

Avaliando Métodos de Aprendizado de Máquina na Concepção de uma Abordagem para Predição Antecipada do Risco de Mortalidade em UTIs

Alexandre Renato Rodrigues de Souza^{1,5}, Fabrício Neitzke Ferreira²,
Rodrigo Blanke Lambrecht^{3,4}, Leonardo Costa Reichow⁴,
Renata Hax Sander Reiser⁵, Adenauer Correa Yamin^{3,5}

¹IFRS, ²IFSUL, ³UCPEL, ⁴LIFEMED, ⁵UFPEL

Abstract. *The main objective of this research is the conception of an approach to predict the risk of mortality in ICUs. A cohort of 17,734 patients was used, from the MIMIC-III Database, considering 10 input predictor variables and 8 Machine Learning methods. The best performance was achieved by the Gradient Boosting Machine (GBM) method, which obtained 0,843 ($\pm 0,015$) of AUC and 0,503 ($\pm 0,048$) of F1 score. The results are promising and, in some cases, superior to those obtained by other proposals identified in the literature review.*

Resumo. *Esta pesquisa tem por objetivo central a concepção de uma abordagem para predição do risco de mortalidade em UTIs. Foi empregada uma coorte de 17.734 pacientes provenientes do Banco de Dados MIMIC-III, sendo consideradas 10 variáveis preditoras de entrada e 8 métodos de Aprendizado de Máquina. A melhor performance foi alcançada pelo método Gradient Boosting Machine (GBM), que atingiu 0,843 ($\pm 0,015$) de AUC e 0,503 ($\pm 0,048$) de F1 score. Os resultados são promissores e, em alguns casos, superiores a outras propostas identificadas na revisão de literatura.*

1. Introdução

As Unidades de Terapia Intensiva (UTIs) dos hospitais tem atraído relevantes esforços de pesquisa, pois os pacientes requerem monitoramento contínuo de seus parâmetros fisiológicos devido à gravidade de sua condição de saúde e também apresentarem elevado risco de rápida deterioração clínica [Purushotham and al. 2018].

Um desfecho primário de interesse nas terapias intensivas é a mortalidade, pois as suas taxas nas UTIs são as mais altas entre as unidades hospitalares, em torno de 10% a 29%, dependendo da idade e da doença. Deste modo, a identificação precoce de quais pacientes estão em maior risco é fundamental para melhorar os desfechos dos tratamentos [Harutyunyan et al. 2019].

Considerando este cenário, o presente trabalho tem por objetivo geral investigar o uso de métodos de classificação de Aprendizado de Máquina no desenvolvimento de um modelo de predição do risco mortalidade em UTIs, para auxiliar os médicos na tomada de decisão. Isto é feito analisando diferentes dados clínicos de pacientes coletados nas primeiras 48 horas após a internação.

Por sua vez, os objetivos específicos deste estudo são: (i) considerar os principais desafios no uso de técnicas de Aprendizado de Máquina para previsão do risco de mortalidade hospitalar de pacientes internados em UTIs; (ii) propor uma abordagem para a previsão do risco de mortalidade para estes pacientes; e, (iii) avaliar e comparar o desempenho de diferentes métodos de Aprendizado de Máquina, empregando um banco de dados construído a partir de informações reais provenientes de UTIs (MIMIC-III).

2. Trabalhos Relacionados Explorando Aprendizado de Máquina para Predição do Risco de Mortalidade em UTIs

Dentre os esforços associados à pesquisa apresentada neste artigo foi realizada uma Revisão Sistemática de Literatura, sendo identificados diversos estudos relacionados à predição do risco de mortalidade em UTIs. A descrição desses trabalhos relacionados, assim como uma comparação contemplando os aspectos centrais considerados na concepção da abordagem proposta por este estudo, estão disponíveis na Internet em um Apêndice¹.

3. Predição do Risco de Mortalidade em UTIs: Concepção da Abordagem

Esta seção descreve a abordagem proposta para lidar com a predição do risco de mortalidade de pacientes internados em UTIs, considerando as decisões de projeto que orientaram sua construção. Uma visão geral desta abordagem está caracterizada na Figura 1.

