



Universidade de Brasília
Instituto de Química

Mariana Moura Moutinho

**ANÁLISES INSTRUMENTAIS E QUIMIOMETRIA NA DISCRIMINAÇÃO DE
CERVEJAS INDUSTRIAIS NACIONAIS:
UM COMBATE À FALSIFICAÇÃO**

Brasília – DF

2026

Mariana Moura Moutinho

MARIANA MOURA MOUTINHO

**ANÁLISES INSTRUMENTAIS E QUIMIOMETRIA NA DISCRIMINAÇÃO DE
CERVEJAS INDUSTRIAIS NACIONAIS:
UM COMBATE À FALSIFICAÇÃO**

Dissertação de Mestrado apresentada ao Programa de Pós-
Graduação em Química da Universidade de Brasília para
obtenção do Título de Mestre em Química
Área de Concentração: Química Analítica e Quimiometria
Orientadora: Prof.^a. Dr.^a. Grace Ferreira Ghesti

Brasília – DF

2026

AGRADECIMENTOS

À Universidade de Brasília (UnB), em especial ao Instituto de Química, pela contínua contribuição na minha formação profissional e acadêmica.

Ao Instituto de Criminalística da Polícia Civil do Distrito Federal, especialmente ao Laboratório de Química e Física Forense (LQFF) por ter se tornado um espaço de constante aprendizado nos últimos anos.

À Coordenação de Aperfeiçoamento Pessoal de Nível Superior (CAPES) pelo apoio financeiro.

Ao perito Dr. Diego M. Souza pela paciência, incentivo, orientação e conhecimento transmitido, que muito contribuíram para a elaboração deste trabalho e do meu crescimento como aluna e pesquisadora.

À professora orientadora Dr. Grace F. Ghesti por me orientar, torcer, e me tranquilizar durante toda elaboração deste trabalho.

À minha mãe, que sempre foi a primeira a acreditar no meu potencial. Seu incentivo constante, seu exemplo de força e sua certeza de que o estudo transforma vidas foram pilares essenciais nesta jornada. Foi abrigo e colo em todos os momentos em que precisei.

Ao meu pai, que desde a infância despertou em mim o gosto pela leitura, pela curiosidade e pelo questionamento que até hoje fazem parte de mim.

À minha família, que sempre foi e continua sendo meu alicerce. Seu apoio incondicional, amor e presença me sustentaram nos dias difíceis e celebraram comigo cada conquista.

Aos que me ouviram falar do mestrado dia e noite, Amanda Rabelo, Ana Valadão, Arthur Lorenzo, Bruna Samy, Felipe Gomes, Gabriel Santos, Geisa Maria, Luiza Diniz e tantos outros amigos que me acompanharam do início ao fim nessa jornada, me viram nos meus melhores e piores dias e sempre me lembraram que eu era capaz quando precisei, vocês foram fundamentais.

Aos que dividiram disciplinas e congressos de química e quimiometria comigo, Thayllan Lima e Luana Lisboa, vocês fizeram da academia um lugar mais leve e as aulas, mais engraçadas.

Aos meus colegas de trabalho do Versa que compreenderam a importância da pesquisa na minha vida de professora e me deram espaço e suporte para fazer as duas coisas que eu amo.

A todos que de alguma forma contribuíram direta e indiretamente para conclusão deste trabalho, meu muito obrigada!

RESUMO

A crescente ocorrência de fraudes por rotulagem indevida em cervejas industriais no Brasil evidencia a necessidade de métodos analíticos acessíveis para autenticação e controle de qualidade. Este trabalho empregou espectroscopia Raman, IV, UV-Vis e análise físico-química via Alcoalyzer Beer Analyzer associadas a técnicas quimiométricas para discriminar cervejas puro-malte e adjuntadas e identificar desvios indicativos de adulteração. Os dados foram pré-processados (SNV, MSC e Savitzky-Golay) e avaliados por PCA, PLS-DA e LDA com validação cruzada. A PCA revelou padrões estruturais e agrupamentos consistentes entre categorias, enquanto PLS-DA e LDA apresentaram métricas robustas de classificação, sendo avaliadas métricas de desempenho como acurácia, F1-macro, curvas AUC/ROC e matrizes de confusão. Destaca-se o desempenho expressivo do UV-Vis, com acurácia de 100% em modelo LDA, capaz de promover separação eficiente com baixo custo e fácil operacionalização, evidenciando seu potencial como ferramenta rápida para triagem e verificação de autenticidade. Os resultados comprovam que a combinação de técnicas espectroscópicas e quimiometria constitui uma abordagem eficaz e viável para o monitoramento da integridade de cervejas comerciais, distinção de diferentes marcas de cerveja e apoio a ações de fiscalização e segurança do consumidor, auxiliando a polícia e demais órgãos no monitoramento e combate à falsificação de cervejas de ampla circulação no Brasil.

Palavras-chave: cerveja; falsificação; espectroscopia; quimiometria; LDA; PCA; PLS-DA.

ABSTRACT

The increasing occurrence of label counterfeit in industrial beers in Brazil highlights the need for accessible analytical methods for authentication and quality control. This study employed Raman, FTIR, and UV-Vis spectroscopy, along with physicochemical analysis using the Alcozyzer Beer Analyzer, combined with chemometric techniques to discriminate between all-malt and adjunct beers and identify deviations indicative of adulteration. Spectral data were pre-processed (SNV, MSC, and Savitzky-Golay) and evaluated through PCA, PLS-DA, and LDA with cross-validation. PCA revealed structural patterns and consistent clustering among beer categories, while PLS-DA and LDA demonstrated robust classification performance, evaluated using accuracy, macro-F1 score, AUC/ROC curves, and confusion matrices. UV-Vis spectroscopy showed remarkable results, achieving 100% accuracy in the LDA model, enabling efficient separation with low cost and simple operation, reinforcing its potential as a rapid screening tool for authenticity verification. These findings demonstrate that combining spectroscopic techniques with chemometrics constitutes an effective and feasible strategy for monitoring the integrity of commercial beers, distinguishing different brands, and supporting regulatory and forensic applications, aiding law-enforcement agencies in monitoring and combating beer fraud in Brazil.

Keywords: beer; counterfeit; spectroscopy; chemometrics; LDA; PCA; PLS-DA.

LISTA DE TABELAS

Tabela 1. Informações contidas nos rótulos de cada marca analisada.....	47
Tabela 2. Resultados Obtidos do Analisador de Cervejas com tratamento estatístico descritivo.....	56
Tabela 3. Métricas de desempenho para cada técnica, a partir de diferentes modelos de classificação.....	70
Tabela 4. Dados de modelagem para parâmetros quantitativos.....	85

LISTA DE FIGURAS

Figura 1. Espectro pré-processado do FTIR com SNV	59
Figura 2. Espectro pré-processado do Raman Resolve com Savitzky-Golay.	60
Figura 3. Espectro pré-processado do Raman Tactid com SNV.....	61
Figura 4. Espectro pré-processado do UV-VIS com MSC.....	62
Figura 5. Modelo bidimensional (2D) para a Análise de Componentes Principais do Analisador de Bebidas.....	63
Figura 6. Modelo bidimensional (2D) para a Análise de Componentes Principais do Raman Tactid	65
Figura 7. Modelo bidimensional (2D) para a Análise de Componentes Principais do Raman Tactid com rótulos com adjunto em azul e puro malte em laranja.	66
Figura 8. Modelo bidimensional (2D) para a Análise de Componentes Principais do FTIR.....	68
Figura 9. Modelo bidimensional (2D) para a Análise de Componentes Principais do Raman Resolve.	69
Figura 10. Modelo bidimensional (2D) para a Análise de Componentes Principais do UV-VIS..	70
Figura 11. Os histogramas das distribuições de LD1 para FTIR, Resolve, Tactid e UV-VIS	78
Figura 12. Gráficos de escores LD1 × LD2 para FTIR, Resolve, Tactid e UV-VIS.....	78
Figura 13. Matrizes de confusão do LDA para FTIR, Resolve, Tactid e UV-VIS.....	79
Figura 14. Curvas ROC/AUC para FTIR, Resolve, Tactid e UV-VIS.....	81
Figura 15. Gráficos de escores LV1 × LV2 para FTIR, Resolve, Tactid e UV-VIS.....	82
Figura 16. Gráficos de escores LV1 × LV2 para FTIR, Resolve, Tactid e UV-VIS.	83
Figura 17. Matrizes de confusão para FTIR, Resolve, Tactid e UV-VIS.	84
Figura 18. Gráficos de regressão de parâmetros físico-químicos para dados do FTIR.	87
Figura 19. Gráficos de regressão de parâmetros físico-químicos para dados do Raman Resolve.	88
Figura 20. Gráficos de regressão de parâmetros físico-químicos para dados do Raman Tactid....	89

LISTA DE ABREVIATURAS E SIGLAS

ADF – Grau aparente de fermentação (*Apparent Degree of Fermentation*)

AMS – Amstel (marca de cerveja)

ANN – Rede neural artificial (*Artificial Neural Network*)

ANT – Antartica (marca de cerveja)

BRA – Brahma (marca de cerveja)

CART – *Classification and Regression Trees* (Árvores de Classificação e Regressão)

CV – Coeficiente de variação

DPR – Desvio padrão relativo

EN – Erro normalizado

ER – Erro relativo

EVD – Decomposição em autovalores (*Eigenvalue Decomposition*)

FTIR – Espectroscopia no infravermelho por transformada de Fourier (*Fourier Transform Infrared Spectroscopy*)

HEI – Heineken (marca de cerveja)

KNN – K-vizinhos mais próximos (*K-Nearest Neighbors*)

LDA – Análise discriminante linear (*Linear Discriminant Analysis*)

LOO – Leave-One-Out

LOOCV – Leave-One-Out Cross-Validation

LV – *Latent Variable* (Variável Latente)

MIR – Espectroscopia no infravermelho médio (*Mid-Infrared Spectroscopy*)

MLR – Regressão linear múltipla (*Multiple Linear Regression*)

MSC – Correção multiplicativa de espalhamento (*Multiplicative Scatter Correction*)

NIPALS – Algoritmo iterativo não linear de mínimos quadrados parciais (*Nonlinear Iterative Partial Least Squares*)

NIR – Espectroscopia no infravermelho próximo (*Near-Infrared Spectroscopy*)

ORI – Original (marca de cerveja)

PCA – Análise por componentes principais (*Principal Component Analysis*)

PCR – Regressão por componentes principais (*Principal Component Regression*)

PET – Petra (marca de cerveja)

PLS – Mínimos quadrados parciais (*Partial Least Squares*)

PLS-DA – Análise discriminante por mínimos quadrados parciais (*Partial Least Squares Discriminant Analysis*)

PRESS – Predicted Residual Error Sum of Squares

QDA – Análise Discriminante Quadrática (*Quadratic Discriminant Analysis*)

RF – *Random Forest* (Floresta Aleatória)

RMN – Ressonância magnética nuclear de ^1H

ROC-AUC – Receiver Operating Characteristic – Area Under the Curve

SIMCA – Modelagem independente suave por analogia de classes (*Soft Independent Modeling of Class Analogy*)

SKO – Skol (marca de cerveja)

SNV – Variável normal padrão (*Standard Normal Variate*)

SPA – Spaten (marca de cerveja)

STE – Stella Artois (marca de cerveja)

SVD – Decomposição em valores singulares (*Singular Value Decomposition*)

SVM – Máquina de vetores de suporte (*Support Vector Machine*)

UNEQ – *UNEQual Dispersions* (Modelagem de Dispersões Desiguais)

UV-VIS – Espectrofotometria no ultravioleta e visível (*Ultraviolet-Visible Spectroscopy*)

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Cerveja	12
1.2	Falsificação de Produtos, Controle de Qualidade e Autenticidade	17
1.3	Análise Instrumental	22
1.3.1	Analisador de Bebidas	24
1.3.2	Espectroscopia do Infravermelho com Transformada de Fourier (FTIR). ...	26
1.3.3	Espectroscopia Raman	27
1.3.4	Espectrofotometria no Ultravioleta-Visível (UV-Vis).....	29
1.4	Pré-Processamento de Dados.....	31
1.5	Quimiometria.....	34
1.5.1	Análise por Componentes Principais (PCA).....	37
1.5.2	Análise Discriminante por Quadrados Parciais (PLS-DA)	39
1.5.3	Validação Cruzada.....	41
1.5.4	Análise Discriminante Linear (LDA)	43
2	OBJETIVOS.....	45
2.1	Objetivo geral	45
2.2	Objetivos específicos	45
3	MATERIAIS E MÉTODOS.....	47
3.1	Obtenção e Armazenamento das Amostras.....	47
3.2	Aquisição dos dados espectroscópicos	48
3.3	Pré-processamento dos dados	50
3.4	Estratégias quimiométricas	51
3.4.1	Exploração inicial: Análise por Componentes Principais (PCA)	51
3.4.2	Modelagem discriminante: Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA) e Análise Discriminante Linear (LDA)	52
3.4.3	Validação e particionamento.....	54
3.4.4	Avaliação e interpretação	54
3.5	Modelagem de Regressão e Ambiente Computacional	55
4	RESULTADOS E DISCUSSÃO.....	56

4.1	Análise dos Parâmetros Físico-Químicos das Cervejas	56
4.2	Dados Pré-Processados.....	58
4.3	Análise PCA dos dados do Analisador de Bebidas.....	62
4.4	Análise PCA dos dados do Raman Tactid.	65
4.5	Análise PCA dos dados de FTIR	67
4.6	Análise da PCA do Raman Resolve.....	68
4.7	Análise da PCA da Espectrofotometria UV-VIS.....	70
4.8	Modelagem e Análises Supervisionadas.....	71
4.9	Resultados de LDA.....	76
4.10	Resultados PLS-DA.....	7
4.11	Resultados Preliminares e Perspectivas Futuras.....	85
5	CONCLUSÃO.....	90
6	REFERÊNCIAS BIBLIOGRÁFICAS.....	92
	MATERIAL COMPLEMENTAR.....	102

1 INTRODUÇÃO

1.1 Cerveja

A cerveja é uma das bebidas alcoólicas mais antigas e amplamente consumidas no mundo, com registros históricos que remontam a aproximadamente 6.000 a.C., na antiga Mesopotâmia. Evidências arqueológicas indicam que as primeiras civilizações, como os sumérios e os egípcios, já produziam uma forma primitiva da bebida a partir da fermentação de cereais, sendo inclusive associada a práticas religiosas, medicinais e sociais (HORNSEY, 2003). Com o passar dos séculos, o conhecimento empírico sobre a fermentação e a produção da cerveja foi sendo transmitido e aprimorado por diferentes culturas. Na Europa medieval, os mosteiros desempenharam papel essencial no refinamento das técnicas de fabricação, bem como na padronização de receitas e controle de qualidade rudimentar, consolidando a bebida como um importante elemento cultural e econômico (NELSON, 2005).

A cerveja é, essencialmente, uma bebida fermentada à base de cereais, elaborada a partir de quatro ingredientes principais: água, malte (geralmente de cevada), lúpulo e levedura. Adicionalmente, podem ser utilizados adjuntos cervejeiros como arroz, milho, sorgo e outros grãos, que alteram o perfil sensorial e reduzem os custos de produção (COELHO et al., 2022). Esses ingredientes e suas proporções variam conforme o estilo da cerveja e os objetivos do produtor. Atualmente, o mercado cervejeiro brasileiro é composto por dois segmentos principais: a cerveja artesanal e a cerveja industrial. Segundo a Associação Brasileira de Cerveja Artesanal (ABRACERVA, 2025), a cerveja artesanal é caracterizada pela produção em pequena escala, limitada a até 5 milhões de litros por ano (cerca de 416 mil litros por mês), sendo elaborada por cervejarias independentes que priorizam a diversidade de estilos, qualidade sensorial e identidade regional. Por outro lado, a cerveja industrial é fabricada em grande escala, com foco na

padronização e competitividade no mercado. Este segmento frequentemente utiliza formulações otimizadas para reduzir custos, como a substituição parcial do malte por cereais não maltados de menor valor agregado. Em 2024, o Brasil registrou 1949 cervejarias, um crescimento de 5,5% em relação ao ano anterior, conforme o Anuário da Cerveja publicado em 2025 pelo Ministério da Agricultura e Pecuária (MAPA).

Em 2024, o estado de São Paulo manteve-se como a unidade da federação com o maior número de cervejarias registradas no Brasil, totalizando 427 estabelecimentos, após crescimento de 4,1% em relação ao ano anterior, correspondente à abertura de 17 novas cervejarias. Santa Catarina destacou-se como o estado com maior crescimento absoluto no número de estabelecimentos em comparação a 2023, apresentando aumento de 25 cervejarias, o que representa crescimento de 11,1%. No que se refere ao crescimento relativo, o Tocantins apresentou a maior variação percentual, com aumento de 50,0%, passando de quatro cervejarias registradas em 2023 para seis em 2024. Por outro lado, Acre, Amazonas, Mato Grosso do Sul, Piauí e Sergipe mantiveram, em 2024, o mesmo número de cervejarias registradas observado no ano anterior. Em sentido oposto, Goiás apresentou o maior decréscimo absoluto, com redução de quatro estabelecimentos, o que corresponde a uma queda de 9,3%, passando de 43 cervejarias em 2023 para 39 em 2024. Ademais, os estados do Amapá, Paraíba e Roraima também registraram diminuição no número de cervejarias, com reduções de 50,0%, 12,5% e 50,0%, respectivamente, o que, em todos os casos, correspondeu à perda de um estabelecimento em relação ao ano anterior (BRASIL, 2025).

Do ponto de vista técnico e regulatório, a cerveja pode ser classificada de diferentes formas conforme suas características físico-químicas e sensoriais. Internacionalmente, guias como o *Beer Judge Certification Program* (BJCP) e o *Brewers Association Style Guidelines* fornecem

descrições detalhadas de estilos reconhecidos, como pilsner, stout, IPA, entre outros. No Brasil, o MAPA estabelece critérios legais para a definição de “cerveja”, “cerveja puro malte”, “cerveja com adjuntos”, além de regulamentar parâmetros como teor alcoólico, uso de aditivos e rotulagem. Essa normatização é essencial para proteger o consumidor, garantir a conformidade entre os produtos e orientar os processos de fiscalização sanitária e tributária.

As cervejas comercializadas mundialmente podem ser classificadas, em termos técnicos, de acordo com o tipo de fermentação, sendo os principais grupos Lager, Ale e Lambic. No Brasil, entretanto, o mercado é amplamente dominado pelas cervejas do tipo Lager, mais especificamente do subestilo Pilsen, caracterizadas por fermentação a baixas temperaturas (cerca de 7 a 13°C), com uso da levedura *Saccharomyces pastorianus*. Essas cervejas possuem coloração clara, corpo leve, baixo amargor e elevada carbonatação, sendo elaboradas para atender ao consumo em larga escala, muitas vezes associadas a ambientes de clima quente e consumo gelado (BAMFORTH, 2009).

A Instrução Normativa nº 65, de 10 de dezembro de 2019, do Ministério da Agricultura, Pecuária e Abastecimento (MAPA), estabelece os Padrões de Identidade e Qualidade (PIQ) para produtos de cervejaria no Brasil, contemplando, entre outros, o estilo Pilsen. Conforme essa normativa, a cerveja Pilsen deve conter, no mínimo, 55% de malte de cevada ou extrato de malte em relação ao extrato primitivo, sendo permitida a utilização de adjuntos cervejeiros, como cereais não maltados, em proporção máxima de 45% do extrato primitivo. Caso esse percentual seja ultrapassado, o produto deve ser rotulado como "cerveja de", seguido da denominação do cereal predominante. Em termos de rotulagem, é autorizada a utilização de expressões internacionalmente reconhecidas, como "Pilsen", desde que essas sejam apresentadas de forma clara e prontamente identificável no rótulo, respeitando as características originais do produto.

Além disso, a Instrução Normativa nº 65/2019 estabelece parâmetros físico-químicos, microbiológicos e sensoriais que visam assegurar a qualidade e a segurança do produto final, contemplando limites para contaminantes, requisitos organolépticos e outros critérios técnicos específicos. Entre as marcas mais populares no Brasil, como Skol, Brahma, Antártica, Original, Heineken, Budweiser e Stella Artois, todas pertencentes à categoria das Lager, possuem variações no grau de amargor e corpo. Skol, Brahma e Antártica são exemplos clássicos de Pilsen brasileiras: leves, pouco amargas e com perfil sensorial suave. A Original, embora também do tipo Pilsen, é comercializada como uma versão mais "premium", com leve incremento no corpo e maior equilíbrio de malte.

De acordo com pesquisa realizada pela Brazil Panels em parceria com a agência Conexão Vasques, as marcas de cerveja mais consumidas no Brasil são: Brahma (43,1%), Heineken (40,6%), Skol (36,6%), Amstel (33,2%), Budweiser (28,8%), Antarctica (27,6%), Itaipava (26,5%), Stella Artois (18%), Petra (16,7%) e Original (16,4%) (PAVAN, 2024). Heineken e Stella Artois, classificadas como *International Pale Lager*, apresentam amargor um pouco mais pronunciado e sabor mais lupulado (HEINEKEN, [s.d.]), enquanto a Budweiser, uma *American Lager*, utiliza arroz como adjunto, resultando em um perfil seco e suave (BUDWEISER, 2025). Apesar de suas origens internacionais, essas cervejas seguem características sensoriais próximas das Pilsen populares, o que favorece sua ampla aceitação no mercado brasileiro. Essa homogeneidade estilística, embora vantajosa para a padronização e comercialização, limita a diversidade de perfis sensoriais, o que tem impulsionado o crescimento do mercado de cervejas artesanais nos últimos anos (COELHO et al., 2022).

Outro problema crescente, especialmente em países com alto consumo interno como o Brasil, é a falsificação de bebidas alcoólicas, incluindo a cerveja. Além dos prejuízos econômicos,

tributários e à reputação das marcas, a presença de produtos falsificados representa riscos diretos à saúde pública, uma vez que podem ser armazenados ou manipulados de forma inadequada, utilizar ingredientes de origem desconhecida ou estar sujeitos à contaminação microbiológica devido à ausência de controle sanitário. Em muitos casos, o consumidor não é capaz de perceber sensorialmente a adulteração, o que agrava a exposição a substâncias potencialmente perigosas. Dados da Associação Brasileira de Bebidas Destiladas (ABBD) revelam que 36% do volume total de bebidas comercializadas no país são adulteradas. Além disso, a produção e distribuição desses produtos ilegais estão frequentemente associadas a facções criminosas, gerando não apenas impactos sanitários, mas também perdas fiscais expressivas, estimadas em R\$ 52,9 bilhões em razão da sonegação de impostos (TERRA, 2025).

As adulterações em cervejas industriais podem provocar alterações sensoriais perceptíveis ao consumidor, ainda que, em alguns casos, sejam sutis e difíceis de identificar sem auxílio analítico. Essas modificações incluem alterações no aroma lupulado, que pode apresentar notas menos frescas ou com características oxidativas indesejáveis, sabor residual adocicado devido à presença de açúcares residuais não fermentados, aumento da acidez provocada pela adição de compostos ácidos, além de mudanças no corpo e na sensação de espuma, afetando a experiência geral de consumo. Tais adulterações, ainda que imperceptíveis a uma avaliação sensorial superficial, podem ser detectadas por meio de técnicas físico-químicas e espectroscópicas, como a espectroscopia no infravermelho, Raman e espectrofotometria UV-Vis, que permitem identificar desvios na composição química da bebida. Nesse sentido, a caracterização instrumental torna-se uma ferramenta essencial para confirmar a autenticidade do produto e proteger o consumidor de fraudes, complementando as análises sensoriais e as inspeções regulatórias (GONZALEZ VIEJO; FUENTES; HERNANDEZ-BRENES, 2021).

Nos últimos anos, o interesse científico e tecnológico pela cerveja vem se intensificando, especialmente no campo das análises químicas e na detecção de fraudes. Tal interesse se justifica pela necessidade de garantir a autenticidade do produto, proteger os consumidores e fortalecer a confiança no mercado formal (BRYAN; HILL, 2024). Casos de adulteração, falsificação por troca de rótulos e comercialização de bebidas fora dos padrões têm sido recorrentes em diversas regiões, exigindo a adoção de ferramentas analíticas robustas para o monitoramento e controle da qualidade. Nesse contexto, compreender a história, os ingredientes e a relevância econômica da cerveja é essencial para embasar estudos que envolvam sua caracterização físico-química, autenticidade e rastreabilidade (DA SILVA et al., 2019; DE ALMEIDA et al., 2018; MAHMOOD; PETRACO; HE, 2012).

1.2 Falsificação de Produtos, Controle de Qualidade e Autenticidade

O Acordo sobre os Aspectos dos Direitos de Propriedade Intelectual Relacionados ao Comércio (TRIPS) define bens de marca registrada falsificados como: "quaisquer bens, incluindo embalagens, que ostentem, sem autorização, uma marca que seja idêntica à marca validamente registrada em relação a esses bens ou que não possa ser distinguida, em seus aspectos essenciais, de tal marca, infringindo assim os direitos do proprietário da marca em questão de acordo com a legislação do país de importação" (OMC, 1994, Artigo 51). Nesse contexto, a falsificação pode envolver substituição, como a colocação de produtos inferiores em embalagens autênticas ou reutilizadas; duplicação, que é a cópia direta de embalagens, produtos e/ou instruções; adulteração, que ocorre ao interferir em embalagens ou rótulos e substituir o produto real por itens alterados,

subtraídos ou roubados; além de fraudes relacionadas a devoluções e garantias (ZADBUKE, SHAHI, GULECHA, PADALKAR, & THUBE, 2013).

Segundo a Organização para a Cooperação e Desenvolvimento Econômico (OECD), a pirataria e a falsificação, amplamente definidas como *rebranding* e fabricação não autorizada de produtos sob falsas premissas, geraram prejuízos estimados em quase dois trilhões de dólares em 2020, com a distribuição online de produtos falsificados alcançando cerca de US\$ 323 bilhões anualmente (OECD, 2021). A Organização Mundial da Saúde (OMS) estima que pelo menos 25% de todas as bebidas alcoólicas sejam ilícitas, com algumas regiões atingindo 40%, segundo a Conferência das Nações Unidas sobre Comércio e Desenvolvimento (OMS, 2018).

Exemplos reconhecidos de produtos alimentícios e bebidas falsificados incluem vinho, leite em pó para bebês (BERMAN, 2008) e álcool (JACK, 2018; KUBALLA et al., 2018). Um estudo conduzido por Snowdon e colaboradores (2012) relata que 10% das garrafas ou latas de cerveja vendidas no Reino Unido são falsificadas. No Brasil, o crime organizado obteve uma receita de R\$56 bilhões em 2022 apenas com a fabricação de bebida falsificada, segundo o Fórum Brasileiro de Segurança Pública (FBSP). O valor é maior do que o faturamento da maior cervejaria do país, a Ambev, que no mesmo período, faturou R\$42,6 bilhões, considerando até os produtos não alcoólicos. A estimativa é que 25,7% do mercado brasileiro de bebida está na ilegalidade. O valor usado como referência para o cálculo foi a produção nacional do setor registrada pela Pesquisa Industrial Anual do IBGE (Folha de São Paulo, 2024).

