



República Federativa do Brasil
Ministério da Educação
Fundação Universidade Federal de Mato Grosso do Sul



UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
CAMPUS DE CHAPADÃO DO SUL

José Felipe Roterotte

**IDENTIFICAÇÃO DA CONCENTRAÇÃO DOS DIFERENTES TIPOS DE
ADULTERANTES DO LEITE POR MEIO DE REFLECTÂNCIA
HIPERESPECTRAL APLICADA À APRENDIZAGEM DE MÁQUINAS**

CHAPADÃO DO SUL, MS

2024



República Federativa do Brasil
Ministério da Educação
Fundação Universidade Federal de Mato Grosso do Sul



UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
CAMPUS DE CHAPADÃO DO SUL

**IDENTIFICAÇÃO DA CONCENTRAÇÃO DOS DIFERENTES TIPOS DE
ADULTERANTES DO LEITE POR MEIO DE REFLECTÂNCIA
HIPERESPECTRAL APLICADA À APRENDIZAGEM DE MÁQUINAS**

Trabalho de Conclusão de Curso
apresentado à Universidade Federal
de Mato Grosso do Sul, como parte
dos requisitos para obtenção do
título de Engenheiro Agrônomo.

Orientador: Prof. MSc. Aldair Felix da Silva

CHAPADÃO DO SUL, MS

2024

SUMÁRIO

RESUMO	1
ABSTRACT	2
INTRODUÇÃO	3
MATERIAL E MÉTODOS	4
RESULTADOS	5
DISCUSSÃO	7
CONCLUSÃO	8
REFERÊNCIAS	9

RESUMO

O leite vem sendo adulterado com diferentes soluções com o intuito de aumentar sua quantidade com menor teor dele puro. A adulteração do leite é uma prática perigosa na indústria de laticínios e extremamente prejudicial aos consumidores, já que o leite é um dos produtos alimentícios mais consumidos. Dessa forma, a utilização de dados espectrais, juntamente com a aprendizagem de máquina, pode ser uma ferramenta essencial na detecção de fraudes no leite por diferentes produtos, especialmente por ser uma metodologia inovadora, já que não havia relatos de algo similar na literatura. O objetivo do trabalho foi identificar quais as concentrações dos diferentes tipos de adulterantes do leite que a distinção espectral e o modelo de ML (máquina de vetores de suporte (SVM)) conseguem identificar. Foram utilizadas amostras de leite pasteurizado, às quais foram comparadas amostras do mesmo leite adulteradas com diferentes substâncias em diferentes concentrações. Especificamente, o leite pasteurizado foi adulterado com soro de leite, água e amido, em concentrações de 5, 10, 15, 50 e 75%. As amostras foram lidas em um espectrorradiômetro duas repetições cada uma. Posteriormente, os dados foram tabulados para análises com aprendizado de máquina (ML). Foi utilizado o seguinte algoritmo: máquina de vetores de suporte (SVM) e regressão logística (RL) que foi utilizada como modelo de controle. A metodologia proposta no trabalho foi para identificar diferenças no comportamento espectral das amostras de leite adulterada com diferentes concentrações dos adulterantes para as de leite pasteurizado, sendo possível identificar de forma mais evidente a distinção entre curvas na região do VIS e o do NIR. Outro objetivo do trabalho foi verificar a acurácia do algoritmo de aprendizagem de máquina na classificação das adulterações do leite e suas concentrações, alcançando acurácias de 80 para classificação correta (CC), próximos a 1 para F-score e acima de 0.8 para coeficiente kappa.

PALAVRAS-CHAVE: fraude, máquina vetor suporte, produtos alimentícios, qualidade do leite.

