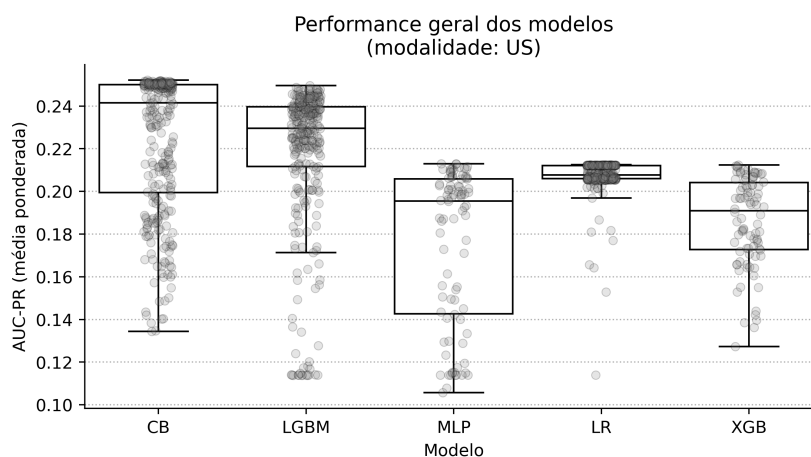


Figura 94 – Performance geral dos modelos de classificação testados para a modalidade de US da unidade de Jundiaí. O modelo com a melhor combinação de hiperparâmetros em termos de média ponderada de AUC-PR foi selecionado. **Fonte:** próprio autor.

Tabela 9 – Informações adicionais sobre a performance geral dos modelos de classificação testados para a modalidade de US da unidade de Jundiaí. O modelo com maior média ponderada de AUC-PR foi selecionado (coluna "Máx." da tabela).

Modelo	Mín.	Mediana	Média	Máx.	Desv. Pad.
CB	0.1345	0.2416	0.2239	0.2522	0.0324
LGBM	0.1139	0.2295	0.2177	0.2497	0.0344
LR	0.1139	0.2078	0.2076	0.2125	0.0085
XGB	0.1273	0.1910	0.1861	0.2124	0.0209
MLP	0.1057	0.1956	0.1752	0.2130	0.0365

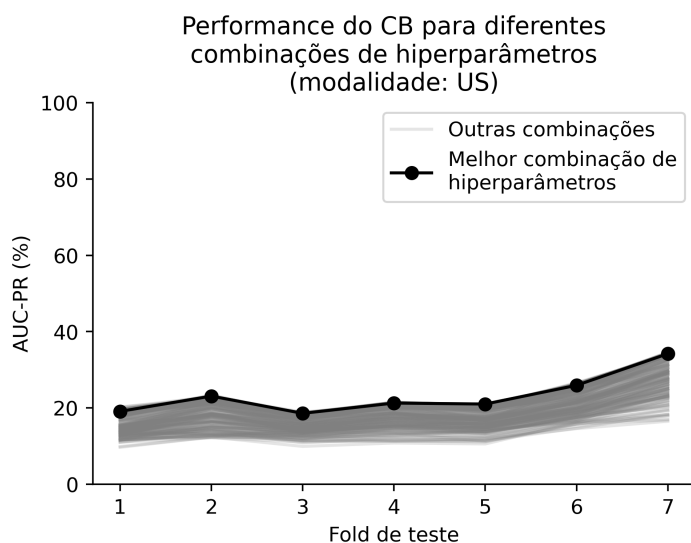


Figura 95 – Performance para diferentes combinações de hiperparâmetros para o modelo de classificação selecionado. **Fonte:** próprio autor.

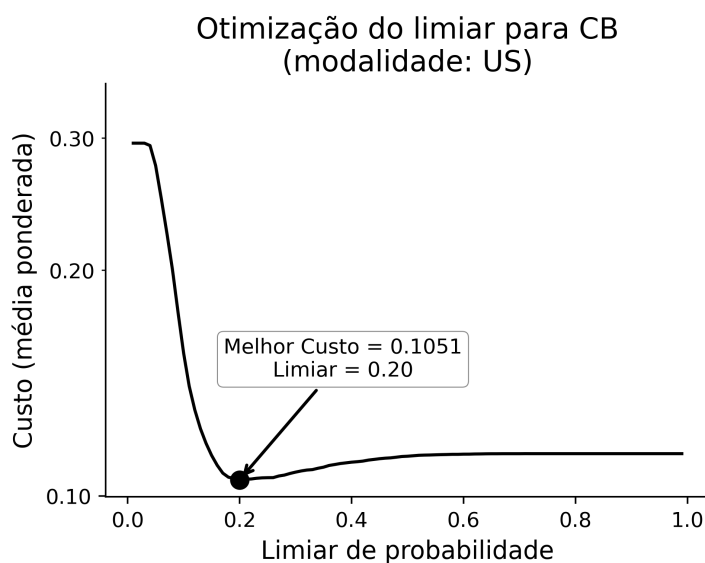


Figura 96 – Ajuste do limiar de probabilidade com base na média ponderada da métrica personalizada Custo para o modelo de classificação selecionado. **Fonte:** próprio autor.

4.5.3 Resultados sobre os dados de teste

Com o treino do modelo selecionado (ajustado com a melhor combinação de hiperparâmetros) sobre os dados entre os anos de 2015 e 2022, ele foi aplicado sobre os dados de teste (ano de 2023) para avaliação de performance. Na Figura 97, é mostrada a matriz de confusão do modelo sobre os dados de teste, a partir da qual, para a classe positiva (paciente não comparece),

foram calculadas as métricas de Precisão = 35.59%, Sensibilidade = 30.87% e F1-score = 33.06%. Tendo em vista que a proporção de comparecimento dos pacientes é de 86.49% nos dados de teste e a Acurácia do modelo foi de 83.14%, isso implica que o modelo está acertando menos do que um modelo que prediz somente a classe positiva, o que é um indicativo ruim em termos de performance, conforme sugerido em (CARRERAS-GARCÍA et al., 2020). Apesar disso, o modelo de ML conseguiu reduzir a métrica Custo de 0.13503 (assumindo um cenário onde é esperado que todos os pacientes compareçam) para 0.11839 (cenário com o uso de ML para prever não-comparecimentos), o que representa uma redução na métrica Custo de 12.2%. Em relação às curvas de performance, a curva ROC (AUC = 0.7131) e Precision-Recall (AUC = 0.3024) estão mostradas nas Figuras 98 e 99, respectivamente. Ambas as curvas indicam que o modelo de ML é melhor do que um classificador aleatório.

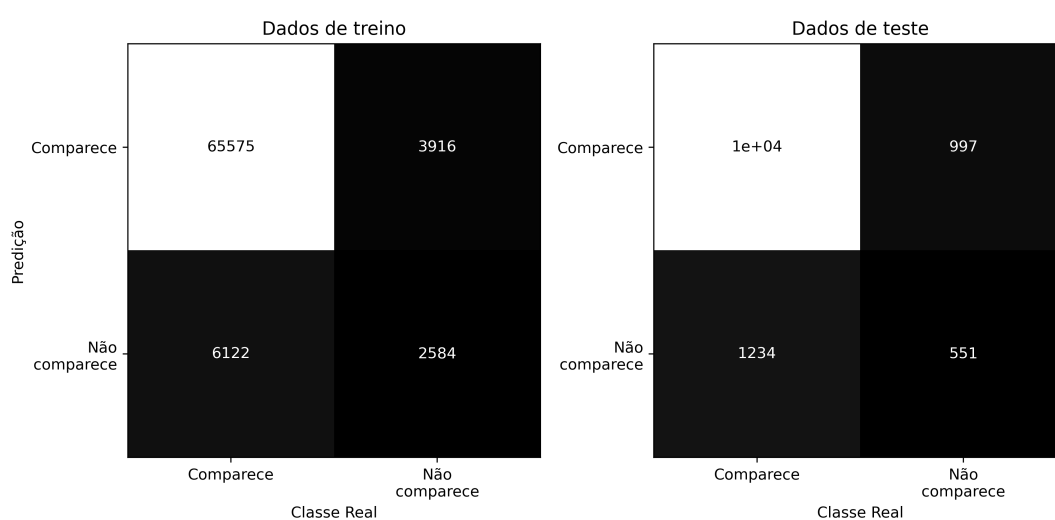


Figura 97 – Matrizes de confusão resultantes da aplicação do modelo selecionado sobre os dados de treino e teste. **Fonte:** próprio autor.

Tabela 10 – Métricas de performance do modelo selecionado conforme calculadas a partir das matrizes de confusão.

Dados	Classe	Precisão (%)	Sensibilidade (%)	F1-score (%)	N
Treino	Comparece	91.46	94.36	92.89	69491
	Não comparece	39.75	29.68	33.99	8706
	Acurácia (Treino)	87.16% (total de 78197)			
Teste	Comparece	89.44	91.29	90.35	11445
	Não comparece	35.59	30.87	33.06	1785
	Acurácia (Teste)	83.14% (total de 13230)			

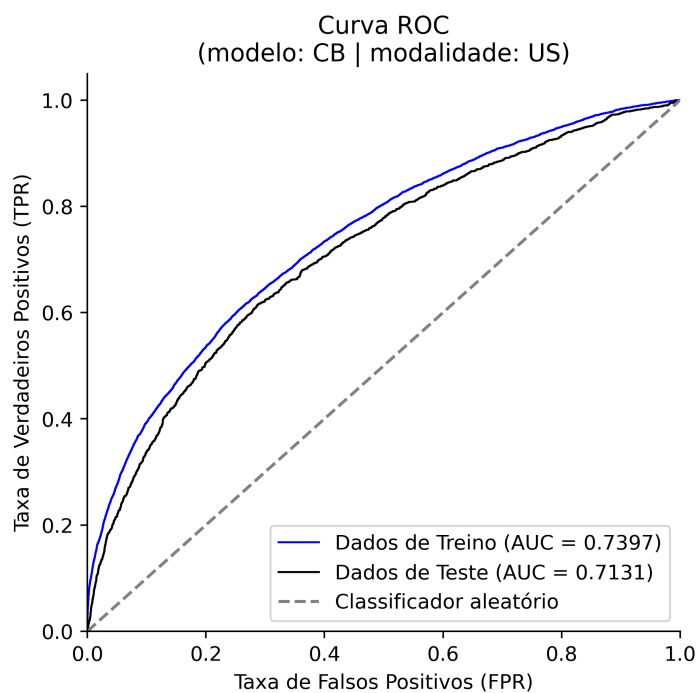


Figura 98 – Curva ROC resultante da aplicação do modelo selecionado sobre os dados de treino e teste. **Fonte:** próprio autor.

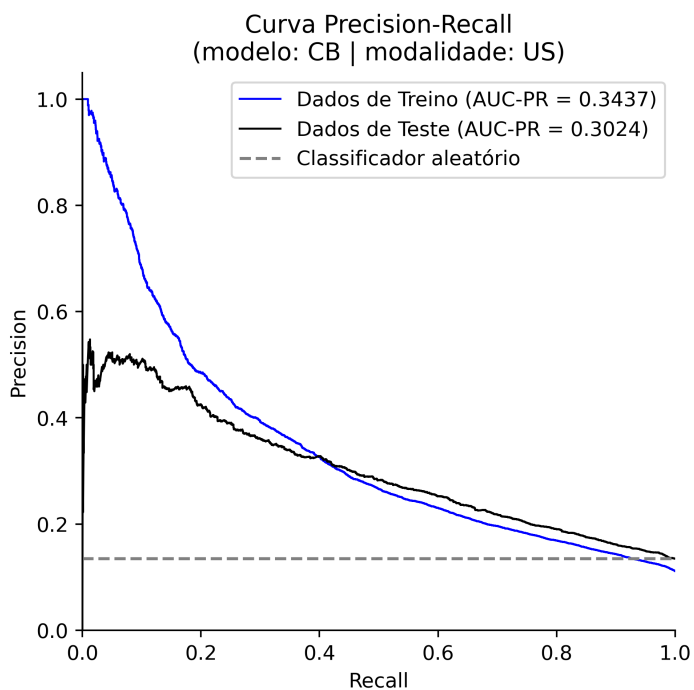


Figura 99 – Curva PR resultante da aplicação do modelo selecionado sobre os dados de treino e teste. **Fonte:** próprio autor.