Como consequência, os impactos das bebidas alcoólicas falsificadas vão além da perda de vendas, gerando uma série de efeitos econômicos conhecidos como "efeitos colaterais", que aumentam os prejuízos financeiros. A substituição de produtos legítimos por falsificados reduz as vendas e as receitas das empresas, além de afetar a produção, diminuindo a demanda por

funcionários e limitando a expansão das fábricas legais. Também há uma queda na arrecadação de impostos, como o Imposto sobre Valor Agregado (IVA), imposto de renda e imposto sobre lucros corporativos, uma vez que produtos falsificados não geram tributos, causando evasão fiscal (BRYAN; HILL, 2024).

Além disso, a redução da produção legítima afeta setores relacionados, como o de serviços e equipamentos industriais, que passam a ser menos utilizados. Isso também reduz a arrecadação de contribuições sociais, como previdência e tributos regionais. Por fim, a menor produção formal impacta as receitas de exportação e os impostos sobre consumo, ampliando as perdas para a economia. Esses efeitos mostram que a falsificação de bebidas alcoólicas causa prejuízos econômicos amplos, afetando não apenas as empresas e os consumidores, mas também a arrecadação fiscal e os setores industriais relacionados (BRYAN; HILL, 2024). No controle de qualidade de alimentos e bebidas, a autenticidade está diretamente relacionada à proteção do consumidor.

A autenticação de produtos é definida como o processo de verificação da conformidade de um alimento com a descrição presente em seu rótulo (ENGEL et al., 2012). Espera-se que os produtos estejam de acordo com as informações do rótulo, incluindo origem geográfica, processo de produção e composição, sem adulterações ou discrepâncias em relação aos padrões oficiais ou às declarações do fabricante (SOON; MANNING, 2019). Esse procedimento é essencial tanto para consumidores quanto para produtores, pois a confiança nas informações declaradas nos rótulos é fundamental para garantir a segurança e a credibilidade junto ao público. Muitas vezes, os consumidores pagam valores elevados por itens rotulados como provenientes de regiões específicas. Para verificar autenticidade, origem, qualidade e possíveis fraudes, utiliza-se a técnica de

fingerprinting (impressão digital), que emprega métodos químicos, biomoleculares ou isotópicos (CUADROS-RODRÍGUEZ ET AL., 2016).

Os fabricantes de bebidas falsificadas frequentemente operam impunemente, pois a detecção analítica exige equipamentos caros, reagentes de alto custo e operadores altamente qualificados (BRYAN; HILL, 2024). Além disso, a identificação de falsificações é dificultada pela falta de um método padrão, já que os processos variam, desde a simples substituição de uma bebida de qualidade superior por uma inferior, até a manipulação com produtos químicos industriais, substituição de grãos e aromatizantes. No caso das cervejas, por exemplo, a troca de rótulos é especialmente difícil de ser detectada, uma vez que a composição química das bebidas falsificadas é muito semelhante à das originais, tornando o processo de diferenciação ainda mais desafiador. As amostras podem ser caracterizadas com base em propriedades físico-químicas e químicas, utilizando instrumentos analíticos como espectroscopia (ultravioleta-visível, Raman, infravermelho próximo e médio, fluorescência), ressonância magnética nuclear de ^1H , espectrometria de massa isotópica e técnicas cromatográficas, entre outros. Esses métodos fornecem dados multivariados que enriquecem a análise. Essas variáveis podem ser direcionadas a objetivos específicos, como a determinação da origem ou classificação do produto. Para identificar padrões, tendências ou informações que auxiliem na discriminação e categorização de alimentos, são aplicados métodos matemáticos e estatísticos. (DA COSTA et al., 2021).

Casos de falsificação de cervejas têm ocorrido no Brasil por meio, principalmente, da troca de rótulos e tampas de garrafas de marcas de baixo valor comercial, pelos de cerveja que apresentam maior preço e volume de comércio (PEREIRA, 2016). Essas fraudes dificilmente são identificadas e percebidas pelos consumidores por conta da similaridade das características sensoriais atreladas a cerveja, por conta disso, ainda existe necessidade de desenvolvimento de

novas metodologias para discriminar e classificar cervejas por tipo, estilo e marca (DA SILVA et al., 2019). Crimes relacionados a esse tipo de falsificação já foram reportados, por exemplo, na Baixada Fluminense/RJ (GRUBERTT; TUFANI, 2022), em Ribeirão Preto/SP (G1, 2023) , em Sete Lagoas/MG (FERREIRA, 2023). Na Serra Gaúcha/RS, a polícia apreendeu 1032 garrafas de cervejas falsificadas e 1,2 mil garrafas vazias, que pelo rótulo era possível ver a falsificação (SOUZA, 2023).

Essas situações exemplificam a particularidade da falsificação que o Brasil enfrenta em relação às cervejas industriais. Pereira e colaboradores (2016) utilizaram PCA e PLS-DA para classificar 141 cervejas brasileiras de maior preço (marcas Antartica, Brahma e Skol) e de menor preço (marcas Cintra, Glacial, Crystal, Lokal e A Outra). A análise por PCA não identificou agrupamentos nos dados. No entanto, o modelo PLS-DA obteve uma precisão de 95% ao utilizar o espectro completo. A legislação brasileira trata a falsificação de bebidas alcoólicas como um grave crime contra a saúde pública, prevista no artigo 272 do Código Penal (BRASIL, 1940), penalizando a prática de corromper, adulterar, falsificar ou alterar substâncias ou produtos alimentícios destinados ao consumo, tornando-os nocivos à saúde ou reduzindo-lhes o valor nutritivo, com pena de reclusão de 4 a 8 anos, além de multa.

A avaliação da autenticidade de alimentos envolve diversas atividades essenciais, como a análise de características físico-químicas, a detecção e quantificação de adulterantes ou contaminantes e a verificação de requisitos técnicos de qualidade. Nesse contexto, as ferramentas quimiométricas, especialmente a análise de dados multivariada e os métodos de reconhecimento de padrões, tornaram-se fundamentais para o controle de qualidade e autenticação de alimentos (JIMÉNEZ-CARVELO, 2019; BRERETN, 2018; GRANATO, 2011). Esses métodos são aplicados para extrair informações relevantes e ocultas dos dados analíticos (como espectros,

cromatogramas e imagens), permitindo caracterizar padrões de qualidade e identificar agrupamentos naturais. Além disso, possibilitam classificar amostras em grupos específicos e prever propriedades de interesse de forma rápida e objetiva (BRERETON, 2018).

Diante do exposto, neste trabalho, busca-se aplicar técnicas instrumentais diversas, como espectroscopia no infravermelho, ultravioleta-visível, espectroscopia Raman, cromatografia gasosa e líquida e espectrometria de massas, combinadas a ferramentas quimiométricas, visando a diferenciação de cervejas de diferentes marcas comercializadas no Brasil. Essa abordagem é inserida em um contexto forense, com o objetivo de detectar falsificações de produtos, garantindo assim a autenticidade e a segurança alimentar.

1.3 Análise Instrumental

Medições quantitativas de atributos de qualidade da cerveja envolvem a determinação precisa da quantidade de parâmetros específicos, fundamentais para identificar adulterações, como a adição de substâncias mais baratas, alterações na concentração de ingredientes ou diluições, desafios que englobam aspectos técnicos, logísticos e financeiros (SØRENSEN et al., 2016). Para atender a essas demandas analíticas, diferentes técnicas instrumentais têm sido aplicadas na caracterização de cervejas, destacando-se as técnicas espectroscópicas e cromatográficas. Em uma revisão conduzida por Da Costa e colaboradores (2021), constatou-se que os métodos de espectroscopia empregados na avaliação de cervejas podem ser classificados em espectroscopia vibracional (infravermelho próximo – NIR, infravermelho médio – MIR e Raman, representando 26,9% dos estudos revisados), espectrofotometria na região do ultravioleta-visível (UV-Vis, 9%) e ressonância magnética nuclear (RMN, 7,5%). Essas técnicas são consideradas práticas de química

analítica verde, pois não geram resíduos tóxicos, dispensam solventes e não oferecem riscos significativos ao operador. Os métodos cromatográficos, como a cromatografia gasosa (GC, 11,9%) e a cromatografia líquida (LC, 7,5%), são utilizados para separar os diferentes componentes de uma amostra, podendo ser combinados com técnicas de espectroscopia ou espectrometria para análises mais detalhadas.

No presente estudo, foram selecionadas as técnicas de espectroscopia FTIR, espectrofotometria UV-Vis e Raman para a análise das amostras de cerveja. Cada uma dessas técnicas apresenta vantagens e limitações específicas em relação ao objetivo central de detecção de falsificação. A espectroscopia FTIR destaca-se por sua alta sensibilidade à presença de compostos orgânicos e pela rapidez de análise, permitindo a detecção de grupos funcionais relacionados a açúcares, álcoois e proteínas (ENGEL et al., 2012). Contudo, a sobreposição de bandas espectrais em matrizes complexas pode dificultar a identificação de adulterações sutis sem o auxílio de ferramentas de quimiometria (GRASSI et al., 2014).

A técnica espectrofotometria UV-Vis apresenta como principais vantagens a simplicidade, o baixo custo e a rapidez na execução, sendo útil para a detecção de compostos fenólicos e pigmentos relacionados à coloração da bebida (MAHMOOD; PETRACO; HE, 2012). Entretanto, sua seletividade é limitada em matrizes complexas, como as cervejas, sendo suscetível a interferências de proteínas e açúcares que podem mascarar alterações na composição. Por sua vez, a espectroscopia Raman oferece alta seletividade para grupos funcionais específicos e menor sensibilidade à interferência da água, o que é particularmente relevante em matrizes líquidas como a cerveja (DUARTE et al., 2004). Contudo, pode apresentar interferências de fluorescência de compostos presentes na bebida, o que compromete a qualidade do espectro e exige cuidados adicionais no pré-processamento (IÑÓN; GARRIGUES; GUARDIA, 2006).

Considerando que um perfil analítico obtido por um único método pode ser insuficiente para discriminar produtos alimentícios (BORRÀS et al., 2015), este estudo propõe a combinação de diferentes técnicas espectroscópicas e ferramentas de quimiometria para explorar a complementaridade entre os métodos, aumentar a sensibilidade e a robustez na detecção de possíveis fraudes, e permitir inferências mais robustas sobre a autenticidade das amostras de cerveja.

1.3.1 Analisador de Bebidas

Para a caracterização físico-química das amostras de cerveja analisadas neste trabalho, utilizou-se o equipamento Analisador de Bebidas Alcolyzer Beer, fabricado pela Anton Paar. Este instrumento é amplamente reconhecido no setor cervejeiro por sua precisão, rapidez e praticidade na determinação de parâmetros fundamentais relacionados à qualidade da bebida. Seu funcionamento baseia-se na combinação da espectroscopia no infravermelho próximo (NIR) com a medição de densidade, permitindo a análise simultânea e não destrutiva de diversas propriedades relevantes da cerveja, sem a necessidade de reagentes químicos ou etapas complexas de preparo da amostra. Dentre os parâmetros avaliados pelo equipamento, destaca-se o teor alcoólico, expresso em porcentagem de volume por volume (% v/v), que representa a quantidade de etanol presente na bebida. Este valor é fundamental tanto do ponto de vista sensorial quanto regulatório, sendo um dos principais indicadores de estilo e conformidade da cerveja com as exigências legais e comerciais (ANTON PAAR, 2025).

Outro parâmetro de grande relevância é a densidade final da cerveja, que corresponde à sua massa por unidade de volume. Este valor está diretamente relacionado à concentração de

substâncias dissolvidas, como açúcares residuais e álcoois, influenciando a percepção de corpo da bebida e permitindo inferências sobre o grau de fermentação alcançado (PALMER, 2017). O equipamento também permite a medição do extrato original da cerveja, ou seja, a concentração de compostos dissolvidos no mosto antes do início da fermentação alcoólica, expressa em graus Plato (°P). Este parâmetro reflete o potencial fermentativo da cerveja, sendo indicativo do teor alcoólico que pode ser alcançado e da intensidade sensorial esperada (ANTON PAAR, 2025).

Após a fermentação, o extrato aparente é utilizado para estimar a quantidade de extrato residual, com base na densidade da cerveja, porém sem considerar a influência do álcool. Para uma estimativa mais precisa, calcula-se o extrato real, que corrige esse valor levando em conta a presença do etanol, fornecendo assim uma representação mais fiel da quantidade de substâncias não fermentadas que permanecem na bebida. Com base na relação entre o extrato original e o extrato aparente, determina-se o grau de fermentação aparente (ADF – Apparent Degree of Fermentation), que expressa o percentual do extrato que foi efetivamente convertido em álcool e dióxido de carbono durante o processo fermentativo. Este índice permite avaliar a eficiência da fermentação e o perfil sensorial resultante da cerveja, sendo que valores mais altos indicam bebidas mais secas e com menor doçura residual (PALMER, 2017).

Além disso, o Alcoalyzer Beer também mensura a concentração de dióxido de carbono dissolvido na cerveja, expressa em gramas por litro (g/L), responsável pela carbonatação da bebida. O teor de CO₂ influencia diretamente na formação e estabilidade da espuma, bem como na sensação de refrescância percebida durante o consumo. A utilização do Alcoalyzer Beer neste estudo justifica-se pela sua capacidade de fornecer dados confiáveis, precisos e rápidos sobre parâmetros fundamentais para o controle de qualidade, a padronização dos produtos e a identificação de possíveis fraudes ou desvios em amostras de cerveja. O equipamento oferece uma análise

abrangente do perfil físico-químico das bebidas, permitindo a comparação entre diferentes estilos e a verificação da conformidade das amostras com os padrões estabelecidos. Dessa forma, a aplicação dessa tecnologia contribui significativamente para a robustez dos resultados obtidos e para a credibilidade das conclusões geradas ao longo deste trabalho.

1.3.2 Espectroscopia do Infravermelho com Transformada de Fourier (FTIR)

A espectroscopia de infravermelho com transformada de Fourier (FT-IR), empregando a técnica de refletância total atenuada (ATR), tem sido amplamente utilizada na identificação e diferenciação de marcas de cerveja (DUARTE et al., 2004; LACHENMEIER, 2007; ENGEL et al., 2012; DE ALMEIDA et al., 2018). Essa abordagem apresenta alta seletividade molecular, permitindo a detecção de grupos funcionais característicos de diversos compostos presentes na matriz cervejeira. O uso da espectroscopia de infravermelho médio (MIR), em particular, oferece vantagens em relação ao infravermelho próximo (NIR), frequentemente empregado em análises de alimentos, por facilitar a interpretação espectral devido à presença de bandas mais específicas e intensas (ENGEL et al., 2012).

Estudos demonstram que a técnica FTIR é eficaz na predição da concentração de componentes como etanol, açúcares fermentáveis e extrato seco, sendo particularmente útil no monitoramento de processos de fermentação (LACHENMEIER, 2007). Na região espectral do infravermelho médio, é possível observar bandas características de compostos presentes na cerveja, como os grupos hidroxila (O–H) dos álcoois e da água, geralmente observados na região de 3200 a 3600 cm^{-1} , as bandas de estiramento C–H dos açúcares (em torno de 2800 a 3000 cm^{-1}), bem como bandas associadas às ligações C=O (em torno de 1700 cm^{-1}) de ácidos orgânicos. A presença de maltose e maltotriose, por exemplo, pode ser indicada por bandas na região de 900 a 1200 cm^{-1} ,

atribuídas a vibrações de deformação C–O e C–C do esqueleto carboidrato. Apesar de sua eficácia analítica, a técnica FTIR-ATR, por si só, não é suficiente para distinguir de forma precisa diferentes marcas de cerveja. Isso se deve à complexidade e à similaridade das composições químicas entre amostras comerciais, que frequentemente pertencem a um mesmo estilo ou compartilham ingredientes semelhantes.

Como resultado, os espectros obtidos apresentam alta sobreposição de bandas, dificultando a distinção visual ou direta entre amostras. Para superar essa limitação, a FTIR é geralmente combinada com ferramentas de quimiometria, como a análise de componentes principais (PCA), análise discriminante ou modelos de regressão multivariada, que permitem explorar variações sutis nos dados espectrais e identificar padrões correlacionados a diferenças entre marcas ou estilos. Diversos estudos exploraram essa abordagem integrada. Duarte, Barros, Almeida, Spraul e Gil (2004), por exemplo, conseguiram diferenciar cinquenta amostras de cerveja (ale, lager e sem álcool) com base no teor alcoólico, utilizando FTIR-ATR acoplado à análise de componentes principais. Iñón, Garrigues e Guardia (2006) empregaram a técnica para estimar os extratos original e real e o teor de etanol em diferentes cervejas, enquanto Grassi et al. (2014a) monitoraram a evolução dos açúcares fermentáveis e do etanol ao longo de seis processos distintos de fermentação, com base nos compostos maltose, maltotriose e etanol.

De Almeida et al. (2018), por sua vez, aplicaram a técnica FTIR-ATR no monitoramento do processo de mosturação, etapa fundamental na produção cervejeira. Assim, a espectroscopia FTIR-ATR representa uma ferramenta poderosa e versátil para a caracterização de cervejas, desde que combinada com métodos estatísticos que ampliem sua capacidade de discriminação e interpretação de dados complexos.

1.3.3 Espectroscopia Raman

A espectroscopia Raman é uma técnica vibracional baseada na interação da radiação eletromagnética com a matéria, permitindo identificar estruturas e ligações químicas por meio das vibrações moleculares. O fenômeno ocorre pelo espalhamento inelástico da luz: quando um feixe monocromático incide sobre uma amostra, a maior parte da radiação é espalhada elasticamente (espalhamento Rayleigh), enquanto uma pequena fração sofre variação de energia, originando o espalhamento Raman. Essa diferença entre a radiação incidente e a espalhada corresponde às transições vibracionais das moléculas e é expressa em número de onda (cm^{-1}). O espalhamento Stokes, mais intenso, é geralmente utilizado nas análises (CLARK; LONG, 1977).

Para que uma vibração seja ativa em Raman, é necessária variação na polarizabilidade molecular durante a vibração, condição complementar à da espectroscopia no infravermelho, que depende da variação do momento de dipolo elétrico. Assim, as duas técnicas são consideradas complementares, permitindo uma caracterização vibracional mais completa das moléculas quando aplicadas de forma conjunta (CLARK; LONG, 1977).

A escolha do laser influencia diretamente a intensidade do sinal e a fluorescência: comprimentos de onda no visível (532 nm) geram sinais mais fortes, enquanto os do infravermelho próximo (785 ou 1064 nm) reduzem interferências, sendo mais adequados para amostras orgânicas. Avanços tecnológicos recentes possibilitaram o desenvolvimento de sistemas portáteis e acoplados a fibras ópticas, que permitem medições *in situ*, rápidas e não destrutivas em diversos tipos de amostras (CLARK; LONG, 1977).

Os espectros Raman apresentam bandas características entre 4000 e 400 cm^{-1} , associadas a vibrações de grupos funcionais como C–H, O–H, N–H e C=O. A análise das posições, intensidades

e formas das bandas fornece informações sobre composição, ambiente químico e estrutura molecular (SMITH; DENT, 2019).

Dessa forma, a espectroscopia Raman consolida-se como uma ferramenta analítica de grande valor, unindo praticidade, seletividade e caráter não destrutivo na caracterização de matrizes complexas. A utilização dos espectrômetros TacticID e Resolve, com diferentes faixas de laser, potencializou essas vantagens ao permitir análises rápidas, precisas e adaptadas às diferentes naturezas das amostras. O TacticID, com laser de 785 nm, pode ser eficiente na detecção de compostos orgânicos como álcoois, açúcares e fenólicos (METROHM, 2025), enquanto o Resolve, que utiliza laser de 830 nm, possibilita medições através de embalagens, ampliando a capacidade de identificação de constituintes minoritários e de possíveis adulterantes sem contato direto com o produto (AGILENT TECHNOLOGIES, 2025).

Assim, a combinação dos dois equipamentos reforça o potencial da espectroscopia Raman como método robusto para a autenticação e rastreabilidade de bebidas alcoólicas, podendo diferenciar amostras a partir de suas assinaturas espectrais e garantindo maior confiabilidade na detecção de falsificações e no controle de qualidade de produtos comerciais.

1.3.4 Espectrofotometria no Ultravioleta-Visível (UV-Vis)

A espectrofotometria na região do ultravioleta-visível (UV-Vis) é uma técnica amplamente empregada na análise química quantitativa e qualitativa, fundamentada na absorção de radiação eletromagnética por espécies químicas que apresentam transições eletrônicas em seus átomos ou moléculas (PERKAMPUS, 2013). Quando a luz incide sobre uma substância, parte da radiação é absorvida, e a intensidade dessa absorção é relacionada à concentração do analito por meio da Lei de Beer-Lambert, expressa pela equação 1:

$$A = \varepsilon \cdot b \cdot c \quad (1)$$

em que A é a absorvância, ε é o coeficiente de absorvidade molar, b é o caminho óptico da célula e c é a concentração da espécie absorvente. Essa relação linear permite a determinação precisa de concentrações em solução, desde que sejam respeitadas condições experimentais adequadas de linearidade e ausência de interferências (PERKAMPUS, 2013).

A espectrofotometria UV-Vis baseia-se em transições eletrônicas que ocorrem quando elétrons são promovidos de orbitais de menor energia para orbitais de maior energia ao absorver radiação na faixa de 190 a 800 nm. As regiões do ultravioleta (190–400 nm) e do visível (400–800 nm) correspondem, respectivamente, às transições de grupos cromóforos, como ligações duplas conjugadas, sistemas aromáticos e grupos carbonílicos, e à absorção responsável pela coloração de substâncias. A posição e a intensidade das bandas de absorção fornecem informações sobre a estrutura molecular, o grau de conjugação e interações intermoleculares, permitindo distinguir variações sutis na composição química das amostras (PERKAMPUS, 2013).

Na indústria cervejeira, essa técnica é amplamente utilizada para a determinação de parâmetros de qualidade, como a cor (EBC) e o amargor (IBU). O valor de EBC (European Brewery Convention) é determinado pela absorvância medida a 430 nm, refletindo a intensidade da coloração da cerveja e a presença de compostos cromóforos, como melanoidinas e pigmentos formados durante o processo de malteação e fervura do mosto. Já o IBU (International Bitterness Units) é obtido pela absorvância a 275 nm, associada às isohumulonas, compostos amargos derivados dos alfa-ácidos do lúpulo (COELHO et al., 2022).

Em matrizes complexas, como alimentos e bebidas, a sobreposição de bandas de diferentes constituintes pode dificultar a interpretação direta dos espectros. Nesse contexto, a aplicação de métodos quimiométricos, como a Análise de Componentes Principais (PCA) e a Regressão por

Mínimos Quadrados Parciais (PLS), torna-se fundamental para extrair informações latentes e realizar classificações ou calibrações multivariadas com elevada robustez. Assim, a espectrofotometria UV-Vis destaca-se como uma técnica versátil, rápida e acessível, amplamente utilizada para a caracterização de compostos orgânicos e inorgânicos, o monitoramento de processos industriais e a avaliação de autenticidade e qualidade em sistemas complexos, como cervejas e outras bebidas comerciais (CONTRERAS et al., 2010).

1.4 Pré-Processamento de Dados

O pré-processamento de dados é uma etapa fundamental na análise de dados multivariados, especialmente em contextos que envolvem técnicas espectroscópicas (LEE; LIONG; JEMAIN, 2017). Trata-se de um conjunto de transformações matemáticas e estatísticas aplicadas aos dados brutos com o objetivo de minimizar interferências indesejadas e realçar as informações químicas relevantes. Em espectros obtidos por técnicas como FTIR, espectrofotometria UV-VIS ou Raman, é comum a presença de ruídos, variações de baseline, deslocamentos de pico e efeitos de espalhamento de luz. Essas variabilidades, embora não relacionadas às características químicas propriamente ditas das amostras, podem comprometer significativamente a qualidade e a interpretabilidade dos modelos quimiométricos (BARNES et al., 1989).

A aplicação do pré-processamento visa, portanto, corrigir essas distorções e tornar os dados mais comparáveis entre si, enfatizando as variações relacionadas às propriedades químicas ou físico-químicas das amostras. Entre os métodos mais comuns de pré-processamento, destacam-se a correção de espalhamento multiplicativo (MSC), a normalização por vetor normal (SNV) e a suavização ou derivação de espectros, como o filtro de Savitzky-Golay. A escolha adequada da técnica depende das características específicas dos dados e dos objetivos da análise, devendo ser

feita de forma criteriosa para evitar distorções ou a remoção indevida de informações úteis (MISHRA et al., 2020).

Do ponto de vista da quimiometria, o pré-processamento é essencial para a construção de modelos robustos e confiáveis. Técnicas como a Análise de Componentes Principais (PCA) e a Análise por Mínimos Quadrados Parciais com Discriminação (PLS-DA) são altamente sensíveis a variações sistemáticas que não estejam relacionadas à composição química das amostras (SOUZA; POPPI, 2012). Sem um pré-processamento adequado, essas variações podem mascarar os padrões reais presentes nos dados, dificultando a visualização de agrupamentos e a construção de modelos preditivos precisos (BARNES et al., 1989).

Dessa forma, o uso de técnicas de pré-processamento contribui para a melhora da qualidade dos dados de entrada, aumentando a capacidade discriminante e a acurácia dos modelos estatísticos aplicados posteriormente. A técnica Standard Normal Variate (SNV) é amplamente utilizada para corrigir efeitos de espalhamento de luz e variações de espessura ou concentração da amostra, que afetam a linha de base dos espectros (BARNES et al., 1989). O método atua individualmente sobre cada espectro, centralizando-o em torno de sua média e normalizando-o pela sua variância. Em termos matemáticos, cada ponto espectral x de um espectro X é transformado conforme a seguinte equação 2:

$$xi(SNV) = \frac{(xi - \bar{x})}{s} \quad (2)$$

em que x_i representa a intensidade de cada ponto espectral, \bar{x} é a média do espectro e s é o desvio padrão dele. Esse procedimento elimina variações de offset e de escala entre os espectros, preservando as características relacionáveis à composição química. A aplicação do SNV é particularmente recomendada quando os dados apresentam variações de dispersão relacionadas a fenômenos físicos e não à composição química das amostras (MISHRA et al., 2020). A correção

de espalhamento multiplicativo (MSC) é outra técnica eficaz para corrigir variações sistemáticas causadas por efeitos físicos, como espalhamento de luz, tamanho de partícula e diferenças de caminho óptico, que não trazem informações químicas, mas afetam a linha de base dos espectros. Diferentemente do SNV, o MSC utiliza um espectro de referência, geralmente o espectro médio do conjunto de dados, e ajusta cada espectro individual por meio de uma regressão linear simples, conforme a equação 3:

$$x_i = a + b \cdot x_{ref} + \varepsilon \quad (3)$$

onde a e b são os coeficientes da regressão entre o espectro da amostra e o espectro de referência, e ε é o erro residual. Em seguida, cada espectro é corrigido subtraindo o offset a e dividindo pela inclinação b . Isso permite alinhar os espectros na mesma escala de intensidade, tornando-os mais comparáveis. A principal vantagem do MSC é sua capacidade de corrigir deslocamentos e inclinações nos espectros, sendo especialmente útil quando há variações significativas entre as amostras por efeitos físicos não controlados (SOUZA; POPPI, 2012). Outro método bastante utilizado é a derivação e suavização dos espectros com o filtro de Savitzky-Golay, que aplica um polinômio de baixa ordem em uma janela móvel sobre os dados. Essa técnica permite, ao mesmo tempo, suavizar o espectro, reduzindo ruídos de alta frequência, e realçar bandas espectrais sobrepostas.

