

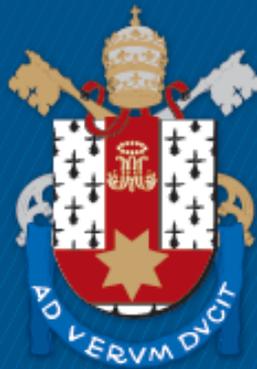
ESCOLA POLITÉCNICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO  
MESTRADO EM CIÊNCIA DA COMPUTAÇÃO

DIEGO CANDIDO DE SOUZA

**MODELOS DE APRENDIZADO DE MÁQUINA PARA  
PREDIÇÃO DE ÓBITO DE PACIENTES EM HEMODIÁLISE**

Porto Alegre  
2024

PÓS-GRADUAÇÃO - *STRICTO SENSU*



Pontifícia Universidade Católica  
do Rio Grande do Sul

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL  
ESCOLA POLITÉCNICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO**

**MODELOS DE APRENDIZADO  
DE MÁQUINA PARA PREDIÇÃO  
DE ÓBITO DE PACIENTES EM  
HEMODIÁLISE**

**DIEGO CANDIDO DE SOUZA**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Ciência da Computação na Pontifícia Universidade Católica do Rio Grande do Sul.

Orientadora: Prof<sup>ª</sup>. Dra. Soraia Raupp Musse

**Porto Alegre  
2024**

## Ficha Catalográfica

D278m de Souza, Diego Candido

Modelo de Aprendizado de Máquina para predição de óbito de pacientes em hemodiálise / Diego Candido de Souza. – 2024.

50 p.

Dissertação (Mestrado) – Programa de Pós-Graduação em Ciência da Computação, PUCRS.

Orientadora: Profa. Dra. Soraia Raupp Musse.

1. Machine Learning. 2. Biomarcadores. 3. Inteligência Artificial. 4. Diretriz Clínica. 5. Diálise. I. Musse, Soraia Raupp. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da PUCRS  
com os dados fornecidos pelo(a) autor(a).

Bibliotecária responsável: Clarissa Jesinska Selbach CRB-10/2051

**DIEGO CANDIDO DE SOUZA**

**MODELOS DE APRENDIZADO DE MÁQUINA  
PARA PREDIÇÃO DE ÓBITO DE PACIENTES EM  
HEMODIÁLISE**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre em Ciência da Computação do Programa de Pós-Graduação em Ciência da Computação, Escola Politécnica da Pontifícia Universidade Católica do Rio Grande do Sul.

Aprovado(a) em 18 de setembro de 2024.

**BANCA EXAMINADORA:**

Prof<sup>a</sup>. Dra. Milene Selbach Silveira (PPGCC/PUCRS)

Prof. Dr. Frederico Orlando Friedrich (PPGM)

Prof<sup>a</sup>. Dra. Soraia Raupp Musse (PPGCC/PUCRS - Orientadora)

## **DEDICATÓRIA**

Dedico este trabalho à minha família, que sempre esteve ao meu lado nos momentos mais difíceis e celebrando cada conquista ao longo dessa jornada. Agradeço especialmente aos meus pais, por terem me mostrado o valor da educação e por todo o apoio incondicional que sempre me deram.

À minha esposa, Fernanda, minha companheira de todas as horas, que com sua paciência, compreensão e amor foi meu porto seguro nos momentos mais desafiadores. Seu apoio e incentivo foram fundamentais para que eu pudesse seguir em frente.

## **AGRADECIMENTOS**

A minha orientadora, Prof<sup>a</sup>. Dr<sup>a</sup>. Soraia Raupp Musse, Prof. Dr. Carlos Eduardo Poli-de-Figueireda, Dr<sup>a</sup>. Isabel Cristina Reinheimer, Prof. Dr. Rafael Heitor Bordini e os alunos Guilherme Vásquez, Henrique Rocha, Lucas Weiss, Rodrigo M. Chaves, Arthur E. Paschoal, Alexius M. Dias e João Pedro T. Carvalho, que apoiaram e contribuíram para o sucesso deste projeto.

# MODELOS DE APRENDIZADO DE MÁQUINA PARA PREDIÇÃO DE ÓBITO DE PACIENTES EM HEMODIÁLISE

## RESUMO

No Brasil, 133.464 pessoas com Doença Renal Crônica (DRC) estão em hemodiálise, com mortalidade. Os biomarcadores obrigatórios para o acompanhamento desses pacientes são determinados pela Diretriz Clínica do Ministério da Saúde de 2014. Mais de trinta biomarcadores são avaliados anualmente. Nenhuma avaliação crítica sobre o valor preditivo desses biomarcadores foi realizada no país, utilizando Aprendizado de Máquina (AM). Assim, o presente projeto tem o objetivo de desenvolver modelos de AM que permitam prever o desfecho de óbito em pacientes em hemodiálise com base nos biomarcadores de rotina. Pretende-se, com isso, gerar instrumentos de monitoramento que possibilitem avaliar a efetividade preditiva dos exames clínicos, a fim de melhorar a qualidade de vida dos pacientes e contribuir na gestão de custos do Sistema Único de Saúde (SUS). Neste trabalho, são utilizados dados de pacientes incidentes em hemodiálise de um estudo de coorte retrospectivo realizado de 2012 a 2019 em 23 unidades de diálise de cinco estados brasileiros. As features utilizadas no desenvolvimento dos modelos são os biomarcadores, variáveis de perfil dos pacientes e desfechos clínicos. Serão testadas diferentes abordagens e algoritmos de AM usando os modelos de *Decision Tree*, *Random Forest*, *Logistic Regression* e *XGBoost ML* no intuito de identificar o melhor modelo preditivo.

**Palavras-Chave:** Machine Learning, Biomarcadores, Inteligência Artificial, Diretriz Clínica, Diálise.

# MACHINE LEARNING MODELS FOR DEATH PREDICTION OF PATIENTS UNDER HEMODIALYSIS

## ABSTRACT

In Brazil, 133,464 people with Chronic Kidney Disease (CKD) are on hemodialysis with mortality of 19.5% per year. The mandatory biomarkers for monitoring these patients are determined by the 2014 Ministry of Health Clinical Guideline. More than thirty biomarkers are evaluated annually, with a third being assessed monthly. No critical assessment of the predictive use value of these biomarkers has been carried out in the country using Machine Learning (ML). Therefore, the present project aims to develop AM models that allow predicting the outcome of death in hemodialysis patients based on routine biomarkers. The aim is to generate monitoring instruments that make it possible to evaluate the predictive effectiveness of clinical examinations in order to improve patients' quality of life and contribute to the cost management of the Unified Health System (SUS). In this work, data from incident hemodialysis patients from a retrospective cohort study carried out from 2012 to 2019 in 23 dialysis units in five Brazilian states are used. The features used in the development of the models are biomarkers, patient profile variables and clinical outcomes. Different ML approaches and algorithms will be tested using the *Decision Tree*, *Random Forest*, *Logistic Regression* and *XGBoost ML* models in order to identify the best predictive model.

**Keywords:** Machine Learning, Biomarkers, Artificial intelligence, Clinical Guideline, Hemodialysis.

## LISTA DE FIGURAS

Figura 2.1 – Visualização do dataset no usando a biblioteca Pandas . . . . .	19
Figura 3.1 – Fluxo processamento dos modelos . . . . .	23
Figura 4.1 – Matriz de Confusão com <i>Decision Tree</i> para o <i>dataset</i> DBRaw . . . . .	31
Figura 4.2 – Matriz de Confusão com <i>Logistic Regression</i> para o <i>dataset</i> DBRaw .	31
Figura 4.3 – Matriz de Confusão com <i>XGBoost</i> para o <i>dataset</i> DBRaw . . . . .	32
Figura 4.4 – Matriz de Confusão com <i>Random Forest</i> para o <i>dataset</i> DBRaw . . . . .	32
Figura 4.5 – Random Forest SHAP. Teste usando todos os biomarcadores no <i>dataset</i> DBRaw . . . . .	33
Figura 4.6 – Matriz de Confusão com <i>Decision Tree</i> para o <i>dataset</i> DBCompleted	35
Figura 4.7 – Matriz de Confusão com <i>Logistic Regression</i> para o <i>dataset</i> DBCompleted . . . . .	35
Figura 4.8 – Matriz de Confusão com <i>XG Boost</i> para o <i>dataset</i> DBCompleted . . . . .	36
Figura 4.9 – Matriz de Confusão com <i>Random Forest</i> para o <i>dataset</i> DBCompleted	36
Figura 4.10 – Logistic Regression. Teste usando todos os biomarcadores no <i>dataset</i> DBCompleted . . . . .	37
Figura 4.11 – Matriz de Confusão com <i>Decision Tree</i> para o DBClean . . . . .	39
Figura 4.12 – Matriz de Confusão com <i>Logistic Regression</i> para o <i>dataset</i> DBClean	39
Figura 4.13 – Matriz de Confusão com <i>XG Boost</i> para o <i>dataset</i> DBClean . . . . .	40
Figura 4.14 – Matriz de Confusão com <i>Random Forest</i> para o <i>dataset</i> DBClean . . . . .	40
Figura 4.15 – Logistic Regression. Teste usando todos os biomarcadores no <i>dataset</i> DBClean . . . . .	41
Figura 4.16 – Métrica AUC ROC para o <i>dataset</i> DBRaw . . . . .	42
Figura 4.17 – Métrica AUC ROC para o <i>dataset</i> DBCompleted . . . . .	42
Figura 4.18 – Métrica AUC ROC para o <i>dataset</i> DBClean . . . . .	43

## LISTA DE TABELAS

Tabela 3.1 – Atributos coletados dos pacientes. ....	22
Tabela 3.2 – Resumo dos <i>datasets</i> .....	24
Tabela 3.3 – DBCompleted: Pacientes que precisam de completudes em suas informações para ficarem completos .....	25
Tabela 3.4 – Visão geral dos <i>datasets</i> .....	26
Tabela 4.1 – Executados 4 modelos de AM no <i>dataset</i> DBRaw. ....	30
Tabela 4.2 – Executados 4 modelos de AM no <i>dataset</i> DBCompleted .....	34
Tabela 4.3 – Executados 4 modelos de AM no <i>dataset</i> DBClean .....	38

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>12</b>
1.1	JUSTIFICATIVA	13
1.2	ORGANIZAÇÃO DO TRABALHO	14
<b>2</b>	<b>CONTEXTUALIZAÇÃO</b>	<b>15</b>
2.1	TRABALHOS RELACIONADOS	15
2.2	INTRODUÇÃO À TECNOLOGIAS USADAS EM APRENDIZADO DE MÁQUINA	17
2.2.1	LINGUAGEM PYTHON	17
2.2.2	JUPYTER NOTEBOOK	18
2.2.3	NUMPY	18
2.2.4	PANDAS	18
2.2.5	MATPLOTLIB	19
2.2.6	SCIKIT-LEARN	19
2.3	O CONTEXTO DO PRESENTE TRABALHO	20
<b>3</b>	<b>METODOLOGIA</b>	<b>21</b>
3.1	DADOS E PRÉ-PROCESSAMENTO	21
3.2	MÉTRICAS	26
3.3	ALGORITMOS DE APRENDIZADO DE MÁQUINA UTILIZADOS	27
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b>	<b>30</b>
4.1	RESULTADOS COM O <i>DATASET</i> DBRAW	30
4.2	RESULTADOS COM O <i>DATASET</i> DBCOMPLETED	33
4.3	RESULTADOS COM O <i>DATASET</i> DBCLEAN	37
4.4	CURVA RECEIVER OPERATING CHARACTERISTIC (ROC) DOS MODELOS	41
4.5	DISCUSSÃO	43
<b>5</b>	<b>CONSIDERAÇÕES FINAIS</b>	<b>45</b>
	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>47</b>

## Lista de Abreviaturas

**DRC** Doença Renal Crônica

**IA** Inteligência Artificial

**ROC** *Receiver Operating Characteristic*

**AUC** *Assets Under Custody*

**RNA** Rede Neural Artificial

**CNN** Rede Neural Convolucional

**GBM** *Gradient Boosting Machine*

**PLN** Processamento de Linguagem Natural

**AM** Aprendizado de Máquina

**SHAP** *SHapley Additive exPlanations*

**TRS** Terapia Renal Substitutiva

**TFG** Taxa de filtração glomerular

**LDL** Lipoproteína de baixa densidade

**Ca** Cálcio

**Na<sup>+</sup>** Íons Sódio

**K<sup>+</sup>** Potássio

**P** Proteína

**Hb** Hemoglobina

**PTH<sub>i</sub>** Hormônio da paratireoide

## 1. INTRODUÇÃO

A hemodiálise é um tratamento de substituição renal frequentemente usado para pacientes com doença renal crônica avançada. A previsão de desfecho desses pacientes é um problema complexo e de grande importância para a saúde pública, pois a taxa de mortalidade é alta e os exames solicitados aos pacientes trazem custos significativos para o sistema de saúde público no Brasil (SUS).

Pacientes com insuficiência renal crônica devem fazer hemodiálise, tratamento indispensável que permite a filtração artificial do sangue e a remoção das toxinas dos rins não saudáveis. No entanto, apesar do desenvolvimento da tecnologia e da medicina, a mortalidade entre os pacientes submetidos a esse tratamento permanece bastante alta [30]. O reconhecimento precoce de indivíduos pertencentes a um grupo com elevado risco de morte permite realizar intervenções preventivas adequadas e personalizadas, o que ajuda a melhorar os resultados clínicos. O Aprendizado de Máquina (AM) começou a atuar como uma ferramenta muito promissora para prever resultados hospitalares tão graves com base em grandes diversidades de dados complexos. Os modelos possuem uma alta capacidade de detectar padrões ocultos em grandes conjuntos de dados, permitindo extrair conhecimento não disponível com métodos estatísticos tradicionais [19].

Para a sobrevivência desses pacientes, é necessário realizar a Terapia Renal Substitutiva (TRS), na modalidade de diálise peritoneal, hemodiafiltração ou hemodiálise [9]. No Brasil, cerca de 144.779 pessoas fazem o TRS, sendo que 92% desses pacientes estão realizando a hemodiálise. A taxa de mortalidade bruta desses pacientes chegou a 24,5% ao ano de 2020 [31]. Nos casos extremos, a mortalidade desses pacientes é de 10 para cada 100 pacientes da mesma idade. Os gastos com procedimentos de hemodiálise, no Brasil, aumentaram de \$1,5 bilhões (em 2009) para \$2,9 bilhões (em 2018), totalizando o montante de \$22,4 bilhões de reais nesse período [1]. O desafio para a computação é lidar com grandes volumes de dados na área da saúde, assim como a escassez de dados fornecidos pelo Sistema Único de Saúde, acarretando dificuldade de unificação dos dados em uso clínico.

Na literatura internacional, observa-se que há diferentes publicações de estudos sobre o desenvolvimento de modelos de aprendizado de máquina (AM) usando *dataset* da DRC, pois se tem pouca quantidade de informação armazenada. Um exemplo, são os inúmeros estudos que utilizam o *Machine Learning Repository* da *University of California* que foi criado lá em 1987 pelo estudante David Aha e seus colegas de graduação. Este *dataset* possui apenas 25 atributos de pacientes com DRC e 400 instâncias com dados faltantes e divergentes, com isso não se tem um resultado preciso usando modelos de AM.

Wickramasinghe et al. (2017) desenvolveram um modelo para analisar e prever planos de dieta para pacientes com DRC utilizando o *Machine Learning Repository*. Foram utilizados 21 atributos, dentre eles a idade, pressão arterial e resultados dos exames realizados nesses pacientes. Quatro algoritmos de aprendizado de máquina foram avaliados (*random forests*, *logistic regression*, *neural network* e *random jungle*) obtendo uma acurácia de 82.5% - 99.17% [41].

