

Utilização de *machine learning* como método de análise para predição de injúria renal aguda de pacientes em pós-operatório de cirurgia cardíaca no HU-UFPI

Bolsista: Antonia Márcia Sousa Cardoso

Orientador: Dr. Ginivaldo Victor Ribeiro do Nascimento

Hospital Universitário, da Universidade Federal do Piauí (HU-UFPI)

Teresina, 21 de agosto de 2024

Introdução

É notório que o uso da inteligência artificial (IA) expandirá a cada dia devido à capacidade de uma máquina executar tarefas cognitivas humanas e trazer diversos benefícios para as áreas de atuação. Neste cenário, o *machine learning* é um vertente com a função de permitir ao computador o reconhecimento de padrões e criação de previsões através de algoritmos, construindo um modelo de aprendizado. Dentro da medicina, a aplicação da tecnologia pode auxiliar na previsão de detecção de disfunção de órgão e patologias, incluindo complicações. O estudo tem como questão norteadora: o desenvolvimento de injúria renal aguda e seus desfechos desfavoráveis, no pós-cirúrgico cardíaco, poderia ser previsto com o auxílio da ferramenta de predição de dados através de *Machine Learning*? Este estudo tem como objetivo desenvolver uma ferramenta de predição para a análise e avaliação da injúria renal aguda em pacientes do Hospital Universitário da Universidade Federal do Piauí (HU-UFPI), bem como fornecer dados relevantes sobre a ocorrência de complicações pós-operatórias em pacientes submetidos à cirurgia cardíaca.

Metodologia

O presente projetou de um estudo de análise preditiva exploratória, uma vez que os dados foram tratados para, então, se buscar um padrão que explique a evolução de complicações dos pacientes pós-cirúrgicos, submetidos a cirurgias cardíacas no Hospital Universitário, da Universidade Federal do Piauí (HU-UFPI), do município de Teresina (PI), operados entre janeiro de 2018 a dezembro de 2022.

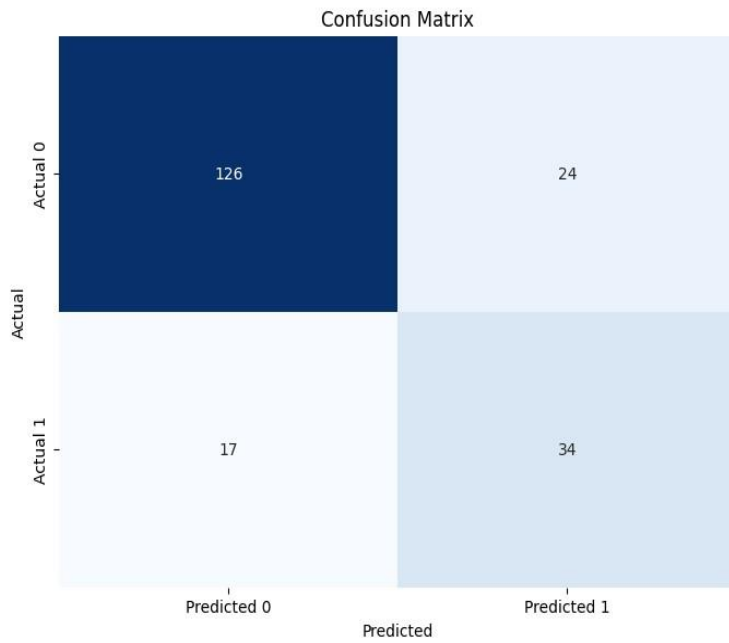
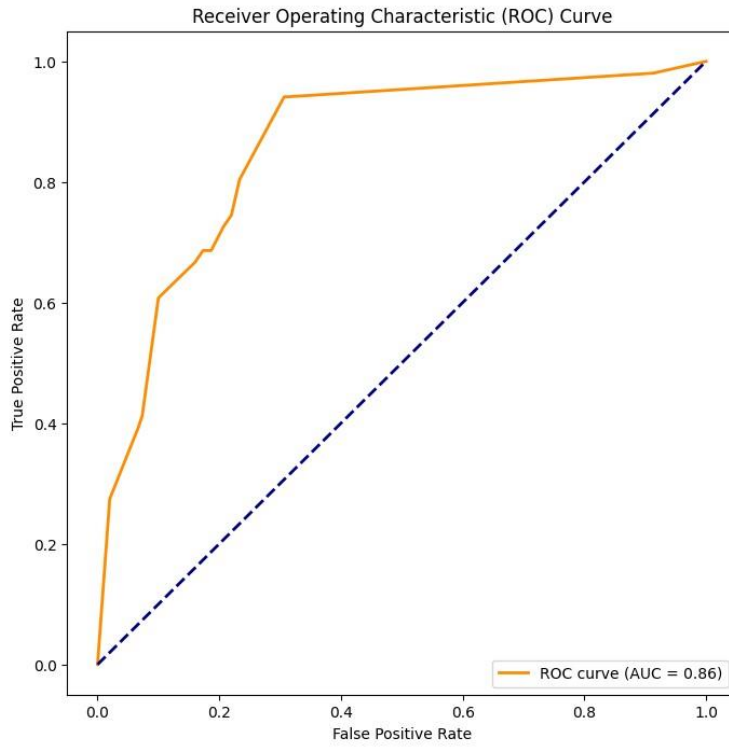
Este projeto foi conduzido por meio da coleta de dados primários, provenientes de pacientes internados no setor de cardiologia do Hospital Universitário da Universidade Federal do Piauí (HU-UFPI). Esses dados foram extraídos do banco de dados do sistema interno AGHU utilizando consultas estruturadas em SQL, realizadas pelo setor de inovação do HU-UFPI. Após a extração, os dados foram anonimizados com o auxílio técnico de uma equipe de tecnologia da informação.