A suavização simples é útil para reduzir o ruído sem distorcer os sinais, enquanto a aplicação da primeira derivada destaca a taxa de variação do sinal e ajuda a eliminar deslocamentos de linha de base. A segunda derivada, por sua vez, destaca inflexões no sinal, permitindo separar picos muito próximos (MISHRA et al., 2020). A escolha da ordem do polinômio e do tamanho da janela de suavização deve ser feita com cautela, pois parâmetros inadequados podem introduzir artefatos ou aumentar o ruído (SOUZA; POPPI, 2012). Considerando-se que a análise química de

alimentos e bebidas, como as cervejas, envolve diferentes técnicas espectroscópicas, incluindo espectroscopia no infravermelho médio com transformada de Fourier (FTIR) e espectroscopia na região do ultravioleta-visível (UV-Vis), torna-se necessária a utilização de métodos de pré-processamento.

Cada técnica apresenta características específicas quanto à forma de interação com a amostra e susceptibilidade a interferências físicas, como variações no comprimento do caminho óptico, intensidade da luz incidente e condições de contato. Tais variações podem introduzir ruídos e distorções que mascaram informações químicas importantes para a correta identificação e análise das amostras (JIAO et al., 2020). Portanto, o uso combinado de técnicas de pré-processamento como SNV, MSC e Savitzky-Golay é fundamental para garantir a comparabilidade dos dados gerados por diferentes instrumentos, aumentar a robustez e a acurácia das análises quimiométricas, como a PCA e a PLS-DA, e, assim, possibilitar a detecção de adulterações, bem como a diferenciação por marca ou origem da bebida (DA COSTA et al., 2021).

1.5 Quimiometria

O termo quimiometria refere-se à aplicação de métodos matemáticos e estatísticos para extrair informações de dados químicos (VARMUZA E FILZMOSER, 2016). Esse campo tem três objetivos principais: descrição, predição de dados e planejamento experimental (ESTEKI et al., 2018b). Nesse cenário, a quimiometria é definida, ainda, pela IUPAC como a ciência que relaciona medições químicas com o estado do sistema por meio de métodos matemáticos e estatísticos (HIBBERT, 2016).

Na quimiometria, diversos algoritmos são usados para descrever e prever dados, incluindo técnicas de pré-processamento (como seleção e extração de variáveis, redução de

dimensionalidade) e modelagem (agrupamento, classificação e regressão). Muitos desses algoritmos, originados na ciência da computação e inteligência artificial, são amplamente aplicados em data mining e aprendizado de máquina. Em estudos de análise de alimentos, geralmente começa-se com uma análise exploratória de dados, utilizando técnicas não supervisionadas, como a análise de componentes principais (PCA) e a análise de agrupamento hierárquico (HCA). Essas ferramentas identificam padrões ocultos e similaridades entre amostras. (DA COSTA ET AL., 2021).

Modelos de regressão, como mínimos quadrados parciais (PLS), regressão de componentes principais (PCR) e regressão linear múltipla (MLR), são amplamente utilizados para calibração e predição, relacionando características químicas ou físicas das amostras a sinais instrumentais de menor custo (BRERETON et al., 2018). Já na classificação, o objetivo é associar cada amostra a uma classe predefinida, utilizando algoritmos discriminantes, como análise discriminante linear (LDA), análise discriminante quadrática (QDA), *k-nearest neighbors* (kNN) e máquinas de vetor de suporte (SVM), ou algoritmos de modelagem de classes, como SIMCA e UNEQ, que avaliam a compatibilidade das amostras com uma classe específica (OLIVERI E DOWNEY, 2012).

Os métodos de reconhecimento de padrões podem ser divididos em dois grandes grupos: os não supervisionados (como análise de componentes principais (PCA) e análise de agrupamentos hierárquicos, HCA) e os supervisionados (como k-vizinhos mais próximos, kNN, análise discriminante por mínimos quadrados parciais, PLS-DA, e modelagem independente por analogia de classes, SIMCA) (BALLABIO; CONSONNI, 2013). Com o aumento do volume de dados analíticos, especialmente em estudos de alimentos, a automação e a informatização tornaram indispensável o uso de ferramentas de análise avançada (PARASTAR; TAULER, 2018). Na autenticação de alimentos, alguns autores sugerem priorizar métodos de modelagem de classes,

como SIMCA, que delimitam regiões de aceitação para cada classe e garantem maior confiabilidade (RODIONOVA et al., 2016).

Além disso, métodos baseados em árvores de decisão (CART e *Random Forest*) são alternativas úteis para dividir dados em subconjuntos hierárquicos, baseando-se em variáveis de entrada (Ai et al., 2014). Diferentemente das práticas convencionais, a combinação de múltiplos métodos para comparação de classes e construção de modelos preditivos pode revelar informações adicionais, indicando a necessidade de utilizar mais de uma abordagem estatística (JIMÉNEZ-CARVELO et al., 2019). Na prática, a análise de dados de alimentos começa com uma exploração inicial, usando PCA e HCA, para revelar padrões e similaridades ocultas (DA COSTA et al., 2021). Posteriormente, técnicas de calibração, como PLS, PCR e MLR, relacionam sinais analíticos a características físico-químicas de interesse (BRERETON et al., 2018).

Para classificação, utiliza-se uma ampla gama de algoritmos discriminantes (LDA, QDA, kNN, SVM) ou de modelagem de classes (SIMCA, UNEQ) (OLIVERI & DOWNEY, 2012). Com o crescimento da complexidade dos dados, métodos de mineração de dados e aprendizado de máquina, como árvores de decisão (CART, Random Forest), SVM e ANN, têm sido cada vez mais aplicados, proporcionando alta eficiência em problemas complexos (JIMÉNEZ-CARVELO et al., 2019).

Neste estudo, buscou-se investigar se as características físico-químicas são suficientes para distinguir cervejas industriais de diferentes marcas. A comprovação dessa hipótese permitirá utilizar essas informações como ferramenta para combater falsificações e adulterações, além de aprimorar processos e controle de qualidade, com potencial aplicação em laboratórios forenses e de perícia. Para isso, diferentes técnicas instrumentais, como espectroscopia no infravermelho, ultravioleta-visível, espectroscopia Raman, cromatografia gasosa e líquida, e espectrometria de

massas, foram combinadas a métodos quimiométricos (PCA, LDA e PLS-DA) para analisar e interpretar os dados obtidos, contribuindo para o desenvolvimento de estratégias analíticas confiáveis.

1.5.1 Análise por Componentes Principais (PCA)

A análise de componentes principais (*Principal Component Analysis, PCA*) é uma das ferramentas mais importantes na Quimiometria, servindo como base para métodos de reconhecimento de padrões, classificação e calibração multivariada (SOUZA; POPPI, 2012).

Essa técnica permite reduzir a dimensionalidade dos dados ao projetar o conjunto de variáveis em um novo sistema de eixos chamados componentes principais (PCs), possibilitando visualizar dados complexos em poucas dimensões. No espaço original, as amostras ocupam posições em um espaço n-dimensional, onde n é o número de variáveis; com a PCA, essas amostras passam a ser representadas em espaços reduzidos de duas ou três dimensões definidos pelas PCs. Matematicamente, a matriz de dados X é decomposta em duas matrizes, escores (T) e pesos (P), e uma matriz de resíduos (E), conforme a equação 4:

$$X = T P^t + E \quad (4) \text{ (JOLLIFE, 2002).}$$

Os escores representam a projeção das amostras no novo sistema de coordenadas, enquanto os pesos indicam o quanto cada variável original contribui para cada componente principal. A primeira componente principal (PC1) é construída na direção de maior variação nos dados, enquanto a segunda (PC2) é ortogonal à primeira e captura a maior parte da variação remanescente, e assim por diante. A quantidade de componentes a ser retida no modelo depende da porcentagem

de variância explicada, sendo comum selecionar o número de PCs que explique a maior parte da variabilidade presente nos dados originais (HAIR et al., 2009). Para o cálculo da PCA, diferentes algoritmos são utilizados, como Nonlinear Iterative Partial Least Squares (NIPALS), Singular Value Decomposition (SVD), Power Iteration e Eigenvalue Decomposition (EVD) (BRERETON, 2018).

Embora a PCA seja uma técnica robusta quanto à normalidade multivariada (SHAW, 2003) e, por ser exploratória, não envolva testes de significância estatística, há preocupações sobre seu uso excessivo, aplicação inconsistente e falta de compreensão de suas limitações e requisitos de dados (WEBSTER, 2001). Na prática, cabe ao usuário garantir que os dados sejam adequados para a aplicação da PCA e que o pré-tratamento necessário tenha sido realizado, o que nem sempre é padronizado, apesar de sua grande importância (ZITKO, 1994). Para garantir a eficiência do modelo, é fundamental que uma parcela significativa da variância do conjunto de dados seja explicada, geralmente 70% ou mais, pelos primeiros componentes principais, que frequentemente se limitam a dois ou três. (GRANATO et al., 2011).

Na análise de cervejas, Alcázar e colaboradores (2012a) classificaram três tipos de cervejas claras (lager, pilsner e de baixo teor alcoólico) utilizando PCA e LDA, com precisão de 92,7%. Rodrigo e colaboradores (2017) aplicaram ICP-MS para classificar sete estilos de cerveja com base no conteúdo mineral, obtendo 81% de precisão com LDA. Estudos semelhantes demonstraram que modelos como redes neurais artificiais (ANN) e PLS-DA podem alcançar até 100% de precisão, como nas análises de espectroscopia no infravermelho médio (MIR) e RMN para classificação de estilos de cerveja (GORDON et al., 2018; JEONG et al., 2017).

Além disso, em uma revisão da literatura realizada por Da Costa e colaboradores (2021), os autores concluíram que diversos métodos de quimiometria e aprendizado de máquina podem ser

aplicados à análise de alimentos. Contudo, no caso da análise de cervejas, nesse mesmo estudo, o método mais utilizado foi a Análise de Componentes Principais (PCA), presente em 56,7% dos estudos revisados, sendo aplicada como ferramenta exploratória combinada com algoritmos de classificação (em 30 artigos) ou de forma isolada (em 8 artigos) (DA COSTA et al., 2021). A segunda técnica mais empregada foi a Análise Discriminante Linear (LDA), utilizada em 37,3% dos artigos, seguida por PLS-DA (23,8%), PLS (22,3%), Redes Neurais Artificiais (ANN, 17,9%), SIMCA (14,92%), KNN (11,9%), SVM (7,4%) e outros algoritmos, cada um utilizado em um único artigo. As porcentagens não somam 100%, pois alguns estudos aplicaram mais de uma técnica de classificação. Entre os modelos de regressão utilizados, a regressão por mínimos quadrados parciais (PLS) é a mais comum, sendo aplicada em 93% dos estudos analisados (DA COSTA et al., 2021).

1.5.2 Análise Discriminante por Quadrados Parciais (PLS-DA)

O *Partial Least Squares – Discriminant Analysis* (PLS-DA) é uma técnica multivariada de redução de dimensionalidade amplamente utilizada na quimiometria há mais de duas décadas. Esse método é especialmente útil para lidar com conjuntos de dados complexos, que costumam apresentar elevado número de variáveis, presença de ruído, valores ausentes, grande volume de informações e, frequentemente, uma quantidade de amostras bem menor do que o número de variáveis disponíveis (RUIZ-PEREZ et al., 2020). Nesse contexto, o PLS-DA surge como uma ferramenta robusta para enfrentar tais desafios.

O PLS-DA pode ser entendido como uma versão supervisionada da Análise de Componentes Principais (PCA). Embora ambas as abordagens tenham como objetivo central a redução da dimensionalidade, a diferença fundamental está no fato de que o PLS-DA utiliza

explicitamente a informação de classe, enquanto a PCA é uma técnica não supervisionada. Em termos práticos, isso significa que, enquanto a PCA projeta os dados em componentes que preservam a máxima variância, o PLS-DA constrói componentes capazes de maximizar a covariância entre os preditores e os rótulos de classe (RUIZ-PEREZ et al., 2020).

Dessa forma, o PLS-DA não apenas resume as informações contidas nas variáveis explicativas, mas também organiza esse resumo de modo a otimizar a separação entre grupos. Sob uma perspectiva matemática, tanto PCA quanto PLS-DA podem ser descritos como processos iterativos de transformação linear da matriz de dados, nos quais o objetivo é projetar vetores de alta dimensão em um espaço reduzido, minimizando o erro dessa transformação. No PLS-DA, a primeira componente é definida de forma a maximizar a covariância entre as variáveis explicativas e as classes, podendo, portanto, apontar direções distintas daquelas obtidas pela PCA.

Apesar de sua relevância, o uso do PLS-DA requer cautela. A presença de variância explicada elevada ou correlação aparente nas primeiras componentes não garante, por si só, que o modelo seja robusto. Existe risco de sobreajuste, já que o método pode identificar separações artificiais nos dados (RUIZ-PEREZ et al., 2020). Por essa razão, a validação cruzada é indispensável, seja o objetivo do modelo a seleção de variáveis, a classificação ou simplesmente a visualização dos dados. É a validação que assegura a robustez do modelo e permite diferenciar padrões reais de ajustes artificiais (SZYMAŃSKA et al., 2012).

O objetivo deste estudo foi avaliar se as características físico-químicas são suficientes para distinguir cervejas industriais de diferentes marcas. Caso demonstrada sua eficácia, a obtenção desses parâmetros pode se tornar uma ferramenta útil para o combate à falsificação e adulteração, assim como desenvolvimento de processos e de controle de qualidade, com potencial aplicação em

campo ou em laboratórios de perícia. Esse conhecimento pode auxiliar peritos e policiais de diversas instituições a identificar uma cerveja falsificada e/ou adulterada.

1.5.3 Validação Cruzada

A validação cruzada é uma técnica estatística amplamente empregada para avaliar o desempenho e a capacidade de generalização de modelos preditivos. Seu objetivo é estimar como o modelo se comporta diante de dados não vistos, evitando o sobreajuste (*overfitting*), situação em que o modelo se ajusta excessivamente às particularidades do conjunto de treinamento e perde poder de previsão. Em estudos quimiométricos, essa etapa é essencial, pois assegura que os modelos desenvolvidos sejam robustos e representativos da variabilidade real das amostras (STONE, 1978).

Entre os diferentes esquemas de validação, o 10-fold *cross-validation* e o *Leave-One-Out Cross Validation* (LOOCV) estão entre os mais utilizados (BERRAR, 2019). No primeiro, o conjunto de dados é dividido em dez partes de tamanhos semelhantes, e o modelo é iterativamente treinado em nove delas e testado na décima, até que todas as partições tenham sido utilizadas. Já no LOOCV, apenas uma amostra é deixada de fora em cada iteração, sendo usada como teste enquanto as demais compõem o conjunto de calibração. O LOOCV aproveita ao máximo o número de amostras disponíveis, sendo ideal para bases pequenas, enquanto o 10-fold oferece menor variância e maior eficiência computacional, adequando-se melhor a conjuntos de dados extensos. A escolha entre eles depende, portanto, do tamanho da amostra e do equilíbrio desejado entre precisão e custo computacional (BERRAR, 2019).

No particionamento dos dados, métodos como o Kennard–Stone são frequentemente empregados para garantir que os subconjuntos de calibração e validação representem

adequadamente a variabilidade multivariada do conjunto total. Essa abordagem é particularmente relevante em estudos com número limitado de amostras, como ocorre com frequência em análises de bebidas, reduzindo vieses e aumentando a representatividade dos modelos (BROWNE, 2000).

Entre as métricas associadas à validação cruzada, destaca-se o PRESS (Predicted Residual Error Sum of Squares), que quantifica o erro de predição ao somar os quadrados dos resíduos obtidos quando cada amostra é prevista por um modelo ajustado sem a sua participação. Valores menores de PRESS indicam melhor desempenho, embora o aumento excessivo do número de variáveis latentes (VLs) possa levar a ganhos marginais e ao risco de sobreajuste. Para complementar essa análise, o teste de Van der Voet verifica se as diferenças entre modelos com diferentes números de VLs são estatisticamente significativas, permitindo a seleção do modelo mais simples sem perda relevante de desempenho, em consonância com o princípio da parcimônia (BROWNE, 2000).

A adoção de estratégias de validação estratificada mostra-se essencial neste estudo, uma vez que o número de garrafas analisadas varia entre as diferentes marcas de cerveja. Essa diferença no tamanho das classes pode afetar diretamente a representatividade dos subconjuntos de calibração e validação, influenciando o desempenho e a confiabilidade dos modelos de classificação (DIAMANTIDIS; KARLIS; GIAKOUMAKIS, 2000). A estratificação garante que a proporção de amostras de cada marca seja preservada em todas as partições, proporcionando uma avaliação mais equilibrada e realista do modelo.

A performance dos modelos é avaliada por meio de métricas derivadas da matriz de confusão, como acurácia, sensibilidade, especificidade e área sob a curva ROC (AUC), que permitem mensurar a capacidade discriminatória e a consistência das predições (NEUNHOEFFER; STERNBERG, 2019). Embora a validação cruzada forneça estimativas internas robustas, a

validação externa, realizada com um conjunto independente de amostras, continua sendo o padrão-ouro para confirmar a capacidade de generalização. Quando inviável devido ao número limitado de amostras, a combinação entre validação cruzada estratificada, métricas de desempenho e testes estatísticos (como PRESS, Van der Voet e VIP) constitui uma alternativa sólida e confiável. Essa abordagem, aliada ao princípio da parcimônia, assegura a construção de modelos mais simples, interpretáveis e estatisticamente robustos, fortalecendo a aplicabilidade prática e a credibilidade das inferências em análises quimiométricas de bebidas.

1.5.4 Análise Discriminante Linear (LDA)

A Análise Discriminante Linear (LDA, *Linear Discriminant Analysis*) é uma técnica estatística multivariada amplamente utilizada para classificação e redução de dimensionalidade, cujo objetivo principal é encontrar combinações lineares das variáveis originais que maximizem a separação entre grupos previamente definidos (MARTINEZ; KAK, 2001). Desenvolvida em 1936, a LDA baseia-se no princípio de que, ao projetar os dados em um espaço de menor dimensão, é possível preservar a variabilidade entre as classes e minimizar a variabilidade dentro de cada classe, obtendo assim uma representação mais discriminante e interpretável dos padrões presentes nos dados (OLIVERI & DOWNEY, 2012).

Matematicamente, a LDA busca determinar vetores discriminantes que maximizem a razão entre a variância entre as classes (*between-class variance*) e a variância dentro das classes (*within-class variance*). Em termos práticos, o método constrói uma ou mais funções discriminantes lineares, combinando as variáveis originais com pesos específicos atribuídos de modo a otimizar essa separação (MARTINEZ; KAK, 2001).

Durante o processo de classificação, cada amostra é projetada nas funções discriminantes obtidas, e sua distância em relação ao centro de cada grupo é avaliada. A amostra é então atribuída à classe cujo centroide apresenta a menor distância ou a maior probabilidade de pertencimento, considerando a hipótese de que os grupos seguem uma distribuição normal multivariada e possuem matrizes de covariância iguais. Embora essa suposição seja idealizada, a LDA mostra-se robusta mesmo em condições moderadamente desviantes, o que justifica sua ampla aplicação em contextos experimentais (GIANSANTE; DI VINCENZO; BIANCHI, 2003).

No campo da quimiometria, a LDA é empregada para discriminar e classificar amostras com base em perfis espectrais, físico-químicos ou composicionais, sendo útil em estudos de autenticidade, rastreabilidade e controle de qualidade. Em particular, quando associada a técnicas espectroscópicas, como FTIR, Raman ou UV-Vis, a LDA permite identificar padrões característicos de diferentes grupos de amostras, como marcas de cervejas, variedades de grãos ou origens geográficas. Além disso, a combinação da LDA com etapas prévias de pré-processamento e redução de dimensionalidade (como PCA ou seleção de variáveis) contribui para eliminar redundâncias, melhorar a robustez dos modelos e reduzir o risco de sobreajuste (OLIVERI & DOWNEY, 2012).

A principal vantagem da LDA reside em sua interpretação direta e baixo custo computacional, o que a torna eficaz para modelos preditivos simples e transparentes (OLIVERI & DOWNEY, 2012). No entanto, seu desempenho depende da adequação das premissas estatísticas e da qualidade dos dados utilizados. Quando corretamente aplicada, a LDA proporciona classificações consistentes e facilmente interpretáveis, sendo uma das abordagens mais consolidadas para análise discriminante em sistemas químicos e espectroscópicos complexos.

2 OBJETIVOS

2.1 Objetivo geral

Investigar a aplicabilidade de métodos quimiométricos combinados com técnicas analíticas instrumentais para caracterizar amostras de cervejas industrializadas nacionais, com o intuito de identificar possíveis fraudes, como falsificações e substituições de rótulos, por meio da análise de parâmetros físico-químicos e espectrais.

2.2 Objetivos específicos

- Realizar análises exploratórias das amostras de cerveja utilizando diferentes técnicas analíticas, incluindo espectroscopia no infravermelho com transformada de Fourier (FTIR), espectroscopia no ultravioleta-visível (UV-Vis), espectroscopia Raman e analisador de bebidas (Alcolyzer Beer).
- Obter perfis espectrais (fingerprints) das diferentes marcas de cerveja, visando à construção de uma base de dados comparativa;
- Determinar parâmetros físico-químicos relevantes, como pH, cor, teor de dióxido de carbono (carbonatação), teor alcoólico, grau de fermentação aparente, presença de antioxidantes, subprodutos da fermentação e densidade;
- Aplicar etapas de pré-processamento de dados, como normalização, correções espectrais e filtragem de ruídos, de forma a melhorar a qualidade e a comparabilidade das informações obtidas.
- Empregar técnicas quimiométricas, como Análise de Componentes Principais (PCA), Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA) e Análise Discriminante Linear (LDA) para verificar a capacidade de discriminação entre marcas e identificar padrões associados a possíveis falsificações.

- Implementar procedimentos de validação cruzada a fim de garantir a robustez e a confiabilidade dos modelos quimiométricos construídos, evitando problemas de sobreajuste e avaliando seu desempenho preditivo.
- Avaliar o potencial das metodologias propostas como ferramentas de triagem analítica voltadas ao controle de qualidade e à separação de cervejas de marcas diferentes.

3 MATERIAIS E MÉTODOS

3.1 Obtenção e Armazenamento das Amostras

Foram analisadas, no laboratório de Química e Física Forense da Polícia Civil do Distrito Federal (LQFF/PCDF), 46 amostras de cervejas industrializadas nacionais, incluindo marcas de grande circulação no mercado brasileiro. As amostras foram adquiridas de forma a contemplar, sempre que possível, diferentes lotes de produção, garantindo maior representatividade dos dados. Para cada marca analisada, foram selecionadas no mínimo três unidades, respeitando o critério de variabilidade de lotes. Enquanto mantidas lacradas, as amostras foram armazenadas à temperatura ambiente, fora da incidência direta de luz solar e em posição vertical, a fim de preservar suas características originais até o momento da análise. Antes das medições instrumentais, realizou-se a degaseificação em banho ultrassônico por 30 minutos, visando à remoção do dióxido de carbono.

Após a realização de todas as etapas analíticas, o volume remanescente das amostras foi acondicionado em seus próprios recipientes de vidro (casco original) e armazenado em congelador, garantindo a conservação das propriedades físico-químicas e permitindo a possibilidade de reanálises, caso necessário.

Tabela 1. Informações contidas nos rótulos de cada marca analisada.

Marca	Distribuidora	Fábrica	Ingredientes	teor alcoólico
Brahma Chopp	Ambev	Ambev - Jaguariúna	Água, malte, milho e lúpulo	4,8%
Antártica	Ambev	Ambev - Jaguariúna	Água, malte, milho e lúpulo	4,9%
Skol	Ambev	Ambev - Jaguariúna	Água, malte, milho e lúpulo	4,7%
Original	Ambev	Ambev - Jaguariúna	Água, malte, milho e lúpulo	5,0%
Stela Artois	Ambev	Ambev - Jaguariúna	Água, malte e lúpulo	5,0%
Spaten	Ambev	Ambev - Jaguariúna	Água, malte e lúpulo	5,2%

Império	Cidade Imperial	Cervejaria Cidade Imperial - Petrópolis	Água, malte e lúpulo	4,5%
Petra	Grupo Petrópolis	Cervejaria Petrópolis - Teresópolis	Água, malte e lúpulo	4,8%
Heineken	Heineken	Heineken - Jacareí	Água, malte e lúpulo	5,0%
Amstel	Heineken	Heineken - Jacareí	Água, malte e lúpulo	4,6%

3.2 Aquisição dos dados espectroscópicos

A metodologia empregada para a identificação de cervejas falsificadas foi baseada na aplicação de diferentes técnicas analíticas, com o objetivo de caracterizar detalhadamente as amostras quanto aos seus parâmetros físico-químicos. Inicialmente, realizou-se a análise espectroscópica por espectroscopia Raman, utilizando o equipamento Resolve™ (Agilent Technologies), que opera com tecnologia de imageamento por espectroscopia hiperespectral Raman, sendo especialmente projetado para análises não invasivas e superficiais por meio de janelas ópticas em materiais. O Raman Resolve é otimizado para realizar análises diretamente através do recipiente de acondicionamento, o que, neste estudo, permitiu a obtenção dos espectros sem a necessidade de abertura das garrafas de cerveja, mantendo a integridade da amostra. O equipamento utiliza um laser de excitação de 785 nm e realiza a varredura espectral na faixa de 200 a 2000 cm^{-1} . As medições foram feitas por contato direto da lente com a superfície externa do vidro, sem preparo prévio da amostra. Cada espectro foi adquirido e os dados posteriormente exportados para as etapas de pré-processamento e análise quimiométrica.

Em seguida, as amostras foram analisadas no sistema de análise de cerveja Acoolyzer Plus (Anton Paar GmbH, Graz, Áustria), instrumento desenvolvido especificamente para a análise de bebidas alcoólicas. Esse equipamento fornece informações sobre densidade, teor alcoólico, graus Plato, extrato real, extrato aparente e extrato primitivo. Para a remoção do gás carbônico, etapa

essencial para evitar interferências em análises subsequentes, as amostras foram submetidas a um banho ultrassônico durante 10 minutos.

Posteriormente, foram realizadas análises por espectroscopia no infravermelho com transformada de Fourier com refletância total atenuada (FTIR-ATR), espectrofotometria na região do ultravioleta-visível (UV-Vis) e nova aquisição de espectros por espectroscopia Raman, desta vez, diretamente do conteúdo.

Os espectros de FTIR foram adquiridos em um espectrômetro de infravermelho com transformada de Fourier (FTIR) modelo ALPHA II com acessório de refletância total atenuada (Platinum ATR), da Bruker Optik GmbH (Ettlingen, Alemanha). Cada amostra de cerveja (aproximadamente 1 mL) foi transferida, após o banho ultrassônico, com uma pipeta, para uma placa de refletância total atenuada (ATR), com cristal de diamante, localizada no compartimento do instrumento. Foi tomado cuidado para minimizar a presença de bolhas. Para cada amostra, foi registrado um espectro de feixe único com 24 varreduras acumuladas em uma resolução nominal de 4 cm^{-1} . Posteriormente, um espectro de fundo de referência do ar foi subtraído. Os espectros foram obtidos na faixa de 600 a 4000 cm^{-1} . Entre as amostras, a superfície do cristal ATR foi limpa com álcool isopropílico e seca com papel macio. O espectro de base subsequente registrado pelo espectrômetro foi examinado visualmente para garantir que nenhum resíduo da amostra anterior permanecesse no cristal ATR.