Gunarathne et al. (2017) publicaram outro estudo usando o mesmo *dataset*, considerando apenas quatorze atributos, desenvolvendo um classificador para ajudar no diagnóstico de DRC [16]. Foram investigados os mesmos quatro algoritmos (*random forests*, *logistic regression*, *neural network* e *random jungle*) e todos tiveram um poder preditivo entre 95.0 e 99.1%. Apesar dos resultados promissores, ambos os estudos precisam de validação externa, ou seja, serem testados com outras bases de dados mais robustas para identificar a acurácia e a relevância desses modelos.

Diante do exposto, o presente projeto possui como objetivo principal desenvolver modelos de Aprendizado de Máquina que permitam prever óbito de pacientes em hemodiálise com base nos biomarcadores de rotina, utilizando um *dataset* de 23 centros de diálise do Brasil.

Os objetivos específicos são: *i*) Identificar quais exames (biomarcadores) têm maior impacto no modelo preditivo de óbito; e *ii*) Determinar qual dos modelos testados possui melhor desempenho preditivo no caso de estudo.

## **1.1 Justificativa**

Para investigar esse problema, modelos de aprendizado de máquina têm sido propostos na literatura [26]. Vários modelos de aprendizado de máquina foram desenvolvidos, por exemplo, Cesconetto [5] sugeriu utilizar aprendizado de máquina para prever a mortalidade nesses pacientes. Os autores usaram uma série de variáveis clínicas como entradas para o modelo, incluindo idade, sexo, tempo desde o início da diálise e vários indicadores laboratoriais. Ainda outros autores da literatura estudaram esse contexto.

Outros estudos utilizaram técnicas como regressão logística e árvores de decisão para prever óbitos em pacientes em hemodiálise [2] e [29]. No entanto, esses estudos também destacaram a necessidade de mais pesquisas para melhorar a qualidade dos resultados destes algoritmos. A utilização de modelos de aprendizado de máquina na saúde é uma tendência que tem se fortalecido cada vez mais. Diversos estudos têm investigado o uso dessas técnicas para otimizar a tomada de decisões clínicas, melhorando a qualidade do atendimento e os resultados dos pacientes [35]. No contexto específico da hemodiálise, a predição é um aspecto crítico que pode beneficiar-se significativamente da aplicação desses modelos.

Gunarathne et al. (2017) [16] publicaram um estudo desenvolvendo um classificador para ajudar no diagnóstico de Doença Renal Crônica, no qual foram executados quatro algoritmos (*random forests*, *logistic regression*, *neural network* e *random jungle*) e todos tiveram um poder preditivo entre 95,0 e 99,1%. Os resultados promissores desses estudos precisam de validação externa, ou seja, serem testados com outras bases de dados mais robustas para identificar a relevância desses modelos. A análise dos biomarcadores pode revelar fatores determinantes na mortalidade desses pacientes, proporcionando uma ferramenta valiosa para profissionais da área de saúde [32].

Estudos anteriores demonstraram que técnicas como redes neurais artificiais podem alcançar alta acurácia na predição de eventos clínicos adversos [39]. Este estudo justifica-se pela necessidade urgente e crescente no campo da nefrologia de métodos eficazes e precisos para o monitoramento e previsão da evolução clínica dos pacientes. A utilização adequada dos modelos preditivos poderá auxiliar médicos a tomarem decisões com informações baseadas em evidências quantitativas robustas, potencialmente salvando vidas ao antecipar problemas graves antes mesmo que eles se manifestem clinicamente.

## **1.2 Organização do Trabalho**

No presente trabalho, a Revisão da Literatura é apresentada no Capítulo 2, enquanto o método desenvolvido é descrito no Capítulo 3, seguido pela apresentação da análise e representação gráfica dos resultados no Capítulo 4. Por fim, o Capítulo 5 apresenta considerações finais e propõe direções futuras para o trabalho.

## 2. CONTEXTUALIZAÇÃO

Géron (2019) [17] definiu o aprendizado de máquina como “um campo de estudo que dá aos computadores a capacidade de operar sem programação explícita”. O aprendizado de máquina é, portanto, definido como o campo da inteligência artificial (IA) que busca desenvolver algoritmos que permitem que as máquinas façam tarefas específicas.

O aprendizado não supervisionado consiste em encontrar relações em dados de entrada não rotulados, viabilizando o encontro de padrões, estruturas e representações sobre os dados de treinamento para identificar o conjunto de dados de treinamento, podendo ser utilizados para atividades, como clusterização, detecção de anomalias, redução de dimensionalidade e visualização de conjuntos de dados [17].

Os algoritmos de aprendizagem supervisionado são categorizados em duas classes, de acordo com sua abordagem: regressão e classificação. Algoritmos de regressão são empregados para estimar o valor de uma variável inserida em um local específico com base nas propriedades de entrada e permitem executar modelos de regressão para projetos, como saber o valor de uma ação e fazer previsões de valores futuros. Métodos de classificação são empregados para determinar o tipo da variável alvo a partir do processamento nos dados de entrada [17].

Os modelos de aprendizado utilizam dados de treinamento para classificar e categorizar respostas, fazendo a transformação da variável alvo em um vetor de entrada. São aplicados em várias tarefas, como classificação de imagens, análise de sentimento e detecção de fraudes, entre outras. Existem vários métodos de classificação que podem ser empregados em um esforço de aprendizado de máquina, o que permite fazer a seleção do algoritmo a ser empregado no modelo, que deve considerar fatores como facilidade de implementação, atributos do conjunto de dados, dificuldade da tarefa a ser empregada, eficiência computacional, precisão desejada e interpretabilidade.

Para Géron (2019) [17], os principais métodos computacionais empregados para classificação incluem regressão logística, *Naive Bayes*, *k-nearest neighbors algorithm*, *Decision Tree*, *Logistic Regression*, *Random Forest* e *Support Vector Machine*. Para avaliar a eficiência da aplicação desses algoritmos, existem diversas métricas, com destaque para o índice de acurácia, a matriz de confusão, curva *Receiver Operating Characteristic (ROC)* e índice *Assets Under Custody (AUC)*.

### 2.1 Trabalhos relacionados

Os modelos de aprendizado de máquina prometeram nos últimos anos prever resultados clínicos em vários contextos médicos, incluindo a hemodiálise. A hemodiálise é

um procedimento crítico para pacientes com insuficiência renal crônica, e a previsão dos resultados hospitalares pode contribuir muito para melhorar o atendimento e o manejo desses pacientes. O objetivo desta revisão da literatura é compilar e analisar as pesquisas mais recentes sobre o uso de aprendizado de máquina para prever resultados hospitalares em pacientes em hemodiálise. Deve-se notar que os modelos de aprendizado de máquina são, por sua natureza, poderosos o suficiente para lidar com a análise de excesso de dados, além de serem capazes de detectar padrões críticos.

Estudos como o de Goto et al. (2020) [14] mostram que algoritmos de redes neurais artificiais (RNAs) são capazes de ser usados de forma eficaz para prever a mortalidade de pacientes em diálise. Além disso, métodos mais sofisticados, como modelos baseados em árvores de decisão, também provaram ser eficazes na previsão de resultados clínicos com alta precisão. A precisão dos modelos de aprendizado de máquina depende muito da qualidade e quantidade dos dados disponíveis. Por exemplo, no trabalho de Montemayor et al. (2019) [12], foi demonstrado que a inclusão de variáveis clínicas refinadas, como parâmetros laboratoriais e históricos médicos completos, melhoram significativamente o desempenho preditivo. Outra área emergente é a utilização da aprendizagem profunda. Isso pode lidar com conjuntos de dados ainda maiores e discernir relações entre variáveis mais complicadas. Um estudo recente de Kim et al. (2021) [25] usou Redes Neurais Convolucionais (CNN) para análise médica baseada em imagens, juntamente com dados tabulares tradicionais, e alcançou maior precisão na previsão de resultados hospitalares.

O uso de modelos de aprendizado de máquina no domínio da saúde ganhou recentemente atenção substancial na previsão de resultados hospitalares para pacientes em hemodiálise. Segundo Kourou et al. (2015) [27], esses modelos podem revolucionar os sistemas de saúde, fornecendo previsões precisas e personalizadas, o que seria de grande ajuda na prestação de cuidados de qualidade aos pacientes. Alguns trabalhos de pesquisa recentes aplicaram diversas metodologias e algoritmos para prever os resultados de pacientes com doença renal crônica. Por exemplo, Chen et al. (2020) [7] combinaram redes neurais profundas com algoritmos tradicionais como Random Forests para prever mortalidade em pacientes em hemodiálise, com boa precisão e taxas de sensibilidade com base em seus resultados. A aplicação de técnicas baseadas em aprendizagem profunda é vantajosa, uma vez que esses modelos podem lidar com excesso de dados em diferentes formas de fontes de informação devido à sua heterogeneidade, como prontuários médicos eletrônicos e dados clínicos históricos [36].

Esses métodos permitem capturar padrões complicados que podem não ser capturados usando métodos estatísticos tradicionais. Outra abordagem adequada é a utilização de *Gradient Boosting Machines* (GBM), que tem demonstrado sua eficácia no tratamento de dados desbalanceados, cenário muito frequente na área médica. De acordo com Ke et al. (2017) [24], os dados podem se ajustar bem a conjuntos de dados complexos e fornecer robustez ao *overfitting*, o que é um problema comum quando não se tem

uma quantidade adequada de dados clínicos. O progresso nas metodologias de processamento de linguagem natural (PLN) também contribuiu para o aumento da precisão das previsões em contextos clínicos.

Huang et al. (2021) [22] mostraram que, após combinar o processamento de linguagem natural com redes neurais, é possível extrair características muito importantes de registros médicos textuais, ajudando assim a melhorar o poder preditivo dos modelos. O desenvolvimento contínuo na área de aprendizado de máquina pode ser uma excelente oportunidade para melhorar as previsões de produção hospitalar com base em pacientes em hemodiálise. No entanto, é igualmente importante observar as preocupações éticas e a transparência dos processos como forma de aumentar a confiança entre os prestadores de cuidados de saúde e os seus pacientes.

O uso de modelos preditivos para resultados hospitalares baseados em algoritmos de aprendizado de máquina vem sendo tema importante na literatura, dada a sua capacidade de melhorar as decisões clínicas e a qualidade da prestação de cuidados nesta época. Sua aplicabilidade pode ser muito útil, principalmente em casos que envolvem pacientes em hemodiálise com comorbidades muito complicadas [13].

De acordo com um estudo mais recente de Zhang et al. (2022) [21], a adição dessas informações pode aumentar significativamente a eficiência dos modelos preditivos. Vários modelos baseados em aprendizado de máquina, como Redes Neurais Artificiais (RNAs) e Florestas Aleatórias, foram extensivamente estudados para a previsão de muitos resultados clínicos. As evidências mostram que eles podem ser melhores do que os métodos tradicionais em termos de precisão e sensibilidade preditiva. No desenvolvimento de tais modelos, a incorporação de variáveis clínicas é um fator determinante para sua excelência.

## **2.2 Introdução à Tecnologias usadas em Aprendizado de Máquina**

Dada a natureza interdisciplinar dessa dissertação, mencionam-se nesta seção brevemente algumas tecnologias atualmente utilizadas na área.

### **2.2.1 Linguagem Python**

Conhecido por sua simplicidade e facilidade de uso, Python se destaca como uma linguagem de programação de alto nível que surgiu em 1991. Ganhou popularidade entre profissionais e acadêmicos de diversas áreas devido à sua acessibilidade e ao apoio da comunidade de que dispõe. A rica biblioteca e o ambiente do Python garantem que a

experiência do usuário/desenvolvedor seja tranquila. Python é amplamente reconhecido no desenvolvimento de software e computação científica [15].

### 2.2.2 Jupyter Notebook

O Projeto Jupyter nasce como parte do ecossistema Python em 2014 e inclui o indispensável Jupyter Notebook. Um empreendimento de código aberto que visa fornecer um ambiente interativo para ciência de dados e computação científica que abrange diferentes linguagens de programação. Os notebooks Jupyter – sendo aplicativos baseados na Web – permitem que os usuários criem e compartilhem documentos que combinam código ao vivo com texto narrativo, equações e visualizações. Esta estrutura binária garante que o processo computacional seja transparente, enquanto a explicação que o rodeia serve fins pedagógicos ou apoia relatórios de análise de dados [15].

### 2.2.3 NumPy

A principal biblioteca de código aberto para computação numérica em Python é NumPy. Seu foco é produzir arrays multidimensionais e um conjunto de funções para operar nesses arrays. O NumPy surgiu em 2005, quando dois estudantes de graduação fundiram as bibliotecas Numeric e Numarray, evoluindo mais tarde para ser a base da rica pilha de software científico do Python. No campo das iniciativas de ciência de dados, o NumPy apresenta capacidades de criação e manipulação de arrays multidimensionais que garantem alto desempenho. Esta biblioteca inclui funções integradas que suportam álgebra e estatística (juntamente com cálculos), permitindo assim aplicações fáceis em todos os níveis com entrada vetorial ou matricial. Assim, o NumPy acaba sendo uma arma obrigatória no arsenal de qualquer pessoa envolvida com trabalhos de ciência de dados [15].

### 2.2.4 Pandas

Uma ferramenta dinâmica e poderosa usada para lidar facilmente com dados de tabelas. Construído em 2008, o Pandas foi desenvolvido e construído com Python. Foi criada com o objetivo principal de permitir a organização e análise de dados de forma eficaz, por meio do fornecimento de um rico conjunto de recursos (funções e métodos), que facilitam as diversas tarefas na análise dos dados. O Pandas é construído em torno do conceito de uma tabela bidimensional chamada *DataFrame*, que é capaz de armazenar

dados de diferentes tipos em colunas. Ele forma a estrutura central do Pandas porque através dele muitas operações podem ser realizadas – por exemplo, importar dados de diferentes formatos de arquivo, como CSV, JSON, Excel ou até mesmo tabelas de banco de dados como SQL [15].

A demonstração do poder e das capacidades do Pandas pode ser feita usando a função `.describe()` para gerar um resumo completo das estatísticas descritivas para o primeiro `DataFrame`. Na Figura 2.1 podemos ver o que este método retornou. Esta função nos fornece diferentes medidas estatísticas para cada característica que incluem: contagem, média, desvio padrão, valor mínimo, quartis (25%, 50% e 75%) e valor máximo - aplicando esta abordagem.

Código	Mês/ano	Sexo	Idade	Índice de Massa Corporal	Anti-HBs (Titulação)	Anti-HBs	HBs Ag	Anti-HIV	Taxa de Redução da Uréia	K+	Na+	HB do mês atual	Ferritina	Saturação de Transferrina	Alb. ln(Vende bromocresol)	Proteína Total	Coolesterol Total	La Total do mês atual	P	Último resultado do PTH	Fosfatase Alcalina Total	desfecho	
33400048	2014-09-01	Masculino	50	28.8	10	Positivo	Negativo	Negativo	75.7	4.8	134	12.8	1283	38	4.1	7100	170.0	10.0	4.7	992	71	Vivo	
33400048	2014-10-01	Masculino	50	28.8	10	Positivo	Negativo	Negativo	75.8	4.8	133	13.2	1283	38	4.1	7100	170.0	9.7	5.5	992	71	Vivo	
33400048	2014-11-01	Masculino	50	28.8	10	Positivo	Negativo	Negativo	59.7	5.8	135	10.3	1431	40	3.8	6900	170.0	9.8	8.0	1228	94	Vivo	
33400048	2014-12-01	Masculino	50	28.8	10	Positivo	Negativo	Negativo	72.7	5.2	148	8.0	1431	40	3.8	6900	170.0	8.4	7.4	1228	94	Vivo	
33400048	2015-01-01	Masculino	51	28.8	10	Positivo	Negativo	Negativo	71.3	3.7	134	8.8	1431	40	3.8	6900	170.0	9.4	0.8	1228	94	Vivo	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
BRV0010053	2016-08-01	Masculino	71	25.3	17	Positivo	Negativo	Negativo	71.9	5.4	134	10.8	59	12	3.3	6300	110.0	8.4	4.2	277	54	Vivo	
BRV0010053	2016-09-01	Masculino	71	25.3	17	Positivo	Negativo	Negativo	79.8	5.3	137	11.0	59	12	3.3	6300	110.0	8.8	5.5	277	54	Vivo	
BRV0010053	2016-10-01	Masculino	71	25.4	17	Positivo	Negativo	Negativo	73.5	5.4	132	9.5	59	12	3.3	6300	110.0	8.3	8.1	277	54	Vivo	
BRV0010053	2016-11-01	Masculino	71	25.5	16	Positivo	Negativo	Negativo	74.4	5.0	139	9.5	13	29	3.8	7800	110.0	9.1	4.9	110	88	Vivo	
BRV0010053	2016-12-01	Masculino	71	25.5	16	Positivo	Negativo	Negativo	71.4	5.3	134	11.1	13	29	3.8	7800	110.0	9.0	4.8	110	88	Vivo	

Figura 2.1 – Visualização do dataset no usando a biblioteca Pandas

## 2.2.5 Matplotlib

O mundo científico passou a adotar o Matplotlib como uma ferramenta de visualização de dados devido ao seu conjunto de recursos e à sua compatibilidade com outras bibliotecas de ciência de dados (NumPy, Pandas). O Matplotlib não é apenas uma ferramenta, é um recurso poderoso que permite aos usuários ter controle total sobre a personalização dos gráficos e os estilos. Matplotlib, de fato, é superior na visualização de dados pela comunidade científica devido à sua ampla gama de recursos que acompanham um robusto pacote e uma conectividade perfeita com outras bibliotecas do Python, como, por exemplo, o NumPy e o Pandas [18].