ABSTRACT

Milk has been adulterated with different solutions in order to produce more milk with less pure milk. Milk adulteration is a dangerous practice in the dairy industry and extremely harmful to consumers, as milk is one of the most consumed food products. Thus, the use of spectral data together with machine learning can be an essential tool in detecting milk fraud from different sources, especially as it is an innovative methodology, since there have been no reports of anything similar in the literature. The aim of this work was to identify which concentrations of the different types of milk adulterants the spectral distinction and the ML model (support vector machine (SVM)) can identify. Samples of pasteurized milk were used, to which samples of the same milk adulterated with different substances were compared. Specifically, the milk pasteurized was adulterated with whey, water and starch in concentrations of 5%, 10%, 15%, 50% and 75% for each adulterant. The samples were read twice each in a spectroradiometer. The data was then tabulated for analysis using machine learning (ML). The following algorithm was used: support vector machine (SVM) and logistic regression (LR), which was used as a control model. The methodology proposed in the work was that it would be possible to identify a difference in the spectral behavior of milk samples adulterated with different concentrations of adulterants compared to pasteurized milk, making it possible to more clearly identify the distinction between curves in the VIS and NIR regions. Another objective of the work was to verify the accuracy of the machine learning algorithm in classifying milk adulterations and their concentrations, achieving accuracies of 80 for correct classification (CC), close to 1 for F-score and above 0.8 for kappa coefficient.

KEYWORDS: fraud, support vector machine, food products, milk quality.

INTRODUÇÃO

Os métodos usados tradicionalmente para obter resultados de adulterantes no leite, são escolhidos pelo baixo custo, porém, alguns métodos utilizam-se reagentes químicos gerando resíduos que afetam o meio ambiente. Além disso há demanda de esforço manual e tempo, e alguns testes pode não ter eficácia quando a quantidade de adulterantes são baixas, comprometendo a segurança alimentar.

Atualmente, a população mundial está mais atenta quanto a adulterações em produtos alimentícios, especialmente nos de uso rotineiro no dia a dia. Essa preocupação ocorre pela facilidade de adulteração em produtos agropecuários, especialmente os que passam por várias etapas de processamento, o que os torna mais vulneráveis às adulterações para aumentar seu volume aparente, com o intuito de ganhos financeiros ilícitos (Sitorus & Lapcharoensuk, 2023).

Dessa forma, é de suma importância a utilização de tecnologias que permitem a rastreabilidade de produtos de forma rápida, versátil e econômica (Piras et al., 2021). O leite possui uma estrutura molecular complexa, que determina sua funcionalidade, assim, a utilização de sensores hiperespectrais podem contribuir com a identificação de fraudes, especialmente em leites adulterados, já que, para fraudar o leite, poucos adulterantes podem ser misturados para aumentar sua quantidade (Sowmya & Ponnusamy, 2021).

Na literatura não se encontram muitas informações sobre a aplicação da reflectância para a avaliação de atributos de funcionalidade de produtos lácteos. A utilização de informação VIS/NIR para prever atributos de produtos lácteos de forma não invasiva já é considerada na literatura, mas com poucas menções práticas (Wang et al., 2021). A utilização dessa técnica oferece mensurações e interpretações de espectros específicos, devido à interação da radiação eletromagnética e da matéria. Dessa forma, se torna possível a constatação de estruturas moleculares diferentes das originais do leite puro (Sowmya & Ponnusamy, 2021). Os dados espectrais, por produzirem uma gama de dados, podem ser explorados usando métodos de aprendizado de máquina (ML), a fim de identificar modelos que representem as características de amostras de leite pasteurizado e adulterado (Neto et al., 2019).

O leite vem sendo adulterado com diferentes soluções, com o intuito de aumentar seu rendimento, porém, com um teor de pureza menor. A adulteração do leite é uma prática perigosa na indústria de laticínios e extremamente prejudicial aos consumidores, já que o leite é um dos produtos alimentícios mais consumidos. Dessa forma, a utilização

de dados espectrais, juntamente com a aprendizagem de máquina, podem ser ferramentas essenciais na detecção de fraudes no leite por diferentes fontes, especialmente por ser uma metodologia inovadora, já que não há relatos de algo similar na literatura. O objetivo do trabalho foi identificar quais as concentrações dos diferentes tipos de adulterantes do leite que a distinção espectral e o modelo de ML (máquina de vetores de suporte (SVM)) conseguem identificar.

