



UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA

Autorizada pelo Decreto Federal nº 77.496 de 27/04/76
Recredenciamento pelo Decreto nº 17.228 de 25/11/2016



PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
COORDENAÇÃO DE INICIAÇÃO CIENTÍFICA

XXIX SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA UEFS **SEMANA NACIONAL DE CIÊNCIA E TECNOLOGIA - 2025**

Utilização de Aprendizado de Máquina na UTI para Predição de Tempo de Permanência

Davi Jatobá Galdino¹; Tiago Amador Coelho²

1.Davi Jatobá Galdino, PVIC/UEFS, ENGENHARIA DE COMPUTAÇÃO, e-mail: ddavijatoba33@gmail.com

2.Tiago Amador Coelho, DEPARTAMENTO DE CIÊNCIAS EXATAS, e-mail: tiago@uefs.br

PALAVRAS-CHAVE: Unidade de Terapia Intensiva; Machine Learning; Tempo de Permanência

INTRODUÇÃO

As Unidades de Terapia Intensiva (UTI), conforme a Resolução CFM nº 2.271/2020, são setores especializados no suporte vital avançado a pacientes críticos, com monitoramento contínuo e equipe multidisciplinar disponível 24 horas por dia. Nesse ambiente, são gerados e armazenados dados detalhados sobre a trajetória dos pacientes, incluindo informações demográficas, medições fisiológicas, observações clínicas, resultados de exames laboratoriais e intervenções terapêuticas (Shillan et al., 2019).

O grande volume e complexidade desses dados tornam as UTIs ambientes propícios para a aplicação de técnicas de aprendizado de máquina (Machine Learning), que utilizam métodos matemáticos, estatísticos e computacionais para construir modelos capazes de prever resultados específicos, classificar informações ou detectar padrões (Paixão et al., 2022). Nesse contexto, muitos estudos têm empregado Machine Learning para prever o tempo de permanência em hospitais, uma informação essencial para o planejamento e a gestão de recursos, impactando diretamente no fluxo de pacientes, no consumo de insumos e na gravidade das condições clínicas (Stone et al., 2022).

No Brasil, os gastos com saúde representaram 4,7% do Produto Interno Bruto (PIB) em 2023 (Tesouro Nacional, 2025), o que evidencia a necessidade de otimizar a gestão hospitalar. Assim, a previsão do tempo de permanência não apenas contribui para aprimorar a experiência do paciente e a qualidade dos serviços de saúde, mas também auxilia na organização do trabalho dos profissionais, na tomada de decisões médicas e no gerenciamento eficiente dos recursos hospitalares (Awad et al., 2017).

MATERIAL E MÉTODOS OU METODOLOGIA

O projeto teve início com uma revisão preliminar da literatura sobre a aplicação de técnicas de Machine Learning na área médica, com foco na predição do tempo de permanência em UTIs. Paralelamente, foi realizado um estudo das principais ferramentas, bibliotecas e algoritmos utilizados no desenvolvimento de modelos

preditivos baseados em Machine Learning. As ferramentas empregadas incluíram VSCode, Python e Anaconda, bem como as bibliotecas Pandas, NumPy, Scikit-Learn e MLflow. Os experimentos foram conduzidos em um notebook com Windows 11, processador Intel Core i5-120U, 16 GB de RAM e SSD de 512 GB.

Foram implementados e avaliados três algoritmos do tipo *ensemble*, voltados à tarefa de regressão e amplamente utilizados na literatura: **XGBoost**, **LightGBM** e **Random Forest**. Em todos os experimentos, os dados foram divididos em 80% para treinamento e 20% para teste, estratégia comumente adotada em estudos de aprendizado de máquina. Apenas no modelo final, essa proporção foi ajustada para 70% de treino e 30% de teste.

As métricas de avaliação utilizadas foram o **Coefficiente de Determinação (R^2)** e a **Raiz do Erro Quadrático Médio (RMSE)**. O R^2 indica a proporção da variância total da variável dependente explicada pelas variáveis independentes; valores próximos de 1 demonstram que o modelo consegue explicar grande parte da variação dos dados (Tatachar, 2021). Já o RMSE representa, em média, o desvio das previsões em relação aos valores reais, estando na mesma unidade da variável de saída; valores menores indicam melhor desempenho do modelo (Tatachar, 2021).