4.5.4 Importância das variáveis

Aplicando o *framework SHAP* sobre as predições do modelo selecionado nos dados de teste para a modalidade de US, tem-se a Figura 100, que mostra as 20 variáveis preditoras que mais contribuem. Pela Figura, é possível perceber que as 10 variáveis mais importantes nesse contexto são idade do paciente, o histórico de não-comparecimento do paciente, o tempo de espera do paciente, distância até a clínica, variável indicadora de se é a primeira vez do paciente na clínica, o plano de saúde do paciente, o sexo do paciente e três variáveis decorrentes de PCA baseadas nas taxas prévias de não-comparecimento e no histórico de demanda da clínica. Todas essas variáveis (exceto as baseadas em PCA) tem sido destacadas em estudos anteriores como sendo relevantes para a predição de não-comparecimentos (DANTAS et al., 2018; CARRERAS-GARCÍA et al., 2020), e isso sugere que o modelo treinado apreendeu padrões importantes nesse contexto. Chama-se a atenção para as variáveis resultantes do PCA aparecendo no top 20 (inclusive no top 10), indicando que, mesmo originalmente servindo para séries temporais, apresentaram contribuição relevante no problema de predição de não-comparecimentos.

Note, entretanto, que a Figura 100 não indica a direcionalidade da contribuição das variáveis, de modo que é preciso abstrá-las parcialmente através da seção de exploração dos dados, que inclui análises univariadas. Ainda assim, devido às interações complexas que existem entre as variáveis, a direcionalidade de contribuição pode não se manter para todas as instâncias de predição, sendo necessário uma alternativa para compreender cada predição. Nesse sentido, a Figura 101 serve como uma ilustração de como o *framework SHAP* pode ser útil nesse sentido, podendo servir como uma ferramenta para interpretar as predições individualmente, para ambas as classes estudadas.

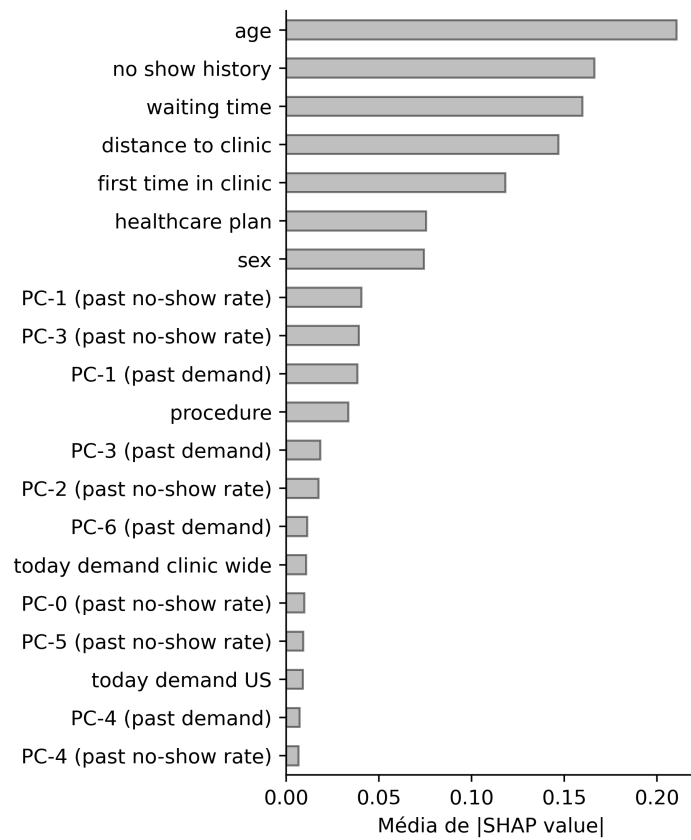


Figura 100 – Média dos valores SHAP absolutos resultantes da aplicação do modelo selecionado sobre os dados de teste. **Fonte:** próprio autor.

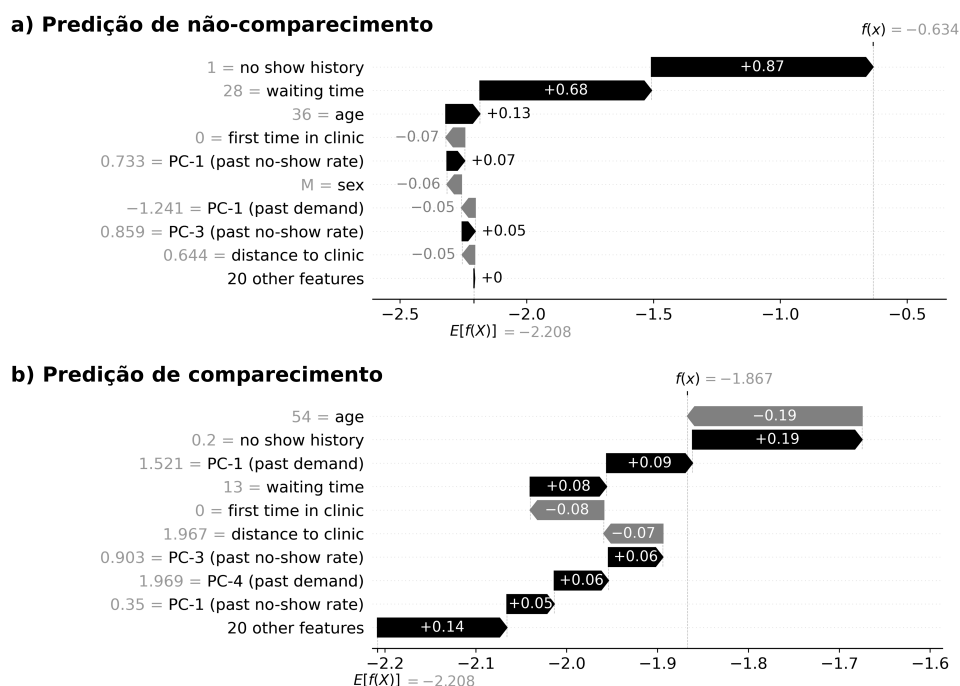


Figura 101 – Exemplo de predições individuais feitas pelo modelo selecionado em instâncias do conjunto de teste, com seus respectivos valores de contribuição das variáveis. O eixo horizontal representa as contribuições em log-odds. As setas pretas indicam contribuições positivas para a probabilidade de não-comparecimento, enquanto as setas cinzas indicam contribuições negativas. O valor base $E[f(x)]$ corresponde à saída média do modelo e serve como ponto de referência. A saída final do modelo $f(x)$ reflete a contribuição cumulativa de todos os valores de atributos para a instância específica. **Fonte:** próprio autor.

4.6 Comparações com a literatura

Nas seções anteriores, foram obtidos resultados acerca da performance dos modelos de classificação para as modalidades de MR, CT, consultas e US. Os valores das métricas isoladamente, entretanto, não são capazes de transmitir a ideia de se eles foram satisfatórios ou não. Nesse sentido, serão feitas comparações com diferentes trabalhos na literatura científica que tiveram uma metodologia ou objetivos semelhantes aos do presente trabalho.

4.6.1 Comparação de performance com um trabalho realizado em um hospital pediátrico no Chile envolvendo múltiplas especialidades médicas (2023)

Em (DUNSTAN et al., 2023), os autores fizeram uso dos modelos *RUS Boost*, *Balanced Bagging*, Regressão Logística, *Ada Boost*, Floresta Aleatória Balanceada (do inglês, *Balanced Random Forest*), *Easy Ensemble*, Floresta Aleatória e Máquinas de Vetores de Suporte (do inglês, *Support Vector Machines*). Os dados, por sua vez, foram provenientes do Hospital Dr. Luis Calvo Mackenna, localizado no Chile e incluindo dados dos anos de 2015 a 2018. Pelo fato de o hospital ser pediátrico, 99.2% da idade dos pacientes estão abaixo dos 18 anos. Ademais, as especialidades médicas incluídas no estudo foram múltiplas (24 no total), e todas na forma de consultas. As variáveis incluídas acerca do paciente foram idade, sexo, comuna de residência, tipo de seguro de saúde, dia da semana, mês, tempo de espera, histórico prévio de não-comparecimento geral (envolvendo todas as especialidades médicas), histórico prévio de não-comparecimento específico (envolvendo separadamente as especialidades médicas) e tipo do agendamento (se é a primeira vez no hospital, uma consulta de rotina ou a primeira vez devido ao encaminhamento por parte dos cuidados primários de saúde). Em relação à metodologia, houve tunagem de hiperparâmetros, seleção dos modelos com base na melhor performance apresentada e ajuste de limiar de probabilidade com base em pesos para os erros do tipo I e II do modelo.

Este trabalho é pertinente pois, em termos de metodologia geral, é muito semelhante com o que foi proposto para este projeto de mestrado. As principais diferenças residem no total e nos tipos de modelos utilizados, no contexto dos dados (um hospital pediátrico de enorme porte do Chile, cujos serviços fornecidos não são exames de imagem, e sim consultas) e na definição da métrica "Custo". No caso, os autores em (DUNSTAN et al., 2023) atribuíram um peso de 1 para o erro do tipo 1 (falsos positivos) e 0.666 para o erro do tipo 2 (falsos negativos) (neste projeto de mestrado, foram aplicados os pesos de 0.333 e 1 para os erros tipo I e II, respectivamente). Ademais, outro fato notável desse trabalho é que os autores foram capazes de aplicar os modelos de classificação na rotina do hospital na forma de um protótipo de solução final, medindo, inclusive, o impacto do uso dos modelos na proporção de não-comparecimento dos pacientes. Mais especificamente, nesse experimento-piloto (de 2 meses e envolvendo as especialidades de oftalmologia, neurologia, traumatologia e dermatologia) foram feitas ligações para os pacientes cujas probabilidades de não-comparecimento eram muito altas, e isso resultou numa redução de 21.0% para 10.3% na proporção de não-comparecimentos, o que é um resultado extremamente relevante.

As principais métricas de performance alcançadas pelos autores para as diferentes especialidades médicas do hospital estão apresentadas na Tabela 11, envolvendo as métricas

Sensibilidade, Precisão e F1-score para as classes 0 (comparecimento do paciente) e 1 (não-comparecimento do paciente). Para uma melhor comparação, essas mesmas métricas de performance são apresentadas novamente para as modalidades incluídas neste projeto na Tabela 13.

Tabela 11 – Métricas de performance do melhor modelo para cada especialidade médica conforme obtido em (DUNSTAN et al., 2023). As métricas Precisão (P), Sensibilidade (S) e *F1-score* (F1) foram separadas para as classes de não-comparecimento (1) e comparecimento (0).