MATERIAL E MÉTODOS

Foram utilizadas amostras de leite pasteurizado, às quais foram comparadas com amostras do mesmo leite adulteradas com diferentes substâncias em diferentes concentrações. Especificamente, o leite pasteurizado foi adulterado com soro de leite, água e amido, em concentrações de 5, 10, 15, 50 e 75% para cada adulterante. Para garantir a representatividade dos dados, foram preparadas cinco amostras distintas para cada nível de adulteração, totalizando 75 amostras adulteradas. Além disso, foram incluídas 25 amostras de leite pasteurizado, sem qualquer adulteração, para fins de controle. Essa configuração experimental resultou em um total de 100 amostras. As amostras foram lidas em um espectrorradiômetro (FieldSpec 4 Jr da Analytical Spectral Devices, Longmont, CO, EUA) duas repetições cada uma. A resolução do equipamento é de 3 nm na faixa de 350 a 700 nm e 30 nm na faixa de 1400 a 2100 nm. O equipamento foi calibrado utilizando uma placa branca de sulfato de bário.

Após a obtenção dos dados espectrais, foram construídas curvas para o leite pasteurizado e para as adulterações realizadas, processando os dados obtidos em um programa próprio do sensor, chamado ViewSpecPro, que permite a conversão dos arquivos em extensão .txt, facilitando a análise e o processamento. Posteriormente, os dados foram tabulados para análises com aprendizado de máquina (ML).

Foram utilizados os seguintes algoritmos: redes neurais artificiais (RNA), árvore de decisão J48 (J48), REPTree (DT), floresta aleatória (RF), máquina de vetores de suporte (SVM) e regressão logística (RL), que foi utilizada como modelo de controle. Os parâmetros dos algoritmos foram definidos conforme configuração padrão do software Weka 3.8.5 (Bouckaert et al., 2010), exceto para RNA, onde foram definidos 10 neurônios na primeira camada e 10 neurônios na segunda camada. Weka é um software de código aberto que reúne diversos algoritmos de aprendizado de máquina para tarefas de

mineração de dados. Torna possível lidar com diversas tarefas, com dados como classificação, regressão, clustering, mineração de regras de associação e visualização.

Para avaliar a precisão da classificação dos algoritmos, foram estimados o percentual de classificações corretas (CC), kappa e F-score. Em seguida, foi realizada análise de variância, adotando-se delineamento inteiramente casualizado (DIC). Para agrupamento das médias do CC e do escore F, adotou-se o teste de Scott-Knott a 5% de probabilidade. O software Rbio versão 162 (Bhering, 2017) e os pacotes ExpDes.pt e ggplot2 do software R versão 4.1.0 (Ohmomo et al., 2022) foram utilizados para todas as análises estatísticas.

RESULTADOS

O comportamento espectral das amostras de leite adulterado teve comportamento diferente entre elas e especialmente do leite pasteurizado. Observou-se que as curvas espectrais do leite pasteurizado e do leite adulterado com soro exibiram um padrão de reflectância mais elevado e similar entre elas. Por outro lado, as curvas correspondentes ao leite adulterado com água e amido demonstraram um padrão de reflectância mais baixo e com maior proximidade entre si. Ao observar as curvas, nota-se que a faixa espectral que possibilita uma distinção mais eficaz entre elas é de 500 e 1300 nm, abrangendo as regiões do visível (VIS, 350-700 nm) e do infravermelho próximo (NIR, 700-1300 nm), nesta faixa, a separação das curvas é mais evidente, permitindo uma distinção maior entre as amostras.