Para a análise na região do ultravioleta-visível (UV-Vis), foi utilizado um espectrofotômetro Spectroquant® Prove 300 (Merck, Alemanha). As amostras foram previamente diluídas em um fator de 1:30 e acondicionadas em cubetas de quartzo com caminho óptico de 10 mm, e as medições foram realizadas na faixa de comprimento de onda de 200 a 450 nm. O

equipamento possui fonte de luz comutável e detector de feixe duplo, garantindo boa estabilidade e reprodutibilidade das medições espectrais.

Além disso, a espectroscopia Raman foi conduzida utilizando o equipamento TacticID®, da B&W Tek (EUA), baseado na tecnologia de dispersão Raman com excitação a 785 nm. Esse sistema permite a análise com penetração mais profunda na amostra, sendo particularmente útil para a detecção de compostos em soluções ou líquidos complexos. As amostras foram analisadas diretamente por meio da janela óptica do dispositivo, sem necessidade de preparo prévio. Os espectros foram adquiridos na faixa de 150 a 2500 cm^{-1} , com tempo de exposição e intensidade do laser ajustados automaticamente pelo equipamento. Os dados espectrais obtidos foram utilizados na etapa de pré-processamento e, posteriormente, nas análises multivariadas.

3.3 Pré-processamento dos dados

Após a coleta dos dados espectrais obtidos pelas técnicas de espectroscopia FTIR, espectrofotometria UV-Vis, Raman Resolve e Raman TacticID, os conjuntos de dados foram organizados e submetidos a etapas de pré-processamento, com o objetivo de corrigir variações não relacionadas às propriedades químicas das amostras, como ruídos de fundo, deslocamentos de linha de base e efeitos físicos de espalhamento.

O tratamento foi conduzido por meio de scripts desenvolvidos em linguagem Python, utilizando bibliotecas específicas para manipulação de dados espectroscópicos e aplicação das rotinas de pré-processamento. As técnicas empregadas incluíram a normalização por vetor normal (SNV) aplicada aos espectros de FTIR e Raman TacticID, a suavização e derivatização pelo filtro de Savitzky–Golay nos dados de Raman Resolve, com seleção de parâmetros (janela, ordem polinomial e derivada) via validação cruzada Leave-One-Out (LOO) e, adicionalmente, a correção

de espalhamento multiplicativo (MSC) quando necessária. Para os dados físico-químicos obtidos pelo analisador de bebidas Anton Paar, não foi empregado pré-processamento espectral, exceto a conversão dos valores para formato numérico. Essa etapa foi fundamental para minimizar os efeitos de baseline, o ruído instrumental e as variações de intensidade, assegurando a qualidade e a comparabilidade dos dados a serem submetidos às análises quimiométricas subsequentes.

3.4 Estratégias quimiométricas

3.4.1 Exploração inicial: Análise por Componentes Principais (PCA)

Cada conjunto de dados (FTIR, UV-Vis, Raman Resolve, Raman TacticID, Anton Paar) foi submetido a PCA para investigar tendências de agrupamento natural, amostras atípicas e possíveis efeitos de lote. A PCA foi escolhida por ser não supervisionada e pela capacidade de condensar a variância em poucas componentes, facilitando inspeções visuais (scores) e a interpretação de fontes de variação (loadings). Esperava-se observar agrupamentos compatíveis com as classes e, se presentes, identificar outliers (via *Hotelling's T²* e *Q-residuals*), os quais podem sinalizar necessidade de reavaliação analítica ou de pré-processamento adicional.

3.4.2 Modelagem discriminante: Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA) e Análise Discriminante Linear (LDA)

A classificação supervisionada é realizada por meio das técnicas PLS-DA e LDA, aplicadas individualmente a cada bloco de dados e, posteriormente, aos conjuntos combinados. O uso conjunto dessas abordagens visa comparar o desempenho de dois métodos amplamente empregados em quimiometria para problemas de discriminação multiclasse.

O PLS-DA é escolhido por lidar adequadamente com alta colinearidade entre variáveis e com situações em que o número de variáveis excede o número de amostras ($p \gg n$), projetando os

preditores em variáveis latentes (LVs) que maximizam a covariância com a matriz de resposta. Já a LDA, também supervisionada, é empregada como método comparativo, por buscar combinações lineares das variáveis originais que maximizam a separação entre classes e minimizam a variabilidade interna.

Para LDA, foram aplicadas duas variações do método: LDA_PCA e LDA_Shrinkage. No LDA_PCA, a Análise de Componentes Principais (PCA) é utilizada previamente para reduzir o número de variáveis, transformando os espectros originais em um conjunto menor de componentes que retêm a maior parte da variância total. O LDA é então ajustado sobre essas componentes principais, o que reduz o ruído instrumental e evita instabilidade numérica em bases de dados com muitas variáveis.

Já o LDA_Shrinkage é uma abordagem regularizada do LDA que dispensa a etapa de redução de dimensionalidade. Nesse modelo, aplica-se um fator de *shrinkage* (ou encolhimento) à matriz de covariância, suavizando suas estimativas e tornando o modelo mais robusto quando há alta correlação entre variáveis ou número limitado de amostras por classe. Essa regularização reduz o risco de sobreajuste e melhora a capacidade de generalização, características especialmente importantes em conjuntos espectrais complexos.

Ambas as abordagens foram implementadas e comparadas com o modelo PLS-DA, de modo a avaliar o impacto das diferentes estratégias de redução e regularização na capacidade discriminante dos dados provenientes das diversas técnicas analíticas utilizadas, e consequente, separação de marcas de cerveja.

O número ótimo de variáveis latentes na PLS-DA é definido pela combinação entre o mínimo valor de PRESS (*Predicted Residual Error Sum of Squares*) e o teste de Van der Voet ($\alpha = 0,05$), que seleciona o modelo mais parcimonioso, não significativamente pior que o de menor erro,

mitigando o overfitting e favorecendo a generalização. Para a LDA, o desempenho é avaliado por validação cruzada estratificada, assegurando que a proporção de amostras por classe seja preservada em cada partição, ou seja, de forma que haja pelo menos uma amostra de cada marca de cerveja no conjunto de calibração, um aspecto essencial neste estudo, dado que o número de garrafas por marca varia consideravelmente. Essa estratégia reduz o viés de amostragem e garante maior representatividade e consistência estatística entre os subconjuntos de calibração e validação.

A avaliação do desempenho dos modelos é conduzida por meio das métricas de acurácia balanceada e F1-Macro, adequadas para conjuntos de dados desbalanceados. A primeira representa a média dos acertos ponderados por classe, enquanto o F1-Macro combina precisão e sensibilidade de forma equilibrada, refletindo o desempenho médio entre todas as classes. Valores elevados dessas métricas indicam boa capacidade discriminante e equilíbrio entre classes, ao passo que baixos desvios-padrão ao longo das rodadas de validação cruzada refletem maior estabilidade e reprodutibilidade do modelo. Essa abordagem assegura uma análise mais justa e robusta da classificação, especialmente diante das diferenças no número de amostras por classe.

3.4.3 Validação e particionamento

O particionamento inicial dos dados em conjuntos de calibração e validação externa é realizado por meio do algoritmo Kennard–Stone estratificado por classe, amplamente empregado em quimiometria por selecionar amostras distribuídas de forma representativa no espaço multivariado dentro de cada estrato. Diferentemente de uma divisão aleatória, esse método assegura que tanto o conjunto de calibração quanto o de validação cobram adequadamente a variabilidade química das amostras, evitando que regiões importantes do espaço químico fiquem sub-representadas.

Como o número total de amostras é moderado, adota-se ainda a estratégia de validação cruzada Leave-One-Out (LOO) dentro do conjunto de calibração, a fim de estimar de forma conservadora os erros preditivos e auxiliar na definição do número ótimo de variáveis latentes e dos hiperparâmetros de pré-processamento (por exemplo, parâmetros do filtro de Savitzky–Golay aplicado aos espectros Raman). Esse desenho metodológico em dois níveis: validação cruzada interna e validação externa independente, minimizam o risco de sobreajuste e permitem estimar de maneira mais realista a capacidade de generalização dos modelos.

3.4.4 Avaliação e interpretação

A avaliação dos modelos combina métricas quantitativas e representações gráficas que permitem interpretar simultaneamente o desempenho estatístico e a coerência química das classificações. As matrizes de confusão evidenciam padrões de erro por classe, indicando confusões sistemáticas entre amostras quimicamente semelhantes ou com espectros de baixa distinção. As curvas ROC/AUC OvR quantificam a capacidade de separação por classe em múltiplos limiares de decisão, sendo úteis em cenários de assimetria nos custos de erro.

3.5 Modelagem de Regressão e Ambiente Computacional

As etapas de calibração, validação e avaliação dos modelos de regressão multivariada foram implementadas em Python (versão 3.10), utilizando o ambiente Anaconda e o editor *Visual Studio Code*. As bibliotecas principais incluem *NumPy* e *Pandas* para manipulação e organização dos dados, *Scikit-learn* para construção e validação dos modelos, *Matplotlib* e *Seaborn* para visualização gráfica e *SciPy* para suporte estatístico.

Os scripts foram desenvolvidos de forma modular, permitindo a leitura automatizada das planilhas experimentais, o pré-processamento espectral, a divisão dos dados em conjuntos de calibração e validação, a otimização de hiper parâmetros e o salvamento dos resultados em arquivos de saída (.xlsx). Essa abordagem garante reprodutibilidade, padronização do fluxo analítico e integração direta com as etapas quimiométricas descritas anteriormente, sendo aplicada à modelagem de regressão quantitativa dos parâmetros físico-químicos das amostras.

4 RESULTADOS E DISCUSSÕES

4.1 Análise dos Parâmetros Físico-Químicos das Cervejas

A análise dos parâmetros físico-químicos das amostras comerciais de cerveja, realizada por espectroscopia NIR utilizando o *Alcolyzer Beer Analyzer* (Anton Paar), revelou conformidade com os limites estabelecidos pela Instrução Normativa nº 65/2022 do Ministério da Agricultura, Pecuária e Abastecimento (MAPA) para o estilo Pilsen. As variáveis analisadas, como teor alcoólico (% v/v), densidade, extrato aparente (Ea), extrato primitivo (P), grau de fermentação aparente (ADF) e extrato real (Er), apresentaram variações relevantes entre as diferentes marcas, refletindo tanto a composição declarada nos rótulos quanto diferenças de formulação e processo produtivo.

Tabela 2. Resultados Obtidos do Analisador de Cervejas com tratamento estatístico descritivo.

classe	teor alcoólico (% v/v)	Densidade (g/cm ³)	CO ₂ (g/L)	Ea (%m/m)	P (%m/m)	ADF (%m/m)	Er (%m/m)
AMS	4,42 ± 0,0	1,0 ± 0,0	4,39 ± 0,09	1,73 ± 0,01	10,15 ± 0,01	82,97 ± 0,10	3,35 ± 0,01
ANT	4,62 ± 0,09	1,01 ± 0,0	4,63 ± 0,32	2,26 ± 0,11	11,01 ± 0,14	79,49 ± 1,02	3,95 ± 0,1
BRA	4,79 ± 0,13	1,01 ± 0,0	4,1 ± 0,97	2,51 ± 0,19	11,47 ± 0,18	78,17 ± 1,67	4,23 ± 0,16
HEI	4,98 ± 0,1	1,01 ± 0,0	4,71 ± 0,04	2,06 ± 0,23	11,46 ± 0,03	82,08 ± 1,97	3,87 ± 0,2
IMP	4,46 ± 0,01	1,01 ± 0,0	5,32 ± 0,16	2,55 ± 0,01	10,97 ± 0,01	76,74 ± 0,09	4,18 ± 0,01
ORI	4,81 ± 0,12	1,01 ± 0,0	4,67 ± 0,09	2,66 ± 0,22	11,71 ± 0,21	77,3 ± 1,74	4,4 ± 0,21
PET	4,6 ± 0,04	1,01 ± 0,0	4,73 ± 0,03	2,28 ± 0,02	10,99 ± 0,06	79,23 ± 0,33	3,96 ± 0,01
SKO	4,73 ± 0,24	1,01 ± 0,0	4,6 ± 0,28	2,53 ± 0,24	11,45 ± 0,52	77,93 ± 1,74	4,25 ± 0,26
SPA	5,13 ± 0,15	1,01 ± 0,0	4,82 ± 0,1	2,03 ± 0,12	11,7 ± 0,34	82,68 ± 0,70	3,89 ± 0,15
STE	4,98 ± 0,09	1,01 ± 0,0	4,83 ± 0,13	1,86 ± 0,09	11,28 ± 0,21	83,48 ± 0,64	3,68 ± 0,11

A análise dos parâmetros físico-químicos das amostras de cerveja revelou variações relevantes entre marcas, mesmo dentro do mesmo grupo industrial, evidenciando que fatores como formulação, eficiência fermentativa e práticas de produção influenciam diretamente o perfil final

da bebida. De acordo com a Instrução Normativa nº 65/2022 do MAPA, o teor alcoólico das amostras, variando entre 4,42% e 5,17% v/v, está dentro dos limites esperados para o estilo Pilsen.

Essa faixa restrita demonstra certo controle de processo nas cervejarias, ainda que pequenas diferenças entre os valores rotulados e medidos instrumentalmente indiquem variações fermentativas, possivelmente associadas ao tipo de levedura, adjuntos ou condições operacionais específicas. As marcas Brahma Chopp (BRA) e Antartica (ANT), ambas da AMBEV, exibiram composições rotuladas semelhantes (água, malte, milho e lúpulo), mas mostraram distinções físico-químicas sutis. A Brahma apresentou maior extrato primitivo (11,47 °P) e maior teor alcoólico médio (4,79%), enquanto a Antartica demonstrou maior eficiência fermentativa (ADF = 79,63% vs. 78,47%), indicando um perfil mais seco e leve. A Brahma, por sua vez, mostrou maior extrato residual (Er = 4,23 °P), sugerindo maior corpo.

A Amstel (AMS) apresentou o maior grau de fermentação aparente entre todas as marcas (ADF = 82,97%), aliado ao menor extrato primitivo (10,15 °P) e ao menor extrato residual (Er = 3,35 °P), o que caracteriza uma cerveja de corpo leve e seca. Esses dados reforçam a coerência entre composição rotulada e desempenho fermentativo.

Já as amostras da Skol (SKO) e Império (IMP) mostraram menor eficiência fermentativa (ADF < 77,5%) e maior variabilidade no extrato real, refletindo perfis menos secos e com maior densidade residual, o que pode ser atribuído tanto à presença de adjuntos quanto a diferenças no processo de fermentação. Por outro lado, marcas como Stella Artois (STE) e Spaten (SPA) apresentaram comportamento mais homogêneo, com ADF médio superior a 83%, teor alcoólico elevado e baixo teor de extrato residual, indicando padronização de qualidade e fermentação eficiente em seus respectivos processos.

Em relação ao teor de CO₂, as amostras variaram entre 2,21 e 5,50 g/L, com destaque para a amostra 5BRA, que apresentou um dos menores valores (2,21 g/L), o que pode impactar negativamente a sensação de frescor e a formação de espuma. Já amostras como 2IMP e 3IMP apresentaram valores de CO₂ superiores a 5,2 g/L, sugerindo carbonatação mais intensa, o que pode afetar tanto a percepção sensorial quanto a estabilidade do envase.

Por fim, a comparação entre os dados físico-químicos e os perfis rotulados revela que, embora todas as amostras atendam à legislação vigente, a rotulagem é insuficiente para descrever integralmente a identidade da bebida. Diferenças na eficiência fermentativa, na quantidade de extrato residual e na carbonatação demonstram que fatores de produção não declarados impactam significativamente o resultado final. Assim, a caracterização físico-química detalhada, aliada à análise estatística e quimiométrica, configura uma ferramenta robusta para monitoramento de qualidade, rastreabilidade e detecção de possíveis fraudes ou inconsistências, contribuindo para a segurança do consumidor e a transparência no mercado cervejeiro.

4.2 Dados Pré-Processados

Para garantir a comparabilidade entre os sinais espectrais obtidos por FTIR, Raman (Resolve e TacticID) e UV-Vis, diferentes métodos de pré-processamento foram testados, incluindo Savitzky-Golay, MSC e SNV. Para cada um deles, foi selecionado o método que apresentasse melhor desempenho na correção de variações de intensidade e na padronização dos espectros, centralizando os dados e minimizando efeitos instrumentais e de dispersão. Embora o FTIR já apresentasse boa definição de bandas no sinal bruto, a aplicação do SNV contribuiu para reduzir

diferenças entre amostras e realçar variações relevantes para análises quimiométricas subsequentes.

A figura com os espectros pré-processados pode ser visualizada na Figura 1.

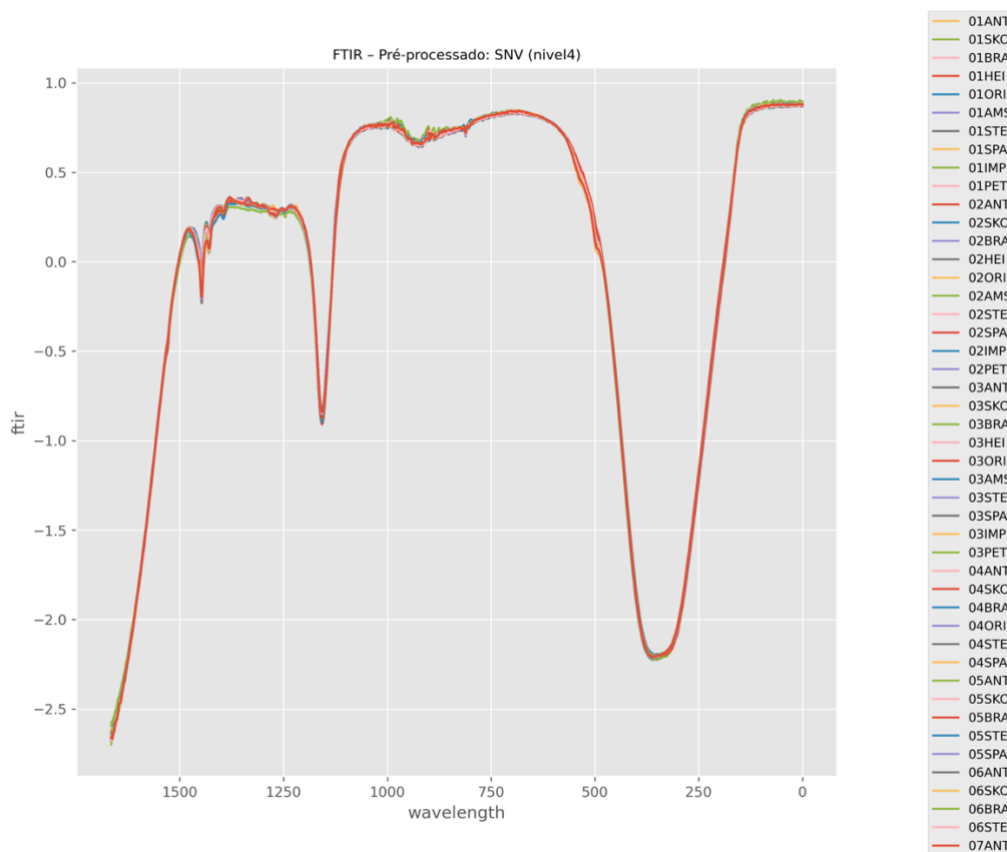


Figura 1. Espectro pré-processado do FTIR com SNV.

Nos espectros provenientes do equipamento Raman Resolve, observou-se a presença de ruídos de fundo mais evidentes, especialmente em regiões afastadas dos principais modos vibracionais. O uso de Savitzky-Golay foi eficaz na suavização desses ruídos, mantendo a integridade dos picos Raman. No entanto, a aplicação do MSC resultou em distorções acentuadas, o que sugere inadequação desse método para dados com elevado nível de ruído ou baixa repetibilidade espectral. O resultado do pré-processamento pode ser observado a seguir, na figura 2.

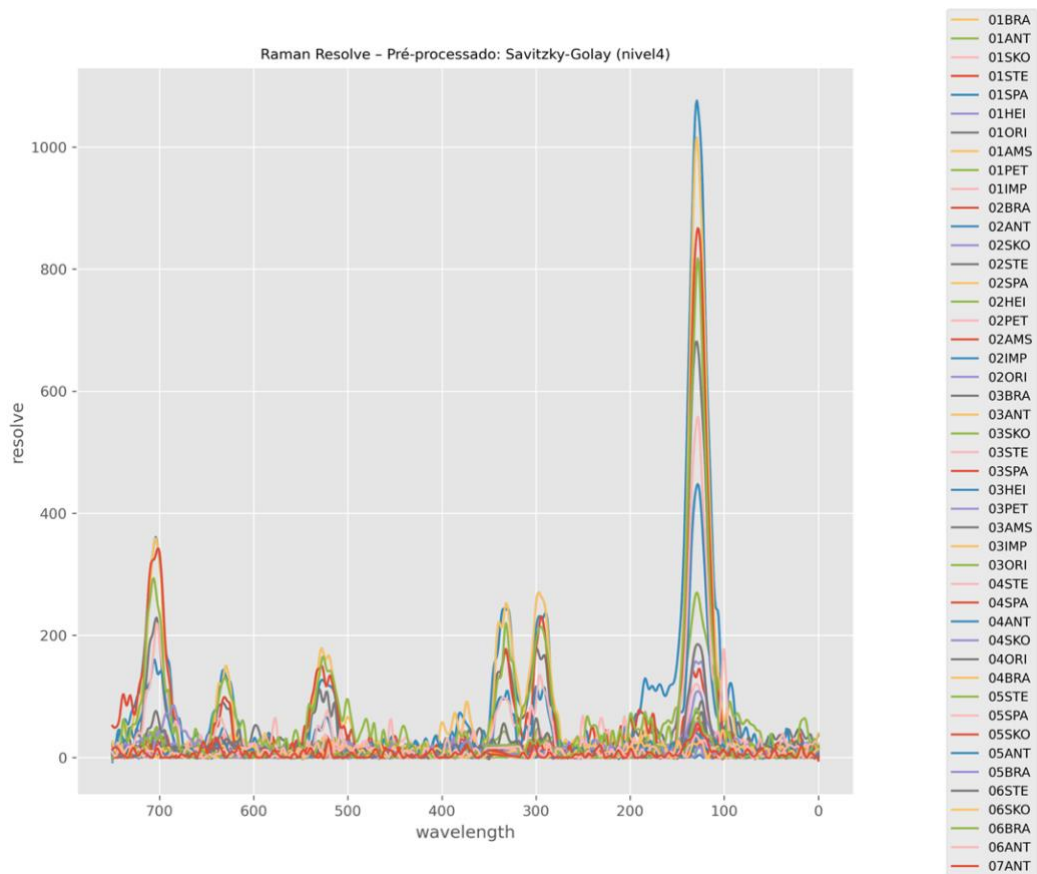


Figura 2. Espectro pré-processado do Raman Resolve com Savitzky-Golay.

Para os dados obtidos com o Raman TacticID, os espectros originais apresentaram boa qualidade, com baixos níveis de ruído e perfis consistentes entre as amostras. Nessa condição, todas as técnicas de pré-processamento mostraram desempenho satisfatório. O Savitzky-Golay promoveu leve suavização, o MSC uniformizou com precisão as curvas espectrais, e o SNV garantiu uma padronização eficaz sem comprometer a forma dos sinais, contribuindo positivamente para a preparação dos dados para as análises multivariadas, sendo o escolhido para o pré-processamento, como observado na figura 3.

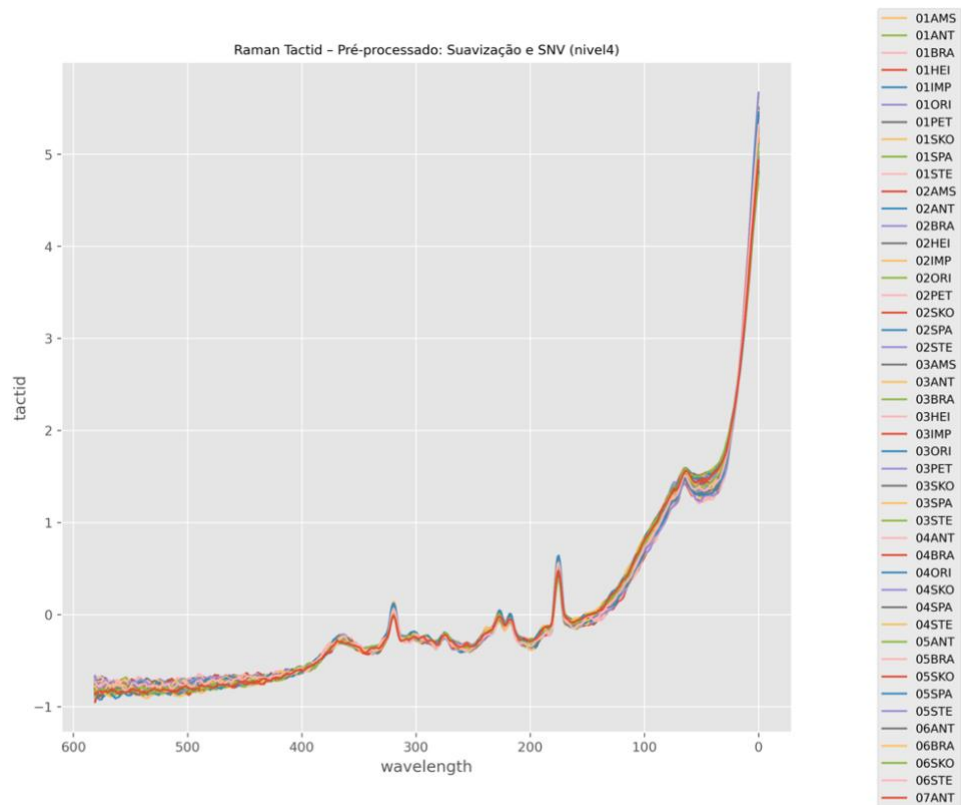


Figura 3 – Espectro pré-processado do Raman Tactid com SNV.

Em relação aos espectros de espectrofotometria UV-VIS, que apresentam curvas de absorção típicas, os dados originais já se mostravam bastante estáveis. Ainda assim, a aplicação dos pré-processamentos proporcionou ajustes úteis. O Savitzky-Golay suavizou pequenas flutuações, enquanto o MSC corrigiu discretas variações de linha de base entre as amostras, melhorando a sobreposição dos espectros. O SNV, por sua vez, aplicou uma padronização que, embora eficaz do ponto de vista estatístico, pode comprometer interpretações baseadas em intensidade absoluta da absorbância. O resultado do pré-processamento com SNV pode ser observado na figura 4.