Discussão do Problema de Pesquisa

Os escores clínicos tradicionais de alerta precoce (EWS, SOFA, APACHE, SAPS, etc.) têm sido empregados para identificar a deterioração do estado de pacientes. Esses sistemas de pontuação, no entanto, levam em consideração somente os dados de saúde de um determinado instante de tempo, sem considerar a tendência de variação dos mesmos no decorrer da internação [Churpek and al. 2016].

Por sua vez, a implantação crescente dos prontuários eletrônicos nos hospitais tem viabilizado o registro de dados históricos dos pacientes, cujos sinais vitais e resultados de exames de laboratório coletados ao longo do tempo podem ser interpretados como séries temporais. Isto tem promovido a aplicação de técnicas computacionais que processam esses dados e permitem produzir previsões da evolução do estado clínico dos pacientes.

A abordagem cuja concepção é discutida neste artigo tem como objetivo prever se pacientes irão morrer ou sobreviver durante sua internação em UTIs. Para esta tarefa de classificação binária será utilizado como base os dados do MIMIC-III. Foram comparadas as performances de vários métodos de Aprendizado de Máquina utilizando dados coletados em uma janela de tempo de 48 horas após a admissão do paciente em uma UTI.

Banco de Dados e População Estudada

A coorte de pacientes relevantes para o desenvolvimento do modelo foi extraída do *Medical Information Mart for Intensive Care* (MIMIC-III) [Johnson and al. 2016]. Para o desenvolvimento do modelo foram incluídos apenas registros de pacientes com 18 anos ou mais e que permaneceram internados em UTIs por um período mínimo de 48 horas. Isso resultou em uma coorte de 17.734 pacientes e 1.456.610 observações. Desses pacientes, 15.328 sobreviveram e 2.406 morreram, resultando em uma taxa de mortalidade de 13,57%.

Análise Exploratória dos Dados e Seleção de Variáveis

As variáveis clínicas selecionadas e extraídas do MIMIC-III para desenvolvimento da abordagem proposta contém informações demográficas dos pacientes, resultados de exames de laboratório, sinais vitais e a escala de coma de *Glasgow*, conforme apresentado na caixa "Variáveis Clínicas" da Figura 1.

