

Modelo de aprendizado de máquina para predição da sepse e avaliação da gravidade com foco em mortalidade

Luiza S. B. Leidemer¹, Cláudio F. R. Geyer¹

¹Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

lsbleidemer@inf.ufrgs.br, geyer@inf.ufrgs.br

Resumo. A sepse é uma resposta inadequada a infecções, com alta taxa de mortalidade, especialmente no Brasil. O estudo visa desenvolver modelos para prever a mortalidade por sepse, superando métodos tradicionais e avaliando o desempenho em bancos de dados de diferentes países, incluindo o brasileiro, além de examinar questões éticas da implementação clínica. Serão utilizados algoritmos como Random Forest e XGBoost, com pré-processamento de dados, como limpeza, normalização e estratificação. Os dados serão coletados de hospitais e bases como MIMIC e eICU. Espera-se identificar o melhor modelo, explorar os vieses entre as bases de dados e avaliar a aplicabilidade clínica dos algoritmos, com a percepção dos profissionais de saúde.

1. Introdução

A sepse, também conhecida como infecção generalizada, é uma resposta inadequada do organismo à uma infecção, resultando em disfunção orgânica progressiva, podendo ser fatal [Instituto Latino-Americano de Sepse (ILAS) 2018]. No Brasil, são registrados cerca de 400 mil casos de sepse em adultos por ano, com 240 mil óbitos. Atualmente, o Brasil tem uma taxa de mortalidade por sepse maior que a de países desenvolvidos, exigindo mais atenção ao problema [Ministério da Saúde 2021].

Algoritmos de aprendizagem de máquina são amplamente utilizados para prever sepse, auxiliar condutas médicas e ações preventivas [Bezerra et al. 2023]. Modelos como árvores de decisão permitem análise preditiva e identificação de padrões em grandes volumes de dados. Eles são avaliados por métricas que garantem previsões confiáveis, evitam *overfitting* e otimizam o desempenho [SILVA et al. 2016]. Sua implementação hospitalar pode reduzir a mortalidade, o tempo de internação e os custos [Islam et al. 2019].

O estudo visa desenvolver algoritmos de *machine learning* para realizar a predição da sepse e sua mortalidade, superando métodos tradicionais, e avaliar o desempenho dos modelos em bancos de dados de diferentes países, incluindo um brasileiro, a fim de verificar possíveis vieses dos algoritmos entre bases distintas. Além das análises técnicas, serão avaliadas as questões éticas da implementação desses modelos no fluxo clínico.

2. Metodologia

A pesquisa usará os algoritmos *Random Forest*, *Decision Tree*, *k-nearest neighbors*, *XGBoost*, *LightGBM* e Regressão Logística para prever a mortalidade de pacientes com sepse, utilizando dados demográficos, sinais vitais, exames laboratoriais e intervenções médicas. Modelos de aprendizado profundo não serão utilizados, pois exigem grandes volumes de dados para treinamento eficaz, e as bases disponíveis são pequenas. Além disso,

embora redes neurais ofereçam alto desempenho, sua falta de interpretabilidade dificulta a aplicação clínica, onde a transparência das decisões é essencial [Sharma et al. 2021]. O pré-processamento incluirá limpeza, imputação de valores ausentes e remoção de inconsistências. A normalização e padronização serão aplicadas às variáveis numéricas. Os dados serão divididos em 70% para treinamento e 30% para teste, com estratificação e uso do SMOTE devido ao desbalanceamento. A atribuição de pesos às classes também será considerada. A avaliação dos modelos será feita por acurácia, sensibilidade, AUC-ROC e F1-score. Os dados serão coletados de um hospital em Porto Alegre, além das bases MIMIC e eICU. A coleta segue as normas éticas e não envolve riscos diretos aos pacientes. O estudo incluirá adultos maiores de 18 anos com indicação para protocolo de sepse, excluindo pacientes terminais sem previsão de medidas invasivas ou reanimação. O plano secundário foca na predição de aspectos adicionais, como a ocorrência de sepse, o tempo de admissão em UTI e o tempo de permanência hospitalar. A aplicabilidade clínica será analisada considerando a infraestrutura hospitalar e a interpretação dos resultados.

3. Resultados Esperados

Os resultados esperados incluem a identificação do melhor modelo para a predição da mortalidade por sepse, superando métodos tradicionais (como SIRS ou SOFA). Além disso, serão analisadas as melhores características e suas correlações, aplicadas técnicas de pré-processamento de dados e abordadas questões éticas no uso de algoritmos de aprendizado de máquina. A análise de diferentes bases de dados, incluindo a brasileira, permitirá identificar potenciais vieses entre as bases e ajustar os modelos conforme necessário. Espera-se que as métricas AUC-ROC, F1-score e sensibilidade mostrem um bom desempenho na detecção de pacientes com risco de mortalidade.

O estudo também avaliará a percepção dos profissionais de saúde sobre o uso do algoritmo, explorando sua integração como ferramenta de apoio à prática clínica. Por fim, a análise das similaridades e diferenças entre os bancos de dados fornecerá *insights* valiosos para a adaptação e validação dos modelos em ambientes hospitalares.

Referências

- Bezerra, A., Maciel, N., Filho, L., Mendes, A., Gois, F., and Silva, L. (2023). Efetividade de algoritmos de inteligência artificial para predição de sepse em adultos de unidades de terapia intensiva: revisão de escopo. *Revista Interfaces*.
- Instituto Latino-Americano de Sepse (ILAS) (2018). Implementação de protocolo gerenciado de sepse.
- Islam, M., Nasrin, T., Walther, B. A., Wu, C.-C., Yang, H.-C., and Li, Y.-C. (2019). Prediction of sepsis patients using machine learning approach: A meta-analysis. *Computer Methods and Programs in Biomedicine*.
- Ministério da Saúde (2021). Dia Mundial da Sepse: Ministério da Saúde alerta para a importância do diagnóstico precoce. Acesso em: [13 Nov 2024].
- Sharma, N., Sharma, R., and Jindal, N. (2021). Machine learning and deep learning applications - a vision. *Global Transitions Proceedings*, 2(1):24–28.
- SILVA, L. A., PERES, J., and BOSCARIOLI, C. (2016). *Introdução à Mineração de Dados com Aplicações em R*. s.n., São Paulo.