O desenvolvimento do projeto seguiu cinco etapas principais:

Exploração da base de dados: A base **MIMIC-III v1.4** foi analisada para compreender sua estrutura, relacionamentos e volume de dados. Foram selecionadas cinco tabelas principais (CHARTEVENTS, ICUSTAYS, ADMISSIONS, PATIENTS, DIAGNOSES-ICD), contendo informações clínicas, administrativas e demográficas.

Pré-processamento dos dados de admissão: Incluiu a seleção de *features* relevantes, unificação das tabelas, filtragem de internações inadequadas o estudo e engenharia de *features*. *Outliers* na variável LOS foram avaliados por meio de testes com diferentes limites de permanência, sendo o melhor resultado obtido com limite de 21 dias.

Pré-processamento dos dados de monitoramento: foram removidos registros inválidos da tabela CHARTEVENTS e agregados valores numéricos por internação, considerando diferentes janelas temporais. As tabelas resultantes foram então integradas aos dados de admissão. Também foi realizada a seleção de *features*, tratando cada tipo de registro clínico como uma variável (totalizando 2.117). A relevância dessas *features* foi avaliada com diferentes técnicas, sendo escolhida a seleção com base nos quartis das *feature importances* obtidas no treinamento dos algoritmos.

Criação e avaliação dos modelos: Foram criados modelos de predição com diferentes configurações dos dados, variando as janelas temporais de monitoramento, a quantidade de *features*, os limites de tempo de permanência e os algoritmos utilizados. Esses modelos foram treinados com as *features* selecionadas e avaliados pelas métricas, para encontrar combinação mais promissora para criação do modelo final.

Melhorias nas *features* e ajuste de hiperparâmetros: Foram realizados experimentos com a expansão das *features*, incorporando diferentes medidas estatísticas além da

média, como mediana, máximo, mínimo e desvio padrão. Em seguida, os hiperparâmetros foram ajustados por meio do GridSearchCV, que possibilitou testar diversas combinações e selecionar uma configuração que otimize o desempenho preditivo.

RESULTADOS E DISCUSSÃO

Nos experimentos realizados o **LightGBM** apresentou o melhor desempenho geral, exemplificado pelos valores de R^2 mostrados na **Figura 1**, referente aos experimentos com diferentes janelas temporais.

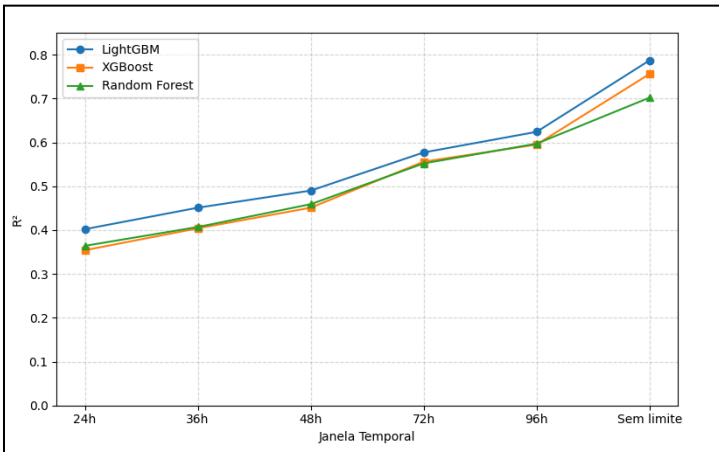


Figura 1: Comparativo dos valores de R^2 dos modelos LightGBM, XGBoost e Random Forest em diferentes janelas temporais.

A **Tabela 1** resume os desempenhos obtidos com o LightGBM em três variações principais: experimentos com diferentes janelas temporais de monitoramento, aplicação da remoção de *outliers* no tempo de permanência e uso de diferentes *thresholds* para seleção de *features*.

Tabela 1: Desempenho no R^2 utilizando o LightGBM em diferentes experimentos.