## 2.2.6 Scikit-learn

Scikit-learn é uma biblioteca do Python (código aberto), sendo uma ferramenta incrivelmente robusta que possibilita que os usuários possam realizar análises preditivas de dados e participar de várias tarefas de aprendizado de máquina. Esta biblioteca abrange todas as funcionalidades necessárias durante todo o ciclo de vida de um projeto

de aprendizado de máquina. Do dimensionamento e normalização de dados à implementação de codificação *one-hot*, aplicação de algoritmos de aprendizagem supervisionados e não supervisionados, otimização de hiper parâmetros, utilização de técnicas de validação cruzada e avaliação de classificações usando métricas, o Scikit-learn cobre tudo [18].

### 2.3 O contexto do presente trabalho

A inteligência artificial vem sendo amplamente utilizada em diversas áreas médicas para promover o desenvolvimento terapêutico, pois tem contribuído para detecção precoce, diagnósticos e conduta clínica. A literatura demonstra que a IA tem *performance* comparável aos médicos *experts*. Possui alta sensibilidade e especificidade na detecção e apresenta vantagens de emitir relatórios imediatos e análise de resultados. Contudo, são os *experts* que informam à IA as predições desejadas. Assim, o uso da inteligência artificial não pode existir na medicina sem humanos e seu sucesso deve-se aos profissionais que treinam e avaliam seus diagnósticos.

Reinheimer et al. (2023) [23] os preditores com impactos significativos nos resultados foram albumina, TFG, creatinina sérica, triglicerídio, colesterol total, idade e LDL. Os algoritmos com a melhor precisão foram *Elastic Net*, *Lasso*, *Ridge*, e regressão linear. Apesar do tamanho amostral, os autores destacaram que, quando se trata de aplicações específicas, os dados podem ser escassos, restringindo o tamanho do *dataset*. Contudo, os resultados obtidos indicaram um poder de análise satisfatório [38].

Recentemente, Nusinovici et al. (2020) [32] compararam um modelo de regressão logística com outros cinco algoritmos de AM (*single-hidden-layer neural network*, *radial basis support vector machine*, *random forest*, *gradient boosting machine* e *k-nearest neighbor*) em 10.333 indivíduos, para prever a incidência de doenças cardiovasculares, doença renal crônica, diabetes e hipertensão. Dentre os modelos, a regressão logística obteve o melhor desempenho ao prever casos de Doença Renal Crônica e Diabetes, ressaltando a importância do uso de modelos clássicos de regressão. Além disso, foi apontado pelo estudo que os valores de Taxa de filtração glomerular, idade e albumina possuem grande importância ao prever casos de Doença renal crônica.

O presente trabalho testará algumas metodologias da literatura em dados reais e coletados no Brasil.

### **3. METODOLOGIA**

A abordagem desta pesquisa foi quantitativa, focada na análise de dados obtidos de prontuários médicos de pacientes em hemodiálise. A pesquisa quantitativa tem como característica o uso de coleta de informações, por meio de técnicas estatísticas e de aprendizado [37].

Na primeira parte do estudo, foi definida a população-alvo, a qual foram os pacientes que estão ou estiveram em tratamento de hemodiálise. Na segunda parte, foi selecionada uma amostra nos prontuários médicos dos pacientes. Esses dados coletados incluíram informações demográficas, como idade e sexo, além dos resultados dos biomarcadores relacionados ao tratamento da hemodiálise. Todos os dados foram anonimizados para preservar a privacidade dos pacientes [17].

Este estudo teve seu projeto registrado junto a Plataforma Brasil através do CAEE 53284121.5.0000.5336 e aprovado pelo Comitê de Ética em Pesquisa (CEP) da PUCRS (Parecer Nº 5.323.158 de 31 de março de 2022). A parceria com a UFF foi formalizada via Emenda no projeto e aprovada pelo parecer Nº 5.424.977 de 23 de maio de 2022 (Versão 3).

#### **3.1 Dados e pré-processamento**

Os dados utilizados neste trabalho são resultados do estudo de coorte retrospectiva com pacientes incidentes em hemodiálise no período de 2012 a 2019 em diferentes localidades do Brasil. Trata-se de um estudo interdisciplinar entre Ciências da Saúde e Ciência da Computação, colaborativo entre a Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS) e a Universidade Federal Fluminense (UFF). Os 23 Centros de Diálise que foram cenários da coleta de dados têm a seguinte distribuição geográfica: Rio de Janeiro (n=14), Distrito Federal (n=3), São Paulo (n=2), Minas Gerais (n=2) e Pernambuco (n=2). Os dados foram coletados mensalmente por meio do Banco de Dados Europeu de Diálise Clínica (EuCliD®) mediante a um prontuário eletrônico padronizado.

A população de pesquisa foram os pacientes incidentes em hemodiálise crônica, maiores de 18 anos e com menos de 3 meses de tratamento dialítico. Os dados coletados incluíram informações demográficas, como idade e sexo, além dos resultados dos biomarcadores relacionados ao tratamento da hemodiálise. Todos os dados são anonimizados para preservar a privacidade dos pacientes [33].

O atributo alvo a considerar é a chance de morte do paciente, chamado desfecho hospitalar. Para a análise, foram considerados 22 biomarcadores, conforme Tabela 3.1. Para cumprir os objetivos iniciais deste estudo, testaram-se os seguintes algoritmos,

bastante usados na literatura: *i) Decision Tree*, *ii) Logistic Regression*, *iii) Random Forest*, e *iiii) XGBoost*. Esses algoritmos são usados para dados estruturados e supervisionados [32].

MENSAL	TRIMESTRAL	SEMESTRAL	ANUAL	EVENTUAL	OUTROS
Cálcio, Taxa de redução de ureia, Fósforo, Potássio, Sódio	Hemoglobina, Fosfatase alcalina, Saturação de transferrina, Paratormônio (PTH), Ferritina, Albumina	AntiHBs, AgHBs, Anti HCV	Colesterol total, Anti-HIV, Vitamina D	Anti- HBC IgM,	Sexo, Idade, Índice de massa corpórea, Comorbidades anteriores

Tabela 3.1 – Atributos coletados dos pacientes.

Os dados contidos nos *datasets* podem conter valores nulos ou ausentes, e isso pode ocorrer por várias razões, como erros na coleta de informações. Assim, é importante que esses valores sejam tratados e sofram processos de curadoria na sua consistência, porque muitos algoritmos de aprendizado de máquina não podem lidar com dados ausentes diretamente. Procedeu-se com a separação das variáveis preditivas (X) e da variável alvo (Y). A X é composta por todas as colunas do *DataFrame*, ou seja, os dados usados no processo de predição. Já a variável Y, por sua vez, contém apenas a coluna desfecho, no caso deste estudo, o óbito. Ainda na fase de pré-processamento dos dados, ou seja, antes da execução dos algoritmos, procede-se com as etapas de limpeza, balanceamento, normalização e imputação de valores ausentes. Estas fases são cruciais para o melhor desempenho e a confiabilidade dos modelos, assegurando que se possa generalizar bem os dados e fornecer previsões confiáveis na análise de mortalidade.

A Figura 3.1 ilustra o fluxograma da execução das fases do presente trabalho. O processo começa com a pergunta que se deseja responder, usando aprendizado de máquina. O próximo passo é coletar os dados necessários para treinar e testar os modelos que serão usados neste trabalho. Os dados coletados geralmente contêm valores ausentes, exigindo assim uma etapa que envolve a limpeza e preparação dos dados, processados para extrair as variáveis a serem usadas pelo modelo. Após o processamento, os dados são divididos em dois conjuntos principais: dados de treinamento e dados de teste. No primeiro, dados de treinamento, são usados para treinar o modelo e é nesta fase que o modelo de aprendizado de máquina é treinado com os dados usando diferentes algoritmos. Após treinar o modelo, ele é testado com um conjunto separado de dados, resultando assim no segundo conjunto, os chamados dados de teste, onde os resultados do modelo são analisados para verificar se ele responde adequadamente à pergunta ini-

cial. Cabe destacar que essa análise pode levar a ajustes no modelo ou até mesmo à reavaliação de toda a abordagem, dependendo dos resultados.

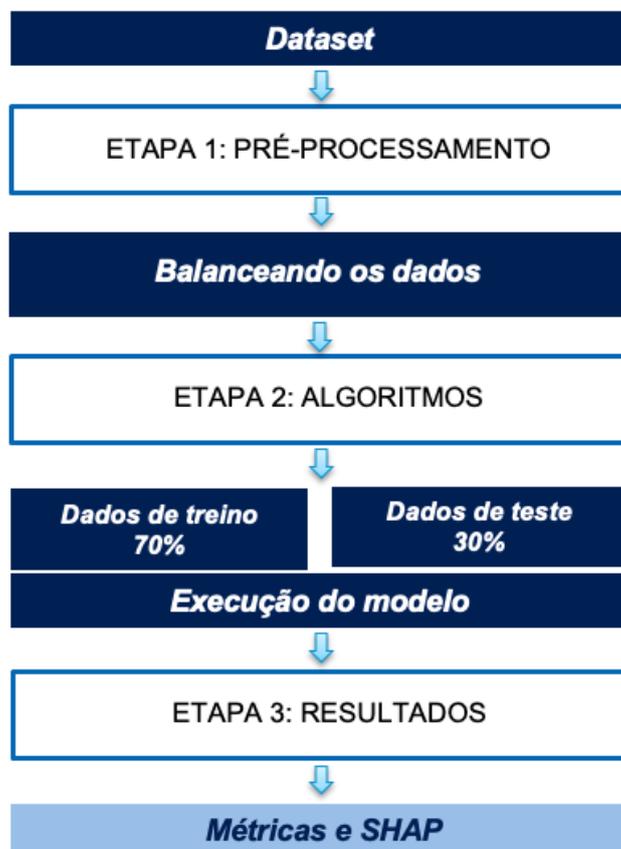


Figura 3.1 – Fluxo processamento dos modelos

Especificamente sobre a limpeza de dados, procedeu-se com um conjunto de regras para detectar e tratar dados faltantes no que tange frequências de exames. O processo de limpeza e completude de informação foram aplicados no *dataset* base para a geração de dados mais consistentes e limpos. A tabela 3.2 fornece uma visão geral dos diferentes *dataset* utilizados no presente trabalho, sendo dois deles gerados a partir das regras de limpeza de dados. A tabela apresenta três versões distintas dos dados: DB Raw, DB Completed e DB Clean. O conjunto de dados bruto (ou base), DB Raw, contém todos os registros originalmente coletados, sendo todos os pacientes maiores de 18 anos. Totaliza 9.367 pacientes, 214.956 exames, 148 biomarcadores, 8.450 pacientes vivos e 917 pacientes falecidos. Esse *dataset* representa os dados base de onde se aplicaram as regras para descartar e também para completar dados, conforme será detalhado a seguir.

<b>dataset</b>	<b>Descrição</b>	<b>Crítérios</b>	<b>Pacientes</b>
DBRaw	Pacientes que foram completados sem seguir regras de completudes.	Informações faltantes em vários registros. Falta de sequencialidade nos registros. Exames mensais/ trimestrais incompletos.	9367
DBClean	Pacientes que não precisam de completude.	Sem informações faltantes.	384
DBCompleted	Pacientes que precisam de completudes em suas informações.	Regras que permitem a completude de informações: <i>i)</i> Nenhum registro no primeiro mês; <i>ii)</i> Um máximo de dois meses faltantes de exames; <i>iii)</i> Dados ausentes no período do exame.	4067

Tabela 3.2 – Resumo dos *datasets*

Neste trabalho, analisaram-se os biomarcadores (apresentados na Tabela 3.1), sendo que cada um possui uma frequência esperada, podendo ser mensal, trimestral, semestral ou anual. Visando tratar os exames ausentes, definiu-se um conjunto de regras para verificação de exames faltantes, conforme segue [10]:

- Quebra de sequencialidade nos meses de registro, i.e., exames mensais que deveriam acontecer, por exemplo, o paciente não possui exame num certo mês.
- Algum exame mensal faltante no período todo, i.e., existem exames no mês determinado, mas falta um exame que deveria existir.
- Algum exame trimestral, semestral ou anual faltando.

Estas regras servem para detectar pacientes que possuem exames faltantes que seriam descartados da construção de um *dataset* limpo e consistente para ser usado nos métodos de aprendizado. Através do processamento destas regras, criou-se o DBClean (Tabela 3.2). Enquanto o DBClean é um *dataset* completo e consistente, observa-se que o número de pacientes baixou muito, de 9367 pacientes no DBRaw para 384 pacientes no DBClean. Esta diminuição de amostras nos dados de treino e teste pode causar um impacto grande nos desempenhos dos métodos de predição, por isso partiu-se para um processamento complementar, onde o objetivo é completar algumas informações faltantes para que mais pacientes possam ser incluídos no processo.

O conjunto DBCompleted foi gerado com base no DBRaw e contém exames de pacientes contidos no DBClean, e ainda flexibilizando as seguintes regras:

- Se o paciente não possuir exame no primeiro mês;
- Aceitam-se exames mensais faltantes desde que não ultrapassem um período de 2 meses;
- Aceitam-se exames trimestrais faltantes desde que não ultrapassem dois períodos;
- Aceitam-se exames semestrais faltantes desde que não ultrapassem dois períodos.

A tabela 3.3 apresenta as regras de completude, que resultaram em 4067 pacientes dos 9367 inicialmente considerados. As regras incluem a presença de informações completas desde o primeiro registro, a sequencialidade temporal dos registros e a inclusão de todos os exames mensais, trimestrais e semestrais, sem falhas nos dados.

<b>Critério</b>	<b>Descrição</b>	<b>Critérios</b>
Informações Faltantes	Pacientes com informações incompletas que podem ser completadas.	Informações completadas usando as regras ajustadas.
Sequencialidade Quebrada	Pacientes com registros não sequenciais que podem ser ajustados.	Sequência ajustada usando registros disponíveis.
Exames Mensais Incompletos	Pacientes com exames mensais incompletos que podem ser completados.	Exames mensais completados usando dados adicionais.
Exames Trimestrais Incompletos	Pacientes com exames trimestrais incompletos que podem ser completados.	Exames trimestrais completados usando dados adicionais.
Exames Semestrais Incompletos	Pacientes com exames semestrais incompletos que podem ser completados.	Exames semestrais completados usando dados adicionais.

Tabela 3.3 – DBCompleted: Pacientes que precisam de completudes em suas informações para ficarem completos

A Tabela 3.4 apresenta os dados resultantes dos processos de limpeza dos *datasets* DBCompleted e DBClean.

