



APLICAÇÃO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA PARA A PREDIÇÃO DE DIABETES TIPO 2 COMO APOIO À SAÚDE

Daniel Martins Ciriaco Nicoli¹, Moisés Barbosa Junior²

¹Academico do Curso de Engenharia de Software, Campus Ponta Grossa-PR, Universidade Cesumar - UNICESUMAR. Bolsista PIBIC/ICETI-UniCesumar. nicolidaniel02@gmail.com

²Orientador, Doutor, Docente no Curso de Engenharia de Software, UNICESUMAR. Pesquisador do Instituto Cesumar de Ciência, Tecnologia e Inovação – ICETI. moises.junior@unicesumar.edu.br

RESUMO

A inteligência artificial (IA) tem se destacado na área da saúde por sua capacidade de extrair padrões complexos de grandes volumes de dados clínicos, apoiando decisões e promovendo a medicina preventiva. Este trabalho objetiva desenvolver e validar um modelo preditivo de risco de diabetes tipo 2, integrando dados secundários públicos do PIMA Indian Diabetes Dataset com dados primários, provenientes de ações de triagem realizadas pela instituição. Serão comparados algoritmos de Regressão Logística e Árvore de Decisão, utilizando validação cruzada e métricas de desempenho como acurácia, precisão e recall. Espera-se como resultado um protótipo funcional capaz de auxiliar em triagens de risco, validado na comunidade acadêmica e passível de replicação em outros contextos.

PALAVRAS-CHAVE: Aprendizado de máquina; Inteligência artificial; Saúde preditiva.

1 INTRODUÇÃO

A aplicação da Inteligência Artificial (IA) na área da saúde tem crescido de forma significativa nos últimos anos, impulsionada pela capacidade dessas tecnologias de processar grandes volumes de dados clínicos de forma rápida e precisa. Sistemas baseados em IA são capazes de identificar padrões complexos, auxiliar na tomada de decisão médica e oferecer suporte ao diagnóstico precoce, o que pode reduzir custos e salvar vidas. Segundo Jiang et al. (2017), soluções de IA têm potencial para transformar o cuidado em saúde, tornando-o mais preditivo, preventivo, personalizado e participativo.

Dentre as condições crônicas mais prevalentes, a diabetes tipo 2 é uma das que mais impacta a saúde pública. Dados da Organização Mundial da Saúde (WHO, 2016) indicam que mais de 460 milhões de pessoas convivem com a doença, e uma parcela significativa desconhece o próprio diagnóstico. Essa situação gera aumento de complicações, internações hospitalares e perda de qualidade de vida. A identificação precoce de indivíduos em risco é, portanto, uma estratégia fundamental para evitar complicações graves, como doenças cardiovasculares e insuficiência renal, e para reduzir a sobrecarga dos sistemas de saúde.

Diversas técnicas de aprendizado de máquina já vêm sendo empregadas para previsão de doenças, demonstrando eficácia na detecção de câncer, doenças cardiovasculares e retinopatia diabética, com resultados comparáveis aos de especialistas humanos (Topol, 2019). Entretanto, grande parte dessas pesquisas utiliza exclusivamente bases de dados internacionais, que podem não refletir as características regionais da população alvo, como hábitos alimentares e perfil socioeconômico.

Este projeto busca preencher essa lacuna ao integrar dados secundários de uso global (PIMA Indian Diabetes Dataset) com dados clínicos primários coletados em triagens de saúde realizadas na própria comunidade acadêmica. Essa integração permitirá avaliar o desempenho de algoritmos em um contexto local e gerar resultados mais aderentes à realidade dos indivíduos atendidos.

Além de seu impacto prático, o projeto possui caráter interdisciplinar, envolvendo Engenharia de Software e Ciências da Saúde, e está alinhado a importantes diretrizes



globais, como o Objetivo de Desenvolvimento Sustentável 3 (Saúde e Bem-Estar), ao promover estratégias tecnológicas de prevenção e triagem precoce de doenças crônicas; ao ODS 4 (Educação de Qualidade), por proporcionar aprendizado aplicado em ciência de dados; e ao ODS 17 (Parcerias), ao fomentar a cooperação entre diferentes áreas do conhecimento.

