

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

**Metodologia para prever a reinternação hospitalar
de pacientes baseada em Inteligência Artificial
Explicável**

Matheus Mendes dos Santos

Monografia - MBA em Ciência de Dados (CEMEAI)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Matheus Mendes dos Santos

Metodologia para prever a reinternação hospitalar de pacientes baseada em Inteligência Artificial Explicável

Monografia apresentada ao Centro de Ciências Matemáticas Aplicadas à Indústria do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Ciências de Dados.

Área de concentração: Ciências de Dados

Orientador: Prof. Dr. Rodrigo Colnago Contreras

Versão original

São Carlos

2023

É possível elaborar a ficha catalográfica em LaTeX ou incluir a fornecida pela Biblioteca. Para tanto observe a programação contida nos arquivos USPSC-modelo.tex e fichacatalografica.tex e/ou gere o arquivo fichacatalografica.pdf.

A biblioteca da sua Unidade lhe fornecerá um arquivo PDF com a ficha catalográfica definitiva, que deverá ser salvo como fichacatalografica.pdf no diretório do seu projeto.

Matheus Mendes dos Santos

Methodology for Predicting Hospital Readmission of Patients Based on Explainable Artificial Intelligence

Monograph presented to the Centro de Ciências Matemáticas Aplicadas à Indústria do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, as part of the requirements for obtaining the title of Specialist in Data Science.

Concentration area: Data Science

Advisor: Prof. Dr. Rodrigo Colnago Contreras

Original version

São Carlos

2023

SUMÁRIO

1	INTRODUÇÃO	9
1.1	Contextualização e Motivação	9
2	REVISÃO BIBLIOGRÁFICA	11
2.1	Trabalhos de classificação na área médica	11
2.2	Técnicas de explicação aplicadas a trabalhos de classificação na área médica	12
3	DESENVOLVIMENTO	15
3.1	Metodologia	15
3.1.1	Definição de domínio do problema	15
3.1.2	Método proposto	15
3.1.2.1	Limpeza, análise exploratória e preparação dos dados	15
3.1.2.2	Treinamento dos classificadores	15
3.1.2.3	Análise de variáveis mais importantes	16
3.1.3	Apresentação dos resultados	16
	REFERÊNCIAS	17

1 INTRODUÇÃO

1.1 Contextualização e Motivação

Nos últimos anos, temos nos deparado com uma imensa quantidade de dados gerados das mais variadas fontes. Desde a utilização de *smartphones* e redes sociais até a realização de transações bancárias e o uso de dispositivos inteligentes em nossas casas, cada ação que realizamos deixa um rastro digital de informações. A internet, sendo um dos pilares dessa revolução digital, desempenha um papel fundamental na geração de dados, permitindo a troca instantânea de informações entre pessoas e sistemas em todo o mundo. Além disso, sensores e dispositivos conectados à Internet das Coisas (IoT) coletam e transmitem dados em tempo real, fornecendo informações sobre o ambiente, a saúde das pessoas e muito mais.

O crescente volume de dados gerados a partir dessas diversas fontes tem impulsionado uma popularização da inteligência artificial (IA), sobretudo com o uso de machine learning (ML), que, de forma sucinta, trata da criação de algoritmos que respondam e se adaptem automaticamente aos dados sem a necessidade de intervenção humana de forma contínua (FILHO, 2015). Aliado a ferramentas poderosas para o processamento massivo e paralelo de dados, esses avanços têm permitido tanto à indústria quanto à comunidade científica desenvolver modelos altamente eficazes e realizar análises mais sofisticadas. A adoção generalizada de técnicas de aprendizado de máquina tem se mostrado fundamental para desvendar o potencial dos dados disponíveis, resultando em descobertas inovadoras e aplicações transformadoras em diversas áreas, incluindo o setor de saúde.