Especialidade (% de 1)	P ₀ (%)	P ₁ (%)	S ₀ (%)	S ₁ (%)	F1 ₀ (%)	F1 ₁ (%)	AUC-ROC (%)
Cardiologia (14.7%)	85%	29%	91%	18%	88%	22%	63%
Dermatologia (28.1%)	76%	42%	90%	22%	82%	28%	65%
Endocrinologia (22.1%)	83%	35%	84%	33%	83%	34%	66%
Gastroenterologia (19.3%)	83%	36%	91%	21%	87%	27%	65%
Cirurgia geral (16.9%)	91%	28%	85%	40%	88%	33%	72%
Genética (24.5%)	78%	31%	84%	24%	81%	27%	57%
Ginecologia (25.1%)	78%	37%	88%	22%	83%	28%	61%
Hematologia (15.8%)	88%	37%	88%	38%	88%	37%	73%
Infectologia (23.7%)	76%	47%	92%	21%	83%	29%	64%
Nefrologia (18.4%)	87%	33%	92%	23%	89%	27%	69%
Neurologia (28.3%)	77%	50%	92%	23%	84%	31%	68%
Nutrição (27.6%)	70%	49%	92%	16%	80%	24%	60%
Oncologia (4.9%)	97%	12%	92%	29%	94%	17%	72%
Oftalmologia (30.3%)	72%	49%	91%	21%	80%	29%	62%
Ortodontia (18.4%)	87%	59%	92%	47%	89%	53%	80%
Otorrinolaringologia (22.7%)	83%	45%	88%	37%	85%	40%	69%
Parasitologia (18.8%)	86%	37%	91%	26%	89%	30%	65%
Odontopediatria (24.9%)	73%	55%	93%	20%	82%	30%	66%
Pediatria (22.6%)	78%	43%	90%	23%	84%	30%	64%
Cirurgia plástica (14.2%)	93%	22%	81%	47%	87%	30%	76%
Psiquiatria (24.0%)	78%	45%	90%	25%	84%	32%	65%
Pneumologia (23.2%)	88%	32%	78%	49%	83%	39%	74%
Reumatologia (20.9%)	80%	33%	91%	16%	85%	22%	60%
Traumatologia (19.9%)	84%	31%	89%	22%	86%	26%	63%
Urologia (19.3%)	83%	33%	89%	23%	86%	27%	63%

Tabela 13 – Métricas de performance do melhor modelo para cada modalidade de procedimentos conforme obtido neste projeto de mestrado. As métricas Precisão (P), Sensibilidade (S) e *F1-score* (F1) foram separadas para as classes de não-comparecimento (1) e comparecimento (0).

Modalidade (% de 1)	P ₀ (%)	P ₁ (%)	S ₀ (%)	S ₁ (%)	F1 ₀ (%)	F1 ₁ (%)	AUC-ROC (%)
MR (8.1%)	93.13%	32.99%	97.02%	17.02%	95.04%	22.46%	65.6%
CT (11.0%)	88.60%	41.43%	87.30%	44.44%	87.95%	42.89%	70.4%
CONSULTAS (17.5%)	89.92%	39.17%	90.70%	37.07%	90.31%	38.09%	72.8%
US (11.5%)	89.44%	35.59%	91.29%	30.87%	90.35%	33.06%	71.3%

Comparando as Tabelas 11 e 13, nota-se que há grande diferença na proporção de não-comparecimentos no conjunto de dados de cada trabalho. Em (DUNSTAN et al., 2023), essas proporções tendem a estar em torno de ou maiores que 15% para todas as especialidades médicas, exceto para oncologia. Já nos dados utilizados para este projeto, somente a modalidade de consultas excede o valor de 15% para a proporção de não-comparecimentos. Apesar disso, em termos de performance dos modelos, os valores obtidos para as métricas Precisão, Sensibilidade e *F1-score* em cada trabalho não apresentaram tantas diferenças. Um detalhe específico é que, para os resultados obtidos em (DUNSTAN et al., 2023), a Precisão tende a ser maior que a Sensibilidade para a classe 1 (exceto para a modalidade de consultas), enquanto que neste projeto de mestrado, a Sensibilidade e a Precisão não foram tão diferentes (exceto para a modalidade MR). Em termos de AUC-ROC, os valores obtidos neste projeto de mestrado foram maiores que grande parte dos valores obtidos para as diferentes especialidades médicas em (DUNSTAN et al., 2023).

Tendo em vista que os valores de performance são parecidos e que em (DUNSTAN et al., 2023) o experimento-piloto resultou em reduções importantes na proporção de não-comparecimentos, isso sugere um potencial relevante de aplicação dos modelos treinados neste projeto de mestrado para a rotina na Clínica Radiológica de Anápolis.

4.6.2 Comparação de performance com um trabalho realizado com dois conjuntos de dados brasileiros (2024)

Em (DEINA et al., 2024), os autores testaram três modelos: K-vizinhos mais próximos (do inglês, *K-Nearest Neighbors*), Máquinas de Vetores de Suporte e Regressão Simbólica (do inglês, *Symbolic Regression*), sendo este último raramente utilizado no contexto de predição do não-comparecimento de pacientes. Para lidar com o desbalanceamento de classes, os autores testaram diferentes técnicas de balanceamento: SMOTE, subamostragem aleatória (*Random under-sampling*, RUS), *NearMiss* (NM) e Limiar de Dificuldade da Instância (*Instance Hardness Threshold*, IHT), sendo as duas últimas as mais recentes e menos exploradas no contexto de predição do não-comparecimento de pacientes. Os dados consistem de dois conjuntos distintos:

1. Um foi coletado do Departamento de Radiologia de um hospital localizado na região Sul do Brasil, para a modalidade de CT envolvendo todo o ano de 2017, com uma proporção de não-comparecimento igual a 6.65% (menor do que foi obtido com os dados deste projeto de mestrado, igual a 11.0%). As variáveis utilizadas foram idade do paciente, sexo do paciente, etnia do paciente, estado civil do paciente, nível de educação do paciente, distância até a clínica, dia do mês, dia da semana, mês do ano, estação do ano, tempo de

espera, histórico de câncer do paciente, total de agendamentos do paciente no ano anterior para a modalidade de CT, total de não-comparecimentos do paciente no ano anterior para a modalidade de CT, total de exames agendados pelo paciente no ano anterior no hospital, total de não-comparecimentos do paciente em exames no ano anterior no hospital;

2. O outro foi coletado do *Kaggle*, envolvendo agendamentos de diferentes centros de saúde. Os autores optaram por realizar o estudo com somente os centros de saúde com o serviço de consulta e localizados no Jardim Camburi, um bairro da cidade de Vitória, no estado do Espírito Santo, Brasil. Esse subconjunto dos dados apresentou uma proporção de não-comparecimentos igual a 19.03%. As variáveis extraídas envolveram idade do paciente, sexo do paciente, auxílio governamental (se o paciente possui ou não), histórico de alcoolismo do paciente, hipertensão (se o paciente possui ou não), diabetes (se o paciente possui ou não), recebimento de SMS (se o paciente recebeu o SMS sobre o agendamento), dia do mês, dia da semana, mês do ano, tempo de espera, total de agendamentos realizados pelo paciente até o momento, total de não-comparecimentos do paciente até o momento e o número de dias entre o último agendamento e o agendamento atual;

Com esses conjuntos de dados, os autores testaram todas as combinações entre técnicas de balanceamento de classes e modelos de classificação, de modo a analisar qual retorna a melhor performance. Ademais, foi feita a tunagem de hiperparâmetros e também de subconjunto de variáveis utilizado para treino. Para selecionar dentre todas as combinações possíveis, foi utilizado o critério de maior valor para o *F1-score*.

A pertinência deste trabalho para comparação consiste no fato de ter sido incluído conjuntos de dados de centros de saúde brasileiros envolvendo as modalidades de CT e consultas, de modo que há semelhança entre os contextos de aplicação. Por outro lado, diferentemente de como ocorreu na comparação da subseção anterior, em (DEINA et al., 2024) foram testados modelos diferentes do conjunto que foi utilizado neste projeto de mestrado, bem como extraídas variáveis preditoras adicionais e alternativas, como as referentes ao histórico de câncer do paciente ou se o paciente possui diabetes, por exemplo. Desse modo, esse trabalho serve como uma base de comparação alternativa e enriquecedora.

Mais especificamente, em (DEINA et al., 2024) foram obtidos resultados variados em termos de AUC-ROC, Acurácia, Sensibilidade, Precisão e F1-score para a classe de não-comparecimento. Essas variações foram decorrentes dos diferentes modelos e técnicas de reamostragem utilizados, porém os autores não deixaram claro que combinações de técnicas e modelos seriam apropriados para uma aplicação prospectiva. Nesse sentido, para que seja possível uma comparação geral, os valores obtidos para essas métricas considerando as diferentes técnicas e modelos são mostrados na Tabela 15 para o conjunto de dados 1 e Tabela 17 para o conjunto de

dados 2. Já os valores correspondentes obtidos para este projeto de mestrado considerando os melhores modelos para as diferentes modalidades de procedimentos são mostrados na Tabela 19.

Tabela 15 – Métricas de performance das diferentes combinações entre técnicas de reamostragem e modelos de classificação testados para o conjunto de dados 1 conforme obtido em (DEINA et al., 2024). As métricas Precisão (P), Sensibilidade (S) e *F1-score* (F1) foram calculadas somente para a classe de não-comparecimento. Também são apresentados a área sobre a curva ROC (AUC-ROC) e a acurácia total (A).

Técnica / Modelo	P (%)	S (%)	F1 (%)	AUC-ROC (%)	A (%)
SMOTE/KNN	9.12%	19.05%	12.17%	53.21%	82.82%
SMOTE/SVM	8.83%	93.20%	16.11%	61.93%	34.83%
SMOTE/SR	3.99%	11.95%	3.99%	51.41%	85.62%
RUS/KNN	9.38%	74.32%	16.64%	61.32%	50.06%
RUS/SVM	8.78%	93.66%	16.03%	61.77%	34.13%
RUS/SR	7.92%	50.99%	11.75%	56.93%	62.08%
NM/KNN	7.02%	75.25%	12.85%	52.11%	32.06%
NM/SVM	6.90%	82.66%	12.73%	51.46%	24.41%
NM/SR	6.73%	73.67%	12.31%	50.45%	30.31%
IHT/KNN	9.21%	89.98%	16.70%	63.02%	39.64%
IHT/SVM	8.81%	94.63%	16.11%	62.04%	33.78%
IHT/SR	8.72%	95.37%	15.96%	61.75%	32.60%

Tabela 17 – Métricas de performance das diferentes combinações entre técnicas de reamostragem e modelos de classificação testados para o conjunto de dados 2 conforme obtido em (DEINA et al., 2024). As métricas Precisão (P), Sensibilidade (S) e *F1-score* (F1) foram calculadas somente para a classe de não-comparecimento. Também são apresentados a área sobre a curva ROC (AUC-ROC) e a acurácia total (A).