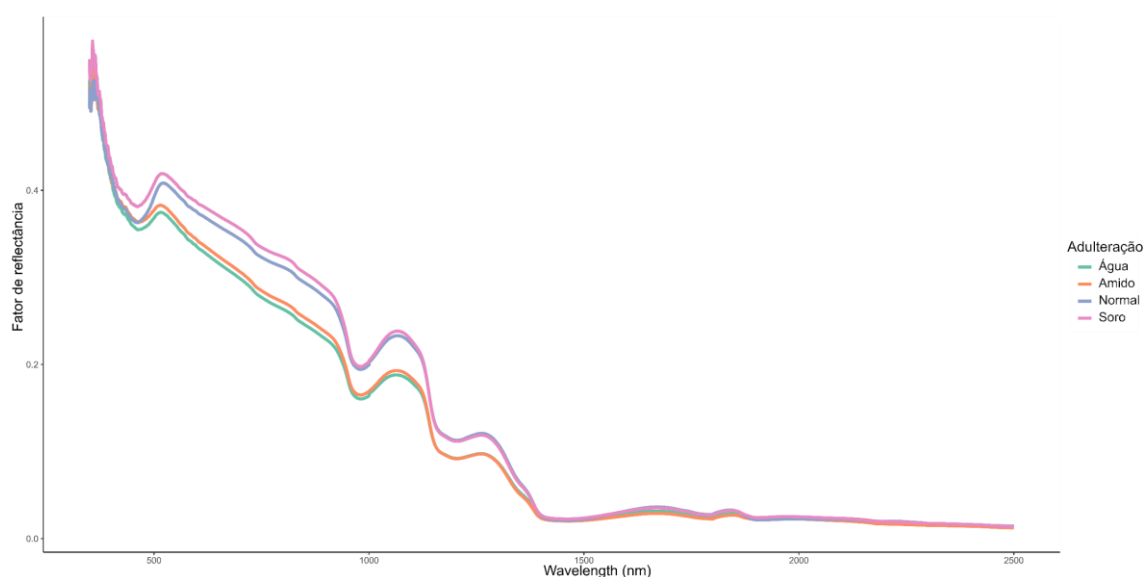


Figura 1 Curva hiperespectral para os diferentes tipos de adulteração realizados no leite.

Aplicar a inteligência computacional na distinção das diferenças espectrais entre as amostras de leite adulteradas ou não, permite uma abordagem mais eficaz, precisa e rápida para a detecção de fraude no leite, contribuindo para a melhoria da qualidade e segurança do produto. Dessa forma, os dados espectrais foram avaliados por meio de modelos de aprendizagem de máquina, buscando os algoritmos que conseguissem distinguir com mais assertividade as adulterações realizadas no leite. Foram utilizados três métricas de taxa de acerto: classificação correta (CC), F-score e Kappa (Figura 2). Em todas as métricas, o algoritmo regressão logística (RL) e máquina de vetor suporte (SVM) apresentaram melhores performances para a distinção das adulterações, com precisões acima de 80% para CC, 1 para F-Score e 0,8% para coeficiente Kappa.

