

Quimiometria aplicada com *software no-code*

*Felipe Rodrigues Nascimento¹(IC), Márcia Helena Cassago Nascimento¹(PG), Matthews Silva Martins²(PG), Bruna Ferro Brun²(PG), Valério Garrone Barauna²(PQ), Paulo Roberto Filgueiras¹(PQ)

*felipe.nascimento.09@edu.ufes.br

¹Departamento de Química, Universidade Federal do Espírito Santo, Av. Fernando Ferrari, 514 - Goiabeiras, Vitória - ES, 29075-910, Brasil,

²Departamento de Ciências Fisiológicas, Universidade Federal do Espírito Santo, Av. Marechal Campos, 1468 – Maruípe CEP 29.040-090, Vitória – ES, Brasil.

Palavras Chave: *no-code*, classificação, quimiometria, adulteração de alimentos, d-dímero

Introdução

O uso de metodologias quimiométricas no tratamento de dados químicos frequentemente envolve programação, mas alternativas *no-code* têm surgido para facilitar esse processo por meio de interfaces gráficas (KIM et al., 2024). Um exemplo é o software *Orange Data Mining*, que, baseado em Python, oferece uma interface amigável para aplicação de diversos métodos de aprendizado de máquina, incluindo análise de dados espectrais. Este trabalho propõe apresentar uma alternativa à alta programação computacional por meio de software *no-code* para abordagens de classificação e de regressão a partir de espectroscopia na região do infravermelho médio.

Resultados e Discussão

Neste trabalho, utilizou-se dois conjuntos de dados de trabalhos anteriores do grupo de pesquisa. São dados de espectroscopia na região do infravermelho médio com reflectância total atenuada (ATR-FTIR) de matrizes distintas. O primeiro conjunto inclui 100 amostras de plasma sanguíneo distribuídas conforme os níveis de concentração de biomolécula D-dímero ($\mu\text{g mL}^{-1}$) (BRUN et al., 2024). O segundo conjunto é composto por 66 espectros de três marcas de *whey protein*, misturadas com diferentes proporções de farinha de trigo (0 a 100 wt%, em intervalos de 10 wt%) (MARTINS et al., 2023).

Para o primeiro conjunto realizou-se uma análise exploratória, por meio de análise de componentes principais (PCA) (PC=10), para avaliar agrupamentos intrínsecos aos dados pré-processados pela segunda derivada. Para classificação, separou-se os conjuntos em treino (70%) e teste (30%), obtendo-se ao final, modelo com exatidão de 0,710 (71%).

No segundo conjunto de dados, os modelos foram treinados com dados pré-processados com correção de linha de base pelo método *rubber band*, média das triplicatas e normalização vetorial. Os modelos de regressão por quadrados mínimos parciais (PLS) foram realizados de maneira prática com o auxílio de um *widget data sampler* que separou os conjuntos de calibração (70%) e validação (30%), seguindo-se para a validação cruzada com o subconjunto calibração por meio do método k-fold ($k=5$), resultando em modelos de alta exatidão (RMSEP de 0,133 a 0,488 wt%, R^2_p de 0,845 a 0,989). Por meio do widget *Test and Score* para comparação simultânea de métodos de classificação, os modelos SVM e Gradient Boosting apresentaram o melhor desempenho no teste externo, cujos resultados estão apresentados na Tabela 1.

Tabela 1- Parâmetros de desempenho de modelos de classificação – conjunto 2, Teste externo (Classe 1 = Whey não adulterado)

Modelo	AUC	EXAT	F1	Prec	Sens	MCC
SVM	1,00	0,923	0,952	1,00	0,909	0,778
Gadient Boosting	1,00	0,923	0,952	1,00	0,909	0,778

AUC: *Area under curve*; EXAT: Exatidão; F1: F1-score; Prec: Precisão; Sens: Sensibilidade; MCC: *Matthews correlation coefficient*

Conclusões

Apesar de limitações na edição de figuras, o *software* mostra-se eficiente para otimização do tempo e uma alternativa viável com diferentes métodos de aprendizado de máquina que poderiam requerer atividade densa de programação. A abordagem *no-code* facilita o acesso e contribui para desmistificar a quimiometria.

Agradecimentos

LABPETRO, UFES, FEST, FAPES, CAPES, CNPq

MARTINS, Matthews S. et al. J. Am. Chem. Soc., v. 172, p. 114161, 2022..

BRUN, M. A. et al. Spectrochim. Acta A Mol. Biomol. Spectrosc., 306, 123873, 2024.

KIM, S.-Y et al. J. Chem. Educ., 101, 1771–1776, 2024.