O número de estudos médicos de IA cresceu de forma exponencial no período de 2005 a 2019 (MESKO; GOROG, 2020), revelando o forte interesse da comunidade científica em buscar métodos cada vez mais eficazes para aprimorar o cuidado e a qualidade de vida dos pacientes. A aplicação de modelos preditivos de aprendizado de máquina possui um potencial significativo para auxiliar na tomada de decisões em diversas etapas do cuidado à saúde, principalmente no diagnóstico, intervenção e acompanhamento de problemas de saúde (OBERMEYER; LEE, 2017). O aumento das pesquisas nessa área acompanha o crescente interesse em métodos que possam facilitar diagnósticos e otimizar o tempo dos profissionais de saúde.

Na área da saúde, as decisões tomadas pelos sistemas de IA podem ter um impacto significativo na vida das pessoas. Se os usuários não conseguem compreender o processo de tomada de decisão, podem não confiar no sistema, chegando até a rejeitá-lo. Portanto, a capacidade de fornecer uma explicação sobre como e por que uma decisão específica foi feita tornou-se uma qualidade essencial em sistemas desse tipo (CONFALONIERI

et al., 2021). Além disso, a explicabilidade também pode auxiliar os desenvolvedores na identificação e correção de erros ou viés nos sistemas de IA, tornando-os mais justos e precisos.

Diante desse contexto, este trabalho se propõe a avaliar diversos modelos de classificação com o objetivo de prever a reinternação hospitalar de pacientes, utilizando um conjunto de dados de cuidados de saúde mental coletados pelo sistema de informação da Coordenação de Internações em Ribeirão Preto, Brasil, de julho de 2012 a dezembro de 2017. A relevância deste tipo de trabalho se dá, pois, por meio dessas previsões, medidas preventivas podem ser adotadas antecipadamente, evitando a necessidade de reinternação e, conseqüentemente, reduzindo os custos hospitalares, visto que o custo médio das internações é significativamente maior que o custo médio dos atendimentos ambulatoriais (CESCONETTO; LAPA; CALVO, 2008). Além da predição, o estudo também utilizará técnicas de inteligência artificial explicável (do inglês: *explainable artificial intelligence* - XAI) a fim de identificar quais variáveis estão mais associadas a casos de reinternação, fornecendo informações relevantes para o desenvolvimento de políticas e medidas preventivas mais eficazes.

2 REVISÃO BIBLIOGRÁFICA

2.1 Trabalhos de classificação na área médica

A utilização de algoritmos de aprendizado de máquina para análises preditivas é uma tarefa que envolve uma série de etapas que vão desde a preparação dos dados, até o lançamento e monitoramento da solução (GÉRON, 2022). O trabalho de Santos *et al.* (2019) se propõe a exemplificar as etapas envolvidas na utilização desses algoritmos para análises preditivas em saúde, com um exemplo de aplicação para prever óbito em idosos de São Paulo. A metodologia empregada envolveu a utilização de diferentes algoritmos de aprendizado de máquina para treinar modelos preditivos a partir de dados do estudo SABE, e a avaliação desses modelos em termos de acurácia, sensibilidade e especificidade. Os resultados mostraram que os modelos tiveram desempenho razoável na predição de óbito em idosos, com destaque para o algoritmo Random Forest. No entanto, os autores ressaltam que a amostra utilizada foi relativamente pequena e que a performance dos modelos pode ser melhorada com mais dados e refinamento dos algoritmos.

Fernandes *et al.* (2021) apresenta uma abordagem multipropósito de aprendizado de máquina para prever o prognóstico negativo da COVID-19 em São Paulo, Brasil. O trabalho envolveu o treinamento de cinco algoritmos de aprendizado de máquina com dados estruturados de pacientes com COVID-19, utilizando técnicas de validação cruzada e otimização bayesiana. O objetivo era prever o risco de desenvolver condições críticas em pacientes com COVID-19. Os resultados mostraram que a abordagem proposta teve um desempenho satisfatório na previsão do prognóstico negativo da COVID-19, o que pode ajudar na tomada de decisões clínicas e na alocação efetiva de recursos de saúde.