A condução da metodologia baseou-se na utilização da linguagem de programação Python, reconhecida por sua eficácia na manipulação e análise de dados. Os dados foram obtidos por meio de consultas SQL ao banco de dados do sistema interno AGHU, conduzidas pelo setor de inovação do HU-UFPI, e posteriormente

submetidos a um processo de anonimização para garantir a confidencialidade dos pacientes.

Resultados

Os resultados foram traduzidos por dois modelos, de predição e de classificação. O primeiro construído utilizando a técnica de ExtraTreesRegressor, foi capaz de prever os valores contínuos da creatinina sérica com alta precisão. Após a otimização dos hiperparâmetros e validação cruzada, o modelo alcançou uma precisão impressionante, com um R^2 de aproximadamente 0.99 no conjunto de teste, indicando que o modelo foi capaz de explicar cerca de 99% da variabilidade nos valores observados da creatinina sérica. Por outro lado, de classificação focou na identificação binária (técnica de DecisionTreeClassifier) de casos de IRA entre os pacientes. Após a otimização dos hiperparâmetros e validação cruzada, o modelo alcançou um recall de aproximadamente 81.59% no conjunto de validação, o que indica sua capacidade de identificar corretamente os casos positivos de IRA. A análise adicional da curva ROC mostrou que o modelo teve um AUC cerca de 0.79, sugerindo um desempenho moderadamente bom na distinção entre verdadeiros positivos e falsos positivos.



PrevINA



EBSERH

Ferramenta de Inferência de Insuficiência Renal Aguda com Machine Learning

Informações do(a) Paciente

Para utilizar o processo de inferência, preencha todos os campos abaixo:

O que é Insuficiência Renal Aguda?

A Insuficiência Renal Aguda (IRA) é uma condição clínica súbita e abrupta da função renal, geralmente em horas ou dias.

Principais Consequências da IRA:

- Acúmulo de toxinas no sangue
- Desequilíbrio de líquidos e eletrólitos
- Inchaço
- Hipertensão
- Convulsões
- Falência renal
- Morte

O que é PrevINA?

PrevINA é uma ferramenta desenvolvida para unir medicina e machine learning. Ela utiliza o modelo de aprendizado de máquina supervisionado "ExtraTreesRegressor" com dados curados da base do HU-UFPI para oferecer auxílio no diagnóstico precoce da IRA.

Benefícios da Ferramenta:

- Identificação de pacientes em risco
- Agilização do diagnóstico
- Melhoria do prognóstico

MAE da Creatinina

0.006

Acuracia da IRA

+90%

Recall da IRA

98%



Sem indicativo de IRA

Foi inferido o valor de 0.8 mg/dL de Creatinina para as próximas 48 horas. O valor inferido não ultrapassa o limiar de variação de 30% da Creatinina atual, dessa forma não há indicativo de IRA.