	<b>DB Raw</b>	<b>DB Completed</b>	<b>DB Clean</b>
Pacientes:	9367	4064	384
Exames:	214956	71091	4587
Biomarcadores:		22	
Vivos:	8450	3815	365
Mortos:	917	249	19

Tabela 3.4 – Visão geral dos *datasets*

Esses pacientes selecionados garantem uma completude nos dados usados ao executar os modelos de *Machine Learning*, proporcionando um cenário ideal para a análise. No entanto, além da limpeza dos dados, procedeu-se com o balanceamento das classes vivos e mortos, a fim de garantir que o modelo não seja tendencioso em relação à classe majoritária. Neste caso, foi utilizada a técnica de subamostragem aleatória. A maneira de obter o balanceamento adequado entre as classes é utilizar um número aleatório fixo de pacientes, assim, garantindo a reprodutibilidade dos resultados. Assim, como pode ser visto na Tabela 3.4, O DBRaw pode utilizar 917 pacientes vivos e 917 mortos, o DBCompleted 249 de cada classe e o DBClean apenas 19 de cada classe.

Os dados utilizados para avaliar o desempenho dos modelos foram divididos em conjuntos de treinamento e teste, seguindo uma proporção de 70% para treinamento e 30% para teste. Esse particionamento é comum em aprendizado de máquina, pois permite que o modelo aprenda a partir da maioria dos dados (treinamento) e, em seguida, seja avaliado em um subconjunto separado (teste), garantindo que a avaliação seja feita com dados que o modelo não viu durante o treinamento.

Os 30% dos dados reservados para teste foram usados para gerar as matrizes de confusão apresentadas nos resultados. A matriz de confusão é uma ferramenta que ajuda a visualizar o desempenho dos modelos, mostrando a quantidade de previsões corretas e incorretas em termos de verdadeiros positivos, falsos positivos, verdadeiros negativos e falsos negativos. Essa avaliação com os dados de teste é essencial para verificar a capacidade do modelo de generalizar para novos dados, assegurando que as previsões feitas pelo modelo sejam confiáveis e aplicáveis a cenários clínicos reais.

### 3.2 Métricas

Os algoritmos foram aplicados sucessivamente aos pacientes em cada conjunto de dados, utilizando métricas padrão, como acurácia, precisão, *recall* e pontuação para avaliar o desempenho. Todos os experimentos foram realizados no ambiente em Python e usando o Jupyter Notebook, proporcionando um ambiente robusto e escalável para a execução dos modelos.

As métricas utilizadas neste trabalho são comumente usadas em métodos de AM e são descritas a seguir:

- **Acurácia (*Accuracy*):** Mede a proporção de previsões corretas (tanto positivas quanto negativas) em relação ao total de previsões feitas pelo modelo. É uma métrica simples, mas pode ser enganosa em conjuntos de dados desbalanceados, onde uma classe pode ser predominante.
- **Precisão (*Precision*):** Mede quantas das instâncias previstas como positivas estão realmente corretas. No contexto deste estudo, a precisão indica a capacidade do modelo de prever corretamente os casos de óbito entre os pacientes. Uma precisão alta é essencial para minimizar falsos positivos, que poderiam causar preocupações desnecessárias nos pacientes.
- **Revocação (*Recall*):** Mede a proporção de instâncias positivas reais que foram corretamente previstas pelo modelo. Aqui, a revocação reflete a habilidade do modelo de identificar corretamente os pacientes que realmente estão em risco de óbito. Um valor alto de revocação é crítico para reduzir falsos negativos, que poderiam deixar de identificar pacientes em risco.
- **F1-Score:** É a média harmônica entre precisão e revocação, equilibrando as duas métricas. O F1-Score é particularmente útil em casos onde há um desbalanceamento entre as classes, como no caso de pacientes vivos e mortos em hemodiálise, fornecendo uma medida mais robusta da eficácia do modelo.

Para assegurar uma avaliação consistente e confiável dos modelos, implementamos a técnica de validação cruzada com 5 *folds*. Essa abordagem divide o conjunto de dados em cinco partes, utilizando quatro delas para o treinamento e uma para teste, repetindo o processo até que todas as partes tenham sido utilizadas para teste. Isso garante que o desempenho do modelo seja avaliado de forma mais robusta e generalizável.

A utilização do método SHAP (*SHapley Additive exPlanations*) também foi essencial para interpretar os resultados e identificar os biomarcadores com maior influência nas previsões de óbito. A combinação das métricas padrão com a análise SHAP forneceu uma visão abrangente do desempenho dos modelos e permitiu ajustes finos para melhorar a precisão preditiva.

### **3.3 Algoritmos de Aprendizado de Máquina Utilizados**

Os algoritmos usados para o desenvolvimento deste trabalho estão apresentados nesta seção. Quatro algoritmos foram testados: *Decision Tree*, *Logistic Regression*,

*Random Forest*, e *XGBoost*. Esses modelos foram escolhidos por sua eficácia comprovada em lidar com dados estruturados e de alta dimensionalidade, características comuns em conjuntos de dados clínicos.

- *Decision Tree*: É um algoritmo baseado em uma estrutura de árvore, onde os nós internos representam condições em um atributo, as ramificações representam o resultado da condição, e os nós finais representam as previsões (classes). O *Decision Tree* é fácil de interpretar e pode lidar com dados categóricos e numéricos, tornando-o uma escolha popular para problemas médicos [4].
- *Logistic Regression*: Este algoritmo é amplamente utilizado para tarefas de classificação binária e modela a probabilidade de um evento ocorrer como uma função logística dos atributos de entrada. No contexto deste estudo, ele permite estimar a probabilidade de óbito em pacientes com base nos biomarcadores. *Logistic Regression* é conhecido por sua simplicidade e interpretabilidade, além de ser uma escolha robusta para conjuntos de dados relativamente pequenos ou quando a relação entre as variáveis é aproximadamente linear [20].
- *Random Forest*: Consiste em uma coleção de árvores de decisão independentes que votam para determinar a classe final de uma amostra. Este algoritmo melhora a precisão em relação a uma única árvore, pois reduz o risco de *overfitting* por meio do uso de múltiplas árvores e técnicas de *bagging* (*bootstrap aggregating*). É frequentemente usado em dados clínicos, onde pode lidar com dados desbalanceados e identificar variáveis de alta importância para a predição [3].
- *XGBoost* (*Extreme Gradient Boosting*): É um algoritmo de *boosting* que cria modelos sequenciais de árvores de decisão, onde cada árvore subsequente corrige os erros da anterior. Ele é conhecido por sua eficiência computacional e alto desempenho em competições de aprendizado de máquina, especialmente quando os dados apresentam características complexas e interações não lineares. No estudo, o *XGBoost* foi utilizado por sua capacidade de lidar com dados clínicos de alta dimensionalidade e por demonstrar excelente desempenho preditivo em cenários médicos [8].

Cada algoritmo foi ajustado para atender às necessidades específicas do problema, buscando um equilíbrio entre precisão e generalização. O ajuste dos hiperparâmetros envolveu técnicas como *grid search* e validação cruzada, otimizando parâmetros como a profundidade e taxa de aprendizado (*XGBoost*), garantindo que os modelos não apenas performassem bem nos dados de treino, mas também fossem capazes de generalizar adequadamente para novos dados.

Após a limpeza de dados, técnicas de balanceamento de classes foram aplicadas utilizando subamostragem aleatória para evitar viés, uma vez que os dados originais

apresentavam um desbalanceamento. Esse processo foi essencial para assegurar que o modelo não ficasse tendencioso para a classe majoritária. Para garantir a validade dos modelos e minimizar o risco de *overfitting*, foi implementada validação cruzada com 5 *folds* em todas as execuções. Essa abordagem permitiu uma avaliação mais robusta, distribuindo os dados de forma equilibrada em cada iteração, e assegurando que as métricas finais refletissem a verdadeira capacidade preditiva dos modelos.

Por fim, para cada modelo, métricas como acurácia, precisão, *recall* e *F1-score* foram calculadas, permitindo uma análise detalhada do desempenho e facilitando a comparação entre os algoritmos testados.

## 4. RESULTADOS E DISCUSSÃO

Uma vez organizados os conjuntos de dados (*datasets*), os algoritmos *Decision Tree*, *Logistic Regression*, *Random Forest* e *XGBoost* foram executados e seus resultados estão apresentados nesta seção.

### 4.1 Resultados com o *dataset* DBRaw

Nesta seção, são apresentados os resultados dos algoritmos executados com 9367 registros de pacientes e 214.956 exames registrados. Para garantir a uniformidade entre os modelos, foram realizados os balanceamentos entre registros de pacientes vivos e mortos, totalizando 917 pacientes mortos e outros 917 vivos selecionados de forma aleatória. Isso permitiu a utilização de todos os modelos de aprendizado de máquina com os mesmos dados de entrada, criando um cenário igualitário para todas as análises.

Os resultados estão apresentados na Tabela 4.1 que demonstra os quatro algoritmos executados: *Decision Tree*, *Random Forest*, *Logistic Regression* e *XGBoost*; os indicadores de *Accuracy*, *Precision*, *Recall* e *F1 Score* foram calculados e informados de cada um dos algoritmos. Dentre eles, os *Random Forest*, *Logistic Regression* e *XGBoost* obtiveram a mesma *Accuracy* de 0.71, o modelo *Random Forest* apresentou melhor resultado na métrica *Precision* de 0.75 em comparação aos demais, que foi 0.73.

Decision Tree		Random Forest		Logistic Regression		XG Boost	
Accuracy:	0.64	Accuracy:	0.71	Accuracy:	0.71	Accuracy:	0.71
Precision:	0.65	Precision:	0.75	Precision:	0.73	Precision:	0.73
Recall:	0.63	Recall:	0.65	Recall:	0.67	Recall:	0.67
F1 Score:	0.64	F1 Score:	0.70	F1 Score:	0.70	F1 Score:	0.70

Tabela 4.1 – Executados 4 modelos de AM no *dataset* DBRaw.

O resultado da Matriz de Confusão alcançou nos modelos *XGBoost* (Figura 4.3) e *Logistic Regression* (Figura 4.2) desempenhos semelhantes. Já o *XGBoost* demonstrou uma leve vantagem em termos de verdadeiros negativos (VN = 202). A *Logistic Regression* apresentou números levemente maiores de falsos positivos (FP = 70). A divergência apresentada sugere que o modelo *XGBoost* possui uma capacidade maior em minimizar falsos positivos que, neste caso, é fator de muita importância, em que a redução de falsos é extremamente importante. Pode-se perceber que o “0” representa o desfecho VIVO e o “1” para o ÓBITO.

*Decision Tree* (Figura 4.1) e o *Random Forest* (Figura 4.4) obtiveram desempenhos idênticos, ambos apresentaram 177 verdadeiros negativos (VN), 178 verdadeiros

positivos (VP), 92 falsos positivos (FP) e 104 falsos negativos (FN). Esses resultados apresentados de falsos positivos e falsos negativos demonstram uma menor eficiência de alguns modelos de classificação.