Algoritmo	Janelas temporais (24h, 36h, 48h, 72h, 96h e sem limite)	Remoção de <i>Outliers</i> (limite de 21 dias / sem remoção)	<i>Thresholds</i> para seleção de <i>features</i> (0,75; 0,80; 0,90; 0,95; 0,975; 0,99)
LightGBM	0,402; 0,451; 0,490; 0,577; 0,624; 0,787	0,401; 0,371	0,375; 0,391; 0,401; 0,403; 0,404; 0,404

Os experimentos com diferentes janelas temporais testaram modelos com dados clínicos coletados em períodos distintos de monitoramento, permitindo avaliar o impacto do tempo de coleta. Janelas maiores forneceram mais informações, melhorando o desempenho. A remoção de *outliers* excluiu internações excessivamente longas consideradas atípicas, e a seleção de *features* eliminou variáveis de baixa importância utilizando diferentes *thresholds*, reduzindo a complexidade e o volume dos dados.

Em destaque na tabela, os valores selecionados foram: **24 horas** para uma predição antecipada, **limite de 21 dias** de permanência e **threshold de 0,95** (remoção de 95% das *features* de menor importância), por apresentarem melhor desempenho e equilíbrio entre R^2 (mostrado na tabela) e RMSE.

A expansão das *features*, utilizando média e desvio padrão normalizados, juntamente com a tunagem de hiperparâmetros e a utilização de 70% dos dados para treino e 30% para teste, elevou o desempenho do modelo. Com essa abordagem, o **R² aumentou de 0,402 para 0,421** e o **RMSE reduziu de 2,658 para 2,569**.

O modelo final adotado foi o **LightGBM**, treinado com dados das primeiras 24 horas de monitoramento, limite de permanência de 21 dias, *threshold* de 0,95 e *features* expandidas. Esse modelo apresentou desempenho de **R² = 0,421** e **RMSE = 2,569**.

CONSIDERAÇÕES FINAIS

A predição do tempo de permanência em UTI continua sendo um desafio, mesmo com a aplicação de diferentes algoritmos e tratamentos de dados. Nenhum conjunto de *features* foi capaz de explicar totalmente a variabilidade da internação, apesar do uso de dados laboratoriais, demográficos e clínicos. O desenvolvimento deste trabalho evidenciou que projetos de Machine Learning exigem inúmeros testes, análises e comparações, tornando essencial a documentação cuidadosa de cada etapa. Nesse contexto, o MLflow mostrou-se uma ferramenta valiosa para organizar e acompanhar os experimentos. Observou-se que janelas temporais maiores e a remoção de *outliers* tendem a melhorar o desempenho do modelo, enquanto alterações nos dados podem gerar grande variabilidade nos resultados. Durante os experimentos, foram identificadas as variáveis mais importantes para prever o tempo de internação: dados da admissão (como unidade inicial, tipo de admissão e diagnóstico), sinais vitais e exames laboratoriais para internações curtas, e a escala de coma de Glasgow para internações mais longas. Apesar das limitações computacionais, o projeto alcançou desempenho competitivo, mostrando que é possível construir modelos precisos mesmo com restrições de *hardware*. Trabalhos futuros podem explorar melhor cada *feature*, integrar dados de diferentes bases e testar técnicas adicionais de tratamento, especialmente com maior capacidade computacional. Por fim, este estudo confirma que a predição do tempo de permanência na UTI é um desafio relevante, mas constitui uma variável crucial para a administração hospitalar, impactando a gestão de recursos, a eficiência do tratamento e a qualidade dos cuidados ao paciente.

REFERÊNCIAS

- Conselho Federal de Medicina. (2020). Código de Ética Médica: Resolução CFM nº 2.271/2020. Disponível em: <https://portal.cfm.org.br>. Acesso em: jul. 2025.
- Shillan, D., Sterne, J. A., Champneys, A., & Gibbison, B. (2019). Use of machine learning to analyse routinely collected intensive care unit data: A systematic review. *Critical Care*, 23(1), 284.
- Tatachar, A. V. (2021). Comparative assessment of regression models based on model evaluation metrics. *International Research Journal of Engineering and Technology (IRJET)*, 8(09), 2395–0056.
- Johnson, A. E., Pollard, T. J., Shen, L., Lehman, L.-W. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3(1), 1–9.