

Modelagem Preditiva De Dados Aplicada À Saúde Nutricional: Uma Proposta De Formação Em Ciência De Dados E Inteligência Artificial No Ensino Médio

Larissa Suellen Gomes Andrade de Lima Aquino^{1,4}, Marcílio Ferreira dos Santos^{1,2}, Cleiton de Lima Ricardo^{1,3}, Felipe Fernandes Miranda Vieira^{1,2}

E-mail: marcilio.santos@ufpe.br

¹ Universidade Federal de Pernambuco (UFPE), Centro Acadêmico do Agreste (CAA),

² Núcleo de Formação de Docentes (NFD), Caruaru, Pernambuco, Brasil.

³ Núcleo Interdisciplinar de Ciências Exatas e da Natureza (NICEN), Caruaru, Pernambuco, Brasil.

⁴ Secretaria de Educação do Estado de Pernambuco (SEE-PE), Recife, Pernambuco, Brasil.

Resumo: A Ciência de Dados e a Inteligência Artificial têm ampliado as possibilidades de investigação e tomada de decisão em diferentes áreas do conhecimento, especialmente na saúde. Paralelamente, a Educação Básica demanda estratégias que promovam o desenvolvimento de competências relacionadas ao pensamento computacional, ao letramento estatístico e à resolução de problemas com dados. Nesse contexto, este trabalho apresenta uma proposta de formação em Ciência de Dados aplicada à saúde nutricional, desenvolvida em parceria entre uma escola pública de ensino médio e um consultório de nutrição. A proposta contempla uma etapa formativa para 40 estudantes, envolvendo programação em Python, estatística, análise exploratória de dados e técnicas de *machine learning*, seguida por uma etapa de pesquisa com 10 bolsistas. Os estudantes participarão da construção e análise de uma base de aproximadamente 150 registros anonimizados de pacientes, buscando identificar fatores clínicos, comportamentais e socioeconômicos associados ao sucesso de programas de emagrecimento. Serão empregados modelos estatísticos clássicos, como regressão linear e regressão logística, bem como algoritmos de aprendizado de máquina, incluindo *Random Forest*, permitindo comparar diferentes abordagens analíticas na identificação das variáveis mais relevantes. Espera-se que a proposta favoreça o desenvolvimento de competências em programação, estatística e análise crítica de dados, além de aproximar os estudantes de problemas reais de pesquisa e inovação. Como produtos, prevê-se a construção de uma base de dados estruturada, o desenvolvimento de modelos preditivos, dashboards interativos e materiais de divulgação científica, fortalecendo a integração entre educação, tecnologia e saúde.

Palavras-chave: Ciência de Dados; Inteligência Artificial; Ensino Médio; Machine Learning; Letramento Estatístico.

INTRODUÇÃO

A Ciência de Dados e a Inteligência Artificial vêm transformando significativamente a forma como problemas complexos são investigados em diferentes áreas do conhecimento, especialmente na saúde. A crescente disponibilidade de dados, aliada ao desenvolvimento de métodos estatísticos e algoritmos de aprendizado de máquina, tem ampliado a capacidade de identificar padrões, construir modelos preditivos e apoiar a tomada de decisões baseada em evidências (HAN; KAMBER; PEI, 2022; HASTIE; TIBSHIRANI; FRIEDMAN, 2009). Nesse contexto, torna-se cada vez mais importante promover a formação de estudantes capazes de

compreender, interpretar e utilizar dados na resolução de problemas reais.

No âmbito da Educação Básica, a Base Nacional Comum Curricular (BNCC) estabelece o desenvolvimento do pensamento computacional, da cultura digital e do letramento estatístico como competências essenciais para a formação cidadã (BRASIL, 2018). Em complemento, às Normas sobre Computação na Educação Básica reforçam a necessidade da inserção de conceitos de programação, análise de dados e inteligência artificial nos processos de ensino e aprendizagem, preparando os estudantes para uma sociedade cada vez mais orientada por dados (BRASIL, 2022). Essa perspectiva dialoga com o conceito de pensamento

computacional proposto por Wing (2006), que o compreende como uma habilidade fundamental para resolver problemas, projetar sistemas e compreender fenômenos utilizando princípios da Computação.

Sob o ponto de vista pedagógico, a Aprendizagem Baseada em Projetos (*Project-Based Learning* – PBL) constitui uma estratégia capaz de aproximar os estudantes de problemas autênticos, favorecendo a construção do conhecimento por meio da investigação, da colaboração e da aplicação prática dos conteúdos (KRAJCIK; BLUMENFELD, 2006). Essa concepção também encontra respaldo nas ideias de Papert (2020), ao defender que a aprendizagem ocorre de forma mais significativa quando os estudantes constroem artefatos e desenvolvem soluções para desafios concretos, utilizando recursos computacionais como instrumentos de criação e investigação.