Técnica / Modelo	P (%)	S (%)	F1 (%)	AUC-ROC (%)	A (%)
SMOTE / KNN	55.92%	53.46%	52.46%	71.05%	81.94%
SMOTE / SVM	45.65%	69.58%	53.27%	74.00%	76.73%
SMOTE / SR	54.09%	60.04%	51.61%	71.40%	78.44%
RUS / KNN	40.19%	73.19%	51.16%	73.17%	73.16%
RUS / SVM	40.45%	64.87%	49.08%	70.57%	74.11%
RUS / SR	46.61%	71.03%	49.76%	71.63%	71.99%
NM / KNN	29.35%	78.11%	41.75%	64.59%	56.22%
NM / SVM	25.32%	82.66%	38.45%	61.55%	48.33%
NM / SR	36.60%	72.80%	46.59%	69.00%	66.65%
IHT / KNN	34.18%	94.18%	50.04%	75.44%	63.84%
IHT / SVM	36.80%	91.20%	52.26%	76.84%	67.95%
IHT / SR	36.61%	94.34%	52.14%	77.34%	66.81%

Tabela 19 – Métricas de performance dos melhores modelos aplicados a diferentes modalidades de procedimentos no CRA, conforme obtido para este projeto de mestrado. As métricas Precisão (P), Sensibilidade (S) e *F1-score* (F1) foram calculadas somente para a classe de não-comparecimento. Também são apresentados a área sobre a curva ROC (AUC-ROC) e a acurácia total (A).

Técnica / Modelo	P (%)	S (%)	F1 (%)	AUC-ROC (%)	A (%)
MR	32.99%	17.02%	22.46%	65.6%	90.67%
CT	41.43%	44.44%	42.89%	70.4%	80.09%
CONSULTAS	39.17%	37.07%	38.09%	72.8%	83.24%
US	35.59%	30.87%	33.06%	71.3%	83.14%

Da Tabela 15, referente aos resultados obtidos em (DEINA et al., 2024) para o conjunto de dados 1, nota-se que apenas para as combinações técnica-modelo SMOTE/KNN e SMOTE/SR houve acurácia maior que 70%. Nessas mesmas combinações de técnica-modelo, foram obtidas baixas Sensibilidades e Precisões para a classe de não-comparecimento, dando a indicar que praticamente somente a classe negativa (de comparecimento) estava sendo corretamente predita por esses modelos. Por outro lado, nas combinações em que a Sensibilidade foi alta, a Acurácia decresceu, sugerindo, portanto, que houve muitos falsos positivos (erro do tipo I) nesses casos. Comparado aos resultados obtidos para a modalidade de CT neste projeto de mestrado, os valores neste último caso são melhores, com diferenças significativas: da Tabela 19, a Sensibilidade resultou baixa, mas o F1-score e a Acurácia mantiveram-se altos, e isso poderia mudar caso fosse necessário alterar os pesos associados aos erros do tipo I e II no cálculo da métrica Custo, que foi usada para ajuste do limiar de probabilidade. Essa diferença, entretanto, pode ser atribuída também à diferença entre as proporções de não-comparecimento dos dois conjuntos de dados. Em (DEINA et al., 2024) há um cenário mais difícil, com maior desbalanceamento de classes (cerca de 6.5% de proporção de não-comparecimento) e mais suscetível a vieses.

Já no conjunto de dados 2 (Tabela 17), os resultados obtidos em (DEINA et al., 2024) foram melhores e mais equilibrados, com valores razoáveis para Acurácia, Sensibilidade e Precisão para a combinação técnica-modelo SMOTE/KNN. Com a proporção de não-comparecimentos maior (cerca de 19%) e com variáveis adicionais e alternativos (como as referentes a informações mais específicas acerca dos pacientes), os autores em (DEINA et al., 2024) conseguiram alcançar 81.94% de Acurácia, 55.92% de Precisão e 53.46% de Sensibilidade para a classe positiva (não-comparecimento). Entretanto, deve-se levar em consideração que neste projeto de mestrado não estão sendo usadas variáveis adicionais referentes a detalhes acerca dos pacientes. Isso poderia agregar nos valores de performance finais.

5 Conclusão

A metodologia proposta, envolvendo dados de grande extensão temporal e consistindo das etapas de pré-processamento dos dados, tunagem de hiperparâmetros, ajuste do limiar de probabilidade, seleção dos modelos de classificação e interpretação das predições, é robusta e envolveu quatro modalidades de procedimentos da Clínica Radiológica de Anápolis: MR (unidade do St. Central), CT (unidade do St. Central), consultas (unidade de Jundiaí) e US (unidade de Jundiaí). Essa robustez pode ser evidenciada pelo fato de o projeto ter sido conduzido, através da aplicação de técnicas como Validação Cruzada com Divisão Temporal e modelos de classificação com diferentes algoritmos, conforme tem sido recomendado como boas práticas no contexto de ML. Ademais, os resultados de redução da métrica personalizada Custo e comparações envolvendo as performances obtidas para os modelos e outros trabalhos da literatura científica reforçam ainda mais essa robustez, pois foi encontrado que há boa concordância entre as performances, deixando clara a possibilidade e o potencial dos modelos servirem como solução final para a predição de não-comparecimentos na Clínica Radiológica de Anápolis para a otimização de tarefas como o *overbooking*.

Para além dos resultados obtidos relacionados às predições dos modelos, o cálculo das contribuições das variáveis utilizadas por meio do *framework SHAP* demonstrou-se extremamente útil para o entendimento/esclarecimento das predições e para comparações com a análise exploratória conduzida no início do trabalho. Apesar de o gráfico do valor absoluto médio da contribuição das variáveis preditoras não indicar o sentido da contribuição, é possível interpretar as predições individualmente, de modo que, num cenário de aplicação real da metodologia proposta, a rotina pode ser otimizada e ao mesmo tempo respaldada pela contribuição das variáveis na instância em questão. Isso pode ser útil em casos onde se faz necessário a intervenção junto ao paciente, pois conhecendo-se o fator que mais afetou a probabilidade de não-comparecimento para ele em específico, torna-se possível ajustar os detalhes de agendamento para maximizar suas chances de comparecimento.

No que se refere ao cronograma de condução do projeto, tudo tem sido feito no prazo esperado. Vale a pena citar também que, durante o desenvolvimento do projeto, foi possível a conquista de certificados importantes de cursos relacionados a Ciência de Dados e ML (certificados disponíveis em anexo). Adicionalmente, foi possível participar do [VI Encontro Goiano de Probabilidade e Estatística \(VI ENGOPE\)](#), realizado pelo IME-UFG, onde foi apresentado um pôster contendo parte dos resultados deste projeto e houve participação do 1º Datathon organizado pelo evento, no qual foi recebido a premiação em primeiro lugar junto à equipe

Embrapeiros.

Em relação a publicações científicas, foram feitas quatro publicações no total, envolvendo áreas distintas de aplicação, não necessariamente ligadas diretamente ao projeto de mestrado. Uma delas que está ligada ao projeto de mestrado e está disponível em anexo é "AVALIAÇÃO DO IMPACTO DO RETREINO PERIÓDICO NA PREDIÇÃO DE NÃO-COMPARECIMENTOS DE PACIENTES EM CONSULTAS MÉDICAS UTILIZANDO BALANCED RANDOM FOREST", e explora diferentes quantidades de dados de treino como parâmetro de otimização da performance do modelo de classificação. O estudo conclui que, sim, esse é um parâmetro significativo, indicando que o *concept drift* é um fenômeno que pode ocorrer no contexto de predição de não-comparecimento de pacientes e que deve ser levado em conta para maximizar a performance do modelo ao longo do tempo.

Em relação a submissões adicionais ligadas ao projeto de mestrado, há uma que está em revisão na revista *Journal of the Operational Research Society*, que é de alto impacto. Essa submissão faz uso direto dos resultados encontrados para a modalidade de Ressonância Magnética.

Referências

ABDI, H.; WILLIAMS, L. J. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, Wiley Online Library, v. 2, n. 4, p. 433–459, 2010. Citado 3 vezes nas páginas 39, 40 e 41.

AGUILERA, A. M.; ESCABIAS, M.; VALDERRAMA, M. J. Using principal components for estimating logistic regression with high-dimensional multicollinear data. *Computational Statistics & Data Analysis*, Elsevier, v. 50, n. 8, p. 1905–1924, 2006. Citado 4 vezes nas páginas 40, 41, 42 e 43.

AHMADI, E. et al. A metaheuristic-based stacking model for predicting the risk of patient no-show and late cancellation for neurology appointments. *IISE Transactions on Healthcare Systems Engineering*, Taylor & Francis, v. 9, n. 3, p. 272–291, 2019. Citado na página 37.

AKINKUNMI, M. Hypothesis testing for single population mean and proportion. In: *Introduction to Statistics Using R*. [S.l.]: Springer, 2019. p. 141–151. Citado na página 84.

ALADEEMY, M. et al. New feature selection methods based on opposition-based learning and self-adaptive cohort intelligence for predicting patient no-shows. *Applied Soft Computing*, Elsevier, v. 86, p. 105866, 2020. Citado na página 37.

ALADSANI, M.; BURTON, H.; SE, S. A. Machine learning driven drift capacity model for reinforced concrete walls. *Special Publication*, v. 350, p. 16–26, 2021. Citado 2 vezes nas páginas 8 e 60.

BERG, B. P. et al. Estimating the cost of no-shows and evaluating the effects of mitigation strategies. *Medical Decision Making*, SAGE Publications Sage CA: Los Angeles, CA, v. 33, n. 8, p. 976–985, 2013. Citado 3 vezes nas páginas 25, 28 e 81.

BISCHL, B. et al. Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Wiley Online Library, v. 13, n. 2, p. e1484, 2023. Citado na página 26.

BOUZGOU, H. *Advanced Methods for the Processing and Analysis of Multidimensional Signals: Application to Wind Speed*. Tese (Doutorado) — Université de Batna 2, 2012. Citado 2 vezes nas páginas 8 e 49.

BREIMAN, L. Bagging predictors. *Machine learning*, Springer, v. 24, p. 123–140, 1996. Citado 2 vezes nas páginas 26 e 56.

BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, p. 5–32, 2001. Citado na página 55.

BREIMAN, L. *Classification and regression trees*. [S.l.]: Routledge, 2017. Citado na página 52.

BRO, R.; SMILDE, A. K. Principal component analysis. *Analytical methods*, Royal Society of Chemistry, v. 6, n. 9, p. 2812–2831, 2014. Citado na página 77.