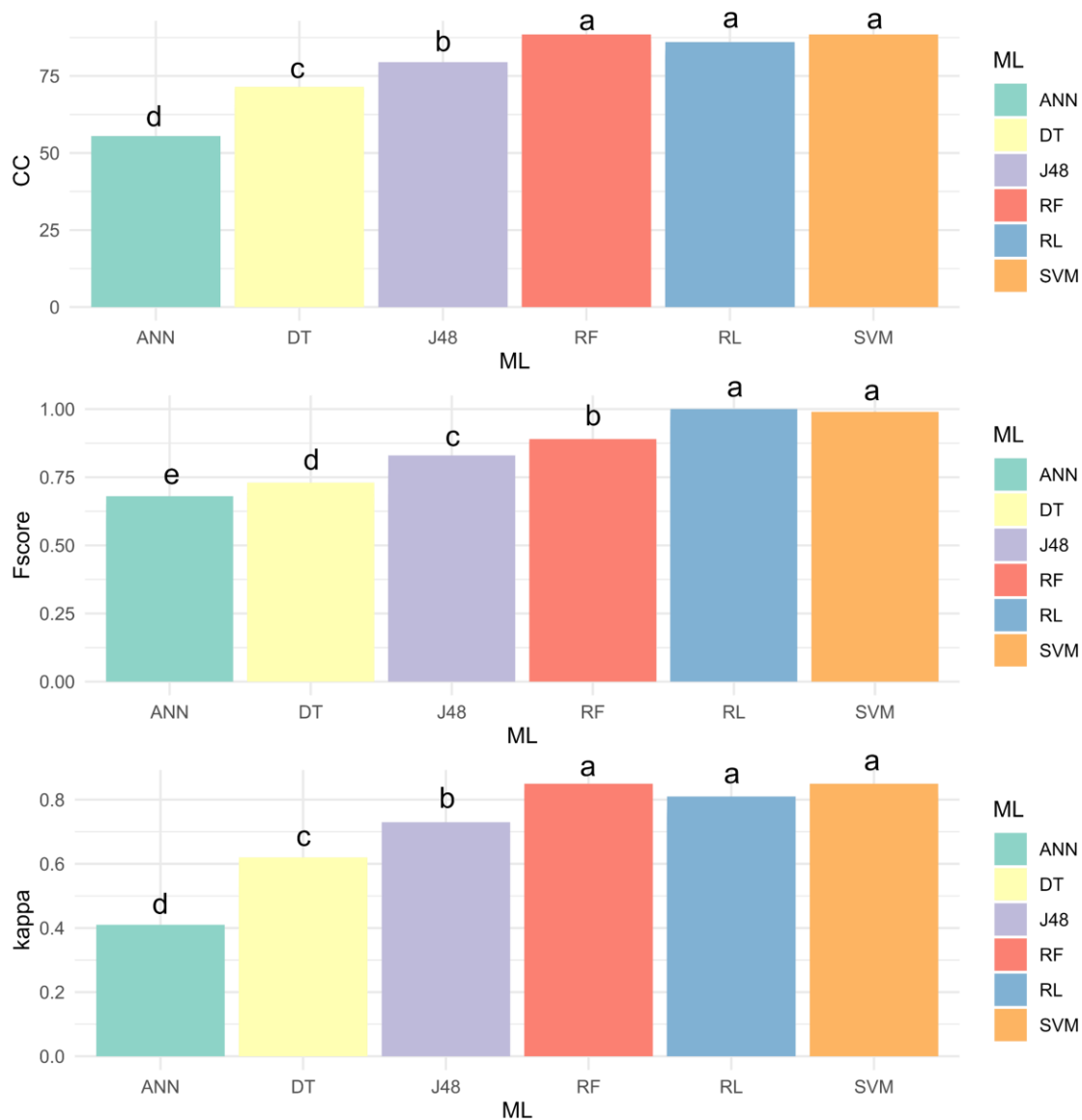


Figura 2. Comparação de médias para os modelos de aprendizagem de máquina para as métricas de acurácia classificação correta (CC), F-score e coeficiente kappa na classificação de diferentes tipos de adulteração realizados no leite.

O comportamento espectral das amostras aplicadas a algoritmos de ML permite a análise rápida e em tempo real de amostras de leite, sendo uma alternativa importante em processos de produção em larga escala e monitoramento contínuo. Isso aumenta a eficiência do controle de qualidade, permitindo intervenções imediatas quando adulterações são detectadas. A utilização dos algoritmos RL e SVM especialmente permitem a distinção das diferenças espectrais de forma mais eficiente, contribuindo para a melhoria da qualidade e segurança do processo de industrialização do leite.