¹<http://ubiq.inf.ufpel.edu.br/arrdsouza/doku.php?id=apendice-semish2022>

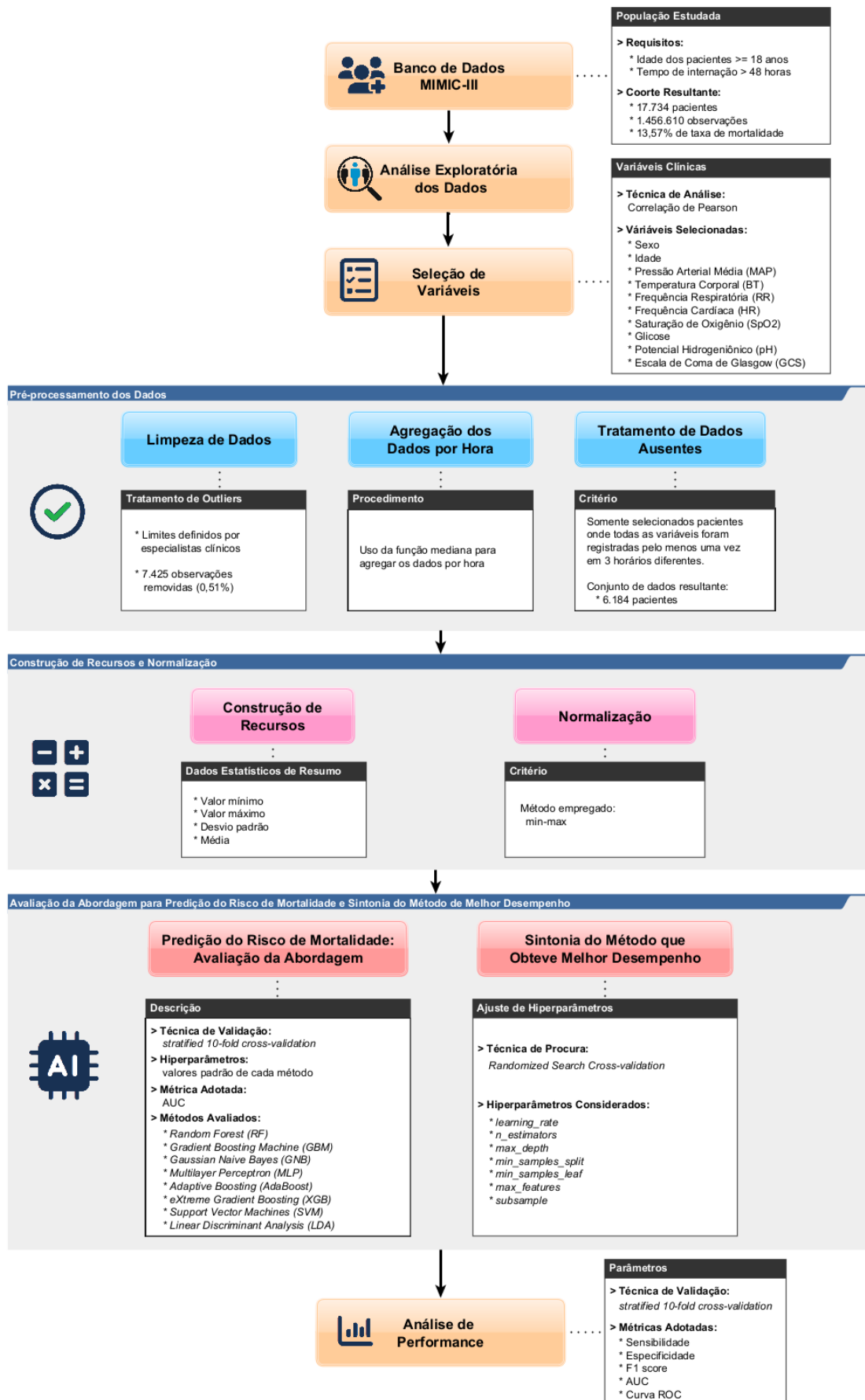


Figura 1. Visão Geral da Abordagem Proposta

Pré-processamento dos Dados

A qualidade e a quantidade de informações úteis são importantes fatores que irão influir na eficiência de um método de Aprendizado de Máquina. Dessa forma é indispensável uma boa preparação dos dados antes de alimentar os modelos. Esta tarefa é realizada pelo pré-processamento, o qual é composto pelas seguintes etapas:

- **Limpeza de Dados.** No modelo proposto, cada variável numérica está associada a limites superior e inferior definidos por especialistas para detectar valores inutilizáveis (*outliers*). O valor observado será excluído se estiver fora desses limites. Ao aplicar essas regras para gerar a coorte do modelo foram removidas 7.425 observações (0,51%) classificadas como valores atípicos extremos.
- **Agregação dos Dados por Hora.** As séries temporais brutas de cada variável oriundas do MIMIC-III são bastante esparsas. Dessa forma, com o objetivo de obter uma representação mais densa dos dados fisiológicos com o intuito de oportunizar uma melhor inferência pelos algoritmos, as observações de cada série temporal foram agregadas em intervalos de hora em hora através do cálculo da mediana [Zhu and al. 2018].
- **Tratamento de Dados Ausentes.** Como estratégia para minimizar o impacto na performance de predição devido aos dados ausentes, foram incluídos apenas pacientes onde todas as variáveis foram registradas pelo menos uma vez em três horários diferentes dentro da janela de tempo de medição de 48h. Este critério resultou em um conjunto de dados com 6.184 pacientes.

Construção de Recursos e Normalização

Para tratamento das informações ausentes nas séries temporais do MIMIC-III, o modelo proposto calcula os dados estatísticos resumidos (valor mínimo, valor máximo, desvio padrão e média) de cada uma das variáveis dentro da janela de tempo de 48 horas. Esta estratégia reduz a complexidade do modelo de Aprendizado de Máquina gerado, pois utiliza informações mais relevantes como entrada para a tarefa de previsão.