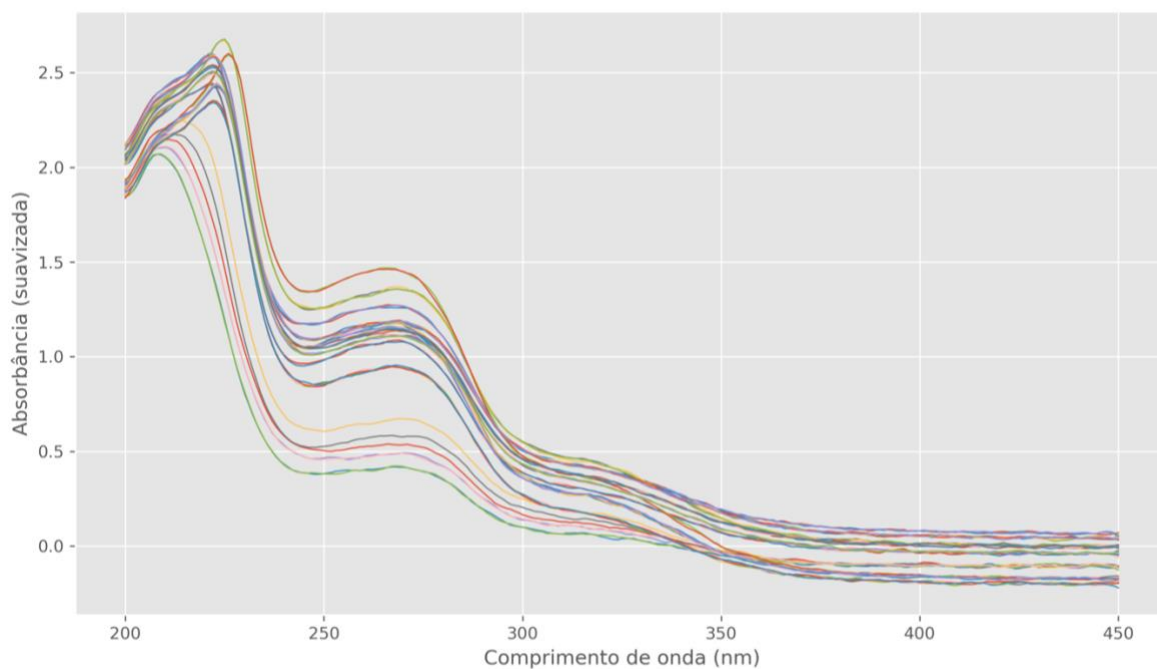


Figura 4 – Espectro pré-processado do UV-VIS com MSC.

A escolha criteriosa do pré-processamento, adaptada à natureza de cada técnica espectroscópica, revelou-se essencial para garantir a robustez e a qualidade das análises subsequentes.

4.3 Análise PCA dos dados do Analisador de Bebidas

A análise exploratória por PCA (Análise de Componentes Principais), baseada nos parâmetros físico-químicos obtidos por espectroscopia NIR utilizando o Alcoalyzer Beer Analyzer (Anton Paar), permitiu visualizar de forma tridimensional a variabilidade entre amostras comerciais de cerveja. Os três primeiros componentes principais explicaram, em conjunto, aproximadamente 90% da variância total dos dados, sendo suficientes para descrever os principais padrões de diferenciação entre as marcas. O gráfico de PC1xPC2 pode ser visualizado abaixo, na figura 5.

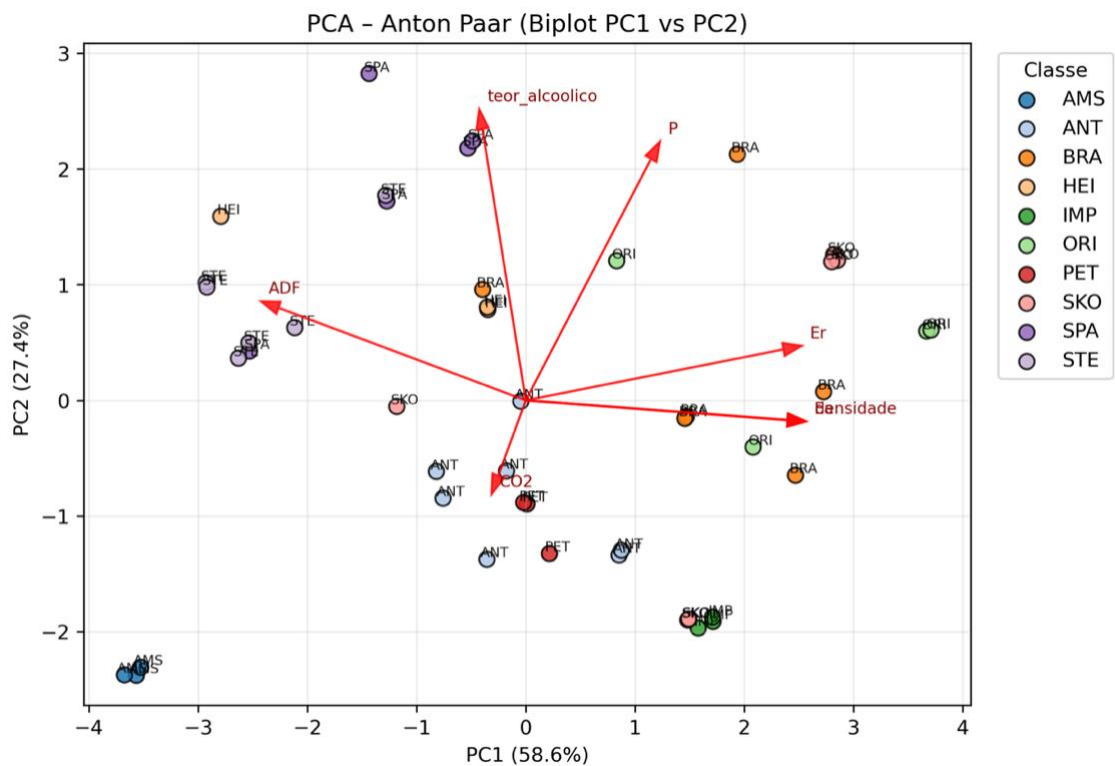


Figura 5. Modelo bidimensional (2D) para a Análise de Componentes Principais do Analisador de Bebidas

O PC1, primeiro componente principal, apresentou carga negativa elevada para o grau de fermentação aparente (ADF) e cargas positivas para os extratos aparente (Ea), real (Er) e primitivo (P). Essa configuração estatística indica que o PC1 representa um gradiente de eficiência fermentativa: amostras com alto ADF (fermentação mais completa) e baixos teores de extratos (menor resíduo de açúcares) tendem a ocupar valores baixos nesse componente. Quimicamente, isso é coerente com o fato de que uma fermentação mais eficiente consome mais açúcares, reduzindo o extrato residual e aumentando a atenuação aparente. Assim, o PC1 pode ser interpretado como um eixo que separa amostras mais secas e atenuadas (à esquerda) daquelas com maior teor de extrato residual (à direita).

O PC2, segundo componente principal, apresentou cargas positivas associadas ao teor alcoólico e ao extrato primitivo (P). Isso sugere que o PC2 reflete a intensidade fermentativa potencial, ou seja, a quantidade de açúcares disponíveis no mosto antes da fermentação e sua consequente conversão em etanol. Amostras com maior extrato primitivo tendem a gerar mais etanol, desde que a fermentação seja eficaz. A correlação positiva indica que, nesse eixo, amostras com maiores teores alcoólicos e maior densidade inicial se posicionam mais positivamente. Portanto, o PC2 pode ser interpretado como um eixo ligado ao poder fermentativo original do mosto, o que pode estar associado tanto ao estilo da cerveja quanto ao nível de adjuntos empregados.

Apesar da padronização aparente nas formulações rotuladas, especialmente entre marcas do grupo AMBEV produzidas na unidade de Jaguariúna (como Brahma, Skol, Antarctica, Original e Stella Artois), a PCA revelou uma dispersão considerável entre essas amostras. A Skol, por exemplo, apresentou valores de ADF inferiores (~77,4%) e extrato residual mais elevado, destacando-se no espaço multivariado. Embora compartilhem os mesmos ingredientes básicos (água, malte, milho e lúpulo), as diferenças observadas indicam que aspectos como a proporção dos ingredientes, tempo de fermentação, temperatura ou variações entre lotes afetam significativamente o produto final. Essa alteração na matriz de carboidratos impacta diretamente o rendimento fermentativo e a densidade final da bebida, justificando sua separação no espaço dos componentes principais. De forma semelhante, Heineken e Amstel, cujos rótulos indicam formulação exclusivamente com malte de cevada, posicionaram-se de maneira distinta na PCA, refletindo perfis mais secos e maior eficiência fermentativa, coerente com a ausência de cereais não maltados.

4.4 Análise PCA dos dados do Raman Tactid.

No espaço bidimensional da PCA, observou-se que marcas como Skol, Brahma, Antarctica e Original concentraram-se no quadrante negativo de PC1, ainda que com dispersões intra-marca visíveis, o que sugere menor consistência espectral. Esse ‘agrupamento’ indica características químicas comuns relacionadas ao uso de adjuntos, em especial milho, que impactam diretamente os perfis vibracionais associados a carboidratos residuais. Em contraste, marcas como Heineken, Bohemia, Petra, Stella Artois, Spaten, Império e Brahma Duplo Malte agruparam-se no quadrante positivo de PC1. Tal posicionamento sugere perfis espectrais associados a formulações puro malte, caracterizadas por maior homogeneidade quanto às contribuições do malte de cevada e às bandas vibracionais ligadas a proteínas, polifenóis e compostos aromáticos, conforme visto na figura 6.

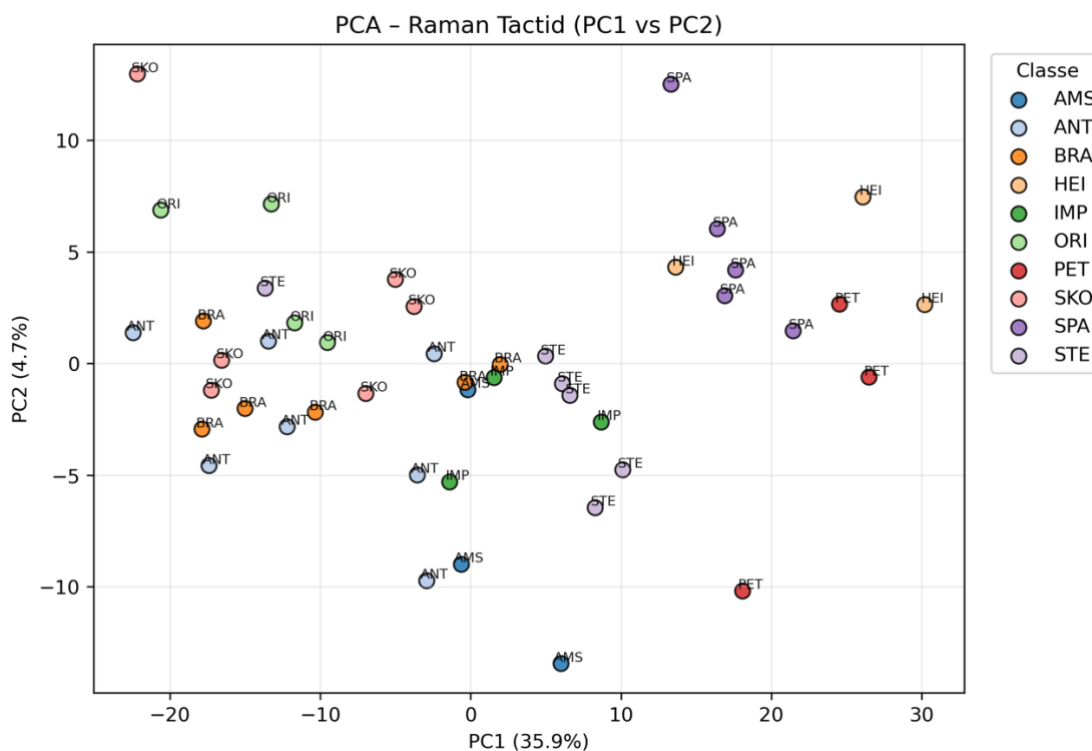


Figura 6. Modelo bidimensional (2D) para a Análise de Componentes Principais do Raman Tactid

Quimicamente, as cervejas puro malte apresentam espectros Raman caracterizados por bandas atribuídas aos constituintes do malte de cevada, como carboidratos complexos (480–500 cm^{-1} , vibrações do anel glicosídico, e 940–960 cm^{-1} , ligações C–O–C da amilopectina), proteínas e peptídeos (1.200–1.350 cm^{-1} , deformações C–N e N–H, e 1.650 cm^{-1} , amida I), além de compostos fenólicos e aromáticos do malte e do lúpulo ($\sim 1.600 \text{ cm}^{-1}$, vibração C=C aromático).

Já as cervejas com adjuntos, como milho, exibem maior contribuição de amido de milho, com assinaturas vibracionais distintas das da cevada e predominância de bandas ligadas a polissacarídeos, enquanto apresentam menor intensidade relativa em regiões associadas a proteínas e fenóis. Essas diferenças químicas, embora sutis e pouco evidentes visualmente, são realçadas pelo pré-processamento (SNV) e exploradas pela PCA, possibilitando a clara distinção entre os dois grupos de amostras no espaço quimiométrico, principalmente quando analisamos em função de PC1. A figura 7 realça bem essa diferença, as amostras em azul possuem em seus rótulos a composição com adjuntos, enquanto as amostras em laranja são puro malte.

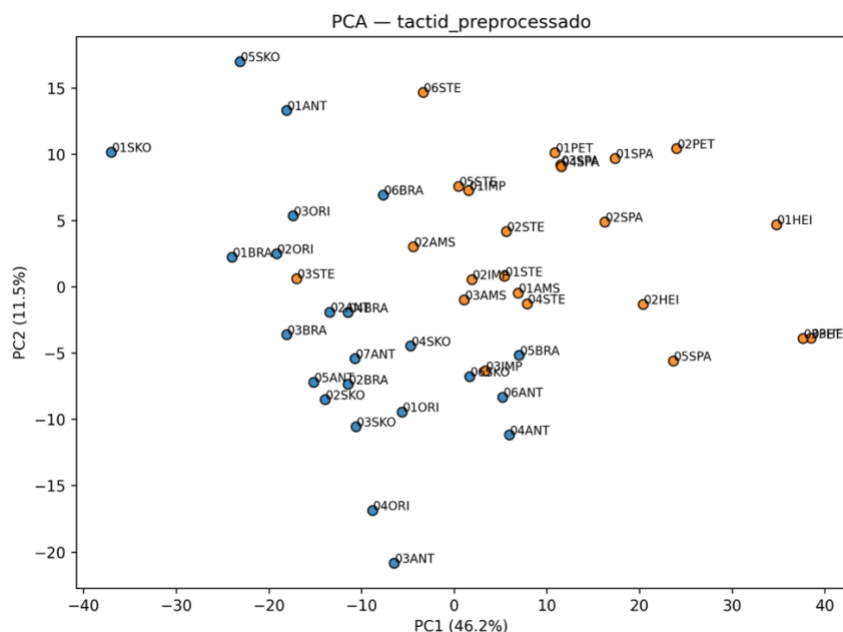


Figura 7. Modelo bidimensional (2D) para a Análise de Componentes Principais do Raman Tactid com rótulos com adjunto em azul e puro malte em laranja.

De modo geral, os resultados sugerem que a PCA aplicada aos espectros Raman é eficaz em discriminar cervejas puro malte das produzidas com adjuntos, evidenciando que diferenças de formulação se refletem em padrões espectroscópicos. As amostras puro malte concentram-se à direita de PC1, enquanto aquelas com adjuntos, como Original, Antartica, Brahma e Skol, localizam-se predominantemente à esquerda.

4.5 Análise PCA dos dados de FTIR

A PCA aplicada aos espectros de FTIR, após pré-processamento por SNV, explicou 75,94% da variância nos dois primeiros componentes (PC1 = 53,91%; PC2 = 22,03%), revelando boa capacidade de redução dos dados. Contudo, diferentemente do observado em outras técnicas, as amostras apresentaram distribuição difusa no espaço bidimensional, sem a formação de agrupamentos bem definidos entre cervejas puro malte e aquelas com adjuntos, como pode ser observado na figura 8. Marcas como Skol, Brahma, Antartica e Original, assim como Heineken, Spaten, Stella Artois, Império e Petra apareceram em ambos os lados do PC1, indicando sobreposição considerável entre os grupos.

Esse comportamento pode ser atribuído ao fato de o FTIR captar intensamente constituintes presentes em qualquer formulação, como água, carboidratos, proteínas e compostos fenólicos, reduzindo o contraste entre amostras. Assim, a variabilidade observada parece refletir tanto diferenças de composição quanto variações de processo e lote, dificultando a separação clara entre puro malte e adjuntos. Dessa forma, embora o FTIR mostre sensibilidade e alto poder explicativo,

seus resultados nesta análise não evidenciam uma discriminação direta entre os grupos, sugerindo a necessidade de métodos supervisionados (como PLS-DA ou LDA) para explorar diferenças mais sutis ou de combinação dos dados com outras técnicas analíticas.

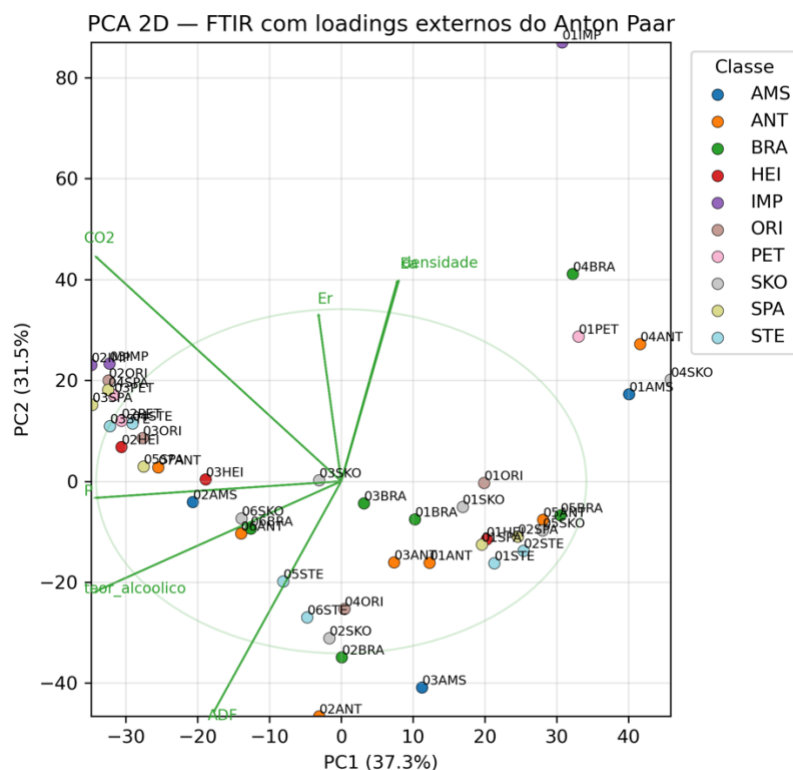


Figura 8. Modelo bidimensional (2D) para a Análise de Componentes Principais do FTIR

4.6 Análise da PCA do Raman Resolve

A PCA dos espectros de Raman Resolve, tratados com Savitzky–Golay, explicou 55,47% da variância nos dois primeiros componentes (PC1 = 46,82%; PC2 = 8,65%) e pode ser visualizada na Figura 9. O gráfico mostra forte dispersão das amostras, sem formação de grupos claros entre cervejas puro malte e com adjuntos, além de grande variabilidade intra-marca. Isso indica que a técnica, neste arranjo, não captura diferenças consistentes de formulação, mas reflete um gradiente

global influenciado por múltiplos fatores, em que a variabilidade dentro de cada marca é maior do que entre categorias.

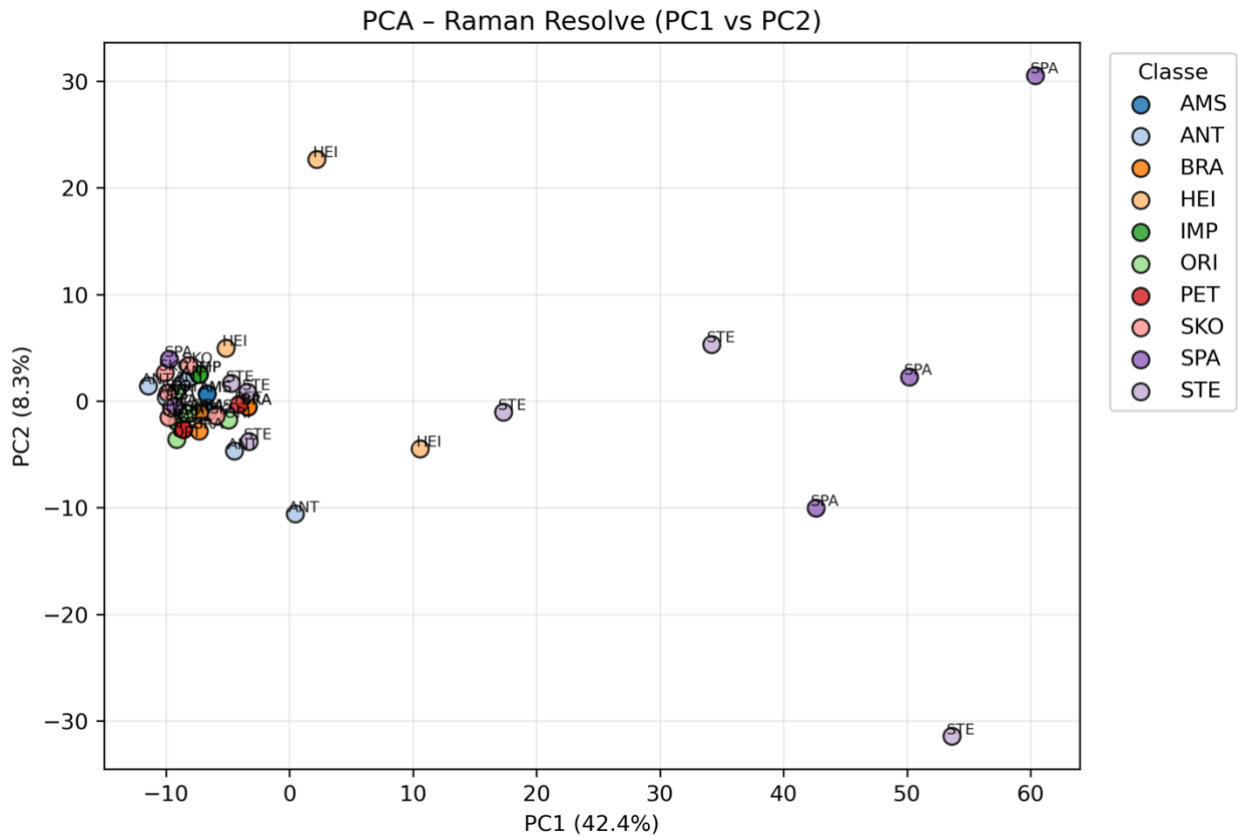


Figura 9. Modelo bidimensional (2D) para a Análise de Componentes Principais do Raman Resolve.

Um dos principais fatores que pode explicar o desempenho limitado da PCA para os dados Raman é o modo de aquisição “através da garrafa”. O vidro atua como interferente no sinal, influenciado por sua cor e composição, além de variações de espessura e curvatura da embalagem. Esses efeitos, associados ainda ao nível de enchimento, turbidez, CO₂ dissolvido e presença de rótulos, alteram o baseline, aumentam a fluorescência e reduzem a razão sinal-ruído, introduzindo variabilidade não relacionada à composição química da cerveja. Como resultado, parte da dispersão observada na PCA reflete interferências experimentais, e não apenas diferenças entre as amostras.

Esse comportamento fica evidente ao observar que garrafas verdes tendem a se dispersar mais no gráfico, enquanto garrafas âmbar permanecem mais agrupadas. Esses achados reforçam a necessidade de padronização das condições de medição ou do uso de recipientes controlados em estudos futuros, a fim de melhorar a reprodutibilidade e a discriminação entre amostras.

4.7 Análise da PCA da Espectroscopia UV-VIS

A PCA construída a partir dos espectros de UV-Vis (PC1 vs PC2) evidencia um padrão muito claro de agrupamento por marca (classe), com réplicas próximas entre si e separações bem definidas entre diferentes grupos. Visualmente, observa-se que a maior parte das classes forma grupos compactos, o que é um indicativo direto de boa repetibilidade instrumental e consistência espectral dentro de cada marca. Isso é particularmente importante no contexto trabalho, porque sugere que o UV-Vis captura uma assinatura espectral estável para cada cerveja, capaz de representar o “perfil” da marca de forma reprodutível.

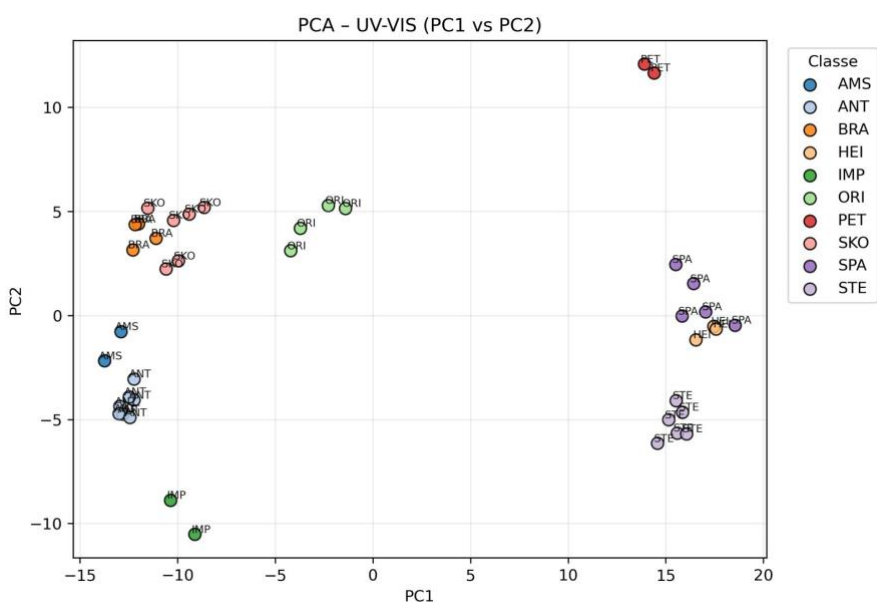


Figura 10. Modelo bidimensional (2D) para a Análise de Componentes Principais do UV-VIS.

Esse resultado é particularmente promissor do ponto de vista aplicado, pois mostra que o UV-Vis, de forma isolada, é capaz de diferenciar marcas e agrupar corretamente amostras autênticas, o que abre caminho para sua utilização como ferramenta de triagem na detecção de possíveis fraudes, como a troca de rótulos. Considerando que se trata de uma técnica acessível, de fácil operação e amplamente disponível em laboratórios, a PCA por UV-Vis se destaca como uma abordagem eficiente e de alto potencial para aplicações em controle de qualidade e fiscalização de cervejas industrializadas.

4.8 Modelagem e Análises Supervisionadas

Inicialmente, foram avaliadas diversas técnicas de classificação supervisionada, incluindo Regressão Logística (LogReg), Máquinas de Vetores de Suporte com kernel linear e radial (LinSVM, RBF-SVM), Análise Discriminante Linear (LDA) e Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA). Essa etapa teve como objetivo identificar os métodos mais adequados para lidar com a estrutura multivariada e colinear dos dados espectroscópicos e físico-químicos. Como os resultados apontaram para um comportamento linear dos dados, as técnicas de LDA e PLS-DA mostraram-se as que mais se adequavam a proposta do trabalho.

Os resultados resumidos das técnicas, apresentados na Tabela 3, resume os valores que as técnicas LDA e PLS-DA apresentaram de desempenho, superior em termos de acurácia média e F1-Macro, além de maior consistência entre rodadas de validação cruzada. A LDA destacou-se pela elevada capacidade de separação em blocos bem condicionados, como UV-Vis e Anton Paar, enquanto a PLS-DA apresentou maior robustez e estabilidade em conjuntos com alta dimensionalidade e colinearidade, como FTIR e Raman. A tabela 3 pode ser visualizada abaixo.

Tabela 3 – Métricas de desempenho para cada técnica, a partir de diferentes modelos de classificação.