Nessa perspectiva, o desenvolvimento de projetos envolvendo análise de dados também contribui para a promoção do letramento estatístico, compreendido como a capacidade de interpretar, avaliar criticamente e comunicar informações baseadas em dados, competências consideradas essenciais para a participação cidadã em uma sociedade orientada pela informação (GAL, 2002). De forma complementar, Skovsmose (2001) argumenta que a Educação Matemática deve favorecer uma postura crítica diante da realidade, estimulando os estudantes a questionarem modelos, interpretações e usos da matemática em contextos sociais. Assim, atividades fundamentadas em projetos e análise de dados ampliam não apenas o domínio técnico dos estudantes, mas também sua capacidade de refletir criticamente sobre problemas reais e tomar decisões fundamentadas.

Inserido nesse contexto, o projeto "Modelagem preditiva de dados aplicada à saúde nutricional", recomendado no resultado preliminar do Edital nº 17/2026 – COMPET Médio-Tec da Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE), propõe integrar educação, saúde e tecnologias digitais por meio da formação de estudantes do ensino médio em Ciência de Dados aplicada a um problema real da saúde nutricional. A proposta prevê a capacitação dos estudantes em programação com Python, estatística e aprendizado de máquina, utilizando dados anonimizados provenientes de um consultório de nutrição para investigar fatores associados ao sucesso de programas de emagrecimento.

Dessa forma, este trabalho tem como objetivo apresentar a concepção metodológica desse projeto, evidenciando seu potencial para promover competências em Ciência de Dados (KELLEHER, 2018; JIAWEI, 2006), Estatística e Inteligência Artificial na Educação Básica, ao mesmo tempo em

que aproxima os estudantes de práticas de pesquisa científica voltadas à resolução de problemas reais.

MATERIAL E MÉTODOS

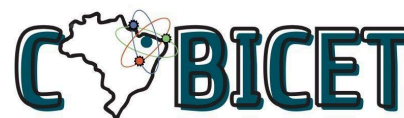
O presente trabalho caracteriza-se como um relato de proposta metodológica de um projeto de pesquisa e extensão tecnológica voltado à formação de estudantes do ensino médio em Ciência de Dados aplicada à saúde nutricional. O projeto será desenvolvido na Escola de Referência em Ensino Médio (EREM) Senador João Cleofas de Oliveira, em parceria com um consultório de nutrição localizado no município de Vitória de Santo Antão-PE.



figura 1: EREM Senador João Cleofas de Oliveira, onde a pesquisadora é lotada e onde será desenvolvido o projeto.

A metodologia está organizada em duas etapas. Na primeira, serão capacitados 40 estudantes do ensino médio em conteúdos de estatística, programação em Python, análise exploratória de dados, visualização de informações e técnicas introdutórias de aprendizado de máquina, utilizando plataformas como Google Colab e Kaggle (GRUS, 2016). Ao término dessa etapa, serão selecionados dez estudantes para participação nas atividades de pesquisa e extensão tecnológica.

Na segunda etapa, será construída uma base de dados composta por aproximadamente 150 registros anonimizados provenientes de pacientes acompanhados pelo consultório parceiro. As informações contemplarão variáveis clínicas, comportamentais e socioeconômicas relacionadas ao tratamento nutricional, observando-se os princípios éticos e de proteção de dados pessoais. Paralelamente, será utilizada uma base pública disponibilizada na plataforma Kaggle para o desenvolvimento inicial dos modelos preditivos. Posteriormente, esses modelos serão aplicados à base



de pacientes do consultório parceiro, permitindo avaliar sua capacidade de generalização para uma população distinta por meio de métricas de desempenho preditivo.

As análises serão realizadas em linguagem Python, empregando técnicas de análise exploratória de dados, regressão linear, regressão logística e o algoritmo Random Forest. Os modelos serão utilizados para identificar os fatores mais associados ao sucesso dos programas de emagrecimento, estimar a probabilidade de perda de peso dos pacientes, comparar o desempenho de abordagens estatísticas clássicas e de aprendizado de máquina e avaliar sua capacidade de generalização quando aplicados à base do consultório parceiro na identificação das variáveis preditoras mais relevantes. Os resultados serão apresentados por meio de indicadores estatísticos, visualizações gráficas e dashboards interativos.

RESULTADOS ESPERADOS

Espera-se que o projeto contribua para o desenvolvimento de competências relacionadas à Ciência de Dados, Estatística, Programação e Inteligência Artificial entre estudantes do ensino médio, promovendo o letramento científico e computacional por meio da investigação de um problema real da área da saúde. A abordagem proposta busca aproximar os estudantes do método científico, favorecendo o desenvolvimento da capacidade de coletar, organizar, analisar e interpretar dados para subsidiar a tomada de decisões baseada em evidências.