DISCUSSÃO

Usar dados espectrais para a detecção de adulteração em alimentos tem ganhado destaque na área de controle de qualidade e segurança alimentar (Qu et al., 2015). Essas técnicas são baseadas na análise da forma de como a luz é refletida por diferentes materiais, permitindo identificar alterações na composição dos alimentos que podem indicar adulterações ou fraudes (Tilley, 2020). O comportamento espectral das amostras de leite adulteradas e puro apresentaram diferenças, possibilitando a identificação das mudanças entre elas, especialmente na faixa do VIS e no NIR, pois fornece uma riqueza de informações para dar suporte ao desenvolvimento de métodos para identificar adulterações (Wang et al., 2021).

O leite é um líquido complexo, constituído por água (87%), gordura (3,7%), proteína (3,4%), lactose (4,8%) e outros elementos (0,7%), como ácido cítrico e minerais, e a luz espalhada e refletida pelo leite possui informações sobre essas partículas e pode ser usada para determinar a qualidade do leite (Gastélum-Barrios et al., 2020), dessa forma, qualquer modificação nessas constituições altera o comportamento espectral. Bogomolov et al. (2012) afirmaram que as dispersões de luz na faixa visível possuem forte relação com gordura e proteína no leite, justificando a maior distinção entre as adulterações na região do VIS.

A espectroscopia de infravermelho próximo (NIR) também já se apresentou ser extremamente útil e eficaz para a análise do leite, sendo utilizado para prever os principais componentes do leite cru (Diaz-Olivares et al., 2020). Sitorus & Lapcharoensuk, (2023), afirmaram que a utilização de espectroscopia no infravermelho próximo (NIR) é um

método não destrutivo e rápido para prever propriedade de produtos agrícolas, sendo sua utilização simples, pois, não requer preparação de amostra e fornece resultados rápidos, podendo ser utilizada para avaliar a adulteração em muitos tipos de alimentos e produtos agrícolas.

Métodos de inteligência computacional são de extrema importância para analisar grandes volumes de dados gerados pelos sensores espectrais, de forma precisa, superando análises manuais. Isso reduz a possibilidade do erro humano na identificação e classificação das amostras de leite adulterado. O uso de algoritmos de aprendizagem de máquina, com a finalidade de qualificação e quantificação em alimentícios, demonstra resultados satisfatórios, como a utilização do modelo SVM para detecção de adulteração em leite de coco (Sitorus & Lapcharoensuk, 2023). Outro exemplo de uso do SVM foi sua aplicação na discriminação de adulterantes no leite, como pseudoproteínas e conteúdos sólidos, utilizando análise espectral de espectros NIR (Piras et al., 2021).

Utilizar dados hiperespectrais para distinguir adulterações realizadas em leite se apresentou eficiente, especialmente com o uso do algoritmo máquina de vetor suporte (SVM). Essas aplicações podem ser abrangidas para outros produtos, como azeite de oliva, mel, leite e sucos de frutas, produtos alimentícios frequentemente fraudados. Técnicas de reflectância aliadas a modelos de ML permitem comparar o espectro do produto testado com um espectro de referência de um produto puro, ajudando a identificar misturas ou substituições fraudulentas.

CONCLUSÃO

A metodologia proposta neste trabalho seria para identificar diferenças no comportamento espectral das amostras de leite adulterado para as de leite pasteurizado. Foi possível identificar, de forma mais evidente, a distinção entre curvas na região do VIS e o do NIR. Outro objetivo foi verificar a acurácia de algoritmos de aprendizagem de máquina na classificação das adulterações do leite, em que houve destaque para máquina de vetor suporte (SVM), atingindo acurácias de 80 para classificação correta (CC), próximos a 1 para F-score e acima de 0.8 para coeficiente kappa.

