



Anais do Simpósio Acadêmico de Engenharia de Produção (SAEPRO) da EEL-USP

IX SAEPRO – 25 e 26 de novembro de 2025

Árvore de Decisão como Método de Identificação de Pessoas com Diabetes

Cristiane da Costa Santana

Universidade Federal do Vale do São Francisco
(cristiane.santana@discente.univasf.edu.br)

Thiago Magalhães Amaral

Universidade Federal do Vale do São Francisco
(thiago.magalhaes@univasf.edu.br)

CONTEXTUALIZAÇÃO: O diabetes é uma das doenças crônicas mais frequentes no mundo, que afeta milhões de pessoas e exige acompanhamento contínuo para a prevenção de complicações (Torres *et. al*, 2009). De acordo com o Atlas do Diabetes (2025), estima-se que até 2050 o número de pessoas com a doença aumente em 45%. No Brasil, segundo a Sociedade Brasileira de Diabetes (2025), em 2022, cerca de 2 mil pessoas morreram em decorrência da falta de diagnóstico. Ainda conforme a instituição, um estudo longitudinal de saúde do adulto revelou que 50% dos indivíduos com diabetes desconheciam da sua condição. Diante desse cenário, a tecnologia surge como uma importante aliada da medicina. Ferramentas de análise de dados, como redes neurais e árvores de decisão, permitem identificar padrões em grandes bases de informações clínicas, contribuindo para diagnósticos mais rápidos e precisos, sendo usada como uma ferramenta de apoio à pesquisa operacional. Esse avanço possibilita não apenas a detecção precoce da doença, mas também o apoio na tomada de decisões médicas, ajudando a salvar vidas e a melhorar a qualidade de vida dos pacientes (Moraes *et al.*, 2023). Entre os diversos algoritmos de aprendizado de máquina aplicados à saúde, a árvore de decisão é um dos mais estudados e utilizados (Zeng, Cheng, 2021). Isso se deve ao fato de que ela apresenta resultados de forma clara e de fácil interpretação, além de possuir um processo de treinamento



Anais do Simpósio Acadêmico de Engenharia de Produção (SAEPRO) da EEL-USP

IX SAEPRO – 25 e 26 de novembro de 2025

relativamente rápido quando comparado a outros modelos (Carvalho, 2014). Segundo Santos (2024), “a aplicação de machine learning para o diagnóstico da diabetes tem o potencial de revolucionar a forma como a doença é detectada e tratada.”

OBJETIVO: Desenvolver um modelo baseado em árvore de decisão para a identificação de pacientes com diabetes, a partir de indicadores de saúde.

MÉTODO: Este estudo é de natureza quantitativa, de cunho descritivo que se utilizou de uma base de dados chamada de “Kaggle”. O objetivo do conjunto de dados é prever diagnosticamente se um paciente tem diabetes, através de variáveis como: Índice de Massa Corporal, Pressão Arterial Sistólica, Pressão Arterial Diastólica, Glicemia em Jejum, Hemoglobina Glicada, Creatinina Sérica, Níveis de Ureia no Sangue, Colesterol Total, Colesterol LDL, Colesterol HDL, Triglicerídeos. Foi utilizado o software Orange para aprendizado de máquina do algoritmo de árvore de decisão. Inicialmente foi feito o carregamento da base de dados no software Orange e, após isso, realizado o pré-processamento para padronizar a escala dos dados. Para a normalização de dados, foi utilizada a normalização do intervalo, que redimensiona os dados em um intervalo entre 0 e 1. Foi feito a divisão dos mesmos em um conjunto de treinamento e um conjunto de teste, utilizando a proporção de 70/30%. Para avaliar o desempenho do modelo foram aplicadas diferentes métricas. Para este estudo, foram utilizadas a acurácia, que corresponde à proporção de classificações corretas em relação ao número total de casos analisados; a precisão, que indica entre as instâncias previstas como positivas, quantas de fato pertencem à classe positiva; O F1-Score, que combina precisão e recall em uma média harmônica, equilibrando as duas medidas; e a Matriz de Confusão, que apresenta de forma detalhada os acertos e erros do classificador, distinguindo entre verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos.