- CARRERAS-GARCÍA, D. et al. Patient no-show prediction: a systematic literature review. *Entropy*, MDPI, v. 22, n. 6, p. 675, 2020. Citado 20 vezes nas páginas 8, 26, 33, 34, 35, 36, 37, 38, 39, 61, 77, 79, 98, 100, 115, 117, 132, 134, 149 e 151.
- CASTRO, L. N. D. *Fundamentals of natural computing: basic concepts, algorithms, and applications*. [S.l.]: Chapman and Hall/CRC, 2006. Citado 3 vezes nas páginas 45, 46 e 50.
- CHARIATTE, V. et al. Missed appointments in an outpatient clinic for adolescents, an approach to predict the risk of missing. *Journal of Adolescent Health*, Elsevier, v. 43, n. 1, p. 38–45, 2008. Citado 2 vezes nas páginas 25 e 33.
- CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. [S.l.: s.n.], 2016. p. 785–794. Citado 2 vezes nas páginas 60 e 79.
- COSTA, L. et al. Multilayer perceptron. *Introduction to Computational Intelligence*, v. 105, 2023. Citado 3 vezes nas páginas 8, 45 e 79.
- CYBENKO, G. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, Springer, v. 2, n. 4, p. 303–314, 1989. Citado na página 48.
- DAGGY, J. et al. Using no-show modeling to improve clinic performance. *Health informatics journal*, SAGE Publications Sage UK: London, England, v. 16, n. 4, p. 246–259, 2010. Citado na página 33.
- DANTAS, L. F. et al. No-shows in appointment scheduling—a systematic literature review. *Health Policy*, Elsevier, v. 122, n. 4, p. 412–421, 2018. Citado 12 vezes nas páginas 8, 25, 26, 29, 30, 33, 34, 36, 100, 117, 134 e 151.
- DEINA, C. et al. Decision analysis framework for predicting no-shows to appointments using machine learning algorithms. *BMC Health Services Research*, Springer, v. 24, n. 1, p. 37, 2024. Citado 5 vezes nas páginas 20, 156, 157, 158 e 159.
- DUNSTAN, J. et al. Predicting no-show appointments in a pediatric hospital in chile using machine learning. *Health Care Management Science*, Springer, v. 26, n. 2, p. 313–329, 2023. Citado 6 vezes nas páginas 19, 26, 81, 154, 155 e 156.
- ELLIS, D. A.; JENKINS, R. Weekday affects attendance rate for medical appointments: large-scale data analysis and implications. *PloS one*, Public Library of Science San Francisco, USA, v. 7, n. 12, p. e51365, 2012. Citado 2 vezes nas páginas 25 e 29.
- FAHLMAN, S. E. et al. *An empirical study of learning speed in back-propagation networks*. [S.l.]: Carnegie Mellon University, Computer Science Department Pittsburgh, PA, USA, 1988. Citado na página 47.
- FREUND, Y.; SCHAPIRE, R. E. et al. Experiments with a new boosting algorithm. In: CITESEER. *icml*. [S.l.], 1996. v. 96, p. 148–156. Citado na página 56.
- FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, NIH Public Access, v. 33, n. 1, p. 1, 2010. Citado na página 44.

- FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, JSTOR, p. 1189–1232, 2001. Citado na página 58.
- FU, G.-H.; YI, L.-Z.; PAN, J. Tuning model parameters in class-imbalanced learning with precision-recall curve. *Biometrical Journal*, Wiley Online Library, v. 61, n. 3, p. 652–664, 2019. Citado 3 vezes nas páginas 63, 64 e 80.
- GHOJOGH, B.; CROWLEY, M. The theory behind overfitting, cross validation, regularization, bagging, and boosting: tutorial. *arXiv preprint arXiv:1905.12787*, 2019. Citado 4 vezes nas páginas 8, 26, 56 e 58.
- GLOWACKA, K. J.; HENRY, R. M.; MAY, J. H. A hybrid data mining/simulation approach for modelling outpatient no-shows in clinic scheduling. *Journal of the Operational Research Society*, Taylor & Francis, v. 60, n. 8, p. 1056–1068, 2009. Citado 2 vezes nas páginas 25 e 28.
- GUO, C.; LU, M.; CHEN, J. An evaluation of time series summary statistics as features for clinical prediction tasks. *BMC medical informatics and decision making*, Springer, v. 20, p. 1–20, 2020. Citado na página 77.
- GUOLINKE, Q. M. et al. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.*, v. 30, p. 52, 2017. Citado na página 60.
- GUYON, I.; ELISSEEFF, A. An introduction to variable and feature selection. *Journal of machine learning research*, v. 3, n. Mar, p. 1157–1182, 2003. Citado na página 36.
- HAIXIANG, G. et al. Learning from class-imbalanced data: Review of methods and applications. *Expert systems with applications*, Elsevier, v. 73, p. 220–239, 2017. Citado 3 vezes nas páginas 60, 61 e 62.
- HASTIE, T. et al. *The elements of statistical learning: data mining, inference, and prediction*. [S.l.]: Springer, 2009. v. 2. Citado na página 60.
- HASVOLD, P. E.; WOOTTON, R. Use of telephone and sms reminders to improve attendance at hospital appointments: a systematic review. *Journal of telemedicine and telecare*, SAGE Publications Sage UK: London, England, v. 17, n. 7, p. 358–364, 2011. Citado 2 vezes nas páginas 25 e 33.
- HAYKIN, S. Neural networks and learning machines. prentice hall. *New York*, v. 547, p. 548, 2008. Citado 6 vezes nas páginas 26, 46, 47, 50, 51 e 52.
- HOCHREITER, D. C. U.; UNTERTHINER, T.; CLEVERT, D. Fast and accurate deep network learning by exponential linear units (elus). In: *Proc. 4th Int. Conf. Learn. Representations*. [S.l.: s.n.], 2016. Citado na página 47.
- HUANG, Y.; HANAUER, D. A. Patient no-show predictive model development using multiple data sources for an effective overbooking approach. *Applied clinical informatics*, Schattauer GmbH, v. 5, n. 03, p. 836–860, 2014. Citado na página 81.
- HUANG, Y.-L.; HANAUER, D. A. Time dependent patient no-show predictive modelling development. *International journal of health care quality assurance*, Emerald Group Publishing Limited, v. 29, n. 4, p. 475–488, 2016. Citado na página 38.

(IBGE), I. B. de Geografia e E. *Censo Demográfico 2022: População e Domicílios*. IBGE, 2022. Disponível em: <<https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2102011>>. Citado na página 68.

JEFFREY, A.; DAI, H. H. *Handbook of mathematical formulas and integrals*. [S.l.]: Elsevier, 2008. Citado na página 46.

JENI, L. A.; COHN, J. F.; TORRE, F. D. L. Facing imbalanced data—recommendations for the use of performance metrics. In: IEEE. *2013 Humaine association conference on affective computing and intelligent interaction*. [S.l.], 2013. p. 245–251. Citado 5 vezes nas páginas 8, 61, 63, 64 e 65.

KE, G. et al. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, v. 30, 2017. Citado na página 79.

KRAWCZYK, B. Learning from imbalanced data: open challenges and future directions. *Progress in artificial intelligence*, Springer, v. 5, n. 4, p. 221–232, 2016. Citado na página 38.

KURASAWA, H. et al. Machine-learning-based prediction of a missed scheduled clinical appointment by patients with diabetes. *Journal of diabetes science and technology*, SAGE Publications Sage CA: Los Angeles, CA, v. 10, n. 3, p. 730–736, 2016. Citado 2 vezes nas páginas 36 e 38.

LAVALLEY, M. P. Logistic regression. *Circulation*, Am Heart Assoc, v. 117, n. 18, p. 2395–2399, 2008. Citado na página 79.

LEMAÎTRE, G.; NOGUEIRA, F.; ARIDAS, C. K. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, v. 18, n. 17, p. 1–5, 2017. Disponível em: <<http://jmlr.org/papers/v18/16-365.html>>. Citado na página 62.

LENZI, H.; BEN, Â. J.; STEIN, A. T. Development and validation of a patient no-show predictive model at a primary care setting in southern brazil. *Plos one*, Public Library of Science San Francisco, CA USA, v. 14, n. 4, p. e0214869, 2019. Citado 2 vezes nas páginas 26 e 33.

LIU, D. et al. Machine learning approaches to predicting no-shows in pediatric medical appointment. *NPJ digital medicine*, Nature Publishing Group UK London, v. 5, n. 1, p. 50, 2022. Citado na página 26.

LIU, J.; CHEN, J.; YE, J. Large-scale sparse logistic regression. In: *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. [S.l.: s.n.], 2009. p. 547–556. Citado 4 vezes nas páginas 26, 41, 42 e 43.

LUNDBERG, S. A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*, 2017. Citado 4 vezes nas páginas 27, 65, 66 e 82.

MAAS, A. L. et al. Rectifier nonlinearities improve neural network acoustic models. In: ATLANTA, GA. *Proc. icml*. [S.l.], 2013. v. 30, n. 1, p. 3. Citado na página 47.

MARBOUH, D. et al. Evaluating the impact of patient no-shows on service quality. *Risk management and healthcare policy*, Taylor & Francis, p. 509–517, 2020. Citado 2 vezes nas páginas 25 e 28.

MENON, A. et al. Characterization of a class of sigmoid functions with applications to neural networks. *Neural networks*, Elsevier, v. 9, n. 5, p. 819–835, 1996. Citado na página 46.

MOCHÓN, F. et al. Machine-learning-based no show prediction in outpatient visits. *International Journal of Interactive Multimedia and Artificial Intelligence*, v. 4, n. 7, p. 29–34, 03/2018 2018. ISSN 1989-1660. Disponível em: <http://www.ijimai.org/journal/sites/default/files/files/2017/03/ijimai_4_7_4_pdf_11885.pdf>. Citado na página 38.

MOURA, E. C. et al. Covid-19: evolução temporal e imunização nas três ondas epidemiológicas, brasil, 2020–2022. *Revista de Saúde Pública*, SciELO Brasil, v. 56, p. 105, 2022. Citado na página 85.

NAHM, F. S. Receiver operating characteristic curve: overview and practical use for clinicians. *Korean journal of anesthesiology*, Korean Society of Anesthesiologists, v. 75, n. 1, p. 25–36, 2022. Citado na página 26.

OZENNE, B.; SUBTIL, F.; MAUCORT-BOULCH, D. The precision–recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. *Journal of clinical epidemiology*, Elsevier, v. 68, n. 8, p. 855–859, 2015. Citado 2 vezes nas páginas 26 e 80.

POTTS, D.; SAMMUT, C. Online nonlinear system identification in high dimensional environments. In: *Proc. of the Australasian Conf. on Robotics and Automation*. [S.l.: s.n.], 2004. Citado 2 vezes nas páginas 8 e 53.

PROKHORENKOVA, L. et al. Catboost: unbiased boosting with categorical features. *Advances in neural information processing systems*, v. 31, 2018. Citado na página 79.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *nature*, Nature Publishing Group UK London, v. 323, n. 6088, p. 533–536, 1986. Citado na página 50.

SAITO, T.; REHMSMEIER, M. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, Public Library of Science San Francisco, CA USA, v. 10, n. 3, p. e0118432, 2015. Citado 2 vezes nas páginas 26 e 80.

SALAZAR, L. H. A. et al. Application of machine learning techniques to predict a patient’s no-show in the healthcare sector. *Future Internet*, MDPI, v. 14, n. 1, p. 3, 2021. Citado na página 33.

SALEHI, F.; ABBASI, E.; HASSIBI, B. The impact of regularization on high-dimensional logistic regression. *Advances in Neural Information Processing Systems*, v. 32, 2019. Citado 2 vezes nas páginas 43 e 44.

SANTO, R. d. E. Principal component analysis applied to digital image compression. *Einstein (São Paulo)*, SciELO Brasil, v. 10, p. 135–139, 2012. Citado 2 vezes nas páginas 39 e 40.

SATIANI, B.; MILLER, S.; PATEL, D. No-show rates in the vascular laboratory: analysis and possible solutions. *Journal of Vascular and Interventional Radiology*, Elsevier, v. 20, n. 1, p. 87–91, 2009. Citado 3 vezes nas páginas 25, 28 e 33.

SCHOOT, R. van de et al. Bayesian statistics and modelling. *Nature Reviews Methods Primers*, Nature Publishing Group UK London, v. 1, n. 1, p. 1, 2021. Citado na página 26.

SHAPLEY, L. S. A value for n-person games. *Contribution to the Theory of Games*, v. 2, 1953. Citado na página 65.

SHERIDAN, R. P. Time-split cross-validation as a method for estimating the goodness of prospective prediction. *Journal of chemical information and modeling*, ACS Publications, v. 53, n. 4, p. 783–790, 2013. Citado 2 vezes nas páginas 26 e 80.

SNOWDEN, S. et al. A neural network to predict attendance of paediatric patients at outpatient clinics. *Neural Computing & Applications*, Springer, v. 3, p. 234–241, 1995. Citado na página 36.