O trabalho de Asif *et al.* (2018) propõe uma metodologia baseada em aprendizado de máquina para identificar genes relacionados a doenças complexas. O modelo de aprendizado de máquina proposto utiliza uma matriz de similaridade funcional entre genes para prever doenças genéticas complexas. Essa matriz é construída a partir de medidas de similaridade entre genes, como perfis de expressão gênica, redes de interação proteína-proteína e Gene Ontology. Em seguida, um classificador de aprendizado de máquina é treinado com essa matriz para identificar genes relevantes para a doença em questão. O modelo proposto foi avaliado em relação à sua capacidade de identificar genes relacionados ao Transtorno do Espectro Autista e mostrou melhorias significativas em relação a outros métodos de identificação de genes. Foi desenvolvido um fluxo de trabalho automatizado para tornar a metodologia acessível a pesquisadores sem conhecimento extenso em programação e aprendizado de máquina.

Rosa *et al.* (2023) apresenta um estudo que utiliza técnicas de aprendizado de

máquina para classificar fatores que influenciam a ocorrência de dermatites ocupacionais. A metodologia empregada envolveu a avaliação de diferentes algoritmos de aprendizado de máquina, a comparação dos resultados obtidos por cada técnica e a identificação dos fatores mais influentes na ocorrência da lesão ocupacional. Os resultados indicaram que todas as técnicas apresentaram acuracidade entre 55% e 69,4%, sensibilidade entre 49,1% e 80,7% e especificidade entre 50% e 66,7%. Os fatores mais influentes identificados foram a exposição a produtos químicos, o uso de equipamentos de proteção individual inadequados e a falta de treinamento adequado.

2.2 Técnicas de explicação aplicadas a trabalhos de classificação na área médica

A explicabilidade já foi identificada como um fator chave para a adoção de sistemas de IA numa vasta gama de contextos (DOSHI-VELEZ; KIM, 2017; LIPTON, 2017; RIBEIRO; SINGH; GUESTRIN, 2016). No estudo conduzido por Abououf *et al.* (2023) é apresentado um sistema de monitoramento de saúde que utiliza inteligência artificial para detectar e classificar eventos e anomalias médicas. A metodologia empregada envolve a utilização de um modelo de detecção de anomalias e eventos, um modelo de classificação e uma técnica de explicabilidade chamada *KernelSHAP*. O sistema foi avaliado em um conjunto de dados de pacientes e obteve resultados promissores na detecção e classificação de eventos e anomalias médicas. Além disso, a técnica de explicabilidade permitiu que os médicos entendessem melhor as decisões tomadas pelo sistema de inteligência artificial.

Com o uso de técnicas de aprendizado profundo e transferência de aprendizado para detectar a COVID-19 a partir de radiografias de tórax, o estudo de Brunese *et al.* (2020) utiliza as camadas de ativação da rede, ou seja, as áreas da radiografia de tórax que o modelo considerou para gerar a predição, para fornecer explicabilidade sobre a predição. Ainda segundo os autores, isso pode representar uma sugestão para o radiologista localizar imediatamente as áreas da radiografia de tórax que podem ser de interesse. Os resultados mostraram que o modelo proposto alcançou uma acurácia de 98,08% na detecção de COVID-19 em radiografias de tórax. O estudo também sugere que essa tecnologia pode ser usada como uma ferramenta de triagem para ajudar no diagnóstico da COVID-19 em configurações clínicas do mundo real.

Em (KIM; KIM, 2022) foi desenvolvida uma técnica de previsão de mortalidade relacionada ao calor em uma unidade espacial detalhada dentro de uma cidade, utilizando um modelo baseado em floresta aleatória e a técnica de explicação SHAP. A metodologia empregada incluiu a coleta de dados meteorológicos, demográficos e socioeconômicos, pré-processamento de dados, divisão em conjuntos de treinamento e teste, construção do modelo, otimização de hiperparâmetros, avaliação de desempenho e interpretação dos resultados. Os resultados mostraram que o modelo proposto é capaz de prever com precisão a mortalidade relacionada ao calor em uma área específica da cidade de Daegu, na

Coreia do Sul. A técnica SHAP permitiu uma interpretação global e local dos resultados, identificando as variáveis mais importantes para a previsão.