Técnica	Modelo	Acurácia Média	F1-macro	Acurácia Balanceada VAL	F1- macro VAL
Analizador de Bebidas	LDA_PCA	42%	36%	60%	49%
Analizador de Bebidas	LDA_Shrinkage	43%	38%	55%	45%
Analizador de Bebidas	PLS-DA	31%	21%	30%	21%
FTIR	LDA_PCA	11%	7%	25%	17%
FTIR	LDA_Shrinkage	12%	6%	15%	5%
FTIR	PLS-DA	5%	3%	35%	28%
Raman TACTID	LDA_PCA	19%	14%	50%	42%
Raman TACTID	LDA_Shrinkage	22%	15%	40%	31%
Raman TACTID	PLS-DA	24%	17%	35%	29%
Raman RESOLVE	LDA_PCA	16%	11%	20%	14%
Raman RESOLVE	LDA_Shrinkage	14%	10%	55%	47%
Raman RESOLVE	PLS-DA	16%	9%	20%	15%
UV-VIS	LDA_PCA	97%	96%	100%	100%
UV-VIS	LDA_Shrinkage	97%	96%	100%	100%
UV-VIS	PLS-DA	84%	77%	87%	83%
COMBINADO	LDA_PCA	30%	24%	75%	71%
COMBINADO	LDA_Shrinkage	54%	47%	65%	63%
COMBINADO	PLS-DA	42%	34%	50%	45%

Das métricas apresentadas na tabela, acurácia média, F1-macro, acurácia balanceada e F1-macro de validação, descrevem o desempenho dos modelos de classificação aplicados aos diferentes conjuntos instrumentais. A acurácia média mostra quantas amostras o modelo classificou corretamente no total, mas pode esconder erros em classes com poucas amostras. O F1-macro é uma média que considera o equilíbrio entre acertos e erros em todas as classes, mostrando se o modelo trata todas com a mesma importância. A acurácia balanceada faz algo parecido, avaliando o desempenho por classe e corrigindo o efeito de desbalanceamentos. Já o F1-macro de validação

mostra como o modelo se comporta com dados novos, indicando sua capacidade de generalizar e não apenas decorar o conjunto de treino.

Nos resultados obtidos para o Analisador de Bebidas, observa-se que os modelos LDA_PCA e LDA_Shrinkage apresentaram desempenhos semelhantes, com acurácia média de 42 a 43% e F1-macro em torno de 36 a 38%. Embora esses valores indiquem uma classificação apenas moderada, as métricas de validação, especialmente a acurácia balanceada (55–60%), sugerem que, quando corrigido o efeito do desbalanceamento entre classes, o modelo é capaz de distinguir parcialmente os grupos de amostras. O modelo PLS-DA, entretanto, apresentou desempenho inferior, com acurácia média de apenas 31%, indicando que, para esse conjunto instrumental, a abordagem baseada em variáveis latentes não capturou adequadamente a variabilidade discriminante dos dados. Essa tendência sugere que o comportamento físico-químico medido pelo analisador de bebidas é mais bem representado por modelos lineares clássicos, como o LDA, especialmente quando combinados à redução de dimensionalidade via PCA.

Para os dados obtidos por FTIR-ATR, os resultados foram notavelmente inferiores em todas as métricas. As acurácias médias ficaram entre 5 e 12%, e o F1-macro abaixo de 10%, com desempenho levemente melhor para o LDA_PCA na validação (acurácia balanceada de 25%). Esses números indicam forte sobreposição espectral entre as amostras ou elevada variação dentro das mesmas classes, dificultando a discriminação. Embora o PLS-DA tenha apresentado um ganho na validação (35% de acurácia balanceada), seu desempenho geral ainda foi considerado insatisfatório. Essa limitação reflete o comportamento já esperado para espectros FTIR de matrizes complexas como a cerveja, nos quais a contribuição de bandas amplas e sobrepostas dificulta a

diferenciação entre amostras similares. Esses resultados reforçam a necessidade de estratégias de pré-processamento mais agressivas ou da fusão de dados com outras técnicas complementares.

Os resultados do Raman TacticID mostraram desempenho intermediário. O modelo LDA_Shrinkage apresentou ligeira vantagem em relação ao LDA_PCA, com acurácia média de 22% e F1-macro de 15%. O PLS-DA teve desempenho similar (24% de acurácia média e 17% de F1-macro), mas os valores de acurácia balanceada em validação (35–50%) revelam que o modelo consegue reconhecer padrões entre as classes de forma equilibrada, ainda que o número total de acertos seja limitado. Essa diferença entre métricas globais e balanceadas indica que o Raman TacticID contém informação discriminante, porém com alta variabilidade instrumental ou interferências de fluorescência, comuns em amostras com pigmentos e compostos orgânicos complexos.

De modo semelhante, o Raman Resolve apresentou resultados modestos, porém com comportamento particular. O LDA_Shrinkage obteve apenas 14% de acurácia média, mas alcançou 55% de acurácia balanceada na validação, revelando um possível efeito de sobreajuste nas classes minoritárias ou variação intensa entre réplicas. O PLS-DA e o LDA_PCA tiveram desempenhos mais consistentes, mas igualmente baixos. Essa discrepância entre treino e validação sugere que, embora o equipamento Raman de alta resolução capture detalhes espectrais relevantes, as diferenças entre amostras podem não ser suficientemente expressivas para sustentar um modelo preditivo robusto sem ajustes adicionais ou fusão de dados.

Em contraste com as demais técnicas, o UV-Vis apresentou desempenho excepcional, sendo capaz de separar as marcas de cerveja e agrupar amostras de uma mesma marca mesmo quando utilizado isoladamente. Os modelos LDA_PCA e LDA_Shrinkage alcançaram acurácia

média de 97% e F1-macro de 96%, atingindo 100% de acurácia balanceada na validação, enquanto o PLS-DA também apresentou resultados elevados (84% de acurácia média e 87% balanceada). Esses achados são particularmente relevantes porque demonstram que o UV-Vis, uma técnica amplamente disponível, de baixo custo operacional e de fácil uso, é capaz de identificar possíveis falsificações decorrentes de troca de rótulos, algo até então associado a métodos mais complexos ou equipamentos especializados, como o analisador de bebidas. A sensibilidade do UV-Vis à absorção de compostos fenólicos, pigmentos e produtos da reação de Maillard confere ao método assinaturas espectrais suficientemente características para discriminar diferentes cervejas com alta precisão. Dessa forma, o UV-Vis se destaca como uma ferramenta acessível, robusta e reprodutível, com grande potencial de aplicação em laboratórios de controle de qualidade, fiscalização e verificação de autenticidade de bebidas.

Apesar de a fusão midlevel multitécnica elevar algumas métricas em relação aos blocos espectroscópicos isolados de menor desempenho (FTIR e Raman), o modelo combinado não supera o obtido com os dados de UV-Vis, que atinge 100% de acurácia; portanto, neste conjunto, a integração diminui a modelagem frente ao melhor bloco individual. Esse efeito pode ocorrer quando a informação adicional traz ruído, colinearidade redundante ou sinais conflitantes entre técnicas, ampliando a dimensionalidade sem ganho proporcional de discriminação. Também podem contribuir diferenças de escala entre blocos, desequilíbrio de classes e a ausência de ponderação/seleção de variáveis por bloco. Em trabalhos futuros, recomenda-se investigar fusão com ponderação baseada em desempenho (block-scaling), seleção de variáveis dirigida por VIP/coeficientes por técnica, alinhamento e normalização por bloco, além de fusão *high level* que combine decisões em vez de concatenar variáveis. Do ponto de vista aplicado, os resultados

indicam que o UV-Vis isolado já oferece poder discriminante suficiente para autenticação neste cenário.

De forma geral, observa-se que o LDA_Shrinkage apresentou desempenho mais estável e robusto em quase todos os conjuntos, superando o LDA_PCA e o PLS-DA na maioria dos casos. A técnica UV-Vis destacou-se como a mais discriminante, seguida pelo Analisador de Bebidas e pela fusão de dados, enquanto o FTIR e os espectros Raman apresentaram desempenhos limitados.

Dessa forma, as análises subsequentes concentram-se nessas duas abordagens, consideradas as mais representativas e interpretáveis dentro do conjunto de técnicas testadas, permitindo uma comparação direta entre modelos lineares clássicos e métodos baseados em projeções latentes.

4.9 Resultados de LDA

A Análise Discriminante Linear (LDA) foi empregada para avaliar a capacidade de separação entre as amostras de cerveja com base nos diferentes blocos de dados físico-químicos e espectroscópicos. O método apresentou comportamento consistente com a natureza de cada conjunto: maior desempenho em blocos de baixa colinearidade e menor dimensionalidade, como os dados físico-químicos (Anton Paar) e o espectro UV-Vis, e resultados mais modestos nas técnicas vibracionais (FTIR e Raman), nas quais o alto grau de correlação entre variáveis limita a separação linear entre classes.

Os melhores resultados foram obtidos para o conjunto UV-Vis, com acurácia e F1-Macro de 100%, indicando separação completa entre as amostras. No bloco físico-químico, a LDA também se destacou (acurácia média de 52%), com teor alcoólico, densidade e extrato real figurando entre as variáveis mais relevantes para a diferenciação entre marcas e formulações. Por outro lado, os

conjuntos FTIR e Raman (Resolve e TacticID) apresentaram acurácias entre 11% e 28%, refletindo a forte sobreposição espectral e a presença de ruído instrumental, fatores que reduzem a capacidade discriminante do modelo.

A análise dos coeficientes discriminantes evidencia as regiões espectrais mais relevantes para a separação. No UV-Vis, destacam-se comprimentos de onda entre 270 e 290 nm, associados à absorção de compostos fenólicos e pigmentos. No FTIR, as regiões 1200–1000 cm^{-1} (ligações C–O) e 1750–1600 cm^{-1} (ligações C=O) apresentaram maior peso, caracterizando carboidratos, álcoois e ésteres provenientes do malte e da fermentação. Nos espectros Raman, as faixas 1000–1600 cm^{-1} corresponderam a vibrações C=C e C–H aromáticas, igualmente relacionadas a compostos fenólicos e pigmentos do lúpulo e do malte. Essas regiões, apesar de produzirem separações menos nítidas, confirmam que o modelo captura variações químicas coerentes entre as amostras.

Enquanto o conjunto UV-Vis exibe picos bem definidos e praticamente sem sobreposição, os conjuntos Raman e FTIR mostram distribuições parcialmente sobrepostas, indicando menor capacidade discriminante. Ainda assim, observa-se coerência interna, pois amostras da mesma classe tendem a agrupar-se em regiões próximas do eixo discriminante, evidenciando reprodutibilidade instrumental e consistência entre réplicas.

Os histogramas das distribuições de LD1 (Figura 10) revelam graficamente a separação das classes ao longo da principal função discriminante. E os gráficos de escores LD1 \times LD2 (Figura 11) permitem visualizar a separação bidimensional das amostras.

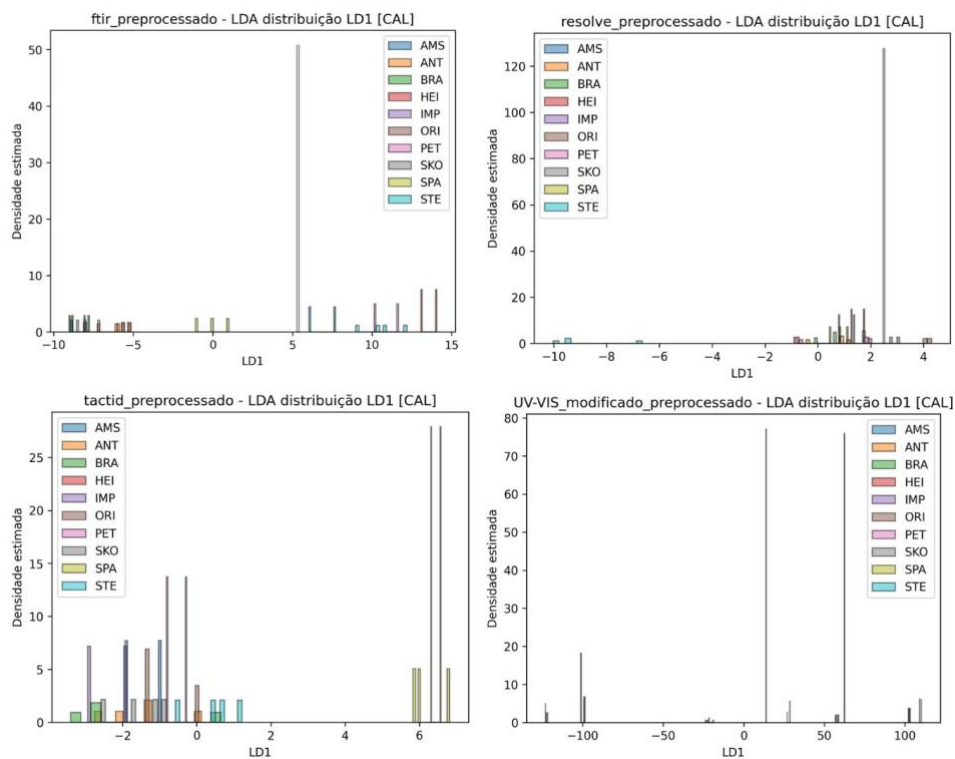


Figura 11. Os histogramas das distribuições de LD1 para FTIR, Resolve, Tactid e UV-VIS.

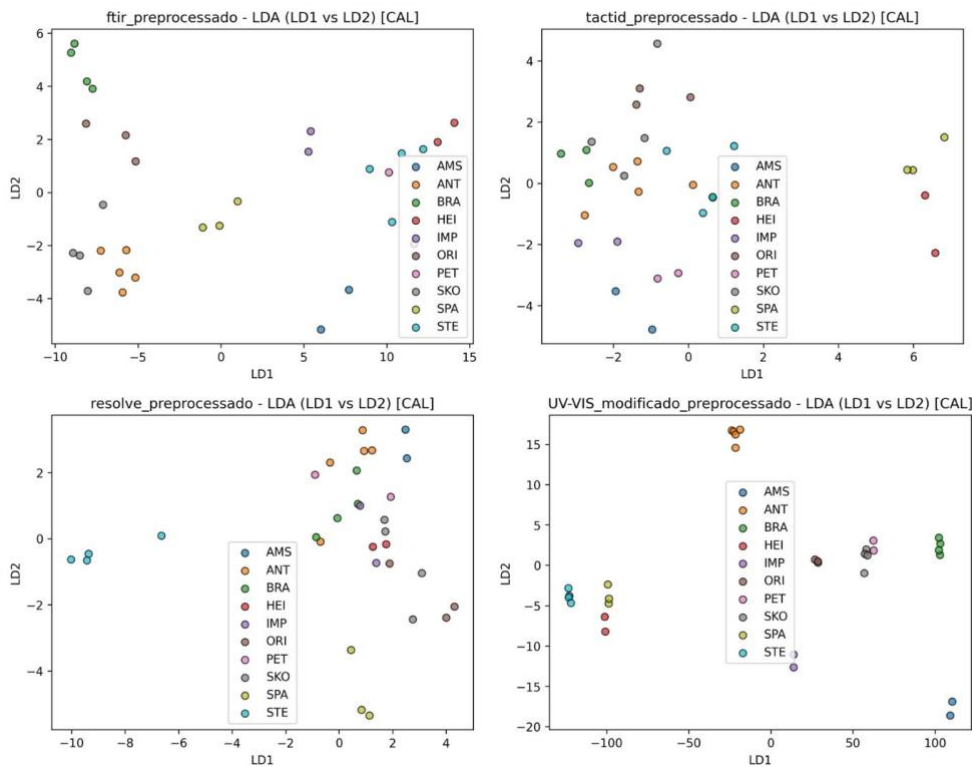


Figura 12. gráficos de escores LD1 × LD2 para FTIR, Resolve, Tactid e UV-VIS.

Nessas projeções, o conjunto UV-Vis apresenta agrupamentos bem definidos e fronteiras claras entre classes, reforçando seu excelente desempenho quantitativo. Já os conjuntos FTIR e Raman mostram maior dispersão e interpenetração de pontos, coerente com suas menores métricas de acurácia. Apesar disso, a separação parcial observada em algumas marcas (por exemplo, SPA e ORI) sugere a presença de assinaturas espectrais específicas, ainda que sutis.

Por fim, as matrizes de confusão (Figura 12) ilustram diretamente o desempenho da LDA na fase de validação.

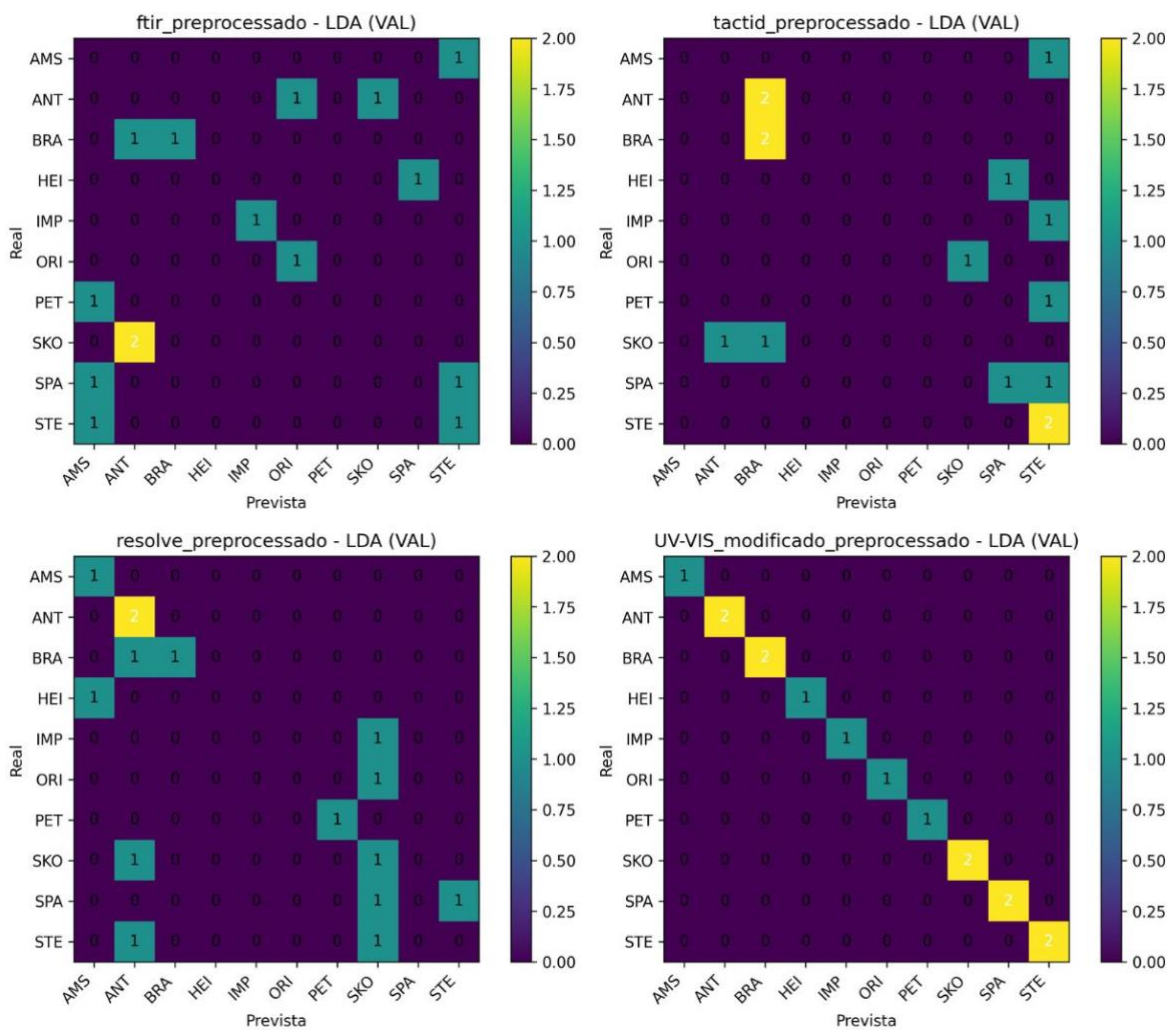


Figura 13. Matrizes de confusão do LDA para FTIR, Resolve, Tactid e UV-VIS.

No conjunto UV-Vis, todas as amostras foram corretamente classificadas, confirmando a eficiência do modelo. Nos demais blocos, predominam acertos nas diagonais, mas também algumas confusões pontuais, especialmente entre classes de perfis físico-químicos semelhantes, como ANT × BRA e SKO × ORI, essas quatro marcas de cerveja caracterizam-se pela presença de adjuntos em sua composição (milho). Esse comportamento é esperado em dados espectroscópicos de elevada complexidade.

De forma geral, os resultados demonstram que a LDA, embora limitada para modelar relações complexas, é uma ferramenta eficiente de análise exploratória e de avaliação preliminar da separabilidade entre classes, principalmente quando se trata da espectrofotometria no ultravioleta-visível. Sua aplicação permitiu identificar padrões consistentes e regiões espectrais de relevância química, estabelecendo uma base sólida para a discussão dos resultados obtidos pela PLS-DA, apresentada a seguir.

4.10 Resultados PLS-DA

A análise PLS-DA foi aplicada aos mesmos blocos de dados utilizados na LDA, com o objetivo de explorar uma abordagem supervisionada mais robusta e capaz de lidar com alta colinearidade entre variáveis e casos em que o número de variáveis excede o número de amostras. O modelo busca maximizar a covariância entre os preditores e as classes de resposta, projetando os dados em variáveis latentes (LVs) que representam direções de máxima discriminação no espaço multivariado.

O desempenho geral dos modelos de PLS-DA não superou o obtido pela LDA, mas apresentou melhoras sutis especialmente para os conjuntos espectroscópicos, nos quais o método

conseguiu capturar variações sutis e associadas à composição química das amostras. O bloco UV-Vis novamente se destacou apresentando acurácia média de 84% e F1-Macro de 77% e curvas ROC/AUC = 1.00 para quase todas as classes, exceto Skol e Original, evidenciando uma separação parcial. Nos blocos Raman Resolve e TacticID, observou-se melhora significativa em relação à LDA, com valores de acurácia média em 16% e 24%, respectivamente, e com valores de AUC superiores a 0,7 para a maioria das classes e poucas confusões residuais entre amostras de perfis espectrais semelhantes (por exemplo, ANT e BRA). O conjunto FTIR, apesar de ainda apresentar sobreposição parcial entre classes, também mostrou aumento de desempenho, com AUCs entre 0,5 e 1,00, indicando maior capacidade do modelo em reconhecer padrões discriminantes nas regiões médias do infravermelho.

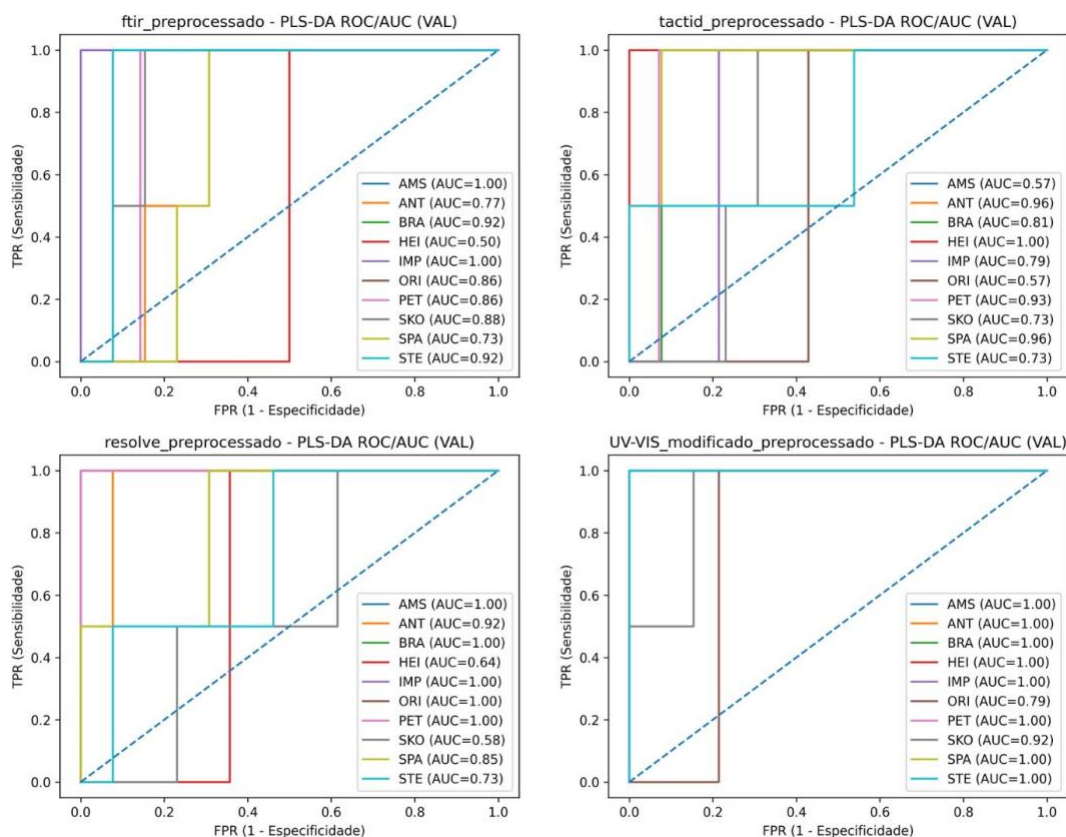


Figura 14. Curvas ROC/AUC para FTIR, Resolve, Tacticid e UV-VIS.

Os gráficos de escores LV1 × LV2 (Figura 14) revelam de forma visual a eficiência da PLS-DA na separação multiclasse.

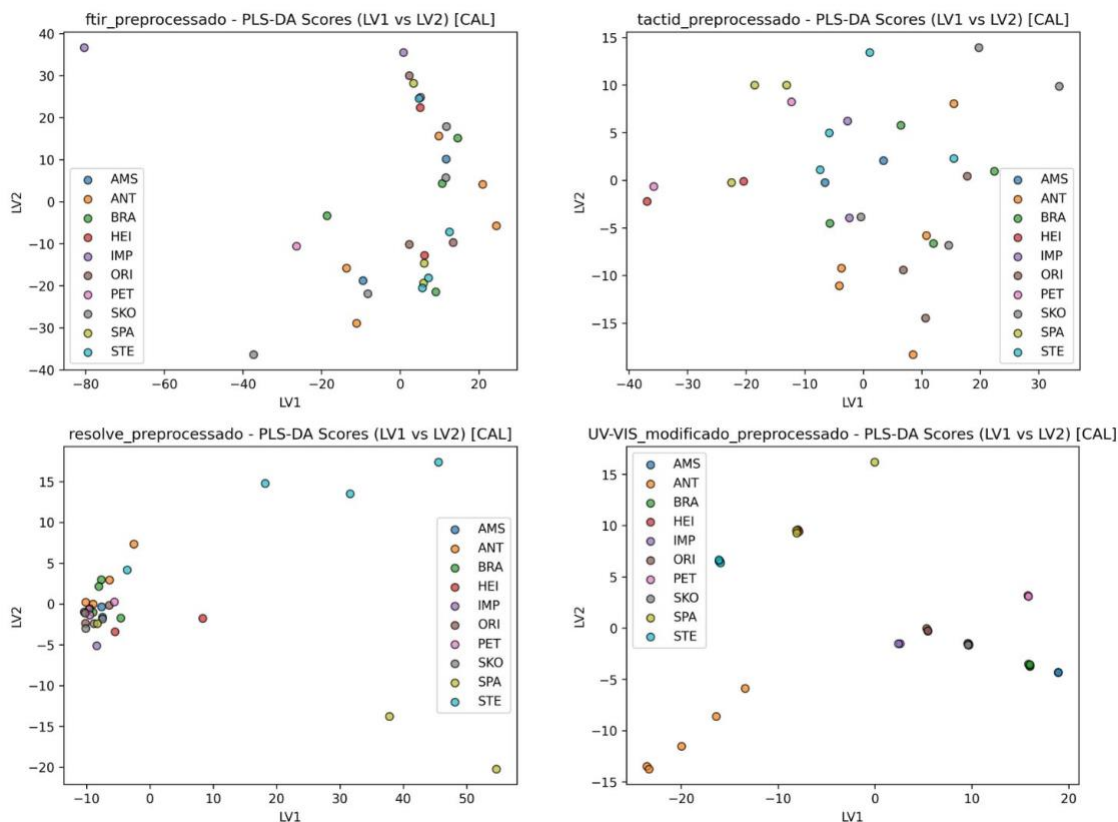


Figura 15. gráficos de escores LV1 × LV2 para FTIR, Resolve, Tactid e UV-VIS.