2 MATERIAIS E MÉTODOS

A pesquisa será conduzida em cinco etapas: (1) Coleta e pré-processamento de dados secundários do PIMA Indian Diabetes Dataset; (2) Integração com dados primários anonimizados coletados em triagens; (3) Padronização de variáveis e unificação das bases; (4) Treinamento de modelos de Regressão Logística e Árvore de Decisão com validação cruzada; (5) Avaliação do desempenho com métricas como acurácia, precisão, recall e matriz de confusão.

Para potencializar o aprendizado e a execução do projeto, foram consideradas as dimensões de estilos de aprendizagem propostas por Silva (2012): Percepção – Concretol, privilegiando fatos e dados concretos; Retenção – Visual, utilizando gráficos, matrizes de confusão e curvas ROC para interpretação; Processamento – Colaborativo, com reuniões de grupo e discussão dos resultados para consolidar o aprendizado.

3 RESULTADOS E DISCUSSÕES

Espera-se obter um banco de dados consolidado, modelos preditivos treinados e otimizados com desempenho mínimo de 70% de acurácia, além de insights sobre as variáveis de maior relevância para o risco de diabetes tipo 2. Os resultados serão documentados em gráficos de desempenho e relatórios técnicos. Esta análise permitirá avaliar a viabilidade do uso do modelo em contextos reais e sugerir melhorias para futuras aplicações.

4 CONSIDERAÇÕES FINAIS

Conclui-se que o desenvolvimento de um modelo preditivo de risco de diabetes tipo 2 é uma ferramenta promissora para apoiar a medicina preventiva. O projeto contribui para o ODS 3 ao propor soluções tecnológicas para triagem precoce de doenças crônicas, e para o ODS 4 e ODS 17 ao fomentar formação prática em ciência de dados e colaboração entre cursos. Os próximos passos incluem a disseminação dos resultados em eventos acadêmicos e a disponibilização do código e base de dados de forma anonimizada para replicação por outros pesquisadores.

REFERÊNCIAS

A DINH, A. A.; MIERTSCHIN, S.; YOUNG, A. W.; MOHANTY, S. K. A review on the applications of machine learning in healthcare. *Journal of Medical Systems*, v. 43, n. 9, art. 305, 2019.

Jiang, F.; Jiang, Y.; Zhi, H.; Dong, Y.; Li, H.; Ma, S.; Wang, Y.; Dong, Q.; Shen, H.; Wang, Y. Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology*, London, v. 2, n. 4, p. 230–243, 2017. Disponível em: <https://doi.org/10.1136/svn-2017-000101>.



Khokhar, F.; Yang, G.; Pereira, F.; Jones, D.; Smith, T. Explainable AI for Diabetes Prediction. *arXiv preprint arXiv:2501.18071*, 2025. Disponível em: <https://arxiv.org/abs/2501.18071>.

Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, Madison, v. 12, p. 2825–2830, 2011. Disponível em: <http://www.jmlr.org/papers/v12/pedregosa11a.html>.

Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, New York, v. 25, p. 44–56, 2019. Disponível em: <https://doi.org/10.1038/s41591-018-0300-7>

World Health Organization. *Global report on diabetes*. Geneva: WHO, 2016. Disponível em: <https://www.who.int/publications/i/item/9789241565257>.

Silva, J.; Oliveira, M.; Santos, R. Aplicações de IA na Saúde. *Revista Brasileira de Informática em Saúde*, São Paulo, v. 5, n. 2, p. 15–28, 2022.

Kaggle. Pima Indians Diabetes Database. *Kaggle Datasets*, 2016. Disponível em: <https://www.kaggle.com/uciml/pima-indians-diabetes-database> .