RESULTADOS E DISCUSSÕES: A árvore de decisão alcançou escores elevados em todas as métricas consideradas. Na acurácia, o resultado foi de 0.84, o F1-Score de 0.84 e a precisão de 0.84, o que sugere uma confiabilidade deste algoritmo na classificação correta dos pacientes. Além disso, a matriz de confusão foi analisada para avaliar a capacidade do algoritmo identificar corretamente os casos de diabetes. O modelo analisou



Anais do Simpósio Acadêmico de Engenharia de Produção (SAEPRO) da EEL-USP

IX SAEPRO – 25 e 26 de novembro de 2025

um total de 2820 pacientes, identificou corretamente 1471 casos de pacientes não diabéticos e o diagnóstico de 908 indivíduos diabéticos. Porém, 219 pacientes que não têm diabetes foram diagnosticados erroneamente como diabéticos e 222 pacientes que têm diabetes foram classificados incorretamente como saudáveis. Resultados semelhantes foram observados no estudo de Santos (2024), cujo modelo, utilizando outras variáveis para a identificação de diabetes, apresentou uma acurácia de aproximadamente 80%. Essa comparação reforça a consistência do desempenho obtido neste estudo e evidencia o potencial das árvores de decisão como ferramenta eficaz para apoio ao diagnóstico da doença.

CONSIDERAÇÕES FINAIS: A análise de dados por meio da árvore de decisão demonstrou ser uma ferramenta promissora para a identificação automatizada de pacientes diabéticos, atingindo uma acurácia geral de aproximadamente 84%. No geral, o desempenho é bom, mas o principal desafio reside na alta ocorrência de falsos negativos. Este é o erro mais crítico, pois implica que um número considerável de pacientes diabéticos pode não ser identificado e, consequentemente, não receber tratamento precoce. Sugere-se que em estudos futuros sejam exploradas técnicas como o ajuste dos parâmetros ou a utilização de outros modelos preditivos, visando a redução desses falsos negativos.

PALAVRAS-CHAVE: diabetes, machine learning, árvore de decisão.

REFERÊNCIAS:

CARVALHO, Hialo Muniz. **Aprendizado de Máquina voltado para Mineração de Dados: Árvores de Decisão**. 2014. Monografia (Bacharelado em Engenharia de Software) — Faculdade UnB Gama, Universidade de Brasília, Brasília, 2014. Disponível em: https://bdm.unb.br/bitstream/10483/9487/1/2014_HialoMunizCarvalho.pdf. Acesso em: 27 set. 2025.

INTERNATIONAL DIABETES FEDERATION (IDF). **Atlas do Diabetes**. 2025. Disponível em: <https://diabetes.org.br/wp-content/uploads/2025/09/atlas-idf-2025.pdf>. Acesso em: 27 set. 2025.



Anais do Simpósio Acadêmico de Engenharia de Produção (SAEPRO) da EEL-USP

IX SAEPRO – 25 e 26 de novembro de 2025

MORAES, Joel Junior de et al. IMPACTO DA TECNOLOGIA DE INTELIGÊNCIA ARTIFICIAL NA MEDICINA DIAGNÓSTICA. **Revista Ibero-Americana de Humanidades, Ciências e Educação**, [S. l.], v. 9, n. 7, p. 1303–1314, 2023. DOI: 10.51891/rease.v9i7.10699. Disponível em:

<https://periodicorease.pro.br/rease/article/view/10699>. Acesso em: 27 set. 2025.

PITITTO, Bianca de Almeida; BAHIA, Luciana; MELO, Karla; et al. Dados epidemiológicos do diabetes mellitus no Brasil. **Departamento de Saúde Pública da Sociedade Brasileira de Diabetes (SBD)**, 2024. Disponível em: https://diabetes.org.br/wp-content/uploads/2025/02/Dados-Epidemiologicos-SBD_comT1Dindex_2024_pdf.pdf. Acesso em: 27 set. 2025.

SANTOS, Vinicius de Souza. Comparison and selection of machine learning algorithms for diabetes prediction: An exploratory quantitative study based on medical data analysis. **Seven Editora**, [S. l.], p. 737–765, 2024. Disponível em: <https://sevenpubl.com.br/editora/article/view/4102>. Acesso em: 6 oct. 2025.

TORRES, Heloisa de Carvalho; FRANCO, Laercio Joel; STRADIOTO, Mayra Alves; HORTALE, Virginia Alonso; SCHALL, Virginia Torres. Avaliação estratégica de educação em grupo e individual no programa educativo em diabetes. **Revista de Saúde Pública**, São Paulo, v. 43, n. 2, abr. 2009. DOI:10.1590/S0034-89102009005000001. Disponível em:

<https://www.scielo.br/j/rsp/a/HHmrMcqQct6BY6y4yPt4VPg/abstract/?lang=pt>

. Acesso em: 6 out. 2025.

ZENG, Yu; CHENG, Fuchao. Medical and Health Data Classification Method Based on Machine Learning. **Journal of Healthcare Engineering**, v. 2021, p. 1–5, 2021. DOI: <https://doi.org/10.1155/2021/2722854>