O UV-Vis apresenta clusters bem definidos e sem sobreposição de classes, confirmando a discriminação total observada nas métricas de desempenho. O FTIR, embora ainda apresente alguma sobreposição, mostra tendência de agrupamento coerente entre réplicas de uma mesma marca, especialmente nas classes ANT, STE e SKO, sugerindo reprodutibilidade das assinaturas moleculares.

A análise dos pesos das variáveis latentes (w) (Figura 15) permite identificar as regiões espectrais que mais contribuíram para a discriminação.

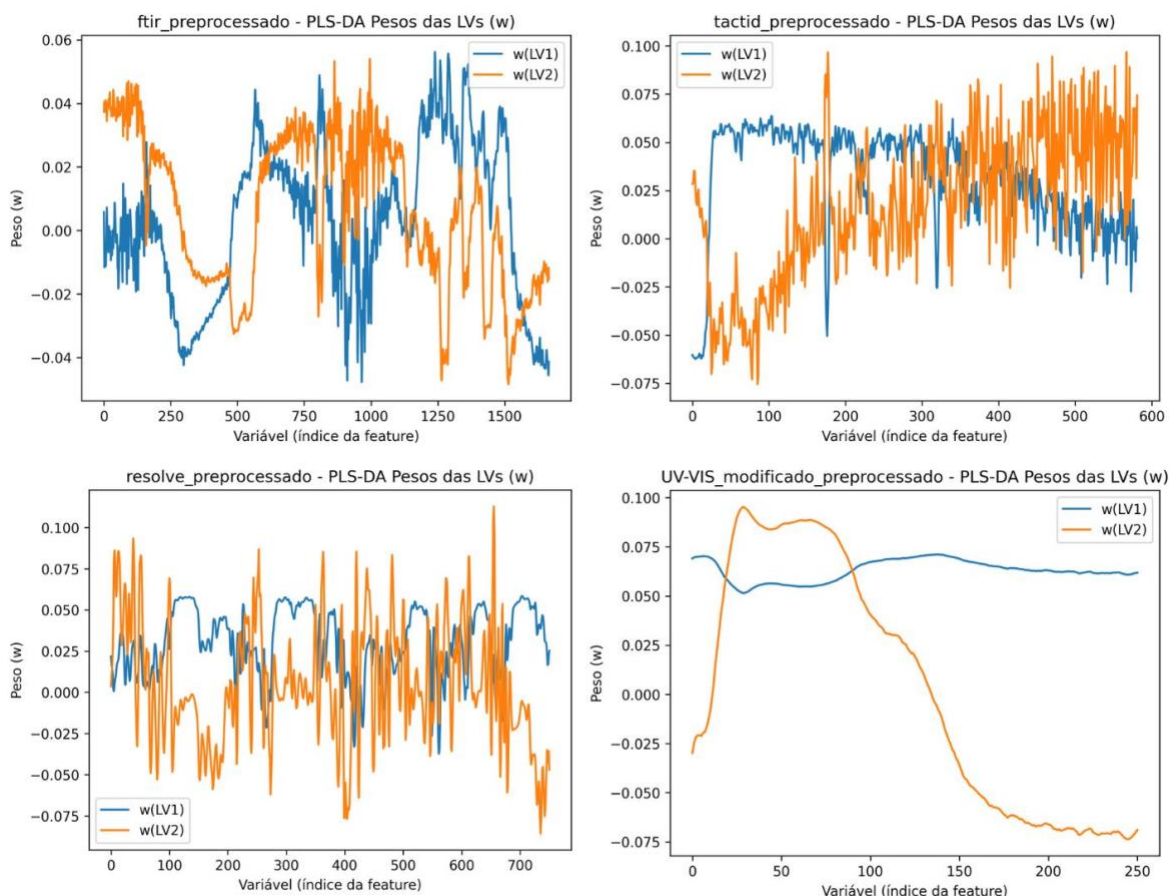


Figura 16. gráficos de escores LV1 \times LV2 para FTIR, Resolve, Tactid e UV-VIS.

No FTIR, os pesos de maior magnitude concentram-se entre 1200 e 1000 cm^{-1} , associados a vibrações C–O e C–C de polissacarídeos e álcoois, e entre 1750 e 1600 cm^{-1} , correspondentes a ligações C=O de ésteres e ácidos orgânicos. Nos espectros Raman, tanto no Resolve quanto no TacticID, as regiões mais relevantes situam-se entre 1000 e 1600 cm^{-1} , abrangendo vibrações aromáticas e deformações C=C, associadas a compostos fenólicos do malte e do lúpulo. Já no UV-Vis, os pesos indicam que as principais variáveis discriminantes se encontram na região do ultravioleta, entre 260 e 300 nm, região de absorção típica de pigmentos e polifenóis, responsáveis pela coloração e estabilidade oxidativa da cerveja. Essa coerência entre as regiões espectrais e os componentes químicos reforça a interpretação físico-química dos modelos e valida o caráter químico das variações observadas.

As matrizes de confusão de validação (Figura 16) mostram o desempenho dos modelos de PLS-DA.

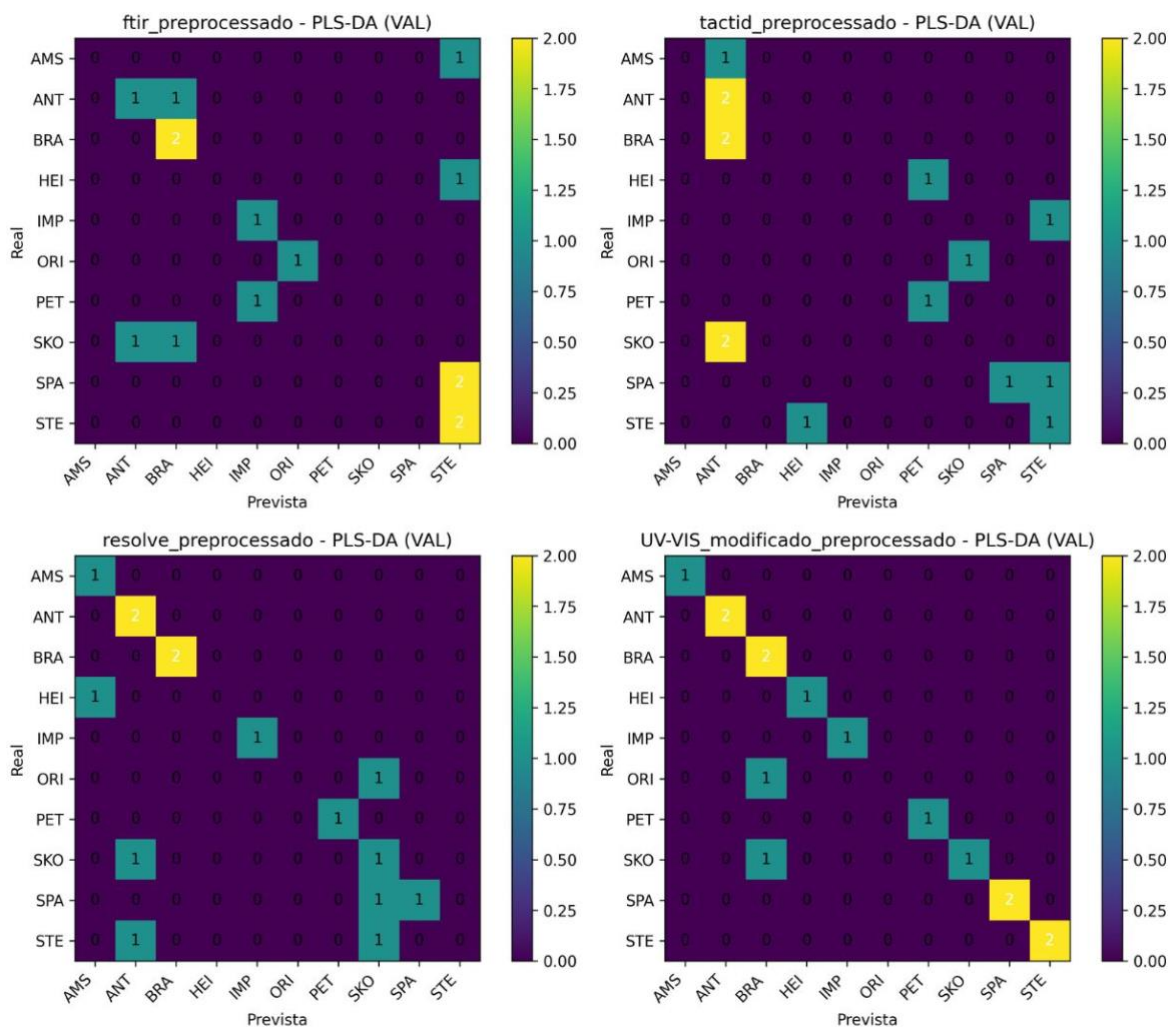


Figura 17. Matrizes de confusão para FTIR, Resolve, Tactid e UV-VIS.

A análise das matrizes de confusão evidenciou que, embora o UV-Vis tenha apresentado elevado desempenho na classificação das amostras, algumas classes ainda foram confundidas, especialmente quando aplicado o PLS-DA. Esse comportamento indica que a modelagem baseada em variáveis latentes nem sempre consegue extrair, de forma eficaz, os padrões discriminantes

necessários, sobretudo em cenários com alta semelhança entre as marcas e presença de ruído experimental.

Nesse contexto, o LDA mostrou-se mais robusto, uma vez que opera diretamente sobre as matrizes de dispersão e maximiza a separação entre grupos, capturando diferenças sutis que se perdem no processo de projeção latente do PLS-DA. Assim, os resultados reforçam que, para este conjunto de dados, estratégias lineares clássicas, quando associadas a pré-processamento adequado, podem superar métodos mais complexos, especialmente na classificação de cervejas por UV-Vis. Conclui-se, portanto, que o LDA se mostrou mais adequado para este cenário, destacando a importância de selecionar o método quimiométrico de acordo com as características do dado, e não apenas pela complexidade da técnica empregada.

4.11 Resultados e Perspectivas

Em uma etapa preliminar, investigou-se o potencial das técnicas espectroscópicas FTIR, Raman TacticID e Raman Resolve para predição quantitativa de parâmetros físico-químicos de cervejas por meio de modelagem quimiométrica. Para cada variável de interesse, incluindo extrato aparente (Ea), extrato real (Er), ADF, densidade, CO₂, graus plato (P) e teor alcoólico, foi selecionado o modelo de regressão com melhor desempenho dentre diferentes algoritmos avaliados, contemplando tanto métodos lineares quanto não lineares. Dessa forma, cada parâmetro e cada técnica foram ajustados individualmente, permitindo identificar a abordagem mais adequada para capturar as relações químicas presentes nos espectros. Os resultados estão resumidos na tabela 4 a seguir.

Tabela 4 – Dados de modelagem para parâmetros quantitativos.

Técnica	Parâmetro	Modelagem	R
FTIR	ADF	ElasticNet	0,67
	Ea		0,81
	Er		0,81
	densidade		0,81
RESOLVE	ADF	SVR_RBF	0,64
	Ea		0,70
	Er		0,68
	densidade		0,70
TACTID	ADF	PCR	0,42
	CO2		0,14
	Ea		0,45
	Er		0,47
	P		0,53
	densidade		0,44
	teor_alcoolico		0,49

No caso dos dados de FTIR, observou-se desempenho expressivo para alguns parâmetros, com destaque para os modelos do tipo Elastic Net, que apresentaram correlações de $R = 0,81$ para Ea, Er e densidade, e $R = 0,67$ para ADF. Entretanto, para CO₂ e graus plato (P), mesmo após pré-processamento, o modelo apresentou $R \approx 0,00$, indicando ausência de relação linear relevante entre os sinais espectrais e esses atributos nessas condições experimentais, como pode ser observado na figura 17.

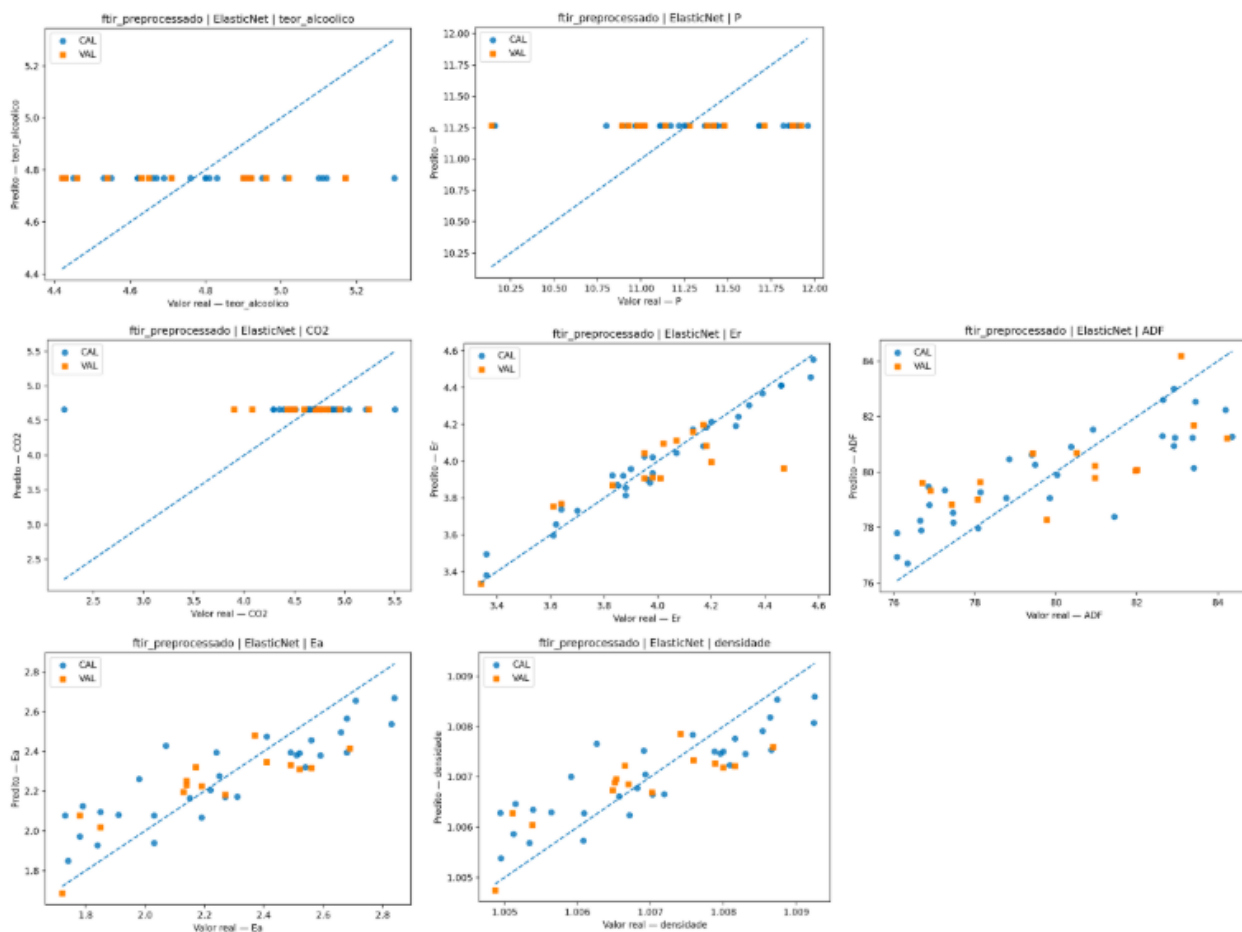


Figura 18. gráficos de regressão de parâmetros físico-químicos para dados do FTIR.

Para o equipamento Raman Resolve, o melhor desempenho foi obtido com o modelo SVR com kernel RBF, sugerindo comportamento não linear dos dados. Foram observadas correlações de $R = 0,70$ para Ea e densidade, $R = 0,68$ para Er e $R = 0,64$ para ADF, valores que indicam potencial preditivo moderado. Esses resultados revelam que, apesar da maior complexidade do sinal Raman, há informação química relevante capaz de sustentar a modelagem quantitativa de alguns parâmetros.

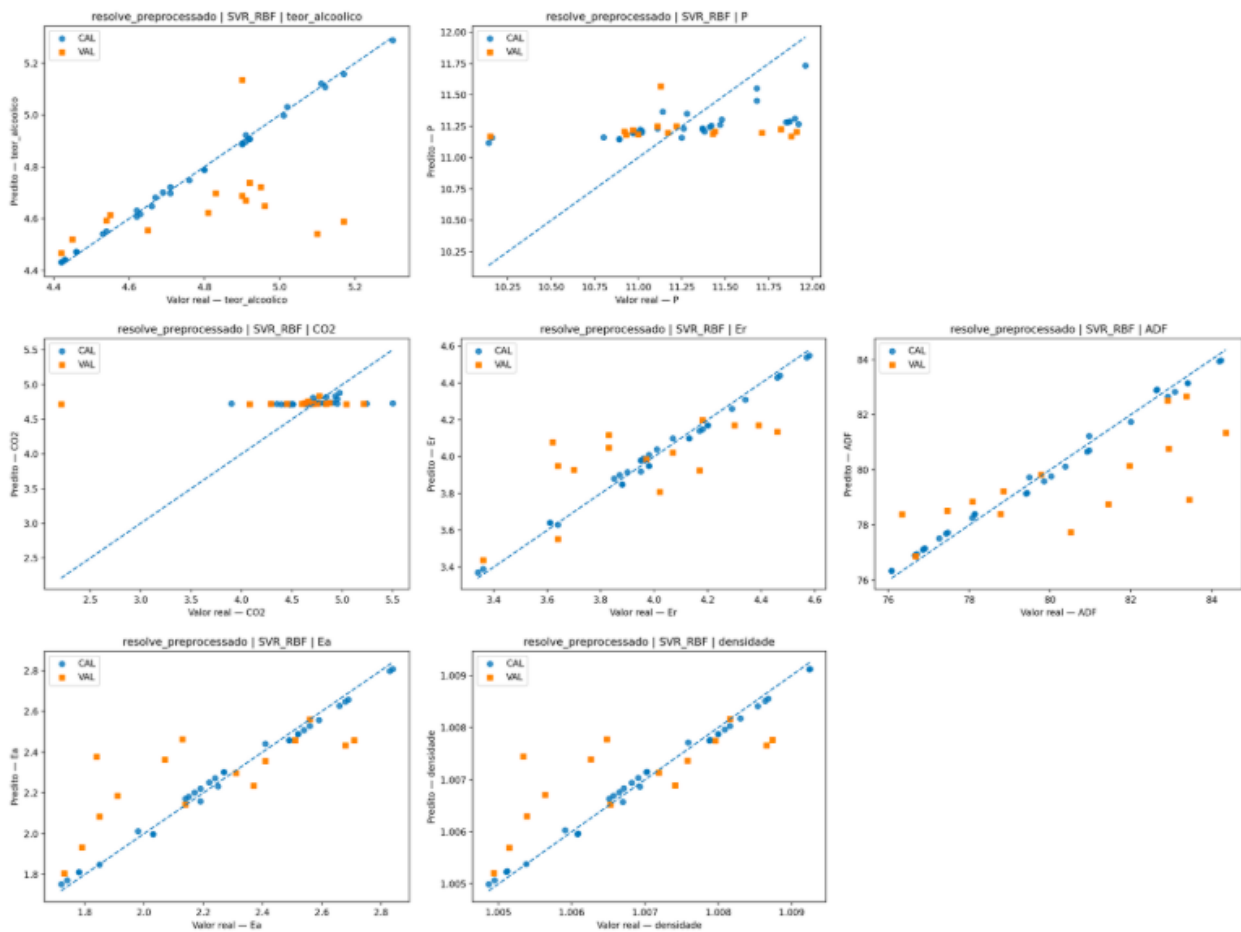


Figura 19. gráficos de regressão de parâmetros físico-químicos para dados do Raman Resolve.

Já para o Raman TacticID, o melhor desempenho foi obtido com PCR, embora com correlações mais modestas. Os valores observados variaram de $R = 0,14$ para CO_2 , $R = 0,42$ para ADF, $R = 0,44$ para densidade, $R = 0,45$ para Ea, $R = 0,47$ para Er, $R = 0,49$ para teor alcoólico e $R = 0,53$ para graus plato (P). Esses resultados indicam que, apesar do potencial qualitativo da técnica para diferenciação de amostras, sua capacidade preditiva quantitativa nesta etapa inicial foi limitada.

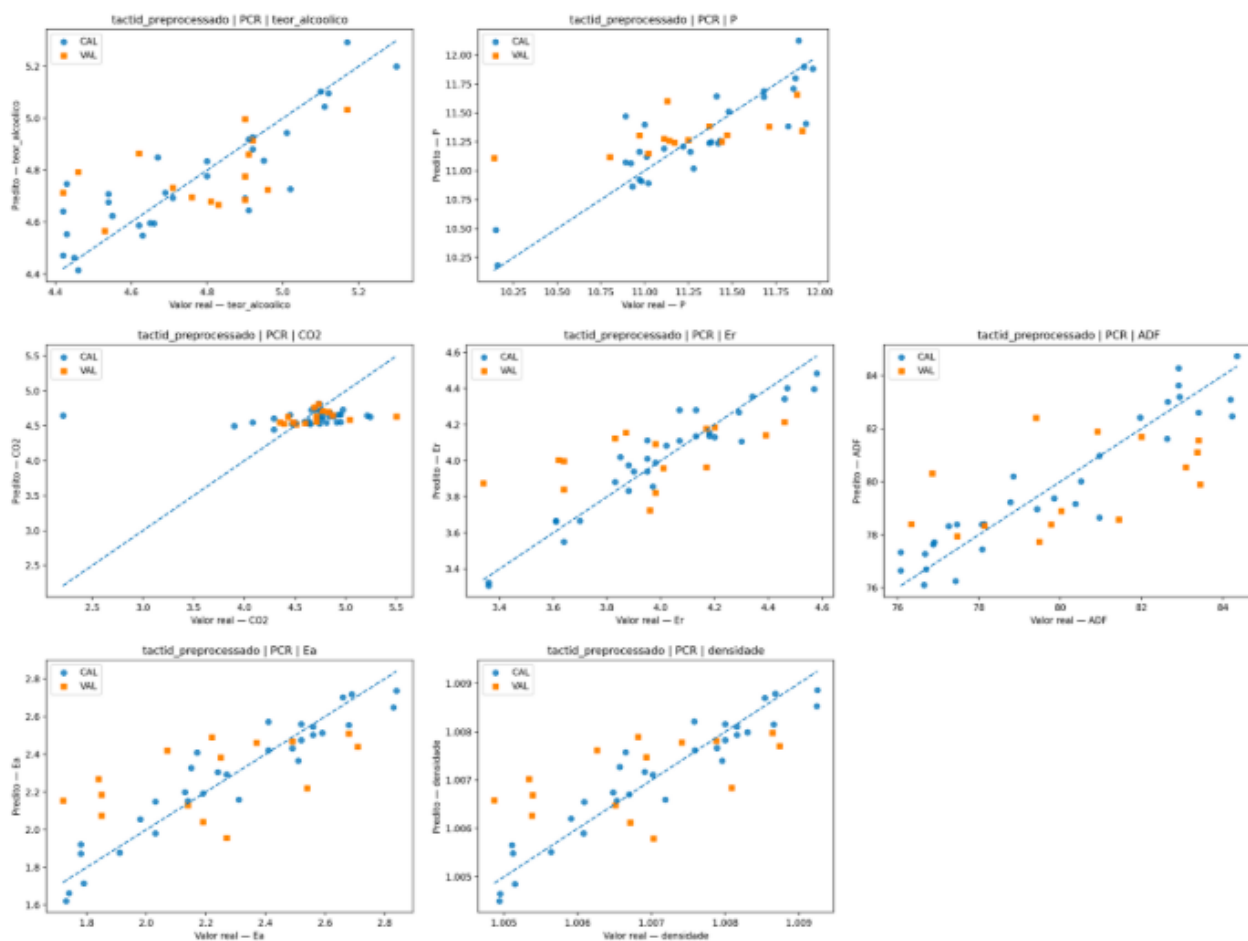


Figura 20. gráficos de regressão de parâmetros físico-químicos para dados do Raman Tactid.

Em conjunto, os resultados mostram que, embora ainda incipientes e com valores de correlação abaixo do ideal para aplicações analíticas, alguns parâmetros apresentaram relações mensuráveis com os espectros, especialmente para FTIR e Raman Resolve. Assim, estes resultados reforçam que a modelagem quantitativa a partir de espectros apresenta potencial, mas exige maior número de amostras, otimização de pré-processamentos, seleção de variáveis e ajustes de hiper parâmetros para amadurecer enquanto abordagem robusta.

De forma geral, essa etapa funcionou como uma avaliação preliminar do potencial das técnicas espectroscópicas e dos diferentes algoritmos de regressão. Embora ainda introdutória, ela fornece

subsídios importantes e indica caminhos promissores para o desenvolvimento futuro de modelos mais robustos, com maior volume de dados.

5 CONCLUSÕES

Este trabalho demonstrou que a combinação de técnicas instrumentais e métodos quimiométricos é uma abordagem eficaz para a caracterização e diferenciação de cervejas industrializadas, com foco na identificação de possíveis fraudes e trocas de rótulos, um problema crescente no mercado brasileiro e relevante para o consumidor, a indústria e órgãos de fiscalização. A análise de dados físico-químicos, espectrais e multivariados permitiu revelar diferenças químicas sutis entre amostras com composição declarada semelhante, evidenciando que informações presentes no produto ultrapassam aquelas fornecidas no rótulo e podem ser exploradas para fins de autenticação.

Entre as técnicas avaliadas, a espectrofotometria UV-Vis apresentou o melhor desempenho isolado, sendo capaz de separar marcas e agrupar amostras de uma mesma marca com elevada precisão, superando plataformas mais complexas, como Raman e FTIR, bem como o analisador de bebidas em alguns cenários. Este achado possui forte relevância prática, pois destaca que laboratórios com infraestrutura simples, incluindo setores de controle de qualidade, universidades e órgãos públicos, podem implementar métodos rápidos e acessíveis de verificação de autenticidade de cervejas. Assim, o UV-Vis se consolida como uma solução robusta, reprodutível, operacionalmente simples e com alto potencial de impacto para detecção de falsificações no mercado nacional.