Entendo

PrevINA



EBSERH

Ferramenta de Inferência de Insuficiência Renal Aguda com Machine Learning

Informações do(a) Paciente

Para utilizar o processo de inferência, preencha todos os campos abaixo:

O que é Insuficiência Renal Aguda?

A Insuficiência Renal Aguda (IRA) é uma condição clínica caracterizada pela perda súbita e abrupta da função renal, geralmente em horas ou dias.

Principais Consequências da IRA:

- Acúmulo de toxinas no sangue
- Desequilíbrio de líquidos e eletrólitos
- Inchaço
- Hipertensão
- Convulsões
- Falência renal
- Morte

O que é PrevINA?

PrevINA é uma ferramenta desenvolvida para unir medicina e machine learning. Ela utiliza o modelo de aprendizado de máquina supervisionado "ExtraTreesRegressor" com dados curados da base do HU-UFPI para oferecer auxílio no diagnóstico precoce da IRA.

Benefícios da Ferramenta:

- Identificação de pacientes em risco
- Agilização do diagnóstico
- Melhoria do prognóstico

MAE da Creatinina

0.006

Acuracia da IRA

+90%

Recall da IRA

98%

Idade:

Digite aqui a idade...

Valor creatinina (atual):

Taxa do coração:

Inferir IRA

A ferramenta PrevINA visa auxiliar profissionais de saúde. Os resultados possuem um grau de erro e devem ser avaliados com cuidado.

Desenvolvido por:

- Antonia Márcia [Lattes](#)
- Givaldo Victor [Lattes](#)
- Francisco Luciani [Lattes](#), [LinkedIn](#)
- Gabriel Coelho [Lattes](#), [GitHub](#), [LinkedIn](#), [Medium](#), [Substack](#), [Instagram](#)

Projeto disponível no [GitHub](#)

Discussão

Ao comparar os resultados dos dois modelos, observamos que ambos apresentaram desempenho promissor na detecção de casos de IRA. Enquanto o modelo de predição se destacou pela precisão na previsão dos valores de creatinina sérica, já o de classificação foi eficaz na identificação binária de casos positivos de IRA. No entanto, é importante considerar as diferentes abordagens e métricas de avaliação utilizadas em cada modelo. Embora ambos os modelos tenham demonstrado um desempenho promissor, é crucial considerar as necessidades e objetivos específicos do contexto clínico do paciente ao escolher entre modelos de predição e classificação. Em alguns casos, pode ser mais relevante prever os valores contínuos de biomarcadores, como a creatinina sérica, enquanto em outros casos pode ser mais importante identificar de forma binária a presença de uma condição clínica, como a IRA.

COMPARAÇÃO A PERFORMANCE DOS MODELOS DE PREDIÇÃO E CLASSIFICAÇÃO:

✓ Modelo de Predição (Regressão)

O modelo de predição, construído utilizando a técnica de ExtraTreesRegressor, foi capaz de prever os valores contínuos da creatinina sérica com alta precisão. Após a otimização dos hiperparâmetros e validação cruzada, o modelo alcançou uma precisão impressionante, com um Coeficiente de Determinação (R^2) de aproximadamente 0.99 no conjunto de teste. Isso indica que o modelo foi capaz de explicar cerca de 99% da variabilidade nos valores observados da creatinina sérica.

Além disso, a análise dos erros mostrou que o modelo teve um desempenho consistente em prever os valores de creatinina sérica, com baixos valores de erro médio absoluto (MAE), erro médio quadrático (MSE) e raiz do erro médio quadrático (RMSE).

✓ Modelo de Classificação (DecisionTreeClassifier)

Por outro lado, o modelo de classificação, desenvolvido utilizando a técnica de DecisionTreeClassifier, focou na identificação binária de casos de IRA entre os pacientes. Após a otimização dos hiperparâmetros e validação cruzada, o modelo alcançou um recall de aproximadamente 81.59% no conjunto de validação, o que indica sua capacidade de identificar corretamente os casos positivos de IRA.

