Detecção de adulteração no leite por meio de inteligência artificial

Autor(res)

Rafael Fagnani
Josiane Ito Eleodoro
Paula Gabriela De Oliveira

Categoria do Trabalho

Pós-Graduação

Instituição

UNOPAR / ANHANGUERA - ARAPONGAS

Introdução

A qualidade do leite cru é um ponto central para garantir a segurança alimentar, a padronização da produção e a competitividade do setor lácteo (Hora et al., 2024). Entre os principais problemas enfrentados pela indústria, destaca-se a adulteração por adição de água, prática ilegal que compromete o valor nutricional, prejudica o processamento industrial e representa risco à saúde do consumidor final.
No Brasil, a Instrução Normativa nº 76/2018 do Ministério da Agricultura, Pecuária e Abastecimento (Brasil, 2018) estabelece parâmetros físico-químicos obrigatórios, como proteína, gordura, lactose, extrato seco desengordurado e sólidos totais, além de limites de crioscopia, utilizados na identificação de fraudes. No entanto, as análises convencionais para detecção de adição de água apresentam limitações em seu uso na indústria e podem não detectar todas as adulterações.
Nesse contexto, o uso de ferramentas de inteligência artificial tem se mostrado promissor no tratamento de grandes volumes de dados e na identificação de padrões complexos. Entre os algoritmos disponíveis, o XGBoost (Extreme Gradient Boosting) destaca-se por sua robustez, alta capacidade de generalização e eficiência no processamento de variáveis correlacionadas. Dessa forma, sua aplicação na detecção de fraudes representa uma alternativa inovadora, com potencial de aumentar a confiabilidade dos controles laboratoriais e contribuir para a qualidade da cadeia produtiva do leite.

Objetivo

Desenvolver e avaliar um modelo preditivo baseado no algoritmo XGBoost para identificar a adição de água em amostras de leite cru a partir de parâmetros físico-químicos, excluindo-se a crioscopia. O estudo busca verificar a eficácia da técnica como ferramenta auxiliar no monitoramento da qualidade do leite.

Material e Métodos

Foi utilizado um banco de dados com 3337 amostras de leite de vaca contendo valores de gordura, proteína, lactose, sólidos totais, extrato seco desengordurado e crioscopia. O índice crioscópico foi utilizado apenas para classificar as amostras quanto à adulteração por adição de água: valores iguais ou superiores a -0,530°H indicaram adulteração (classe 1), enquanto valores entre -0,531°H e -0,555°H indicaram ausência de adulteração (classe 0). Após limpeza e padronização, os dados foram divididos em 70% para treinamento e 30% para teste,

com divisão estratificada para manter a proporção das classes. Devido à predominância de amostras não adulteradas, foi aplicado o balanceamento dos dados para evitar que o modelo tivesse tendência a prever apenas a classe maioritária. Para isso, utilizou-se a técnica de oversampling, aumentando artificialmente a quantidade de amostras adulteradas no conjunto de treinamento. As análises foram realizadas em Python, utilizando o algoritmo XGBoost, comumente aplicado em casos de predição. O modelo foi treinado para predizer a presença de adulteração com base nos demais parâmetros do leite, excluindo a crioscopia, para avaliar a capacidade de detectar adulteração sem essa variável. O desempenho foi avaliado por meio das métricas de precisão, recall e F1-score.

Resultados e Discussão

O modelo apresentou acurácia geral de 80%, indicando desempenho satisfatório na classificação global das amostras. No entanto, ao analisar separadamente as classes, observou-se uma diferença relevante entre o leite adulterado e o não adulterado. Para a classe adulterada, os resultados foram de 59% de precisão, 64% de recall e 62% de F1-score, sugerindo que o modelo ainda enfrenta dificuldades em identificar todas as ocorrências de fraude de forma consistente. Já para a classe não adulterada, o desempenho foi consideravelmente superior, alcançando 88% de precisão, 86% de recall e 87% de F1-score, o que demonstra maior facilidade do algoritmo em reconhecer amostras normais.

Esses achados evidenciam um desafio comum em conjuntos de dados desbalanceados, nos quais a classe maioritária tende a influenciar o aprendizado do modelo. Apesar da aplicação de técnicas de balanceamento, ainda houve diferença de desempenho entre as classes, reforçando a necessidade de ajustes adicionais. Mesmo diante dessas limitações, os resultados obtidos são relevantes, uma vez que demonstram que o XGBoost foi capaz de identificar padrões nos parâmetros físico-químicos do leite e diferenciar amostras adulteradas de não adulteradas com desempenho global satisfatório. Há um destaque para o recall da classe adulterada que, mesmo moderado, indica a capacidade do modelo em detectar uma proporção significativa das fraudes, demonstrando que o modelo tem potencial para a triagem preliminar de amostras suspeitas de adulteração reduzindo o risco de que amostras comprometidas passem despercebidas.

Conclusão

O algoritmo XGBoost mostrou-se uma ferramenta promissora para identificar a adição de água em amostras de leite cru a partir de parâmetros físico-químicos, excluindo a crioscopia. Embora o desempenho tenha sido melhor na classificação de amostras não adulteradas, o modelo apresentou resultados satisfatórios e potencial para ser aprimorado, podendo contribuir como suporte no monitoramento da qualidade do leite.

Agência de Fomento

CAPES-Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

Referências

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa nº 76, de 26 de outubro de 2018. Estabelece os critérios para a composição, contagem microbiológica e parâmetros físico-químicos do leite. Diário Oficial da União, Brasília, 26 out. 2018.

HORA, R. N. S. da; FAGUNDES, I. L. C.; RIBEIRO, J. C. B. Avaliação da qualidade do leite e eficácia das ações de boas práticas de manejo entre agricultores familiares no município de Irajuba - Bahia. Revista Macambira, v. 8,

n. 1, p. 1-20, 2024. doi: https://doi.org/10.35642/rm.v8i1.1348