STEINBERG, D.; COLLA, P. Cart: tree-structured non-parametric data analysis. *San Diego, CA: Salford Systems*, 1995. Citado 2 vezes nas páginas 26 e 52.

SUTTON, C. D. Classification and regression trees, bagging, and boosting. *Handbook of statistics*, Elsevier, v. 24, p. 303–329, 2005. Citado 5 vezes nas páginas 53, 54, 55, 57 e 58.

TOPUZ, K. et al. Predicting pediatric clinic no-shows: a decision analytic framework using elastic net and bayesian belief network. *Annals of Operations Research*, Springer, v. 263, p. 479–499, 2018. Citado na página 38.

WUJEK, B.; HALL, P.; GÜNES, F. Best practices for machine learning applications. *SAS Institute Inc*, p. 3, 2016. Citado na página 78.

YANG, L.; SHAMI, A. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, Elsevier, v. 415, p. 295–316, 2020. Citado na página 78.

YOUNG, H. P. Monotonic solutions of cooperative games. *International Journal of Game Theory*, Springer, v. 14, n. 2, p. 65–72, 1985. Citado na página 67.

Anexos

ANEXO A – Artigo publicado

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](https://doi.org/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 (Marbough *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 não comparecimento 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).

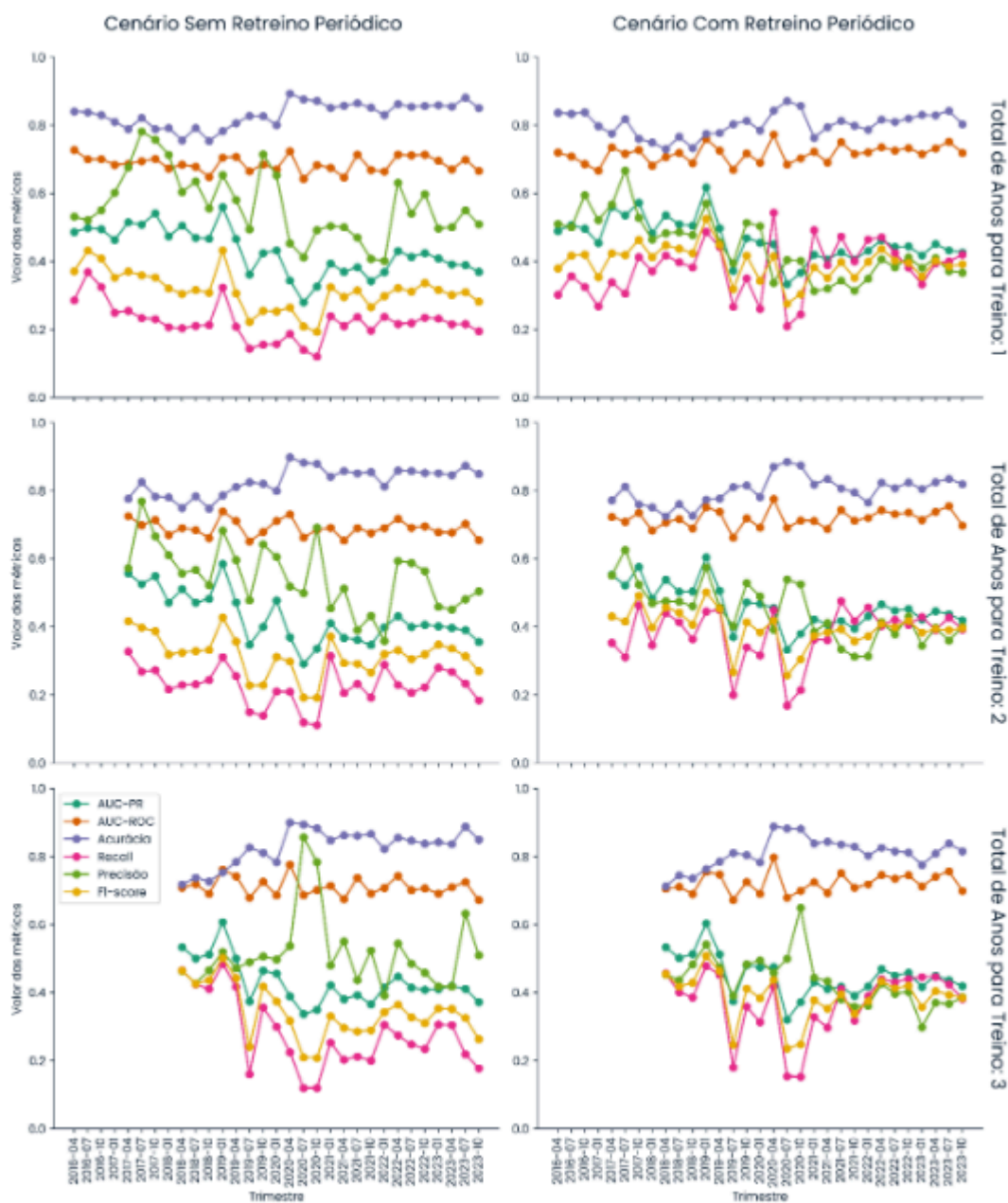


Tabela 2 – Média e desvio padrão (D.P.) das métricas de desempenho, considerando diferentes janelas históricas (1, 2, 3 anos) e os dois cenários

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 falsopositivos. 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. /], 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. /], 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. /], 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. /], 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. /], 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. /], 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. /], 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. /], 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. /], , 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](#)

RevistaFT

A RevistaFT têm 29 anos. É uma **Revista Científica Eletrônica Multidisciplinar Indexada de Alto Impacto e Qualis “B2”**.

Periodicidade mensal e de acesso livre. Leia gratuitamente todos os artigos e publique o seu também clicando aqui,



Contato

Queremos te ouvir.

WhatsApp RJ:

(21) 99451-7530

WhatsApp SP:

(11) 98597-3405

e-Mail:

contato@revistaf
t.com.br

ISSN: 1678-0817

CNPJ:

48.728.404/0001-

22

Fator de

impacto FI=

5.397 (muito alto)

Turismo

Acadêmico

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/expandente

Venha
fazer parte de
nosso time de
revisores
também!

Copyright © Revista ft Ltda. 1996 -
2025

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

ANEXO B – Carta de anuência

Carta de anuência

A quem possa interessar,

1. Informo para os devidos fins e efeitos legais, objetivando atender as exigências para obtenção de parecer do Comitê de Ética em Pesquisa com Seres Humanos, e como representante legal da Instituição, estar ciente do projeto de pesquisa “**Machine Learning aplicado à predição de absenteísmo de pacientes em Medicina Nuclear**”, sob a responsabilidade do Pesquisador Principal Emerson Nobuyuki Itikawa, docente do Instituto de Física da Universidade Federal de Goiás e orientador do discente Carlos Eduardo Gonçalves de Oliveira, discente pelo Programa de Pós-Graduação em Engenharia Elétrica e da Computação pela EMC/UFG.
2. Declaro ainda conhecer e cumprir as orientações e determinações fixadas na Resolução nº 466, de 12 de dezembro de 2012, do Conselho Nacional de Saúde, e demais legislações complementares.
3. No caso de não-cumprimento, por parte do pesquisador, das determinações éticas e legais, a administração do Clínica Radiológica de Anápolis (CRA) tem a liberdade de retirar a anuência a qualquer momento da pesquisa, sem penalização nenhuma.
4. Considerando que esta instituição tem condição para o desenvolvimento deste projeto, autorizo sua execução nos termos propostos mediante a plena aprovação do CEP competente.

_____, 07 de novembro de 2023.



Assinatura: _____

ANEXO C – Termo de compromisso dos pesquisadores

Escola de Engenharia Elétrica, Mecânica e de Computação (EMC) da Universidade Federal de Goiás (UFG)
em parceria com
Clínica Radiológica de Anápolis (CRA)

Declaro que cumprirei os requisitos da *Resolução CNS n.º 466/12* e/ou da *Resolução CNS n.º 510/16*, bem com suas complementares, como pesquisador(a) responsável e/ou pesquisador participante do projeto intitulado “**Machine Learning aplicado à predição de absenteísmo de pacientes em Medicina Nuclear**”. Comprometo-me a utilizar os materiais e os dados coletados exclusivamente para os fins previstos no protocolo da pesquisa acima referido e, ainda, a publicar os resultados, sejam eles favoráveis ou não. Aceito as responsabilidades pela condução científica do projeto, considerando a relevância social da pesquisa, o que garante a igual consideração de todos os interesses envolvidos.

Data: 08/07/2024

Nome do(a) Pesquisador(a)	Assinatura Manuscrita ou Digital
1. Carlos Eduardo Gonçalves de Oliveira	 Documento assinado digitalmente CARLOS EDUARDO GONCALVES DE OLIVEIRA Data: 08/07/2024 15:23:42-0300 Verifique em https://validar.it.gov.br
2. Emerson Nobuyuki Itikawa	

ANEXO D – Parecer consubstanciado do Comitê de Ética em Pesquisa da UFG

PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: MACHINE LEARNING APLICADO À PREDIÇÃO DE ABSENTEÍSMO DE PACIENTES EM MEDICINA NUCLEAR

Pesquisador: Emerson Nobuyuki Itikawa

Área Temática:

Versão: 1

CAAE: 81394924.0.0000.5083

Instituição Proponente: Instituto de Física

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 7.045.338

Apresentação do Projeto:

A proposta envolve a predição de absenteísmo de pacientes que agendam exames em medicina nuclear, utilizando o machine learning. A pesquisa será desenvolvida em duas clínicas radiológicas: uma de Anápolis/GO e outra de Jundiá/SP. Serão incluídos no estudo: pacientes que agendaram/realizaram procedimentos diagnósticos (Medicina Nuclear/Radiologia) ou consultas nas unidades de Jundiá e do Centro da Clínica Radiológica de Anápolis (CRA); pacientes cujo desfecho de agendamento tenha sido interpretado como abstenção, pacientes que cancelaram o agendamento com menos de 24 horas antes da realização do procedimento; e excluídos aqueles cujo "Status do procedimento" não indique a finalização do procedimento ou a abstenção do paciente e na variável "Observações do procedimento", com valores que não sirvam para esclarecer a variável "Status do procedimento"; pacientes com valores não-razoáveis em alguma das variáveis extraídas (por exemplo, data de realização do exame dias antes ou dias depois da data agendada); pacientes que tenham realizado o agendamento do procedimento no mesmo dia de realização do mesmo (por exemplo, no dia 15/03 o paciente agendar um exame às 14:00hs do mesmo dia); pacientes que tenham efetuado o cancelamento do agendamento com o prazo de no mínimo 24 horas antes da realização do procedimento. Os dados serão retrospectivos e coletados na forma tabular junto aos responsáveis pela Equipe de Tecnologia da Informação (TI) de cada Clínica participante. A fonte de dados será o banco disponível em cada uma das Clínica. Serão extraídas as seguintes

Endereço: Rodovia R2, n. 3.061, Parque Tecnológico Samambaia, Edifício K2,sala 110, piso 1

Bairro: Campus Samambaia

CEP: 74.690-970

UF: GO

Município: GOIANIA

Telefone: (62)3521-1215

Fax: (62)3521-2045

E-mail: cep.prpi@ufg.br

Continuação do Parecer: 7.045.338

informações: ID do paciente anonimizado por remapeamento aleatório (para que seja possível extrair variáveis relacionadas ao histórico de absenteísmo dos pacientes); data de nascimento, sexo, CEP/Cidade/UF, distância a percorrer até a clínica, plano de saúde do paciente (sim/não, qual), status do procedimento (cancelado ou se o paciente se absteve), observações acerca do procedimento (campo com texto escrito pelos atendentes; pode conter informações sensíveis como CPF e número de telefone e, por este motivo, todos os números serão substituídos por "X" para devida anonimização. Status do procedimento (modalidade ζ Ultrassom/PET-CT/Tomografia), nome específico do procedimento diagnóstico, cidade do paciente, data em que foi feito o agendamento do procedimento, data e horário agendados para o procedimento, data e horário de realização do procedimento, comparecimento ou não do paciente ao procedimento diagnóstico. Será feito um pré-processamento que envolverá a transformação das variáveis acima em variáveis numéricas assim como estimativas de interação entre elas, de modo iterativo com a análise exploratória dos dados ou com os resultados preliminares. Os dados serão explorados por auxiliar no entendimento dos valores das variáveis preditoras e ocorrerá concomitantemente às etapas de pré-processamento e de aquisição de dados. Serão empregadas as devidas técnicas de visualização para os dados disponíveis, de modo a maximizar o entendimento da correlação entre os valores das variáveis preditoras e a variável a ser predita, que é o comparecimento ou não dos pacientes ao exame. Espera-se que o modelo possa auxiliar a clínica no uso devido dos recursos financeiros e pessoais para que seja prestado o melhor serviço possível para os pacientes. Ademais, pretende-se extrair quantitativamente a importância das diferentes variáveis preditoras que culminam no absenteísmo dos pacientes.