A análise adicional da curva ROC mostrou que o modelo teve uma área sob a curva (AUC) de aproximadamente 0.79, o que sugere um desempenho moderadamente bom na distinção entre verdadeiros positivos e falsos positivos.

Conclusão

Embora ambos os modelos tenham demonstrado um desempenho promissor, é crucial considerar as necessidades e objetivos específicos do contexto clínico ao escolher entre modelos de predição e classificação. Em alguns casos, pode ser mais relevante prever os valores contínuos de biomarcadores, como a creatinina sérica, enquanto em outros casos pode ser mais importante identificar de forma binária a presença de uma condição clínica, como a IRA.

Portanto, a escolha entre modelos de predição e classificação deve ser guiada pela natureza da tarefa de predição, as características dos dados disponíveis e as necessidades clínicas

Referências

- ANACONDA Inc. (2021). Miniconda. Recuperado de <https://docs.conda.io/en/latest/miniconda.html>
- ABADI, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... & Zheng, X. (2016). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.
- BUDTZ-JØRGENSEN, E., Keiding, N., Grandjean, P., Weihe, P. (2019). Confounder selection in environmental epidemiology: Assessment of health effects of prenatal mercury exposure. *Annals of Epidemiology*, 20(4), 226-233.
- CHOLLET, F., et al. (2015). Keras. <https://github.com/fchollet/keras>
- FRIEDMAN, J., Hastie, T., & Tibshirani, R. (2019). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
- GAMEIRO, J., et al., Acute Kidney Injury Definition and Diagnosis: A Narrative Review. *J Clin Med*, 2018. 7(10).
- GUO, H., Wu, C., Wu, D., & Liu, Z. (2016). Exploratory undersampling for class-imbalance learning. *Pattern Recognition*, 60, 649-664.

HARRINGTON, P. Predictive modeling with Python: A hands-on approach to solving real-world predictive problems. O'Reilly Media, Inc. (2019).

HUNTER, J. D. (2007). Matplotlib: A 2D graphics environment. IEEE Annals of the History of Computing, 9(03), 90-95.

KHWAJA, A., KDIGO Clinical Practice Guidelines for Acute Kidney Injury. Nephron Clinical Practice, 2012. 120(4): p. c179-c184.

Little, R. J., & Rubin, D. B. (2019). Statistical analysis with missing data. John Wiley & Sons.

McKinney, W. (2010). Data structures for statistical computing in Python. In Proceedings of the 9th Python in Science Conference (Vol. 445, No. 445, pp. 51-56).

Mohamadlou, H., et al., Prediction of Acute Kidney Injury With a Machine Learning Algorithm Using Electronic Health Record Data. Can J Kidney Health Dis, 2018. 5: p. 2054358118776326.

Ostermann, M. and M. Joannidis, Acute kidney injury 2016: diagnosis and diagnostic workup. Critical care (London, England), 2016. 20(1): p. 299-299.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. Journal of machine learning research, 12(Oct), 2825-2830. Plotly Technologies Inc. (2015). Plotly. Recuperado de <https://plotly.com/python/>

THONGPRAYOON, C., et al., Diagnostics, Risk Factors, Treatment and Outcomes of Acute Kidney Injury in a New Paradigm. Journal of clinical medicine, 2020. 9(4): p. 1104. Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. CreateSpace.

Waskom, M., Botvinnik, O., Hobson, P., Warmenhoven, J., Cole, J. B., Halchenko, Y., ... & Augspurger, T. (2021). mwaskom/seaborn: v0.11.1 (January 2021). Zenodo. <https://doi.org/10.5281/zenodo.4462225>

WEI, C., et al., Machine learning model for predicting acute kidney injury progression in critically ill patients. BMC medical informatics and decision making, 2022. 22(1): p. 17-17.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ... & Yutani, H. (2019). Welcome to the Tidyverse. Journal of Open Source Software, 4(43), 1686.

REIS, J. C.; et al. Avaliação da incidência e preditores de injúria renal aguda após cirurgia cardíaca. *Revista Brasileira de Cirurgia Cardiovascular*, v. 37, n. 3, p. 223-230, 2022.

Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. *Computational and Structural Biotechnology Journal*, 15, 104-116.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794).

Pirracchio, R., Petersen, M. L., & Carone, M. (2019). Machine learning for clinical prediction models. *Jama*, 322(20), 2007-2008.