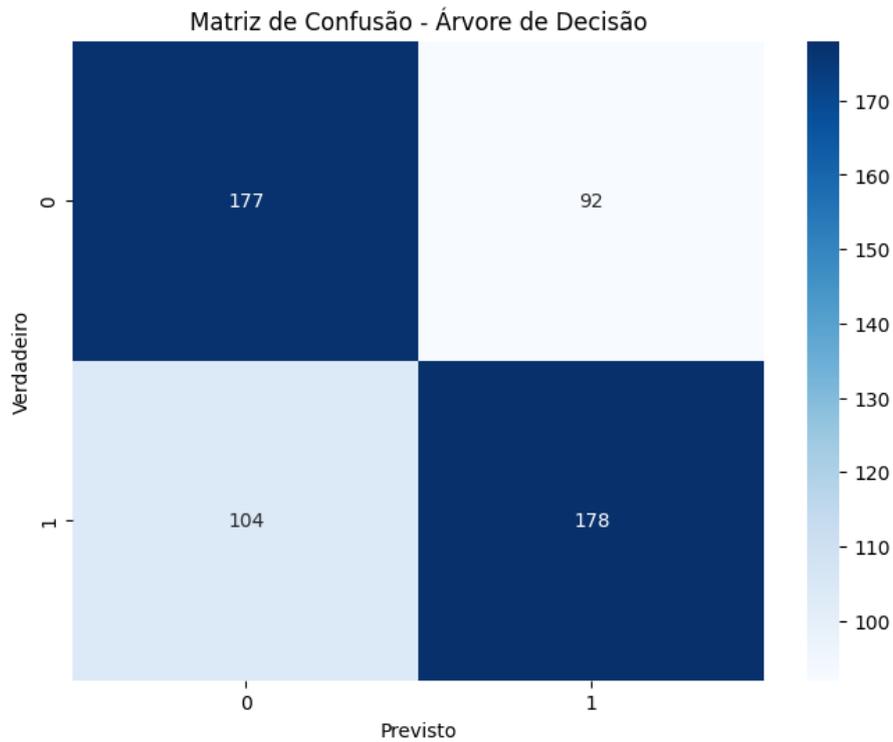


Figura 4.1 – Matriz de Confusão com *Decision Tree* para o *dataset* DBRaw

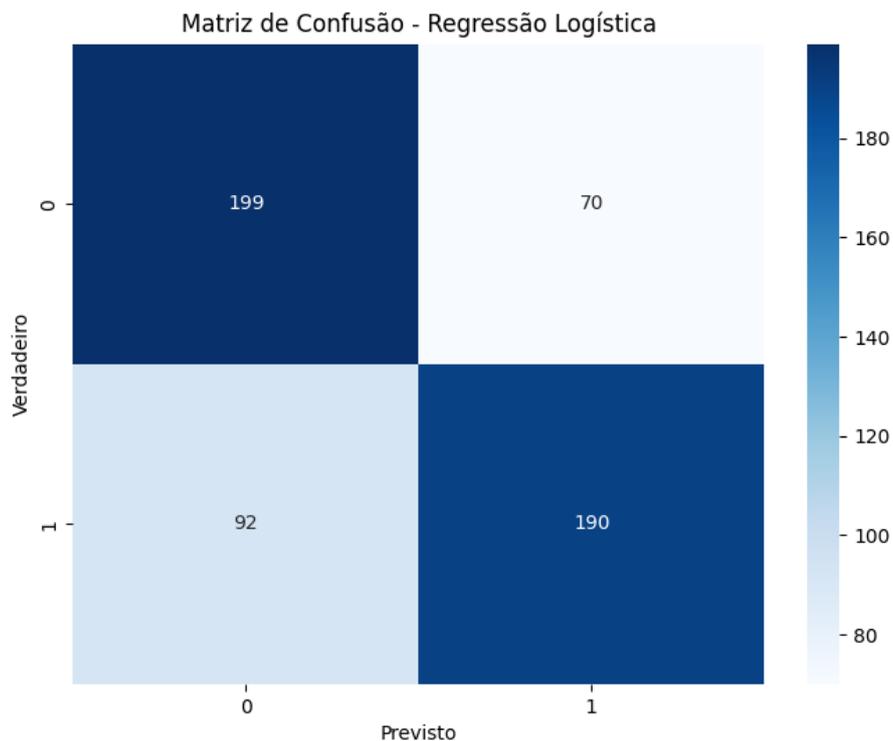


Figura 4.2 – Matriz de Confusão com *Logistic Regression* para o *dataset* DBRaw

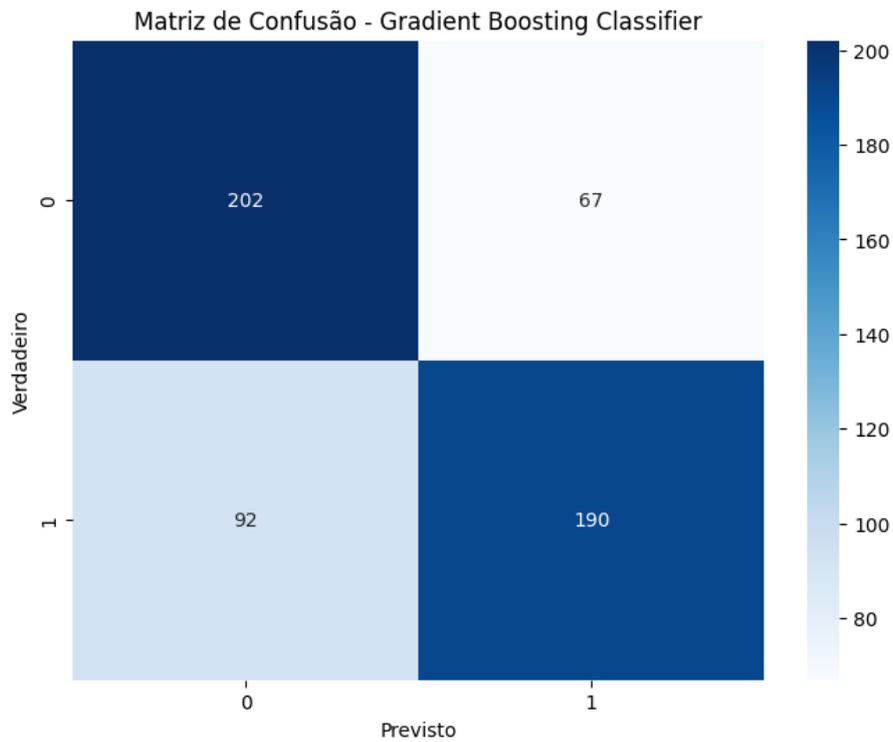


Figura 4.3 – Matriz de Confusão com *XGBoost* para o *dataset* DBRaw

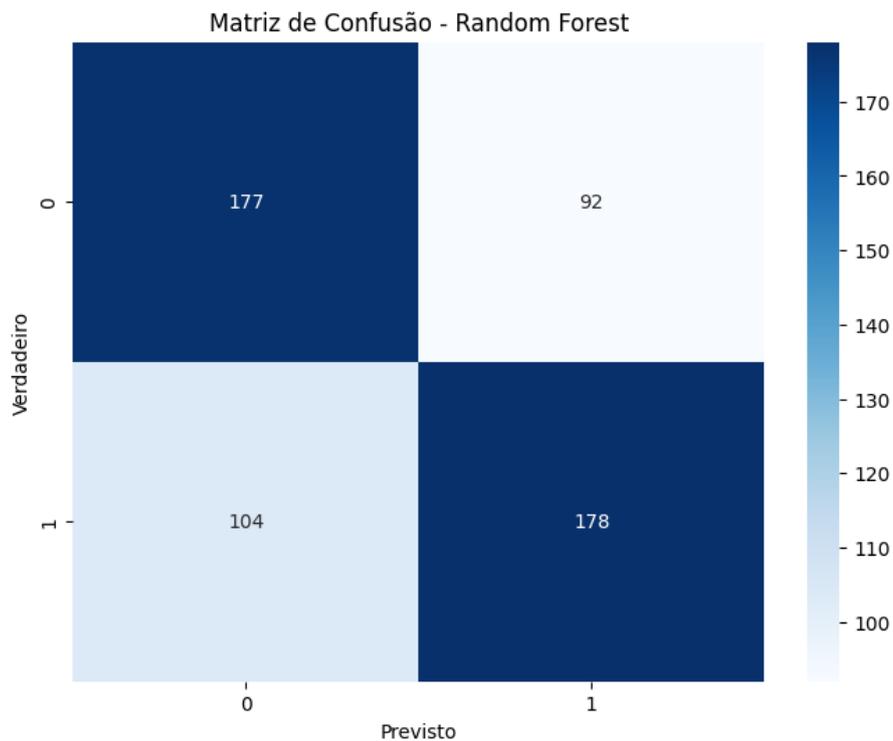


Figura 4.4 – Matriz de Confusão com *Random Forest* para o *dataset* DBRaw

O modelo SHAP do *Random Forest* os três biomarcadores mais relevantes para este modelo foram: Albumina, Idade e Taxa de Redução de Ureia, representada na Figura 4.5.

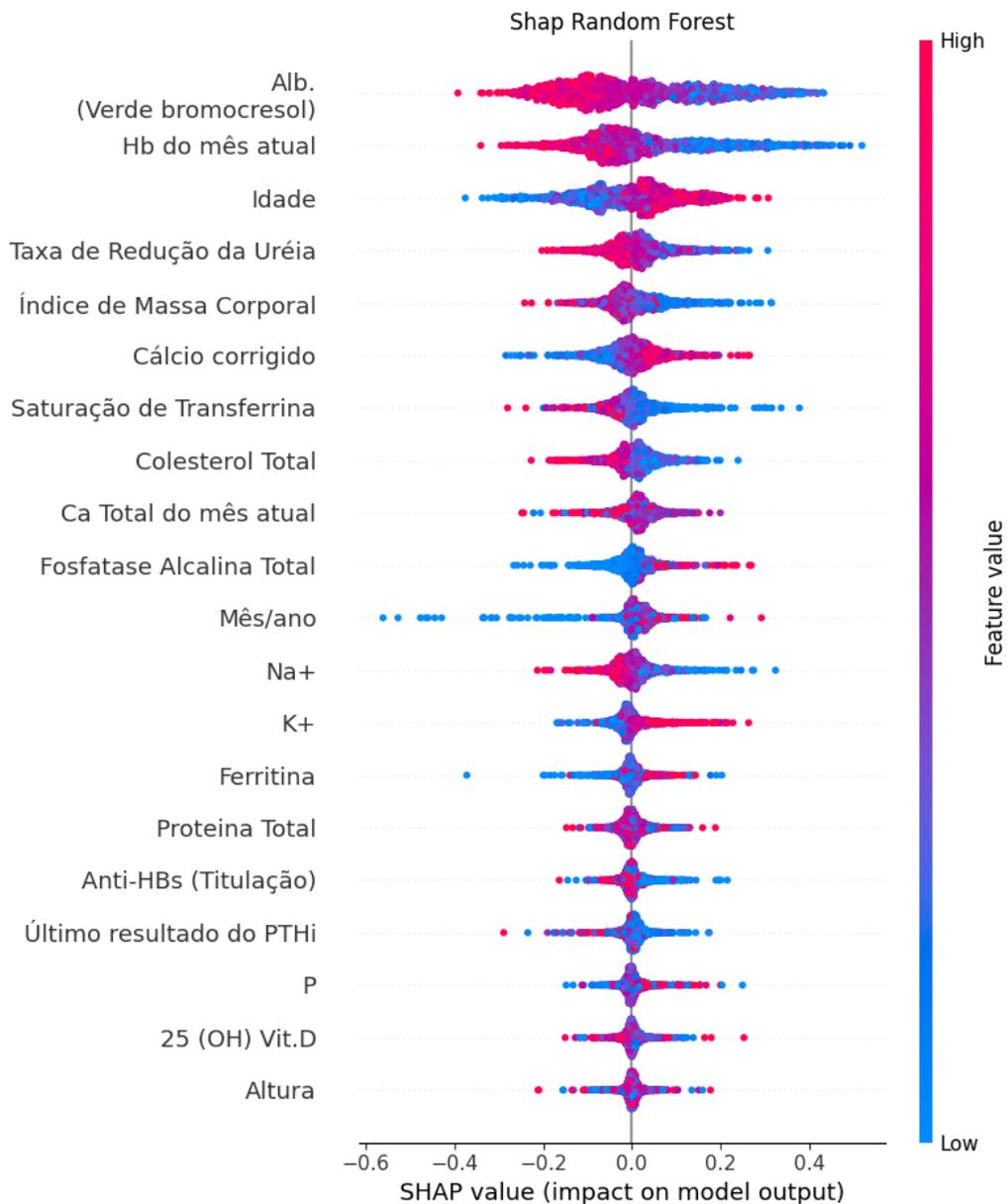


Figura 4.5 – Random Forest SHAP. Teste usando todos os biomarcadores no *dataset* DBRaw

## 4.2 Resultados com o *dataset* DBCompleted

Nesta seção, serão apresentados os resultados dos algoritmos executados com 4.064 registros de pacientes e 71.091 exames registrados. Para garantir a uniformidade entre os modelos, foram realizados os balanceamentos entre registros de pacientes vivos e mortos, totalizando 249 pacientes mortos e outros 249 vivos selecionados de forma aleatória. Isso permitiu a utilização de todos os modelos de AM com os mesmos dados de entrada, criando um cenário igualitário para todas as análises.

Os resultados estão apresentados na Tabela 4.2 que demonstra os quatro algoritmos executados: *Decision Tree*, *Random Forest*, *Logistic Regression* e *XGBoost*; os

indicadores de *Accuracy*, *Precision*, *Recall* e *F1 Score* foram calculados e informados de cada um dos algoritmos. Dentre eles, os *Random Forest* e *Logistic Regression* obtiveram a mesma *Accuracy* de 0.64, o modelo *Logistic Regression* apresentou melhor resultado na métrica *Precision* de 0.65 em comparação ao *Random Forest* que indicou 0.63.

Decision Tree		Random Forest		Logistic Regression		XG Boost	
Accuracy:	0.60	Accuracy:	0.64	Accuracy:	0.64	Accuracy:	0.63
Precision:	0.60	Precision:	0.63	Precision:	0.65	Precision:	0.64
Recall:	0.56	Recall:	0.62	Recall:	0.58	Recall:	0.58
F1 Score:	0.58	F1 Score:	0.63	F1 Score:	0.61	F1 Score:	0.60

Tabela 4.2 – Executados 4 modelos de AM no *dataset* DBCompleted

Na Matriz de Confusão, o modelo de *Logistic Regression* (Figura 4.7) apresentou desempenho muito semelhante ao *XGBoost*(Figura 4.8), com 53 verdadeiros negativos (VN) e 43 verdadeiros positivos (VP), enquanto o *XGBoost* demonstrou 23 falsos positivos (FP) e 31 falsos negativos (FN), havendo uma pequena diferença em comparação com o *XGBoost*, podendo sugerir que o *Logistic Regression* tem uma ligeira vantagem nos resultados.

*Random Forest* (Figura 4.9) revela 48 verdadeiros negativos (VN) e 42 verdadeiros positivos (VP), juntamente com 28 falsos positivos (FP) e 32 falsos negativos (FN), indicando desempenho inferior ao modelo de *Logistic Regression* e ao modelo *XGBoost*, com especial atenção para os menos verdadeiros negativos que mostram sua ineficácia nesta classe. Um valor relativamente mais alto de falsos negativos também indicou menor sensibilidade, o que pode ter efeitos negativos em sua utilidade em aplicações em que a sensibilidade é importante na previsão correta de mortes. A *Decision Tree* (Figura 4.6) era idêntica à *Random Forest*, possuindo 48 verdadeiros negativos (VN), 42 verdadeiros positivos (VP), 28 falsos positivos (FP) e 32 falsos negativos (FN).

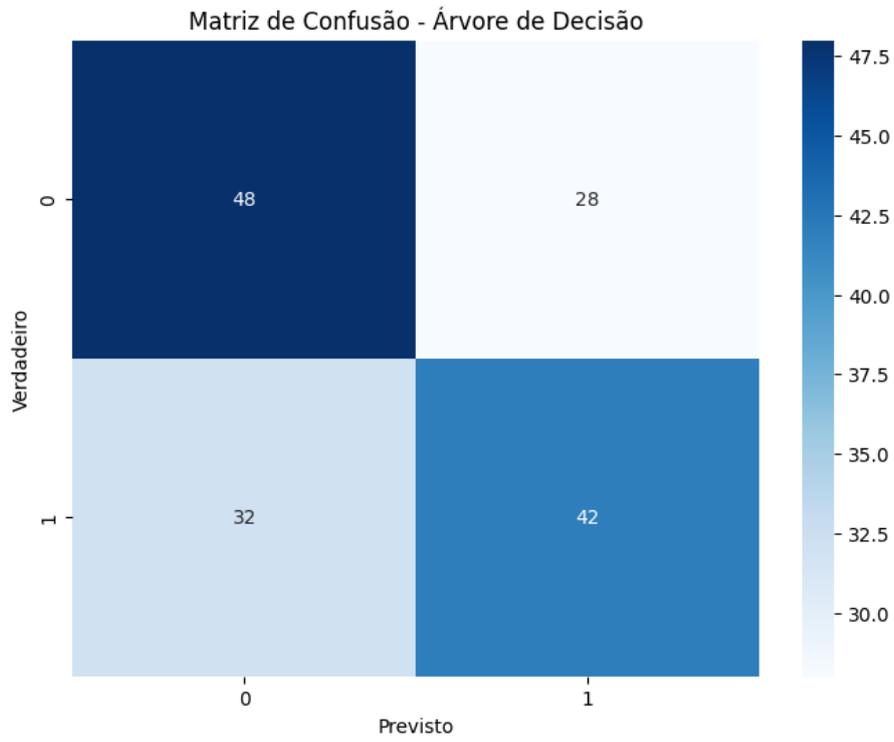


Figura 4.6 – Matriz de Confusão com *Decision Tree* para o *dataset* DBCompleted

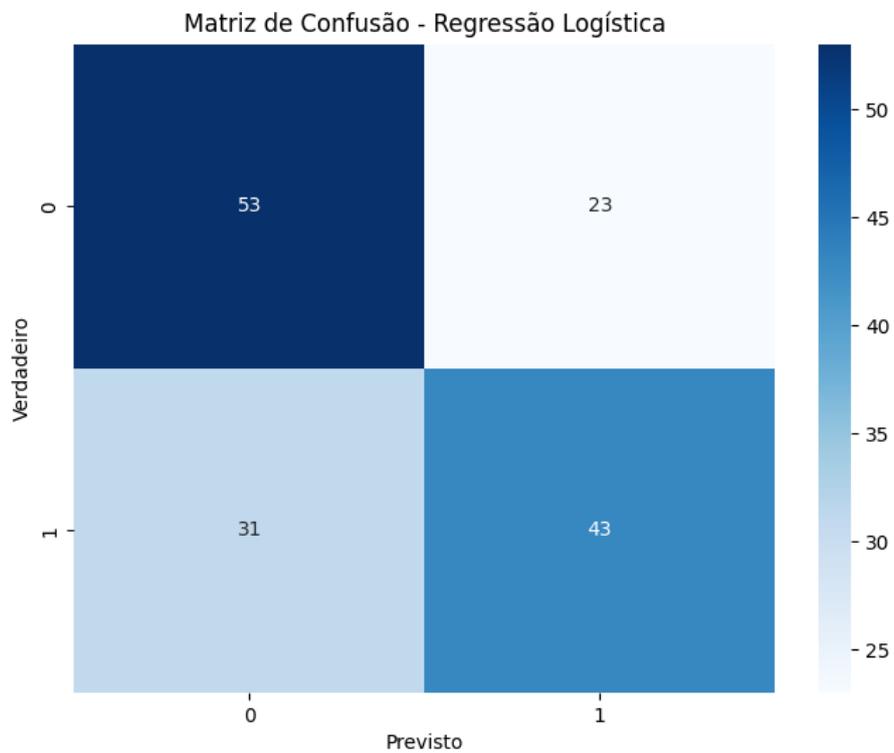


Figura 4.7 – Matriz de Confusão com *Logistic Regression* para o *dataset* DBCompleted