Muitos métodos de Aprendizado de Máquina exigem que as variáveis selecionadas estejam na mesma escala para um melhor desempenho [Faceli et al. 2021]. O método de normalização chamado “min-max” foi escolhido, através do qual os valores foram transformados para uma escala mínimo-zero e máximo-um.

4. Predição do Risco de Mortalidade em UTIs: Avaliação da Abordagem

O conjunto de dados de treinamento foi utilizado para avaliar a performance de predição do risco de mortalidade de 8 métodos de Aprendizado de Máquina. A Tabela 1 apresenta os resultados da AUC média e desvio padrão, os quais foram quantificados empregando o ajuste *default* dos hiperparâmetros de cada algoritmo. A melhor performance foi alcançada pelo método GBM, que atingiu o valor 0,841 ($\pm 0,024$) de AUC. Os valores de desempenho foram obtidos por validação cruzada para evitar avaliações tendenciosas.

A Figura 2 apresenta as curvas ROC dos métodos de predição avaliados. O canto inferior esquerdo das curvas ROC mostram que o algoritmo GNB apresentou menor Sensibilidade em baixas taxas de falsos positivos do que os demais métodos, o que indica menor capacidade em prever corretamente os pacientes que morreram.

Sintonia do Método de Aprendizado de Máquina que Obteve Melhor Desempenho

Após identificado que o método de Aprendizado de Máquina GBM alcançou a melhor performance de predição, foi feito o ajuste de seus hiperparâmetros para a aumentar o

desempenho do modelo. A sintonia do método foi realizada empregando a técnica *Randomized Search Cross-Validation* [Bergstra and Bengio 2012], a qual tem como objetivo encontrar a combinação de valores de hiperparâmetros que resultem no maior desempenho do modelo, medido através de validação cruzada.

Tabela 1. Performances de Classificação

Método	AUC Média	Desvio Padrão
RF	0.830	0.025
GBM	0.841	0.024
GNB	0.786	0.029
MLP	0.838	0.025
Adaboost	0.815	0.031
XGBoost	0.823	0.019
SVM	0.821	0.025
LDA	0.830	0.025

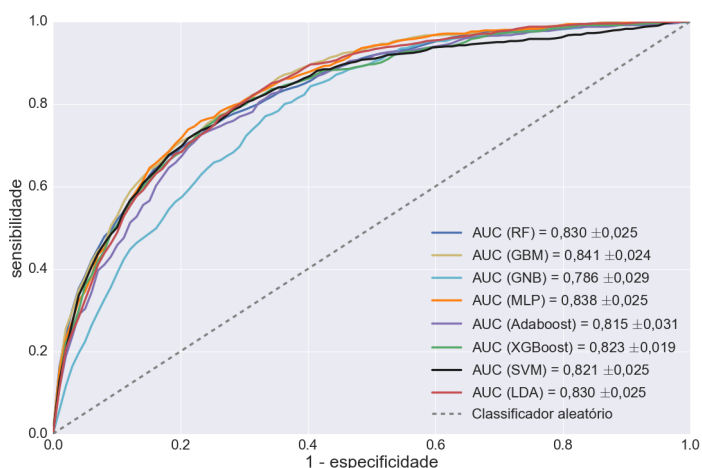


Figura 2. Curvas ROC dos Métodos de Predição

Análise de Performance do Método que Obteve Melhor Desempenho

Após encontrar os valores de hiperparâmetros otimizados para o método GBM, foi utilizado o conjunto de dados de teste para realizar a avaliação final de performance. A validação cruzada foi empregada para investigar a variabilidade no desempenho do modelo por meio de 10 diferentes partições de dados.

Conforme apresentado na Tabela 2, após a sintonia do método GBM por meio do ajuste de seus hiperparâmetros, foi alcançada uma performance AUC de 0,843 ($\pm 0,015$), obtida através de validação cruzada (*10-folds*). Na Figura 3 é apresentada a curva ROC média, também obtida através de validação cruzada. A região em cinza mostra a variância da curva ROC quando os dados são divididos em diferentes subconjuntos para treinamento e teste.