Objetivo da Pesquisa:

OBJETIVO PRIMÁRIO: Estudar, compreender a aplicabilidade e comparar os resultados de diferentes modelos de machine learning para o problema de predição de absenteísmo de pacientes em procedimentos diagnósticos no serviço de medicina nuclear (com a possibilidade de incluir o serviço de radiologia) do CRA. **OBJETIVOS SECUNDÁRIOS:** (1) Efetuar a aquisição dos dados retrospectivos relevantes através do banco de dados do CRA;(2) Fazer o devido pré-processamento dos dados, com foco na solução de possíveis outliers ou valores faltantes;(3) Efetuar a exploração dos dados adquiridos através de sumários estatísticos ou técnicas de visualização;(4) Averiguar a melhor forma de lidar com o desbalanceamento de classes do conjunto de dados (já que a proporção de não-comparecimento de pacientes tende a ser baixa comparado ao número total de procedimentos marcados);(5) Averiguar os modelos de

Endereço: Rodovia R2, n. 3.061, Parque Tecnológico Samambaia, Edifício K2,sala 110, piso 1
Bairro: Campus Samambaia **CEP:** 74.690-970
UF: GO **Município:** GOIANIA
Telefone: (62)3521-1215 **Fax:** (62)3521-2045 **E-mail:** cep.prpi@ufg.br

Continuação do Parecer: 7.045.338

classificação aplicáveis aos dados em questão;(6) Implementar a otimização de hiperparâmetros dos modelos de classificação selecionados usando o conjunto de dados de treino;(7) Usar as melhores combinações de hiperparâmetros para os respectivos modelos de classificação e analisar os resultados sobre os dados de teste;

Avaliação dos Riscos e Benefícios:

Riscos:

Os riscos para o paciente advindos deste projeto é a possibilidade de vazamento de dados que possam de alguma forma possibilitar o rastreamento ou a identificação do paciente. Nesse sentido, os pesquisadores deste projeto, assim como a equipe de T.I. do CRA, implementarão a anonimização dos IDs dos pacientes através de remapeamento aleatório, para que não seja possível o rastreamento de informações sensíveis do paciente no banco de dados original. Ademais, no que se refere aos campos de texto que serão coletados para esclarecimento da variável "Status do procedimento"serão anonimizados TODOS os números por haver a possibilidade de estarem representando a identificação de documentos ou números de telefone. Todas essas tarefas serão feitas junto à equipe de T.I. com o maior cuidado e transparência possíveis.

Comentários e Considerações sobre a Pesquisa:

Proposta atual, interessante, que poderá incluir na predição de absenteísmo de pacientes em clínicas de medicina nuclear, evitando custos diretos e indiretos aos pacientes e serviços que utilizam e oferecem os exames.

Considerações sobre os Termos de apresentação obrigatória:

Projeto - ok

Informações Básicas do Projeto - ok (5 - Aquisição dos dados 02/09/2024 a 31/03/2025 e em outros locais (cronograma jan a jun de 2024)

Dispensa do TCLE - ok

Folha de rosto devidamente assinada pelo pesquisador responsável e diretor do Instituto de Física da UFG.

Termo de Compromisso dos Pesquisadores (3) devidamente assinado (retiraram a Beatriz)

Termo de Anuência do CRA (Jundiá?)

Recomendações:

Padronizar o cronograma de acordo com as informações contidas no documento: "CRONOGRAMA"

Endereço: Rodovia R2, n. 3.061, Parque Tecnológico Samambaia, Edifício K2,sala 110, piso 1

Bairro: Campus Samambaia

CEP: 74.690-970

UF: GO

Município: GOIANIA

Telefone: (62)3521-1215

Fax: (62)3521-2045

E-mail: cep.prpi@ufg.br

Continuação do Parecer: 7.045.338

Termo de Anuência do CRA Jundiáí

Conclusões ou Pendências e Lista de Inadequações:

Salvo melhor juízo deste Comitê e segundo as resoluções CNS n.º 466/12 / Resolução CNS nº 510/16, bem com suas complementares, considera-se o projeto aprovado.

Determina-se que os pesquisadores anexem o Termo de Anuência do Centro de Jundiáí, caso a coleta de dados seja feita no local. Ademais, padronizar o cronograma com a data de aquisição de dados, informada para o período de 02/09/2024 a 31/03/2025, como consta no arquivo Informações Básicas do Projeto.

Considerações Finais a critério do CEP:

Informamos que o Comitê de Ética em Pesquisa/CEP-UFG considera o presente protocolo APROVADO. A pesquisa foi considerada em acordo com os princípios éticos vigentes. Reiteramos a importância deste Parecer Consubstanciado, e lembramos que o(a) pesquisador(a) responsável deverá encaminhar ao CEP-UFG os relatórios parciais e o Relatório Final baseado na conclusão do estudo e na incidência de publicações decorrentes deste, de acordo com o disposto na Resolução CNS n. 466/12 e Resolução CNS n. 510/16. O prazo para entrega do Relatório é de até 30 dias após o encerramento da pesquisa, previsto para abril de 2026.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_2262869.pdf	08/07/2024 19:29:40		Aceito
Outros	Esclarecimento_pendencias_documentais.pdf	08/07/2024 19:28:43	Carlos Eduardo Gonçalves de Oliveira	Aceito
Outros	Instrumento_de_coleta_de_dados.pdf	08/07/2024 19:16:43	Carlos Eduardo Gonçalves de Oliveira	Aceito
Declaração de Pesquisadores	Termo_Compromisso.pdf	08/07/2024 19:13:20	Carlos Eduardo Gonçalves de Oliveira	Aceito
Folha de Rosto	folhaDeRostoCarlos.pdf	24/05/2024 12:38:39	Emerson Nobuyuki Itikawa	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	DispensaTCLE.pdf	11/12/2023 15:10:49	Emerson Nobuyuki Itikawa	Aceito

Endereço: Rodovia R2, n. 3.061, Parque Tecnológico Samambaia, Edifício K2, sala 110, piso 1
Bairro: Campus Samambaia **CEP:** 74.690-970
UF: GO **Município:** GOIANIA
Telefone: (62)3521-1215 **Fax:** (62)3521-2045 **E-mail:** cep.prpi@ufg.br

UNIVERSIDADE FEDERAL DE
GOIÁS - UFG



Continuação do Parecer: 7.045.338

Orçamento	Despesas.pdf	11/12/2023 15:06:17	Emerson Nobuyuki Itikawa	Aceito
Declaração de Pesquisadores	CompromissoPesquisadores.pdf	11/12/2023 15:05:59	Emerson Nobuyuki Itikawa	Aceito
Declaração de Instituição e Infraestrutura	AnuenciaCRA.pdf	11/12/2023 15:05:09	Emerson Nobuyuki Itikawa	Aceito
Cronograma	cronograma.pdf	11/12/2023 15:04:00	Emerson Nobuyuki Itikawa	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_CEGO.pdf	11/12/2023 14:53:00	Emerson Nobuyuki Itikawa	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

GOIANIA, 31 de Agosto de 2024

Assinado por:
Rosana de Moraes Borges Marques
(Coordenador(a))

Endereço: Rodovia R2, n. 3.061, Parque Tecnológico Samambaia, Edifício K2,sala 110, piso 1
Bairro: Campus Samambaia **CEP:** 74.690-970
UF: GO **Município:** GOIANIA
Telefone: (62)3521-1215 **Fax:** (62)3521-2045 **E-mail:** cep.prpi@ufg.br

ANEXO E – Certificado: Cientista de Dados

STATEMENT OF ACCOMPLISHMENT

#456,046

HAS BEEN AWARDED TO

Carlos Eduardo Gonçalves de Oliveira

FOR SUCCESSFULLY COMPLETING

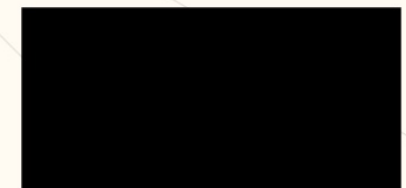
Associate Data Scientist in Python

LENGTH

86 HOURS

COMPLETED ON

MAR 25, 2024



Jonathan Cornelissen
CEO, DataCamp



ANEXO F – Certificado: Cientista de Machine Learning

STATEMENT OF ACCOMPLISHMENT

#460,400

HAS BEEN AWARDED TO

Carlos Eduardo Gonçalves de Oliveira

FOR SUCCESSFULLY COMPLETING

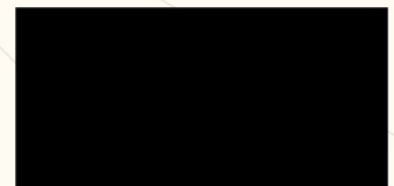
Machine Learning Scientist in Python

LENGTH

85 HOURS

COMPLETED ON

APR 08, 2024



Jonathan Cornelissen
CEO, DataCamp



ANEXO G – Certificado: The Science of Uncertainty and Data

Verified Certificate

This is to certify that

Carlos Eduardo Goncalves de Oliveira

successfully completed and received a passing grade in

6.431x: Probability - The Science of Uncertainty and Data

a course of study offered by MITx, an online learning initiative of the Massachusetts Institute of Technology.



Verified Certificate
Issued December 26, 2024

Valid Certificate ID
[a1516fa9a88e4b759ec99d7d77da6d6d](#)



John Tsitsiklis

Professor, Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology



Eric Grimson

Interim Vice President for Open Learning
Massachusetts Institute of Technology



Chris Capozzola

Senior Associate Dean for Open Learning; Professor of History
Massachusetts Institute of Technology

ANEXO H – Certificado: Machine Learning with Python - From Linear Models to Deep Learning

Verified Certificate

This is to certify that

Carlos Eduardo Goncalves de Oliveira

successfully completed and received a passing grade in

6.86x: Machine Learning with Python-From Linear Models to Deep Learning

a course of study offered by MITx, an online learning initiative of the Massachusetts Institute of Technology.