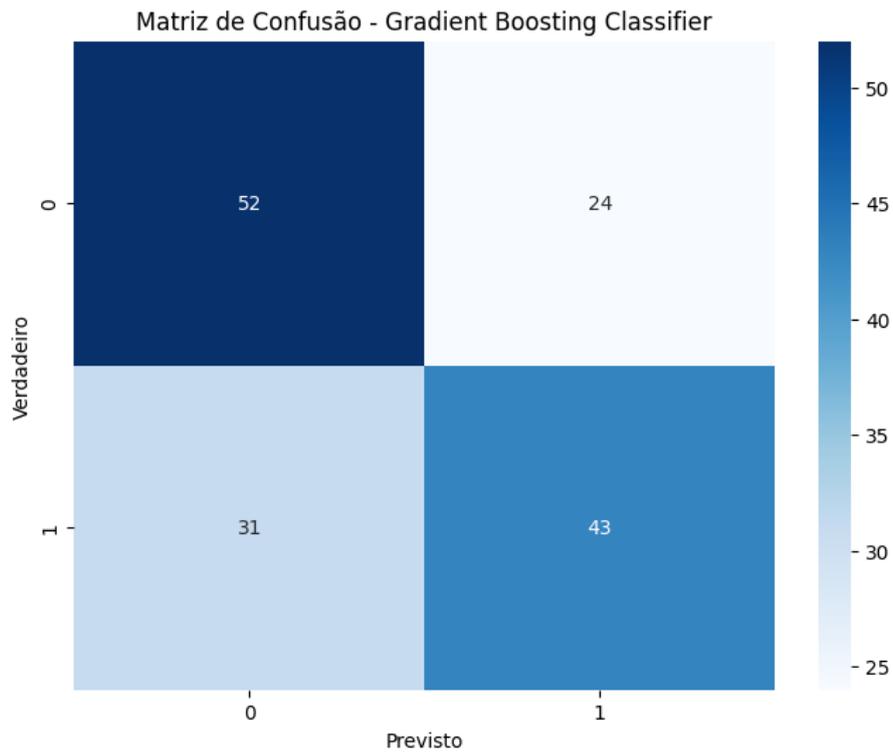


Figura 4.8 – Matriz de Confusão com *XG Boost* para o *dataset* DBCompleted

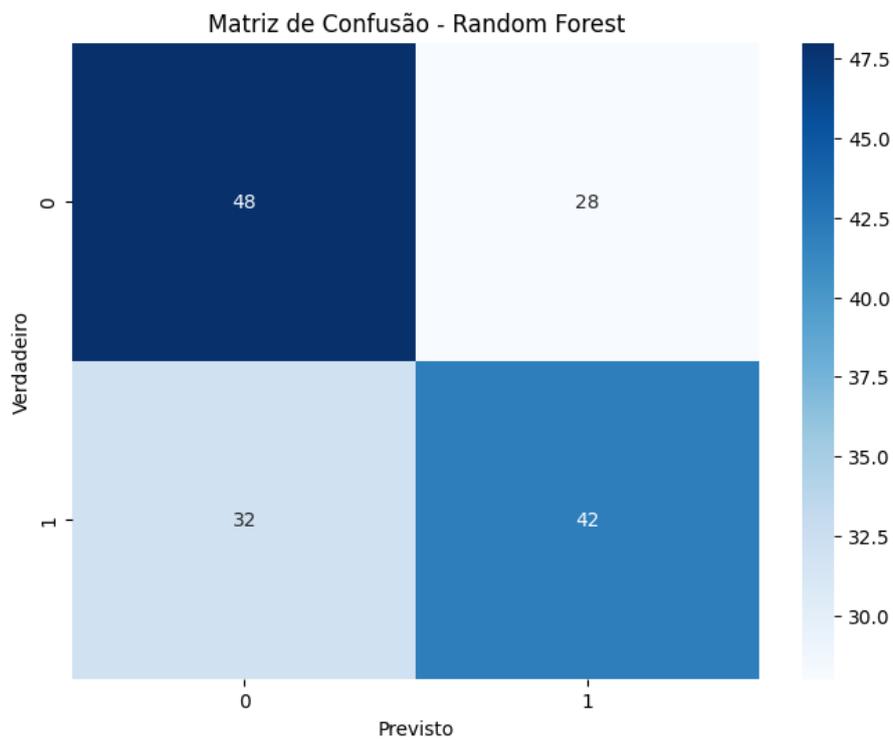


Figura 4.9 – Matriz de Confusão com *Random Forest* para o *dataset* DBCompleted

O modelo SHAP do *XG Boost* os três biomarcadores mais relevantes para este modelo foram: Albumina, Hemoglobina do mês atual (Hb) e Idade, representada na Figura 4.10.

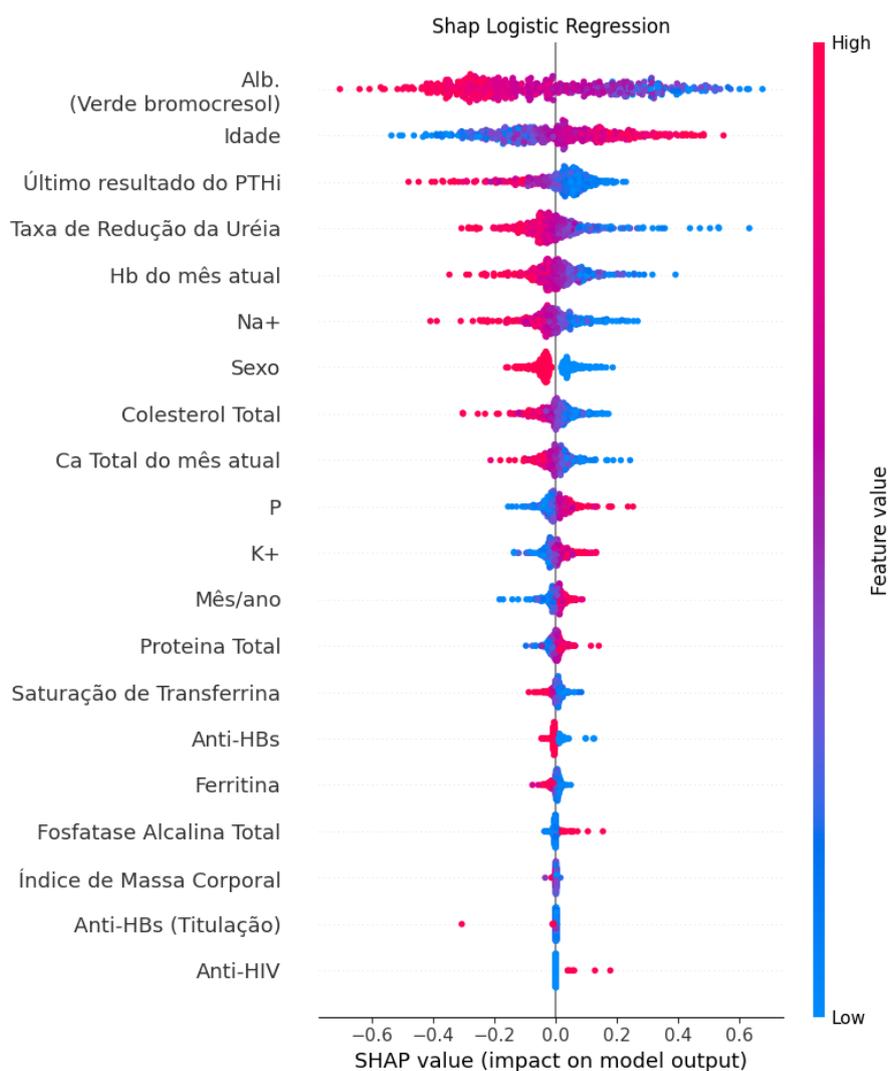


Figura 4.10 – Logistic Regression. Teste usando todos os biomarcadores no *dataset* DB-Completed

### 4.3 Resultados com o *dataset* DBClean

Nesta seção, serão apresentados os resultados dos algoritmos executados com 384 registros de pacientes e 4.587 exames registrados. Para garantir a uniformidade entre os modelos, foram realizados os balanceamentos entre registros de pacientes vivos e mortos, totalizando 19 pacientes mortos e outros 19 vivos, selecionados de forma aleatória. Isso permitiu a utilização de todos os modelos de Aprendizado de Máquina com os mesmos dados de entrada, criando um cenário igualitário para todas as análises.

Os resultados estão apresentados na Tabela 4.3 que demonstra os quatro algoritmos executados: *Decision Tree*, *Random Forest*, *Logistic Regression* e *XGBoost*; os indicadores de *Accuracy*, *Precision*, *Recall* e *F1 Score* foram calculados e informados de cada um dos algoritmos. Dentre eles, os *Logistic Regression* obtiveram a *Accuracy* de 0.92, o

modelo *XGBoost* apresentou um resultado parecido de 0.91, ambos na métrica *Precision* indicaram o mesmo resultado de 1.00.

Decision Tree		Random Forest		Logistic Regression		XG Boost	
Accuracy:	0.50	Accuracy:	0.66	Accuracy:	0.92	Accuracy:	0.91
Precision:	0.50	Precision:	0.66	Precision:	1.00	Precision:	1.00
Recall:	0.66	Recall:	0.66	Recall:	0.83	Recall:	0.83
F1 Score:	0.66	F1 Score:	0.66	F1 Score:	0.91	F1 Score:	0.90

Tabela 4.3 – Executados 4 modelos de AM no *dataset* DBClean

A Matriz de Confusão dos modelos *Random Forest* (Figura 4.14) e *Decision Tree* (Figura 4.11) apresentou desempenho bastante semelhante. Ambos classificaram dois deles como verdadeiros negativos (VN) e 4 verdadeiros positivos (VP). Mas ambos tiveram saída de 4 falsos positivos (FP) e 2 falsos negativos (FN). Há uma limitação nesses modelos que apresentaram uma taxa de erro muito maior que a *Logistic Regression* (Figura 4.12), que foi o número desse resultado. O elevado número de falsos positivos é especialmente problemático porque pode induzir uma tomada de decisões de forma errada nesse contexto.

*XG Boost* (Figura 4.13) o resultado foi o seguinte, 6 verdadeiros negativos (VN) e 5 verdadeiros positivos (VP). Semelhante à *Logistic Regression*, esse modelo apresentou apenas 1 falso negativo (FN) e nenhum falso positivo (FP). A diferença entre esses dois modelos indica que, para ambas as técnicas, pode haver sensibilidade e especificidade ótimas na classificação correta das instâncias.

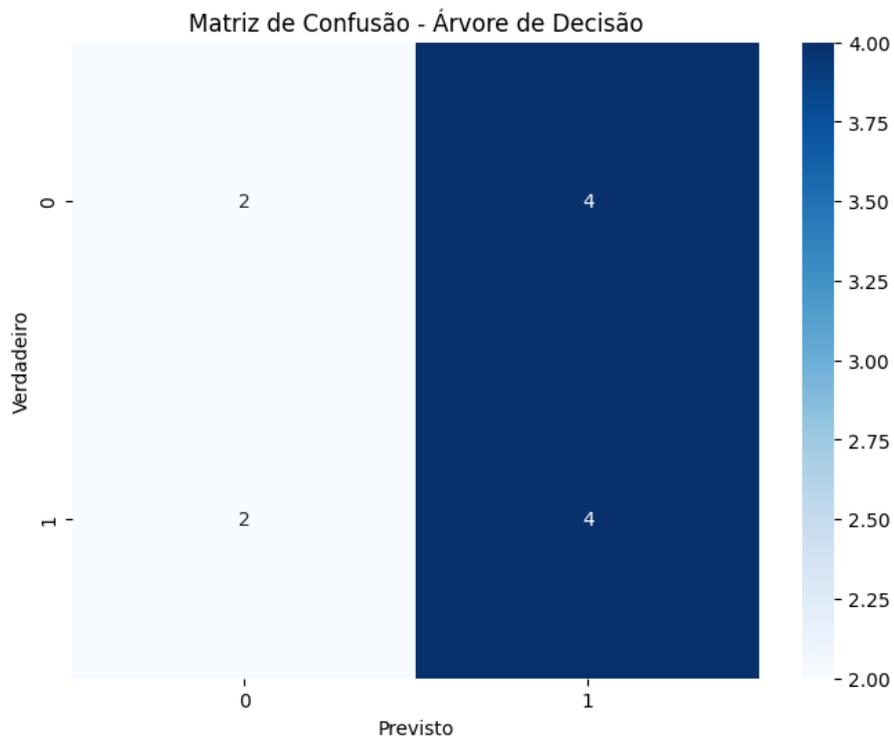


Figura 4.11 – Matriz de Confusão com *Decision Tree* para o DBClean

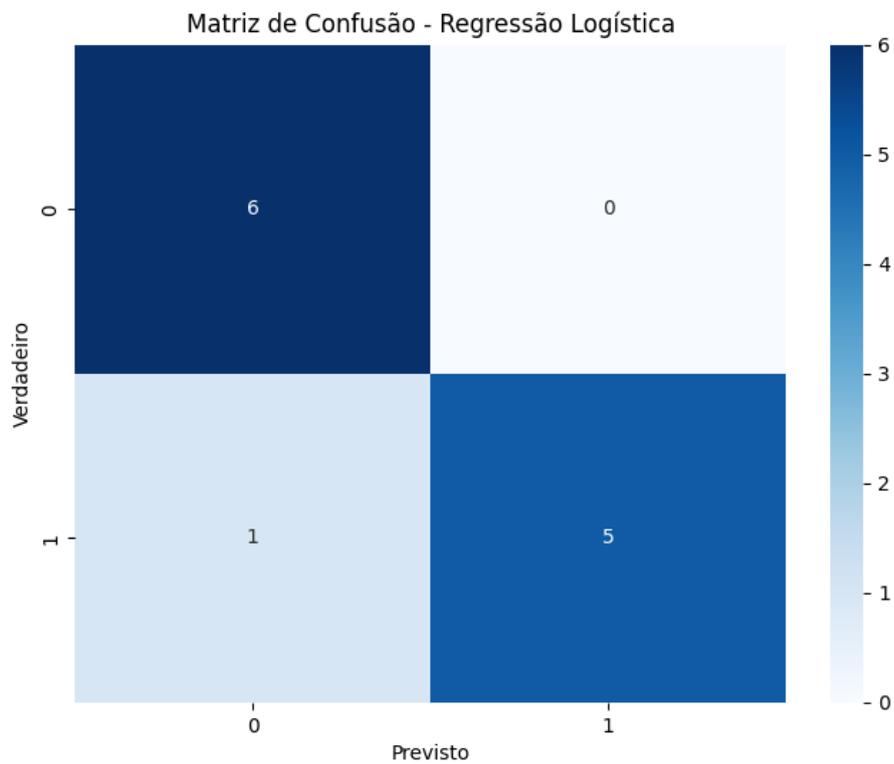


Figura 4.12 – Matriz de Confusão com *Logistic Regression* para o dataset DBClean