A performance AUC de 0,84 alcançada pela abordagem proposta é superior ao trabalho de [Alghatani et al. 2021], que obteve AUC de 0,78. Por sua vez, os trabalhos [Purushotham and at al. 2018] (AUC de 0,87), [Pirracchio and at al. 2015] (AUC de 0,88) e [Harutyunyan et al. 2019] (AUC de 0,87) apresentam melhor performance, entretanto exigem ao total 17 variáveis preditoras, dependem de até 6 exames laboratoriais ou requerem os valores individuais dos 4 critérios da escala de *Glasgow*, da qual muitas vezes é registrado somente o valor total.

Por fim, oportuno enfatizar que a performance obtida pela abordagem discutida neste artigo é superior aos escores tradicionais SAPS-II (AUC de 0,78) e SOFA (AUC de 0,71) [Pirracchio and at al. 2015], o que constitui um indicador bastante significativo para continuidade das pesquisas.

5. Considerações Finais

Este trabalho teve como contribuição central à concepção de uma abordagem para predição do risco de mortalidade hospitalar, explorando Aprendizado de Máquina e utilizando

dados coletados durante as primeiras 48 horas de internação nas UTIs. O MIMIC-III foi empregado como banco de dados de informações clínicas coletadas do mundo real.

Comparando aos escores clínicos tradicionais, a abordagem proposta utiliza técnicas de mineração de dados e Aprendizado de Máquina que geram modelos mais sofisticados, robustos e capazes de detectar padrões ocultos e ter maior poder de discriminação na classificação do risco de mortalidade em UTIs.

Tabela 2. Performance do modelo GBM

Métrica	Valor Médio	Desvio Padrão
AUC	0,843	0,015
Sensibilidade	0,754	0,031
Especificidade	0,761	0,035
F1 Score	0,503	0,048

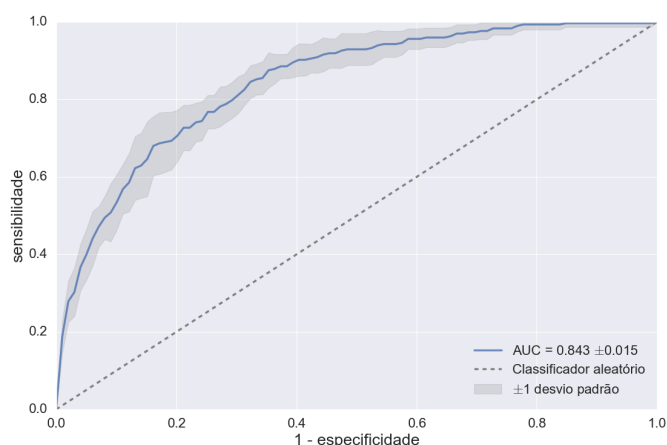


Figura 3. Curva ROC média obtida através de validação cruzada (10-folds)

Agradecimentos

O presente trabalho foi realizado com apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 e Instituto Federal do Rio Grande do Sul (IF-RS).

Referências

- Alghatani, K., Ammar, N., Rezgui, A., and Shaban-Nejad, A. (2021). Predicting intensive care unit length of stay and mortality using patient vital signs: Machine learning model development and validation.
- Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13:281–305.
- Churpek, M. M. and at al. (2016). Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. *Critical Care Medicine*, 44(2):368–374.
- Faceli, K., Lorena, A. C., Gama, J., de Carvalho, A. C. P. d. L. F., and de Almeida, T. A. (2021). *Inteligência Artificial - Uma Abordagem de Aprendizado de Máquina*. 2a edição edition.
- Harutyunyan, H., Khachatrian, H., Kale, D. C., Ver Steeg, G., and Galstyan, A. (2019). Multitask learning and benchmarking with clinical time series data. *Scientific Data*, 6(1).
- Johnson, A. E. and at al. (2016). MIMIC-III, a freely accessible critical care database. *Scientific Data*.
- Pirracchio, R. and at al. (2015). Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): A population-based study. *The Lancet Respiratory Medicine*, 3(1):42–52.
- Purushotham, S. and at al. (2018). Benchmarking deep learning models on large healthcare datasets.
- Zhu, Y. and at al. (2018). Predicting ICU mortality by supervised bidirectional LSTM networks.