O trabalho revelou que com o uso de dados hiperespectrais, somado com a aprendizagem de máquinas, é possível identificar se há ou não presença de adulterantes

no leite a partir de 5%, seja água, soro ou amido, mostrando um grande avanço na detecção de fraudes no leite, sendo extremamente útil para as indústrias alimentícias.

REFERÊNCIAS

- Bhering, L. L. (2017). Rbio: A tool for biometric and statistical analysis using the R platform. *Crop Breeding and Applied Biotechnology*, 17(2), 187–190.
<https://doi.org/10.1590/1984-70332017V17N2S29>
- Bogomolov, A., Dietrich, S., Boldrini, B. & Kessler, R. W. (2012). Quantitative determination of fat and total protein in milk based on visible light scatter. *Food Chemistry*, 134(1), 412–418.
- Bouckaert, R. R., Holmes, G., Pfahringer, B., Reutemann, P. & Witten, I. H. (2010). WEKA- Experiences with a Java Open-Source Project Eibe Frank Mark A. Hall. *Journal of Machine Learning Research*, 11, 2533–2541. <http://www.cs.waikato.ac.nz/ml/weka>.
- Diaz-Olivares, J. A., Adriaens, I., Stevens, E., Saeys, W. & Aernouts, B. (2020). Online milk composition analysis with an on-farm near-infrared sensor. *Computers and Electronics in Agriculture*, 178, 105734. página???
- Gastélum-Barrios, A., Soto-Zarazúa, G. M., Escamilla-García, A., Toledano-Ayala, M., Macías-Bobadilla, G. & Jauregui-Vazquez, D. (2020). Optical methods based on ultraviolet, visible and near-infrared spectra to estimate fat and protein in raw milk: A review. *Sensors*, 20(12), 3356. página???
- Neto, H. A., Tavares, W. L. F., Ribeiro, D. C. S. Z., Alves, R. C. O., Fonseca, L. M. & Campos, S. V. A. (2019). On the utilization of deep and ensemble learning to detect milk adulteration. *BioData Mining*, 12, 1–13.
- Ohmomo, H., Harada, S., Komaki, S., Ono, K., Sutoh, Y., Otomo, R., Umekage, S., Hachiya, T., Katanoda, K., Takebayashi, T. & Shimizu, A. (2022). R: a language and environment for statistical computing. *r foundation for statistical computing*. (No Title), 31(1), 269–279. <https://doi.org/10.1158/1055-9965.EPI-21-0444>
- Piras, C., Hale, O. J., Reynolds, C. K., Jones, A. K., Taylor, N., Morris, M. & Cramer, R. (2021). Speciation and milk adulteration analysis by rapid ambient liquid MALDI mass

spectrometry profiling using machine learning. *Scientific Reports*, 11(1), 3305.

página???

Qu, J.-H., Liu, D., Cheng, J.-H., Sun, D.-W., Ma, J., Pu, H. & Zeng, X.-A. (2015).

Applications of near-infrared spectroscopy in food safety evaluation and control: A review of recent research advances. *Critical Reviews in Food Science and Nutrition*, 55(13), 1939–1954.

Sitorus, A. & Lapcharoensuk, R. (2023). A rapid method to predict type and adulteration of coconut milk by near-infrared spectroscopy combined with machine learning and chemometric tools. *Microchemical Journal*, 195, 109461. página???

Sowmya, N. & Ponnusamy, V. (2021). Development of spectroscopic sensor system for an IoT application of adulteration identification on milk using machine learning. *IEEE Access*, 9, 53979–53995. página???

Tilley, R. J. D. (2020). *Colour and the optical properties of materials*. John Wiley & Sons. páginas???

Wang, C., Reis, M. G., Waterhouse, G. I. N., Hemar, Y. & Reis, M. M. (2021). Prediction of dairy powder functionality attributes using diffuse reflectance in the visible and near infrared (Vis-NIR) region. *International Dairy Journal*, 117, 104981. página???