Do ponto de vista técnico-científico, espera-se construir uma base de dados estruturada contendo aproximadamente 150 registros anonimizados de pacientes, permitindo a aplicação de técnicas estatísticas e algoritmos de aprendizado de máquina para identificar fatores clínicos, comportamentais e socioeconômicos associados ao sucesso de programas de emagrecimento. Além disso, pretende-se comparar o desempenho de modelos estatísticos clássicos e de aprendizado de máquina, bem como avaliar sua capacidade de generalização quando aplicados à base de pacientes do consultório parceiro, contribuindo para a compreensão das potencialidades dessas abordagens em aplicações na área da saúde e no apoio à tomada de decisão clínica.

Como produtos do projeto, espera-se desenvolver modelos preditivos em Python, elaborar dashboards para visualização dos resultados, produzir materiais didáticos e relatórios técnico-científicos, capazes de subsidiar a nutricionista na identificação de perfis com maior probabilidade de adesão e sucesso em programas de emagrecimento, bem como realizar

ações de divulgação científica junto à comunidade escolar. Espera-se, ainda, que a experiência contribua para estimular o interesse dos estudantes pelas áreas de Ciência, Tecnologia, Engenharia e Matemática (STEM), fortalecendo a integração entre escola, pesquisa e inovação tecnológica.

CONCLUSÃO

A proposta apresentada evidencia o potencial da integração entre Ciência de Dados, Inteligência Artificial e Educação Básica como estratégia para promover aprendizagens contextualizadas e o desenvolvimento de competências alinhadas às demandas da sociedade orientada por dados. Ao utilizar um problema real da saúde nutricional como objeto de investigação, o projeto favorece a articulação entre conhecimentos de Matemática, Estatística, Programação e Ciência de Dados, aproximando os estudantes das práticas de pesquisa científica e da tomada de decisão baseada em evidências.

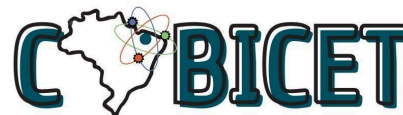
Espera-se que a execução do projeto contribua para o fortalecimento do letramento científico, estatístico e computacional dos participantes, além de estimular o interesse pelas áreas de Ciência, Tecnologia, Engenharia e Matemática (STEM). Do ponto de vista aplicado, espera-se que os modelos desenvolvidos possam auxiliar na identificação de fatores associados ao sucesso de programas de emagrecimento e que a avaliação de sua capacidade de generalização para a base de pacientes do consultório parceiro forneça evidências sobre seu potencial como ferramenta de apoio à tomada de decisão clínica. A iniciativa também poderá fornecer subsídios para a adoção de metodologias semelhantes em outras instituições de ensino, ampliando a inserção da Ciência de Dados e da Inteligência Artificial na Educação Básica e fortalecendo a integração entre escola, pesquisa, inovação e sociedade.

AGRADECIMENTOS

Os autores agradecem à Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE), pelo apoio concedido por meio do Edital nº 17/2026 – COMPET Médio-Tec: Habilidades de Futuro para Ensino Médio e Técnico, bem como à Escola de Referência em Ensino Médio Senador João Cleofas de Oliveira e à MVC Barros Nutrição pelo apoio institucional à execução do projeto.

REFERÊNCIAS

BRASIL. **Ministério da Educação**. *Base Nacional Comum Curricular*. Brasília, DF: MEC, 2018.



Disponível em:
<https://basenacionalcomum.mec.gov.br/>. Acesso em:
16 jul. 2026.

BRASIL, M. Normas sobre computação na educação básica-complemento à bncc. **Brasília-DF: MEC, Secretaria de Educação Básica**, 2022.

GRUS, Joel. **Data science do zero**. Rio de Janeiro: Alta books, 2016.

JIAWEI, Han; MICHELINE, Kamber. **Data mining: concepts and techniques**. Morgan kaufmann, 2006.

HASTIE, Trevor. The elements of statistical learning: data mining, inference, and prediction. 2009.

KELLEHER, John D.; TIERNEY, Brendan. **Data science**. MIT press, 2018.

KRAJCIK, Joseph; BLUMENFELD, Phyllis. 19. project-based learning. **The Cambridge handbook of the learning sciences**, p. 317-333, 2006.

PAPERT, Seymour A. **Mindstorms: Children, computers, and powerful ideas**. Basic books, 2020.

WING, Jeannette M. Computational thinking. **Communications of the ACM**, v. 49, n. 3, p. 33-35, 2006.