Também utilizando modelos SHAP, o trabalho de GhoshRoy, Alvi and Santosh (2023) utiliza técnicas para prever a fertilidade masculina, com o intuito de melhorar a transparência, responsabilidade e explicabilidade de sete modelos de IA padrão da indústria. A metodologia empregada consistiu em coletar dados de pacientes com problemas de fertilidade masculina e aplicar os modelos de IA para prever a fertilidade. Em seguida, foram utilizadas técnicas de amostragem e validação cruzada para avaliar a robustez e estabilidade de cada modelo. Por fim, a técnica de XAI foi utilizada para explicar o desempenho de cada modelo e os resultados mostraram que o SHAP foi capaz de interpretar as previsões dos sistemas de IA e fornecer explicações claras e compreensíveis para os usuários.

3 DESENVOLVIMENTO

Neste capítulo, apresentaremos uma exposição ordenada e detalhada do desenvolvimento do presente trabalho. O início deste capítulo se dá com uma explanação da metodologia empregada, delineando os passos e abordagens utilizados para alcançar os objetivos propostos.

3.1 Metodologia

3.1.1 Definição de domínio do problema

O problema abordado neste estudo refere-se à previsão de reinternação hospitalar de pacientes que receberam cuidados de saúde mental no sistema de informação da Coordenação de Internações em Ribeirão Preto, Brasil, no período de julho de 2012 a dezembro de 2017. Este domínio envolve a aplicação de modelos de classificação de aprendizado de máquina em um conjunto de dados específico, com o propósito de antecipar eventos de reinternação hospitalar. Além disso, o estudo incorpora técnicas de inteligência artificial explicável, notadamente o SHapley Additive exPlanations (SHAP), para identificar e quantificar as variáveis que apresentam maior influência nos casos de reinternação, contribuindo assim para uma compreensão mais profunda dos fatores associados a esses eventos. A análise deste domínio visa aprimorar a capacidade de tomada de decisões e políticas de saúde, com ênfase na prevenção eficaz e na otimização dos recursos hospitalares.

3.1.2 Método proposto

O método proposto para abordar o problema de previsão de reinternação hospitalar de pacientes em cuidados de saúde mental é composto por três etapas.

3.1.2.1 Limpeza, análise exploratória e preparação dos dados

Inicialmente, a limpeza dos dados será realizada para garantir a consistência e integridade do conjunto de dados. Em seguida, uma análise exploratória detalhada dos dados será conduzida, com a possibilidade de criação de novas variáveis que possam proporcionar insights valiosos. Um exemplo é a variável "tempo de internação", definida como a diferença entre a data de entrada e a data de saída, que poderá ser criada para uma compreensão mais precisa do tempo de permanência dos pacientes.

3.1.2.2 Treinamento dos classificadores

Posteriormente, a fase de treinamento dos classificadores será iniciada, explorando diferentes algoritmos de aprendizado de máquina, ajustando seus hiperparâmetros e empregando a validação cruzada para garantir que os modelos sejam robustos e generalizáveis.

3.1.2.3 Análise de variáveis mais importantes

Por fim, será utilizada a biblioteca **shap** para avaliar o impacto das variáveis no processo de tomada de decisões dos modelos. Isso permitirá a identificação de quais características estão mais fortemente associadas a casos de reinternação hospitalar, fornecendo uma visão detalhada das relações entre variáveis e resultados.

3.1.3 Apresentação dos resultados

Os resultados obtidos após o término dos processos mencionados na seção 3.1.2 serão apresentados de maneira clara e concisa, utilizando tabelas e gráficos informativos. As tabelas destacarão métricas de desempenho, como precisão, recall, F1-score e área sob a curva ROC, para cada modelo testado. Além disso, serão fornecidos gráficos que ilustrarão a importância relativa das variáveis no processo de previsão, com base nas análises SHAP.