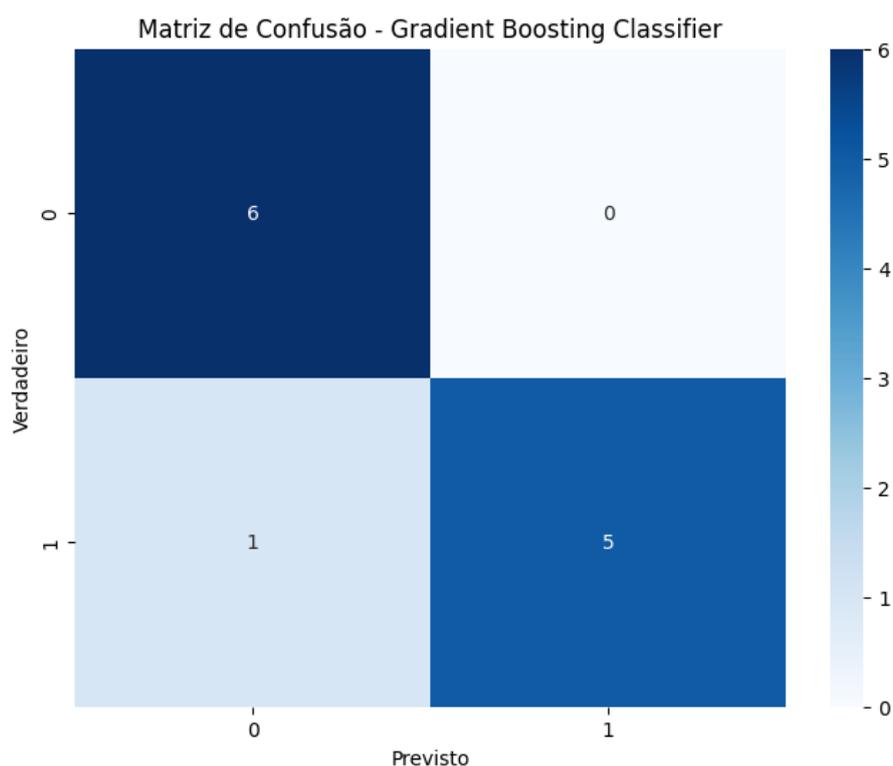


Figura 4.13 – Matriz de Confusão com *XG Boost* para o *dataset* DBClean

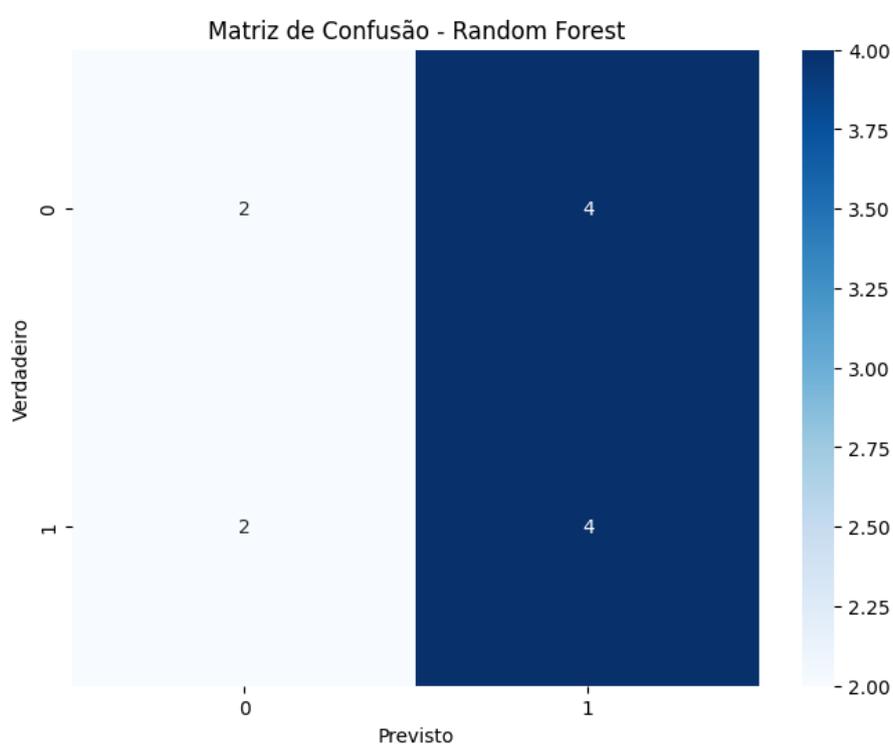


Figura 4.14 – Matriz de Confusão com *Random Forest* para o *dataset* DBClean

O modelo SHAP do *Logistic Regression* os três biomarcadores mais relevantes para este modelo foram: Ferritina, Albumina e Na<sup>+</sup>, representada na Figura 4.15.

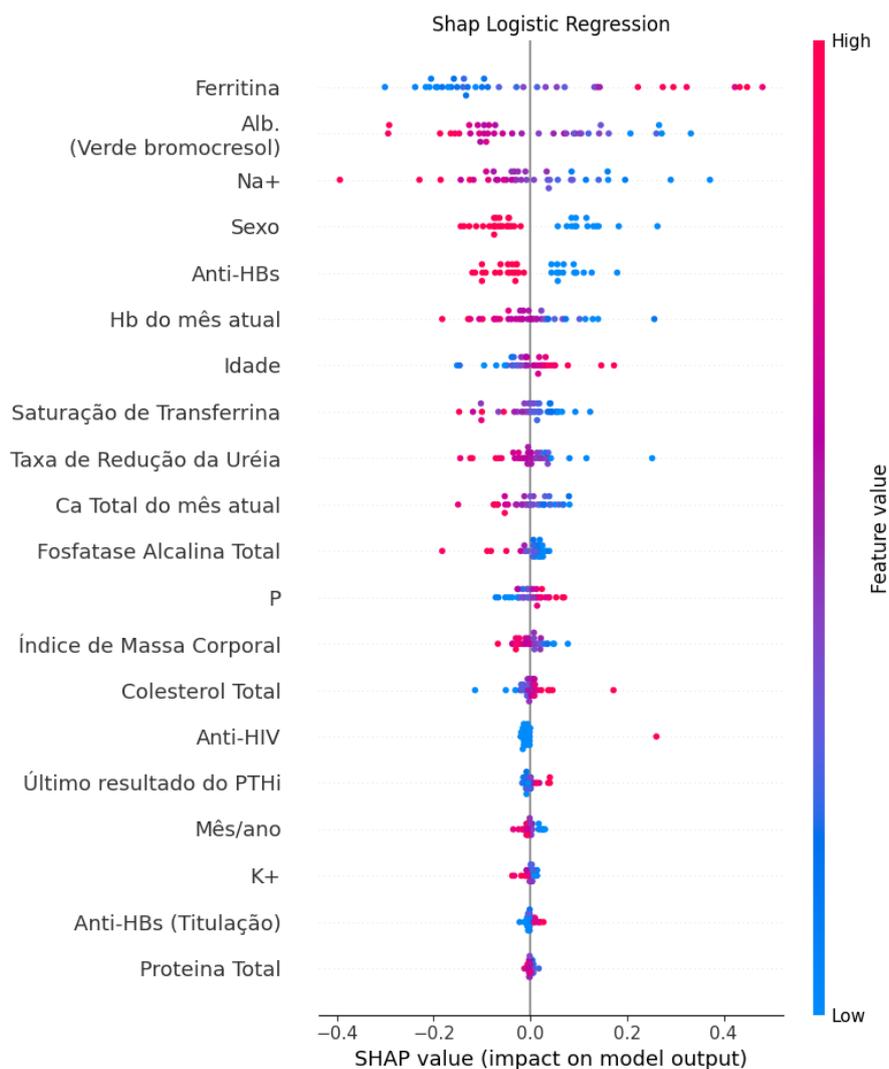


Figura 4.15 – Logistic Regression. Teste usando todos os biomarcadores no *dataset* DB-Clean

#### 4.4 Curva Receiver Operating Characteristic (ROC) dos modelos

A métrica AUC ROC mede o quão bem o modelo pode diferenciar entre verdadeiro positivo e falso positivo, em que a área sobre a curva indica a probabilidade de o caso ser corretamente classificado. Quanto mais a linha dos testes se afastar da linha central (linha pontilhada), mais preciso será o resultado do teste. A Figura 4.16 mostra que o modelo XGBoost apresenta maior valor para a área sob a curva (AUC), indicando melhor capacidade de discriminação de classes (vivos ou mortos). Por outro lado, o modelo de *Decision Tree* tem sua curva mais próxima de uma linha diagonal AUC; portanto, ele tem um desempenho inferior.

Na Figura 4.17, o desempenho foi equilibrado entre os modelos *XGBoost* e *Logistic Regression* com curvas mais próximas. Já a Figura 4.18, o conjunto de dados mais

restrito e completo, tanto o *XGBoost* quanto a *Logistic Regression* atingem uma AUC próxima de 1, mostrando excelente desempenho. Os modelos *Decision Tree* e *Random Forest* apresentam curvas ROC um pouco menos acentuadas.

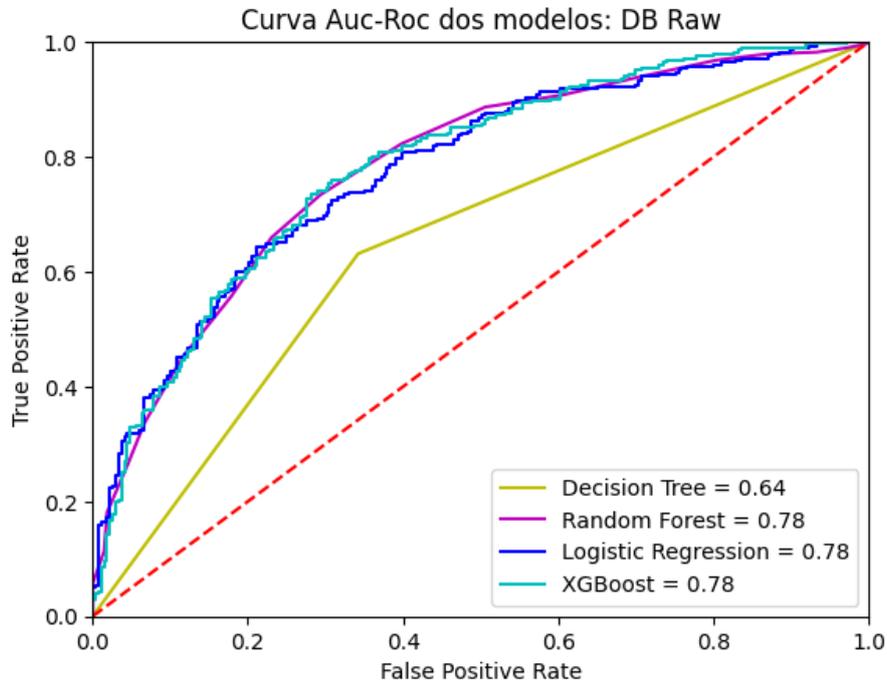


Figura 4.16 – Métrica AUC ROC para o *dataset* DBRaw

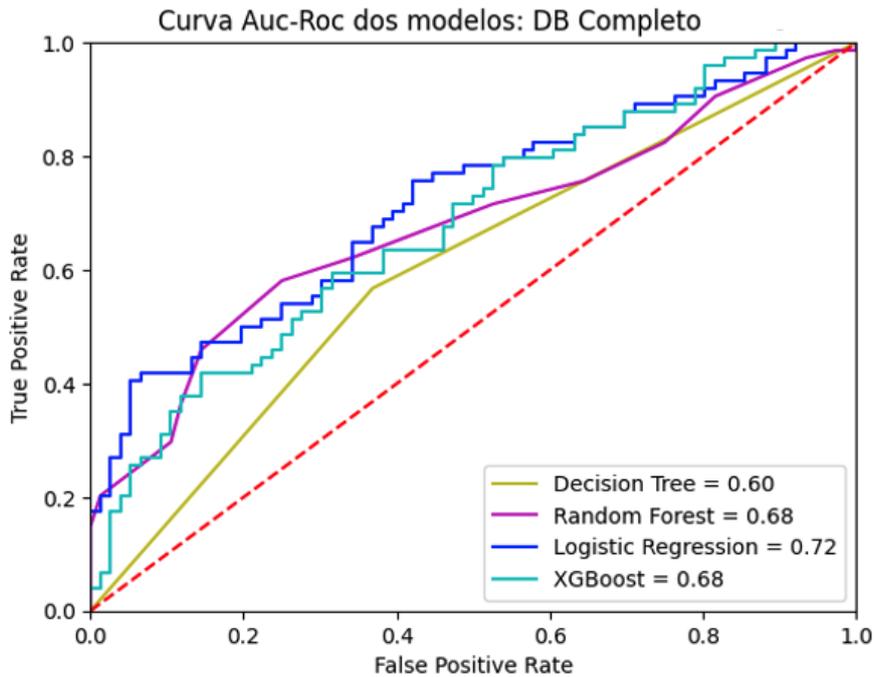


Figura 4.17 – Métrica AUC ROC para o *dataset* DBCompleted

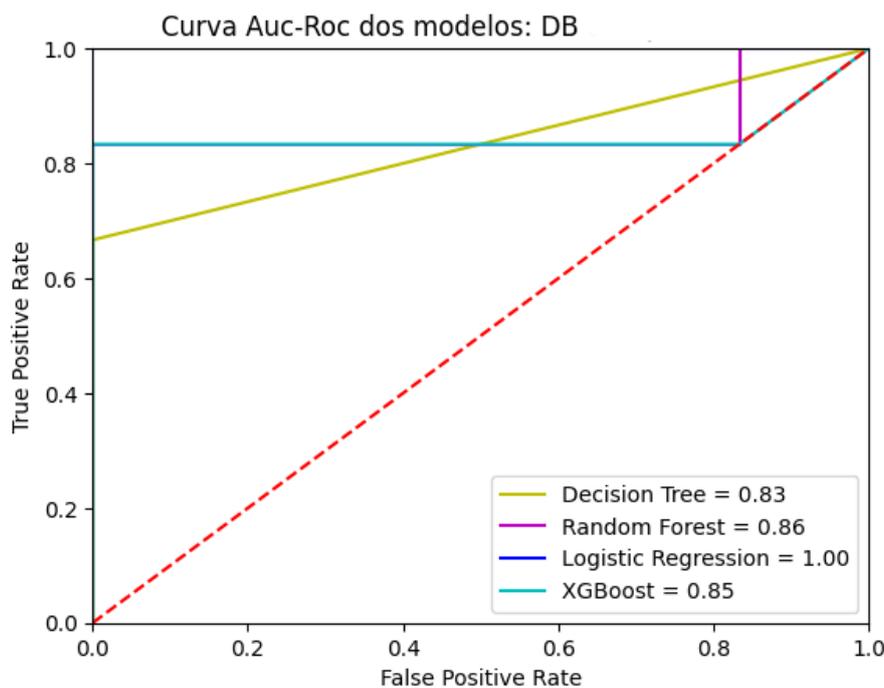


Figura 4.18 – Métrica AUC ROC para o *dataset* DBClean

#### 4.5 Discussão

Os resultados obtidos no estudo revelam visões significativas que contribuem para a literatura existente. Primeiramente, observou-se que os modelos *Logistic Regression* e *XG Boost* apresentaram maior acurácia na predição. A acurácia média desses modelos variou entre 60% e 90%, conforme métricas como a curva *Receiver Operating Characteristic* (ROC) e *F1-Score*. Esses valores são coerentes com estudos anteriores, que indicam a capacidade robusta desses modelos em cenários clínicos complexos [40].

A capacidade de fazer previsões facilita ações mais rápidas e direcionadas com mais precisão, o que pode eventualmente reduzir as taxas de reinternação e melhorar a qualidade do atendimento ao paciente [35]. Além disso, se for possível fazer previsões sobre esses resultados críticos, seria possível adaptar estratégias clínicas pessoais para estes pacientes em hemodiálise [28]. Assim, é igualmente importante a viabilidade de implantação desses modelos nos sistemas hospitalares atuais. Os resultados indicam que algoritmos podem ser incorporados em sistemas de registros médicos eletrônicos existentes com limpeza e estruturação adequada de dados, sem grandes requisitos adicionais de hardware. Isso facilita a sua adoção em ambientes clínicos reais [34]. No entanto, é preciso ter em mente as limitações incorporadas nos próprios dados. Qualquer viés que possa existir nos dados históricos pode ser refletido nas previsões do modelo. Por último, mas não menos importante, existem questões éticas relacionadas com a utilização da aprendizagem automática na medicina. Embora os benefícios sejam evidentes, ainda existem

desafios em termos de transparência dos algoritmos utilizados e privacidade dos dados dos pacientes [6].

Os resultados obtidos demonstraram robustez com estudos prévios que indicam a alta capacidade discriminativa dos algoritmos de aprendizado de máquina em na área médica. Por exemplo, um estudo recente realizado por Nusinovici et al. (2020) [32] também concordou com a eficácia obtida no presente trabalho com a aplicação do *Decision Tree* no DBClean em prever desfechos hospitalares com uma precisão comparável à descoberta deste estudo. A aplicação desses modelos na prática clínica pode potencialmente melhorar o gerenciamento e as intervenções precoces para pacientes em hemodiálise. Esses resultados são importantes por dois motivos: em primeiro lugar, apoiam o interesse emergente na literatura médica em aplicar a aprendizagem automática para melhorar o cuidado de pacientes crônicos; em segundo lugar, com base no estudo de Esteva et al. (2017) [11], o uso de tais tecnologias pode revolucionar a medicina personalizada ao permitir previsões precisas baseadas em abundância de dados.