Verified Certificate
Issued May 18, 2025

Valid Certificate ID
[e6029b355a3a44b59d2bbc67b4bff890](#)



Regina Barzilay

Delta Electronics Professor of Electrical Engineering and Computer Science

Massachusetts Institute of Technology



Tommi Jaakkola

Thomas Siebel Professor of Electrical Engineering and Computer Science and the Institute for Data, Systems, and Society

Massachusetts Institute of Technology



Dimitris Bertsimas

Vice Provost for Open Learning

Massachusetts Institute of Technology



Chris Capozzola

Senior Associate Dean for Open Learning; Professor of History


Massachusetts Institute of Technology

ANEXO I – Certificado: Participação no 1º Datathon, VI ENGOPE, IME/UFG

CERTIFICADO

Certificamos que Carlos Eduardo Gonçalves de Oliveira participou do(a) Competição de dados - Datathon do VI ENGOPE, realizado em 15/10/2024 à 17/10/2024, no VI Encontro Goiano de Probabilidade e Estatística com carga horária de 32 hora(s).

Goiânia, 17 de outubro de 2024.


Amanda Buosi Gazon Milani
Coordenadora do VI ENGOPE

PROEC
PRO-REITORIA DE
EXTENSÃO E CULTURA



UFG
UNIVERSIDADE
FEDERAL DE GOIÁS

ANEXO J – Prêmio: Campeão do 1º Datathon, VI ENGOPE, IME/UFG



 **ENGOPE**
Encontro Goiano de Probabilidade e Estatística
Comemorando 15 anos
do curso de Estatística!

15, 16 e 17
de outubro de 2024
8h às 21h

Destques desta edição:

- Datathon (competição de dados com premiação)
- confraternização Esportiva (evento beneficente)
- Coffee Break todos os dias

Inscrição gratuita e mais informações:
engope.ime.ufg.br



I DATATHON (IME/UFG)

**LAPIG**

1º lugar

**LAPIG**

Vale 1 vaga em um geocurso do LAPIG – UFG*

*simbólico: haverá uma lista com o nome dos ganhadores para acesso ao prêmio.

ANEXO K – Pôster apresentado, VI ENGOPE, IME/UFG

PREDIÇÃO DE ABSENTEÍSMO DE PACIENTES EM PROCEDIMENTOS DE RESSONÂNCIA MAGNÉTICA

Carlos E. G. de Oliveira, Ana B. M. de J. Teixeira, Mateus Simão, Emerson N. Itikawa
Instituto de Matemática e Estatística - IME

c.e.g.o@discente.ufg.br, emersonitikawa@ufg.br

Introdução

O aumento da demanda por serviços de saúde, aliado às pressões por redução de custos, torna crucial a otimização dos recursos nas instituições de saúde. Um dos desafios mais significativos nesse contexto são o não-comparecimento de pacientes, que ocorre quando o paciente não comparece ao procedimento agendado sem aviso prévio. Esse problema gera desperdício de recursos e prejudica a eficiência dos serviços, particularmente em departamentos de radiologia, onde equipamentos como os de ressonância magnética (RM) são caros e não podem ficar ociosos.

Para mitigar os impactos dos não-comparecimentos, diferentes estratégias são empregadas, incluindo lembretes e overbooking, embora essas abordagens também envolvam custos adicionais. Nesse sentido, para otimizar o balanço custo-benefício dessas estratégias, propomos o uso do modelo Balanced Random Forest (BRF) para a predição do não-comparecimento de pacientes em procedimentos de ressonância magnética (RM), utilizando dados da Clínica Radiológica de Anápolis. A análise inclui a validação cruzada e a otimização de hiperparâmetros, com foco em aumentar a interpretabilidade do modelo a partir do framework SHAP para oferecer insights práticos e otimizar o uso de recursos.

Materiais e métodos

Os dados utilizados neste estudo consistem em 28.991 procedimentos de ressonância magnética agendados entre 2015 e 2023, na Clínica Radiológica de Anápolis (Anápolis, Goiás, Brasil). As informações extraídas incluíram idade, sexo, CEP, tipo de exame, detalhes de data e horário do agendamento e histórico de faltas dos pacientes. Apenas instâncias marcadas com ao menos um dia de antecedência foram consideradas, e valores ausentes foram devidamente indicados e tratados com técnicas de imputação de dados, como o método K-Nearest Neighbors.

Resultados

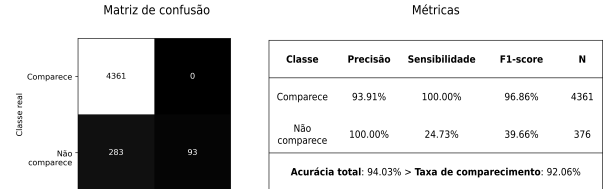


Figura 3: Matriz de confusão sobre os dados de teste (ano de 2023) e respectivas métricas.

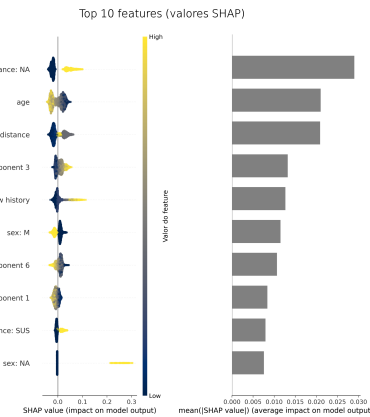


Figura 4: Top 10 features mais importantes e suas respectivas contribuições para a probabilidade de não-comparecimento.

Conclusão

Neste estudo foi desenvolvido um modelo preditivo para faltas de pacientes em exames de ressonância magnética (RM) utilizando o modelo BRF, com dados de 2015 a 2023 de uma clínica radiológica no Brasil. O modelo obteve uma acurácia de 94.03% e precisão de 100%, embora a sensibilidade tenha sido de 24.73%. A análise de interpretabilidade indicou que a ausência de informações completas de contato dos pacientes, idade e distância foram os principais preditores de faltas. Com base nos resultados, é possível sugerir ações práticas para reduzir as faltas, como a melhoria na coleta de dados de contato e a implementação de estratégias de overbooking otimizadas.

Agradecimentos

Este estudo foi financiado em parte pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Os autores também são gratos pela parceria com a Clínica Radiológica de Anápolis, sem a qual este estudo não teria sido possível.

Referências

- [1] B. P. Berg, M. Murr, D. Chermak, J. Woodall, M. Pignone, R. S. Sandler, and B. T. Denton. Estimating the cost of no-shows and evaluating the effects of mitigation strategies. *Medical Decision Making*, 33(8):976–985, 2013.
- [2] D. Carreras-García, D. Delgado-Gómez, F. Llorente-Fernández, and A. Arribas-Gil. Patient no-show prediction: a systematic literature review. *Entropy*, 22(6):675, 2020.
- [3] L. F. Dantas, J. L. Fleck, F. L. C. Oliveira, and S. Hamacher. No-shows in appointment scheduling—a systematic literature review. *Health Policy*, 122(4):412–421, 2018.

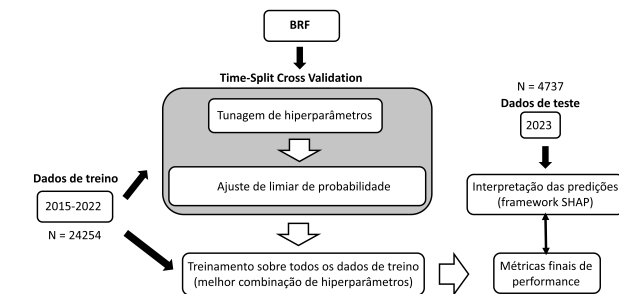


Figura 1: Visão geral da metodologia.

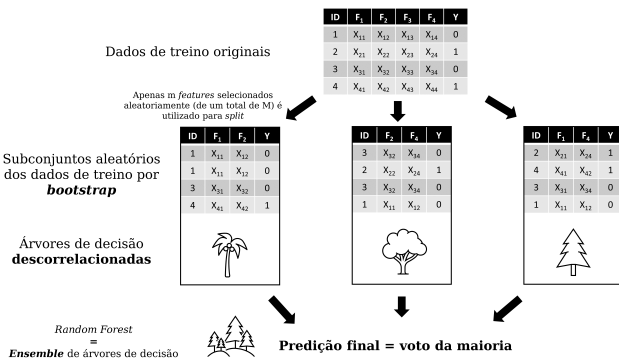


Figura 2: Modelo de classificação Random Forest.

ANEXO L – Indicação, prêmio FAPEG
N° 27/2024: pós-graduando destaque

**CHAMADA PÚBLICA FAPEG Nº 27/2024
PRÊMIO GOIANO DE CIÊNCIA, TECNOLOGIA E INOVAÇÃO - EDIÇÃO 2024**

Resultado da Seleção Interna

CATEGORIA PESQUISADOR(A) DESTAQUE		
SUB CATEGORIA	SUBÁREA DO CONHECIMENTO	NOME DO(A) INDICADO(A)
Ciências da Vida	Ciências Biológicas	José Alexandre Felizola Diniz Filho
	Ciências da Saúde	Marize Campos Valadares
	Ciências Agrárias	Abadia dos Reis Nascimento
Ciências Exatas	Ciências Exatas e da Terra	Carolina Horta Andrade
	Engenharias	Flávio Henrique Teles Vieira
	Tecnologia da Informação e Comunicação	Laura Vilela Rodrigues Rezende
Ciências Humanas	Ciências Sociais Aplicadas	Marcos Inácio Severo de Almeida
	Ciências Humanas	Márlon Herbert Flora Barbosa Soares
	Artes, Letras e Linguística	Francisco José Quaresma de Figueiredo

CATEGORIA PESQUISADOR(A) INOVADOR(A)	
SUBCATEGORIA	NOME DO(A) INDICADO(A)
Setor Empresarial	Wendell Karlos Tomazelli Coltro
Setor Público	Telma Woerle de Lima Soares

CATEGORIA PROFISSIONAL DE COMUNICAÇÃO	
Meio de Veiculação	NOME DO(A) INDICADO(A)
Internet	Carolina Abbadia Melo

CATEGORIA PÓS-GRADUANDO(A) DESTAQUE		
SUBCATEGORIA	MESTRADO OU DOUTORADO	NOME DO INDICADO
Ciências da Vida	Doutorado	Juliana Marques Ferreira
	Mestrado	Carlos Eduardo Gonçalves de Oliveira
Ciências Exatas	Doutorado	Uender Barbosa de Souza
	Mestrado	Aryadine Fernandes de Sousa
Ciências Humanas	Doutorado	Izaias de Souza Silva
	Mestrado	Marcus Vinícius Turíbio de Brito

CATEGORIA PESQUISADOR(A) GOIANO(A) DESTAQUE EM OUTROS ESTADOS		
NOME DO INDICADO	IES/ICT de origem em Goiás	IES/ICT de atuação em outro estado
Karla Maria Longo	Universidade Federal de Goiás (UFG)	Instituto Nacional de Pesquisas Espaciais (INPE)

CATEGORIA PESQUISADOR(A) GOIANO(A) DESTAQUE NO EXTERIOR		
NOME DO INDICADO	IES/ICT de origem em Goiás	IES/ICT de atuação no exterior
Simone Antoniaci Tuzzo	Universidade Federal de Goiás (UFG)	Universidade da Maia – Umaia – Porto - Portugal

Comissão Avaliadora

Fabíola Souza Fiaccadori
Luizmar Adriano Junior
Marinaldo Divino Ribeiro