[Incluir imagem com o fluxograma do processo descrito na metodologia]

REFERÊNCIAS

- ABOUOUF, M. *et al.* Explainable ai for event and anomaly detection and classification in healthcare monitoring systems. **IEEE Internet of Things Journal**, p. 1–1, 2023.
- ASIF, M. *et al.* Identifying disease genes using machine learning and gene functional similarities, assessed through gene ontology. **PLOS ONE**, Public Library of Science, v. 13, n. 12, p. 1–15, 12 2018. Available at: <https://doi.org/10.1371/journal.pone.0208626>.
- BRUNESE, L. *et al.* Explainable deep learning for pulmonary disease and coronavirus covid-19 detection from x-rays. **Computer Methods and Programs in Biomedicine**, v. 196, p. 105608, 2020. ISSN 0169-2607. Available at: <https://www.sciencedirect.com/science/article/pii/S0169260720314413>.
- CESCONETTO, A.; LAPA, J. dos S.; CALVO, M. C. M. Avaliação da eficiência produtiva de hospitais do sus de santa catarina, brasil. **Revista Saúde Pública**, 2008.
- CONFALONIERI, R. *et al.* A historical perspective of explainable artificial intelligence. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, John Wiley Sons Ltd., v. 11, n. 1, p. 21, 2021. ISSN 1942-4787.
- DOSHI-VELEZ, F.; KIM, B. **Towards A Rigorous Science of Interpretable Machine Learning**. 2017.
- FERNANDES, F. T. *et al.* A multipurpose machine learning approach to predict covid-19 negative prognosis in são paulo, brazil. **Scientific reports**, Nature Publishing Group, v. 11, n. 1, p. 1–11, 2021.
- FILHO, A. D. P. C. Uso de big data em saúde no brasil: perspectivas para um futuro próximo. **Epidemiologia e Serviços de Saúde**, SciELO Public Health, v. 24, p. 325–332, 2015.
- GÉRON, A. **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd Edition**. [*S.l.: s.n.*]: O'Reilly Media, Inc., 2022. ISBN 9781098125974.
- GHOSHROY, D.; ALVI, P. A.; SANTOSH, K. Unboxing industry-standard ai models for male fertility prediction with shap. **Healthcare**, v. 11, n. 7, 2023. ISSN 2227-9032. Available at: <https://www.mdpi.com/2227-9032/11/7/929>.
- KIM, Y.; KIM, Y. Explainable heat-related mortality with random forest and shapley additive explanations (shap) models. **Sustainable Cities and Society**, v. 79, p. 103677, 2022. ISSN 2210-6707. Available at: <https://www.sciencedirect.com/science/article/pii/S2210670722000117>.
- LIPTON, Z. C. **The Mythos of Model Interpretability**. 2017.
- MESKO, B.; GOROG, M. A short guide for medical professionals in the era of artificial intelligence. **npj Digital Medicine**, Nature Publishing Group, v. 3, p. 126, 2020.
- OBERMEYER, Z.; LEE, T. H. Lost in thought—the limits of the human mind and the future of medicine. **New England Journal of Medicine**, Mass Medical Soc, v. 377, n. 13, p. 1209–1211, 2017.

RIBEIRO, M. T.; SINGH, S.; GUESTIN, C. **Model-Agnostic Interpretability of Machine Learning**. 2016.

ROSA, A. C. F. d. *et al.* Uso de técnicas de aprendizado de máquina para classificação de fatores que influenciam a ocorrência de dermatites ocupacionais. **Revista Brasileira de Saúde Ocupacional**, Fundação Jorge Duprat Figueiredo de Segurança e Medicina do Trabalho - FUNDACENTRO, v. 48, 2023. ISSN 0303-7657. Available at: <https://doi.org/10.1590/2317-6369/31620pt2023v48e4>.

SANTOS, H. G. d. *et al.* Machine learning para análises preditivas em saúde: exemplo de aplicação para prever óbito em idosos de são paulo, brasil. **Cadernos de saúde pública**, Escola Nacional de Saúde Pública Sergio Arouca, Fundação Oswaldo Cruz, v. 35, n. 7, 2019. ISSN 0102-311X.