## 5. CONSIDERAÇÕES FINAIS

Neste trabalho, discutimos a eficácia e a relevância da aplicação de técnicas de AM avançadas no campo da saúde. Ao longo deste estudo, foram desenvolvidos e avaliados diferentes modelos de aprendizado de máquina, como *Decision Tree*, *Random Forest*, *Logistic Regression* e *XGBoost*. Os resultados indicaram que os modelos *Random Forest* e *XGBoost* foram superiores aos outros, principalmente quando aplicados ao dataset DB Clean, em que os modelos demonstraram precisão ótima, demonstrando ainda mais suas fortes capacidades preditivas para resultados clínicos. Este resultado foi de alguma forma inesperado, pois se considera que com tão poucas amostras a validade dos resultados pode possuir um viés.

Ao utilizar a metodologia *SHapley Additive exPlanations* (SHAP), foram apresentados de forma abrangente os vários biomarcadores e uma compreensão mais precisa nos fatores que têm grande impacto na mortalidade dos pacientes. De pacientes submetidos à hemodiálise entre 2012 e 2019 em 23 unidades de diálise em cinco estados brasileiros, foram coletados esses dados, que abrangem vários aspectos, dentre eles detalhes demográficos, resultados de biomarcadores e aspectos clínicos, para, depois, separar em três conjuntos de dados, chamados de DBRaw, DBCompleted e DBClean.

As descobertas indicaram que altos níveis de precisão e sensibilidade poderiam ser alcançados pelos modelos na previsão de mortalidade. A sua robustez foi destacada por meio da validação cruzada e métricas. O impacto na prática clínica e na gestão hospitalar parece ser relevante, porque a implementação de sistemas preditivos baseados em aprendizagem automática pode permitir que médicos e administradores sejam proativos, tomando decisões informadas com bastante antecedência, o que tem o potencial de melhorar os resultados dos pacientes, bem como de garantir o desempenho ideal com a utilização de recursos hospitalares para uma prestação eficaz de cuidados ao paciente. Em resumo, este trabalho contribui para a base de evidências sobre aprendizado de máquina na medicina personalizada. A implementação dessas tecnologias pode revolucionar não apenas o resultado clínico dos pacientes em hemodiálise, mas também a prática da medicina de forma mais precisa e eficaz. Em síntese, os resultados obtidos neste estudo corroboram a hipótese de que modelos de aprendizado de máquina são ferramentas eficazes para a predição de desfechos hospitalares em pacientes submetidos à hemodiálise. Utilizando um conjunto diversificado de dados clínicos e laboratoriais, os modelos desenvolvidos demonstraram uma precisão significativa na identificação de pacientes com maior risco de mortalidade. Do ponto de vista clínico, a capacidade de prever desfechos adversos pode auxiliar médicos na tomada de decisões mais informadas, permitindo intervenções precoces e personalizadas que podem melhorar a qualidade do atendimento ao paciente [27].

Em suma, os achados deste trabalho estão alinhados à literatura recente sobre o tema e demonstram a importância do uso de técnicas avançadas de aprendizado de máquina na área da saúde. Futuros estudos poderão explorar outros algoritmos para melhorar ainda mais a precisão. Portanto, os modelos de aprendizado de máquina demonstraram um potencial significativo para a predição de desfecho hospitalar. Esses resultados adicionam-se à crescente literatura que destaca o valor do aprendizado de máquina na medicina preditiva.

## REFERÊNCIAS BIBLIOGRÁFICAS

- [1] “Valor apresentado de produção ambulatorial do sus para procedimento hemodiálise no período 2009-2018 no brasil.”, 2021.
- [2] Al-Jaghbeer, M., D. D. B. A. A. R. . K. J. A. “Clinical decision support for in-hospital aki. journal of the american society of nephrology”, vol. 28, 2017, pp. 654–660.
- [3] Breiman, L.; Cutler, A. “Random Forests”. CRC press, 2001.
- [4] Breiman, L.; Friedman, J. H.; Olshen, R. A.; Stone, C. J. “Classification and Regression Trees”. Wadsworth International Group, 1984.
- [5] Cesconetto, D. M., L. J. B. P.-d.-F. C. E. T. G. F. D. . C. d. S. P. M. “Prediction of hospitalization in chronic dialysis patients by machine learning techniques: A case–control study. plos one”, 2020.
- [6] Char, D. S.; Shah, N. H.; Magnus, D. “Implementing machine learning in health care - addressing ethical challenges”, *N Engl J Med*, vol. 378–11, mar. 2018, pp. 981–983.
- [7] Chen, J. H.; Asch, S. M. “Machine learning and prediction in medicine - beyond the peak of inflated expectations”, *N Engl J Med*, vol. 376–26, jun. 2017, pp. 2507–2509.
- [8] 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, 2016, pp. 785–794.
- [9] Couser, W. G.; Remuzzi, G.; Mendis, S.; Tonelli, M. “The contribution of chronic kidney disease to the global burden of major noncommunicable diseases”, *Kidney international*, vol. 80–12, 2011, pp. 1258–1270.
- [10] de Dados de Exames de Pacientes em Hemodiálise, C., 2024.
- [11] Esteva, A.; Kuprel, B.; Novoa, R. A.; Ko, J.; Swetter, S. M.; Blau, H. M.; Thrun, S. “Dermatologist-level classification of skin cancer with deep neural networks”, *Nature*, vol. 542–7639, jan. 2017, pp. 115–118.
- [12] Garcia-Montemayor, V.; Martin-Malo, A.; Barbieri, C.; Bellocchio, F.; Soriano, S.; Pendon-Ruiz de Mier, V.; Molina, I. R.; Aljama, P.; Rodriguez, M. “Predicting mortality in hemodialysis patients using machine learning analysis”, *Clin Kidney J*, vol. 14–5, ago. 2020, pp. 1388–1395.
- [13] Garcia-Montemayor, V.; Martin-Malo, A.; Barbieri, C.; Bellocchio, F.; Soriano, S.; Pendon-Ruiz de Mier, V.; Molina, I. R.; Aljama, P.; Rodriguez, M. “Predicting mortality

in hemodialysis patients using machine learning analysis”, *Clin Kidney J*, vol. 14–5, ago. 2020, pp. 1388–1395.

- [14] Goto, T.; Camargo, Jr, C. A.; Faridi, M. K.; Freishtat, R. J.; Hasegawa, K. “Machine Learning-Based prediction of clinical outcomes for children during emergency department triage”, *JAMA Netw Open*, vol. 2–1, jan. 2019, pp. e186937.
- [15] Grus, J. “Data Science do Zero - 2ª Edição: Noções Fundamentais com Python”. Alta Book, 2021.
- [16] Gunarathne, W.; Perera, K.; Kahandawaarachchi, K. “Performance evaluation on machine learning classification techniques for disease classification and forecasting through data analytics for chronic kidney disease (ckd)”. In: 2017 IEEE 17th international conference on bioinformatics and bioengineering (BIBE), 2017, pp. 291–296.
- [17] Géron, A. “Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems”. O’Reilly Media, 2019.
- [18] Géron, A. “Mãos à Obra: Aprendizado de Máquina com Scikit-Learn TensorFlow”. O’Reilly Media, 2019.
- [19] Habebh, H.; Gohel, S. “Machine learning in healthcare”, *Curr Genomics*, vol. 22–4, dez. 2021, pp. 291–300.
- [20] Hosmer Jr, D. W.; Lemeshow, S.; Sturdivant, R. X. “Applied Logistic Regression”. John Wiley & Sons, 2013.
- [21] Hsieh, W.-H.; Ku, C. C.-Y.; Hwang, H. P.-C.; Tsai, M.-J.; Chen, Z.-Z. “Model for predicting complications of hemodialysis patients using data from the internet of medical things and electronic medical records”, *IEEE J Transl Eng Health Med*, vol. 11, jan. 2023, pp. 375–383.
- [22] Huang, Y.; Talwar, A.; Chatterjee, S.; Aparasu, R. R. “Application of machine learning in predicting hospital readmissions: a scoping review of the literature”, *BMC Med Res Methodol*, vol. 21–1, mai. 2021, pp. 96.
- [23] IC Reinheimer, MA Beer, M. L. d. C. I. T. J. B. d. S. M. V. B. F. H. R. W. R. R. T. P. M. “Machine learning models using biomarkers to predict mortality in hemodialysis”, *KIDNEY INTERNATIONAL REPORTS*, vol. 8, 2023, pp. S298–S299.
- [24] Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. “Lightgbm: A highly efficient gradient boosting decision tree”. In: *Advances in Neural Information Processing Systems*, Guyon, I.; Luxburg, U. V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; Garnett, R. (Editores), 2017.

- [25] Kim, H. W.; Heo, S.-J.; Kim, M.; Lee, J.; Park, K. H.; Lee, G.; Baeg, S. I.; Kwon, Y. E.; Choi, H. M.; Oh, D.-J.; Nam, C.-M.; Kim, B. S. "Deep learning model for predicting intradialytic hypotension without privacy infringement: A retrospective Two-Center study", *Front Med (Lausanne)*, vol. 9, jul. 2022, pp. 878858.
- [26] Kourou, K.; Exarchos, T. P.; Exarchos, K. P.; Karamouzis, M. V.; Fotiadis, D. I. "Machine learning applications in cancer prognosis and prediction", *Comput Struct Biotechnol J*, vol. 13, nov 2014, pp. 8–17.
- [27] Kourou, K.; Exarchos, T. P.; Exarchos, K. P.; Karamouzis, M. V.; Fotiadis, D. I. "Machine learning applications in cancer prognosis and prediction", *Comput Struct Biotechnol J*, vol. 13, nov. 2014, pp. 8–17.
- [28] Ling, Y.; Wang, L.; Liu, X.; Wang, K.; Ma, Z.; Yu, Y.; Liu, W.; Liang, W.; Qian, K.; Xu, Y.; Zuo, X.; Ge, S.; Yao, Y. "Development and validation of prediction model for technique failure in peritoneal dialysis patients: An observational study", *Nephrology (Carlton)*, vol. 29–7, fev. 2024, pp. 383–393.
- [29] Liu Y.C.J, C. C.; Y.J., T. "Big data analysis using artificial intelligence in health care: a systematic review protocol", *BMJ Open*, 2019.
- [30] Mahalingasivam, V.; Su, G.; Iwagami, M.; Davids, M. R.; Wetmore, J. B.; Nitsch, D. "COVID-19 and kidney disease: insights from epidemiology to inform clinical practice", *Nat Rev Nephrol*, vol. 18–8, abr. 2022, pp. 485–498.
- [31] Neves, P. D. M. d. M.; Sesso, R. d. C. C.; Thomé, F. S.; Lugon, J. R.; Nasicmento, M. M. "Brazilian dialysis census: analysis of data from the 2009-2018 decade", *Brazilian Journal of Nephrology*, vol. 42, 2020, pp. 191–200.
- [32] Nusinovici, S.; Tham, Y. C.; Yan, M. Y. C.; Ting, D. S. W.; Li, J.; Sabanayagam, C.; Wong, T. Y.; Cheng, C.-Y. "Logistic regression was as good as machine learning for predicting major chronic diseases", *Journal of clinical epidemiology*, vol. 122, 2020, pp. 56–69.
- [33] Nweke, H. F. "Clinical text classification research trends: Systematic literature review and open issues", 2018.
- [34] Obermeyer, Z.; Emanuel, E. J. "Predicting the future — big data, machine learning, and clinical medicine", *New England Journal of Medicine*, vol. 375–13, 2016, pp. 1216–1219, <https://www.nejm.org/doi/pdf/10.1056/NEJMp1606181>.
- [35] Rajkomar, A.; Oren, E.; Chen, K.; Dai, A. M.; Hajaj, N.; Hardt, M.; Liu, P. J.; Liu, X.; Marcus, J.; Sun, M.; Sundberg, P.; Yee, H.; Zhang, K.; Zhang, Y.; Flores, G.; Duggan, G. E.; Irvine, J.; Le, Q.; Litsch, K.; Mossin, A.; Tansuwan, J.; Wang, D.; Wexler, J.; Wilson, J.; Ludwig, D.; Volchenboum, S. L.; Chou, K.; Pearson, M.; Madabushi, S.; Shah, N. H.;

- Butte, A. J.; Howell, M. D.; Cui, C.; Corrado, G. S.; Dean, J. "Scalable and accurate deep learning with electronic health records", *NPJ Digit Med*, vol. 1, may 2018, pp. 18.
- [36] Rojas, J. C.; Carey, K. A.; Edelson, D. P.; Venable, L. R.; Howell, M. D.; Churpek, M. M. "Predicting intensive care unit readmission with machine learning using electronic health record data", *Ann Am Thorac Soc*, vol. 15–7, jul. 2018, pp. 846–853.
- [37] Saran, R.; Robinson, B.; Abbott, K. C.; Agodoa, L. Y. C.; Albertus, P.; Ayanian, J.; Balkrishnan, R.; Bragg-Gresham, J.; Cao, J.; Chen, J. L. T.; Cope, E.; Dharmarajan, S.; Dietrich, X.; Eckard, A.; Eggers, P. W.; Gaber, C.; Gillen, D.; Gipson, D.; Gu, H.; Hailpern, S. M.; Hall, Y. N.; Han, Y.; He, K.; Hebert, P.; Helmuth, M.; Herman, W.; Heung, M.; Hutton, D.; Jacobsen, S. J.; Ji, N.; Jin, Y.; Kalantar-Zadeh, K.; Kapke, A.; Katz, R.; Kovesdy, C. P.; Kurtz, V.; Lavalee, D.; Li, Y.; Lu, Y.; McCullough, K.; Molnar, M. Z.; Montez-Rath, M.; Morgenstern, H.; Mu, Q.; Mukhopadhyay, P.; Nallamothu, B.; Nguyen, D. V.; Norris, K. C.; O'Hare, A. M.; Obi, Y.; Pearson, J.; Pisoni, R.; Plattner, B.; Port, F. K.; Potukuchi, P.; Rao, P.; Ratkowiak, K.; Ravel, V.; Ray, D.; Rhee, C. M.; Schaubel, D. E.; Selewski, D. T.; Shaw, S.; Shi, J.; Shieu, M.; Sim, J. J.; Song, P.; Soohoo, M.; Steffick, D.; Streja, E.; Tamura, M. K.; Tentori, F.; Tilea, A.; Tong, L.; Turf, M.; Wang, D.; Wang, M.; Woodside, K.; Wyncott, A.; Xin, X.; Zang, W.; Zepel, L.; Zhang, S.; Zho, H.; Hirth, R. A.; Shahinian, V. "US renal data system 2016 annual data report: Epidemiology of kidney disease in the united states", *Am J Kidney Dis*, vol. 69–3 Suppl 1, mar. 2017, pp. A7–A8.
- [38] Shen, J.; Zhang, C. J.; Jiang, B.; Chen, J.; Song, J.; Liu, Z.; He, Z.; Wong, S. Y.; Fang, P.-H.; Ming, W.-K.; et al.. "Artificial intelligence versus clinicians in disease diagnosis: systematic review", *JMIR medical informatics*, vol. 7–3, 2019, pp. e10010.
- [39] Shickel, B.; Tighe, P. J.; Bihorac, A.; Rashidi, P. "Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis", *IEEE J Biomed Health Inform*, vol. 22–5, out. 2017, pp. 1589–1604.
- [40] Shillan, D.; Sterne, J.; Champneys, A.; Gibbison, B. "Use of machine learning to analyse routinely collected intensive care unit data: A systematic review", *Critical Care*, vol. 23, 12 2019.
- [41] Wickramasinghe, M.; Perera, D.; Kahandawaarachchi, K. "Dietary prediction for patients with chronic kidney disease (ckd) by considering blood potassium level using machine learning algorithms". In: 2017 IEEE Life Sciences Conference (LSC), 2017, pp. 300–303.



Pontifícia Universidade Católica do Rio Grande do Sul  
Pró-Reitoria de Pesquisa e Pós-Graduação  
Av. Ipiranga, 6681 – Prédio 1 – Térreo  
Porto Alegre – RS – Brasil  
Fone: (51) 3320-3513  
E-mail: [propesq@pucrs.br](mailto:propesq@pucrs.br)  
Site: [www.pucrs.br](http://www.pucrs.br)