A aplicação de LDA (especialmente com shrinkage) destacou-se entre os modelos supervisionados, frequentemente superando o PLS-DA na classificação das amostras. Este comportamento reforça que métodos lineares clássicos, quando bem pré-processados e aplicados a

dados com diferenças químicas discretas, podem oferecer desempenho superior a técnicas baseadas em variáveis latentes. A etapa exploratória com regressão multivariada também revelou correlações relevantes entre espectros e parâmetros físico-químicos, especialmente para FTIR e Raman Resolve, indicando que há potencial para, futuramente, desenvolver modelos quantitativos não destrutivos, expandindo o papel das técnicas espectroscópicas para além da autenticação qualitativa.

Reconhece-se como limitação deste trabalho a medição Raman através das embalagens, cuja interferência óptica do vidro (especialmente âmbar e verde) introduziu variabilidade experimental e afetou a separação das classes. No entanto, essa limitação prática é também uma contribuição metodológica, pois orienta protocolos futuros e ressalta a importância da padronização das condições de aquisição espectral para aplicações forenses e industriais.

Como perspectivas, destaca-se a ampliação da base amostral e da diversidade de rótulos avaliados, padronização de recipientes ou aplicação direta em cubetas para minimizar interferências ópticas, otimização e validação externa dos modelos quimiométricos e continuidade do desenvolvimento de modelos quantitativos baseados em espectroscopia e aprendizado de máquina.

Assim, este trabalho contribui tanto para o avanço científico no campo da quimiometria aplicada a bebidas, quanto para o fortalecimento de estratégias de controle de qualidade e combate à fraude no mercado cervejeiro brasileiro, reforçando que ciência acessível, padronizada e bem aplicada pode gerar ferramentas práticas, eficientes e socialmente relevantes.

6 REFERÊNCIAS BIBLIOGRÁFICAS

AGILENT TECHNOLOGIES. Resolve handheld spectrometer. Disponível em: <https://www.agilent.com/en/product/molecular-spectroscopy/raman-spectroscopy/handheld-raman-chemical-detection-systems/resolve-handheld-spectrometer>. Acesso em: 28 out. 2025.

ALCÁZAR, A. et al. Multivariate characterisation of beers according to their mineral content. *Talanta*, v. 57, n. 1, p. 45-52, 2002.

ANDRÉS-IGLESIAS, C. et al. New trends in beer flavor compound analysis. *Journal of the Science of Food and Agriculture*, v. 95, n. 8, p. 1571–1576, 1 jun. 2015. DOI: <https://doi.org/10.1002/jsfa.6905>.

ANTON PAAR. Beer Monitor. Disponível em: <https://www.anton-paar.com/br-pt/produtos/detalhes/beer-monitor/>. Acesso em: 28 out. 2025.

ASSOCIAÇÃO BRASILEIRA DE CERVEJA ARTESANAL (ABRACERVA). O que é cerveja artesanal. Disponível em: <https://abrase.com.br/noticias/noticias/afinal-o-que-define-se-uma-cervejaria-e-artesanal/>. Acesso em: 6 jun. 2025.

BALLABIO, D.; CONSONNI, V. Classification tools in chemistry. Part 1: linear models. PLS-DA. *Analytical Methods*, v. 5, n. 16, p. 3790-3798, 2013. Acesso em: 6 jun. 2025.

BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy*, v. 43, n. 5, p. 772-777, 1989.

BERMAN, B. Strategies to detect and reduce counterfeiting activity. *Business Horizons*, v. 51, n. 3, p. 191-199, 2008.

BERRAR, D. Cross-validation. In: RANGARAJAN, A.; DAS, S.; LIU, J. (org.). *Encyclopedia of bioinformatics and computational biology*. Amsterdam: Elsevier, 2019. p. 542–545. DOI: <https://doi.org/10.1016/B978-0-12-809633-8.20349-X>.

BORRÀS, E. et al. Authenticity assessment of coffee blends using untargeted HPLC–UV fingerprinting and chemometrics. *Analytica Chimica Acta*, v. 870, p. 68-76, 2015.

BRASIL. Decreto-Lei nº 2.848, de 7 de dezembro de 1940. Código Penal. Disponível em: https://www.planalto.gov.br/ccivil_03/decreto-lei/del2848compilado.htm. Acesso em: 2 maio 2025.

BRASIL. Ministério da Agricultura e Pecuária. Anuário da Cerveja 2024. Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/noticias/com-1-847-cervejarias-registradas-no-brasil-setor-cresce-6-8-em-2023>. Acesso em: 6 jun. 2025.

BRASIL. Ministério da Agricultura e Pecuária. Secretaria de Defesa Agropecuária. Departamento de Inspeção de Produtos de Origem Vegetal. Anuário da cerveja 2025: ano referência 2024. Brasília: MAPA, 2025. 68 p. il. color., gráficos, mapas, tabelas. Formato digital (PDF). ISBN 978-85-7991-355-6. Acesso em: 7 jan. 2026.

BRERETON, R. G. et al. Chemometrics in analytical chemistry—part II: modeling, validation, and applications. *Analytical and Bioanalytical Chemistry*, v. 410, p. 6691–6704, 2018. DOI: 10.1007/s00216-018-1283-4.

BROWNE, M. W. Cross-validation methods. *Journal of Mathematical Psychology*, v. 44, n. 1, p. 108–132, 2000. DOI: <https://doi.org/10.1006/jmps.1999.1279>.

BRYAN, M. A.; HILL, A. E. Worldwide illicit and counterfeit alcoholic spirits: problem, detection, and prevention. *Journal of the American Society of Brewing Chemists*, v. 82, n. 3, p. 187–203, 2024.

DOI: <https://doi.org/10.1080/03610470.2024.2319934>.

BUDWEISER. Produtos. Disponível em: <https://www.budweiser.com/pt-br/produto/produtos#269ml>. Acesso em: 12 maio 2025.

CASADO, A. Chromatographic fingerprinting: an innovative approach for food “identification” and food authentication – a tutorial. *Analytica Chimica Acta*, v. 915, p. 1–23, 2016. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0003267016300204>. Acesso em: 31 maio 2025.

CLARK, R. J. H.; LONG, D. A. *Raman spectroscopy*. New York: McGraw-Hill, 1977.

COELHO NETO, D. M. et al. Study of the chemical profile of Brazilian beers: an assessment between artisanal and industrial drinks. *Química Nova*, v. 45, n. 5, p. 518–530, 2022. DOI: <https://doi.org/10.21577/0100-4042.20170857>.

CONTRERAS, U. et al. Screening method for identification of adulterate and fake tequilas by using UV–VIS spectroscopy and chemometrics. *Food Research International*, v. 43, n. 10, p. 2356–2362, 2010. DOI: <https://doi.org/10.1016/j.foodres.2010.08.012>.

CUADROS-RODRÍGUEZ, L. et al. Authenticity assessment of coffee blends using untargeted HPLC–UV fingerprinting and chemometrics. *Analytica Chimica Acta*, v. 870, p. 68–76, 2015.

DA COSTA, N. L. et al. A review on the application of chemometrics and machine learning algorithms to evaluate beer authentication. *Food Analytical Methods*, v. 14, n. 1, p. 136–155, 1 jan. 2021. DOI: <https://doi.org/10.1007/s12161-020-01864-7>.

DA SILVA, L. A. et al. ¹H NMR spectroscopy combined with multivariate data analysis for differentiation of Brazilian lager beer according to brewery. *European Food Research and Technology*, v. 245, n. 11, p. 2365–2372, 1 nov. 2019. DOI: <https://doi.org/10.1007/s00217-019-03354-5>.

DE ALMEIDA, F. S. et al. Use of Fourier transform infrared spectroscopy to monitor sugars in the beer mashing process. *Food Chemistry*, v. 263, p. 112–118, 15 out. 2018. DOI: <https://doi.org/10.1016/j.foodchem.2018.04.109>.

DIAMANTIDIS, N. A.; KARLIS, D.; GIAKOUMAKIS, E. A. Unsupervised stratification of cross-validation for accuracy estimation. *Artificial Intelligence*, v. 116, n. 1–2, p. 1–16, 2000. DOI: [https://doi.org/10.1016/S0004-3702\(99\)00095-3](https://doi.org/10.1016/S0004-3702(99)00095-3).

DUARTE, I. F. et al. Multivariate analysis of NMR and FTIR data as a potential tool for the quality control of beer. *Journal of Agricultural and Food Chemistry*, v. 52, n. 5, p. 1031–1038, 2004. DOI: <https://doi.org/10.1021/jf034790p>.

ENGEL, J. et al. Confirmation of brand identity of a Trappist beer by mid-infrared spectroscopy coupled with multivariate data analysis. *Talanta*, v. 99, p. 426–432, 15 set. 2012. DOI: <https://doi.org/10.1016/j.talanta.2012.06.005>.

ESTEKI, M. et al. A review on the application of chromatographic methods, coupled to chemometrics, for food authentication. *Food Control*, v. 93, p. 165–182, 2018. DOI: 10.1016/j.foodcont.2018.06.015.

FERREIRA, A. C. PM descobre fábrica de falsificação de cervejas em MG e cinco pessoas são presas. 25 ago. 2023. Disponível em: <https://g1.globo.com>. Acesso em: 14 abr. 2025.

FOLHA DE SÃO PAULO. Facções tiveram receita maior que da Ambev com bebida falsificada em 2022, estima entidade. 28 dez. 2024. Disponível em: <https://www1.folha.uol.com.br>. Acesso em: 14 abr. 2025.

G1. Polícia detém 12 suspeitos de falsificação de cerveja em galpão de Ribeirão Preto, SP. 16 jan. 2023. Disponível em: <https://g1.globo.com>. Acesso em: 14 abr. 2025.

GIANSANTE, L.; DI VINCENZO, D.; BIANCHI, G. Classification of monovarietal Italian olive oils by unsupervised (PCA) and supervised (LDA) chemometrics. *Journal of the Science of Food and Agriculture*, v. 83, n. 9, p. 905–911, 2003. DOI: <https://doi.org/10.1002/jsfa.1508>.

GONZALEZ VIEJO, C.; FUENTES, S.; HERNANDEZ-BRENES, C. Smart detection of faults in beers using near-infrared spectroscopy, a low-cost electronic nose and artificial intelligence. *Fermentation*, v. 7, n. 3, p. 117, 2021. Disponível em: <https://doi.org/10.3390/fermentation7030117>.

GORDON, R.; CHAPMAN, J.; POWER, A. C.; COZZOLINO, D. Unfrazzled by fizziness: identification of beers using attenuated total reflectance mid-infrared spectroscopy and multivariate analysis. *Food Analytical Methods*, v. 11, p. 1–9, 2018.

GRANATO, D. et al. Characterization of Brazilian lager and brown ale beers based on color, phenolic compounds, and antioxidant activity using chemometrics. *Journal of the Science of Food and Agriculture*, v. 91, n. 3, p. 563–571, fev. 2011. DOI: <https://doi.org/10.1002/jsfa.4222>.

GRUBERTT, B.; TUFANI, V. Galpão de falsificação de cerveja é fechado no Rio; rótulos e até tampinhas eram trocados. 21 jun. 2022. *G1*. Disponível em: <https://g1.globo.com>. Acesso em: 14 abr. 2025.

HAIR, J. F. et al. *Análise multivariada de dados*. 6. ed. Porto Alegre: Bookman, 2009.

HEINEKEN. No fake news. Heineken Brasil, [s.d.]. Disponível em: <https://www.heineken.com/br/pt/no-fake-news>. Acesso em: 12 maio 2025.

HIBBERT, D. B. Vocabulary of concepts and terms in chemometrics (IUPAC Recommendations 2016). *Pure and Applied Chemistry*, v. 88, n. 4, p. 407–443, 2016. Disponível em: <https://doi.org/10.1515/pac-2015-0605>. Acesso em: 6 jun. 2025.

JACK, B. Food fraud: protecting European consumers through effective deterrence. *European Public Law*, v. 24, n. 1, 2018.

JEONG, S. et al. High-resolution NMR spectroscopy for the classification of beer. *Journal of Food Science*, v. 82, n. 5, p. 1234–1240, 2017.

JIAO, Y. et al. Preprocessing methods for near-infrared spectrum calibration. *Journal of Chemometrics*, v. 34, n. 11, p. e3306, 2020. DOI: <https://doi.org/10.1002/cem.3306>.

JIMÉNEZ-CARVELO, A. M. et al. PLS-DA vs sparse PLS-DA in food traceability. *Talanta*, v. 202, p. 349–356, 2019. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0039914020311954>.

JIMÉNEZ-CARVELO, A. M.; GONZÁLEZ-CASADO, A.; BAGUR-GONZÁLEZ, M. G.; CUADROS-RODRÍGUEZ, L. Alternative data mining/machine learning methods for the analytical evaluation of food quality and authenticity – a review. *Food Research International*, v. 122, p. 25–39, 2019. DOI: <https://doi.org/10.1016/j.foodres.2019.03.063>.

JOLLIFFE, I. T. *Principal component analysis*. 2. ed. New York: Springer, 2002. (Springer Series in Statistics).

KNUDSON, E. J.; SIEBERT, K. J. Application of ion chromatography to beer, wort, and brewing water. *Journal of the American Society of Brewing Chemists*, v. 42, n. 2, p. 65–70, abr. 1984. DOI: <https://doi.org/10.1094/asbcj-42-0065>.

KUBALLA, T. et al. Detection of counterfeit brand spirits using ¹H NMR fingerprints in comparison to sensory analysis. *Food Chemistry*, v. 245, p. 112–118, 15 abr. 2018. DOI: <https://doi.org/10.1016/j.foodchem.2017.10.065>.

LACHENMEIER, D. W. Rapid quality control of spirit drinks and beer using multivariate data analysis of Fourier transform infrared spectra. *Food Chemistry*, v. 101, n. 2, p. 825–832, 2007. DOI: <https://doi.org/10.1016/j.foodchem.2005.12.030>.

LEE, L. C.; LIONG, C.-Y.; JEMAIN, A. A. A contemporary review on data preprocessing (DP) practice strategy in ATR-FTIR spectrum. *Chemometrics and Intelligent Laboratory Systems*, v. 163, p. 64-75, 2017. DOI: <https://doi.org/10.1016/j.chemolab.2017.02.008>.

MAHMOOD, N.; PETRACO, N.; HE, Y. Elemental fingerprint profile of beer samples constructed using 14 elements determined by inductively coupled plasma-mass spectrometry (ICP-MS): multivariation analysis and potential application to forensic sample comparison. *Analytical and Bioanalytical Chemistry*, v. 402, n. 2, p. 861–869, jan. 2012. DOI: <https://doi.org/10.1007/s00216-011-5452-y>.

MARTINEZ, A. M.; KAK, A. C. PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 23, n. 2, p. 228–233, 2001. DOI: <https://doi.org/10.1109/34.908974>.

METROHM. Tactic ID. Disponível em: https://www.metrohm.com/pt_br/produtos/espectroscopia-raman/tactic-id.html. Acesso em: 28 out. 2025.

MINISTÉRIO DA AGRICULTURA E PECUÁRIA (MAPA). *Anuário da Cerveja 2022*. [S.l.]: MAPA, 2022.

MISHRA, P. et al. New data preprocessing trends based on ensemble of multiple preprocessing techniques. *TrAC Trends in Analytical Chemistry*, v. 132, p. 116045, 2020. DOI: <https://doi.org/10.1016/j.trac.2020.116045>.

NEUNHOEFFER, M.; STERNBERG, S. How cross-validation can go wrong and what to do about it. *Political Analysis*, v. 27, n. 1, p. 101–106, 2019. DOI: <https://doi.org/10.1017/pan.2018.48>.

OECD. *Global trade in fakes*. [S.l.]: OECD, 2021. (Illicit Trade). DOI: 10.1787/74c81154-en. Disponível em: https://www.oecd.org/en/publications/global-trade-in-fakes_74c81154-en.html. Acesso em: 14 abr. 2025.

OLIVERI, P.; DOWNEY, G. Multivariate class modeling for the verification of food-authenticity claims. *Trends in Analytical Chemistry*, v. 35, p. 74–86, 2012. Disponível em: <https://doi.org/10.1016/j.trac.2012.02.005>. Acesso em: 6 jun. 2025.

ORGANIZAÇÃO MUNDIAL DO COMÉRCIO. Acordo sobre os Aspectos dos Direitos de Propriedade Intelectual Relacionados ao Comércio (TRIPS), Artigo 51. 1994.

PALMER, J. J. *How to brew: everything you need to know to brew great beer every time*. Boulder: Brewers Publications, 2017.

PAVAN, B. Pesquisa lista as marcas de cerveja preferidas dos brasileiros e as mais consumidas. *IstoÉ Dinheiro*, São Paulo, 19 nov. 2024. Disponível em: <https://istoedinheiro.com.br/pesquisa-lista-as-marcas-de-cerveja-preferidas-dos-brasileiros-e-as-mais-consumidas/>. Acesso em: 12 maio 2025.

REVISTA BEER ART. Cervejarias brasileiras: crescimento e exportação em 2023. Disponível em: <https://revistabeerart.com/news/cervejarias-brasil>. Acesso em: 6 jun. 2025.

RODRIGO, S.; YOUNG, S. D.; TALAVERANO, M.; BROADLEY, M. R. The influence of style and origin on mineral composition of beers retailing in the UK. *Journal of European Food Research and Technology*, v. 243, p. 931-939, 2017.

RUIZ-PEREZ, D. et al. So you think you can PLS-DA?. *BMC Bioinformatics*, v. 21, supl. 1, p. 2, 2020. DOI: <https://doi.org/10.1186/s12859-019-3310-7>.

SHAW, P. J. Multivariate statistics for the environmental sciences. London: Arnold, 2003.

SNOWDON, C. *Drinking in the shadow economy*. Disponível em: https://iea.org.uk/wp661content/uploads/2016/07/drinking%20in%20the%20Shadow%20Economy_0.pdf. Acesso em: 5 dez. 2024.

SMITH, E.; DENT, G. *Modern Raman spectroscopy: a practical approach*. 2. ed. Hoboken: John Wiley & Sons, 2019.

SOON, J. M.; MANNING, L. Developing anti-counterfeiting measures: the role of smart packaging. *Food Research International*, v. 123, p. 135-143, 2019.

SØRENSEN, C. A. G.; KHAKIMOV, B.; ENGELSEN, S. B. The use of rapid spectroscopic screening methods to detect adulteration of food raw materials and ingredients. *Food Research International*, v. 89, p. 74-90, 2016.

STONE, M. Cross-validation: a review. *Statistics: A Journal of Theoretical and Applied Statistics*, v. 9, n. 1, p. 127-139, 1978. DOI: <https://doi.org/10.1080/02331887808801414>.

SZYMAŃSKA, E. et al. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics*, v. 8, supl. 1, p. 3–16, 2012. DOI: <https://doi.org/10.1007/s11306-011-0330-6>.

TERRA. 36% das bebidas no Brasil são falsificadas: saúde e economia em risco. *Terra*, 15 abr. 2025. Disponível em: <https://www.terra.com.br/economia/36-das-bebidas-no-brasil-sao-falsificadas-saude-e-economia-em-risco,d273acf32011b06d86fb12876b87dca138yai9wv.html>.

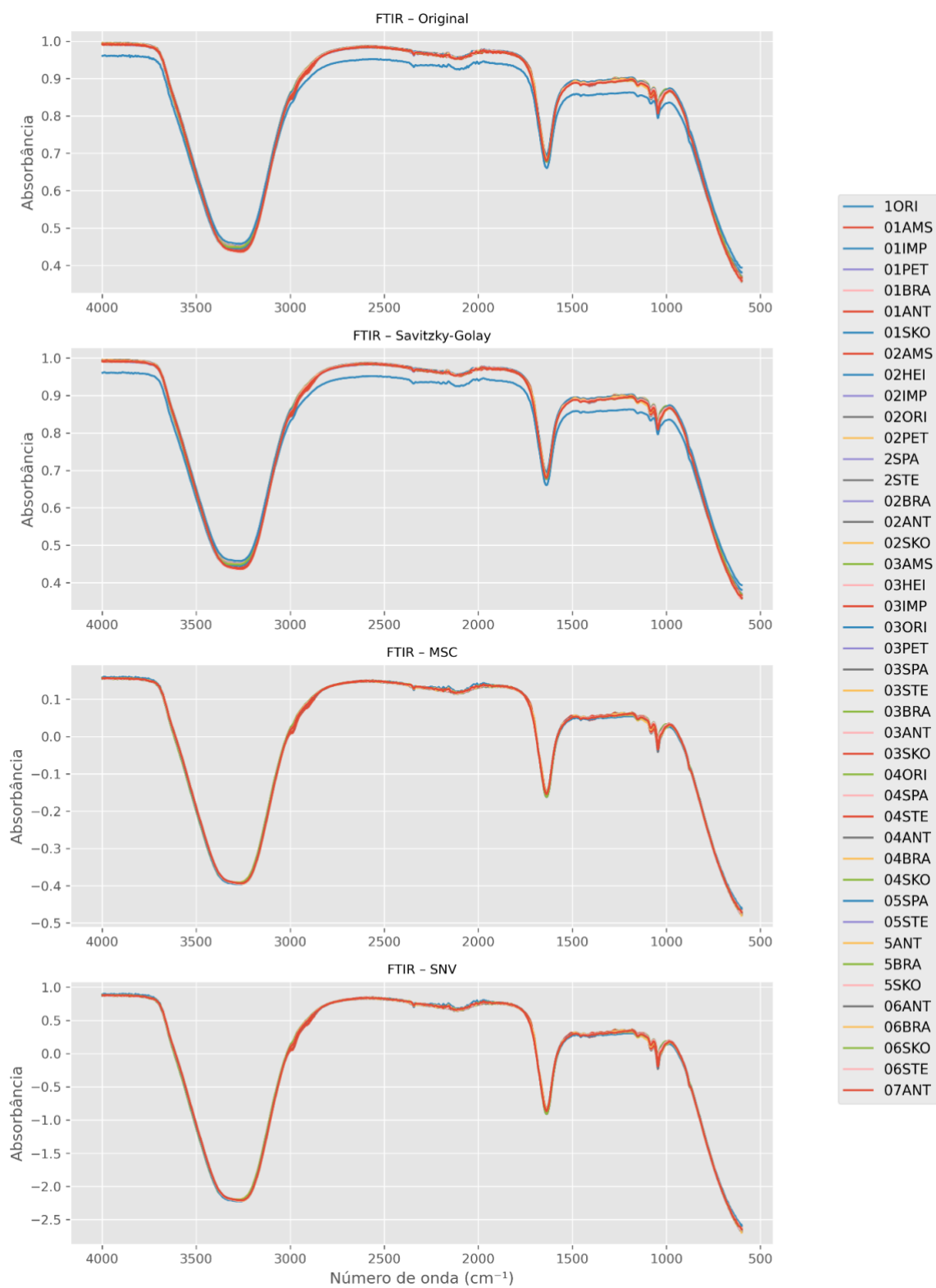
Acesso em: 12 maio 2025.

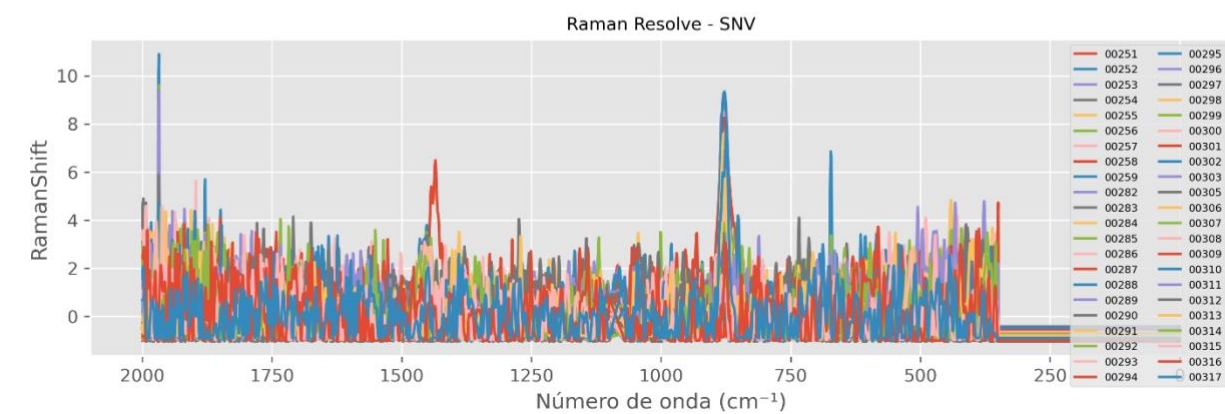
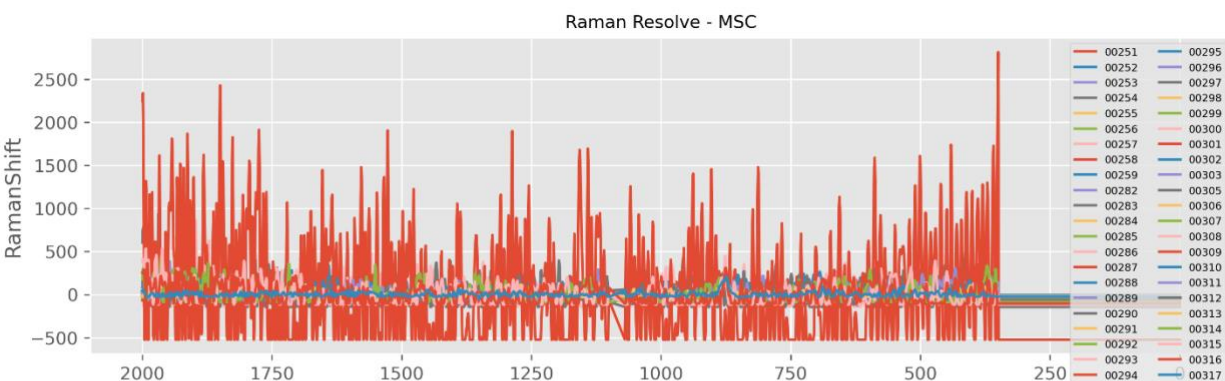
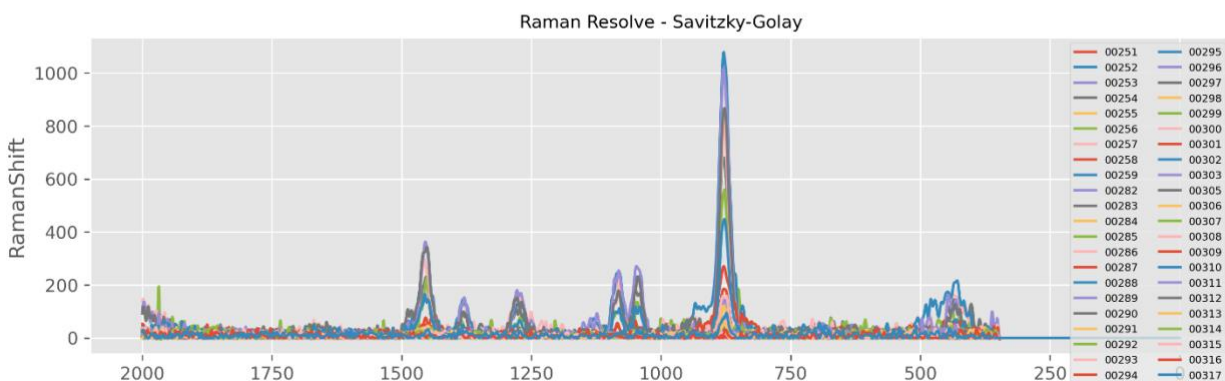
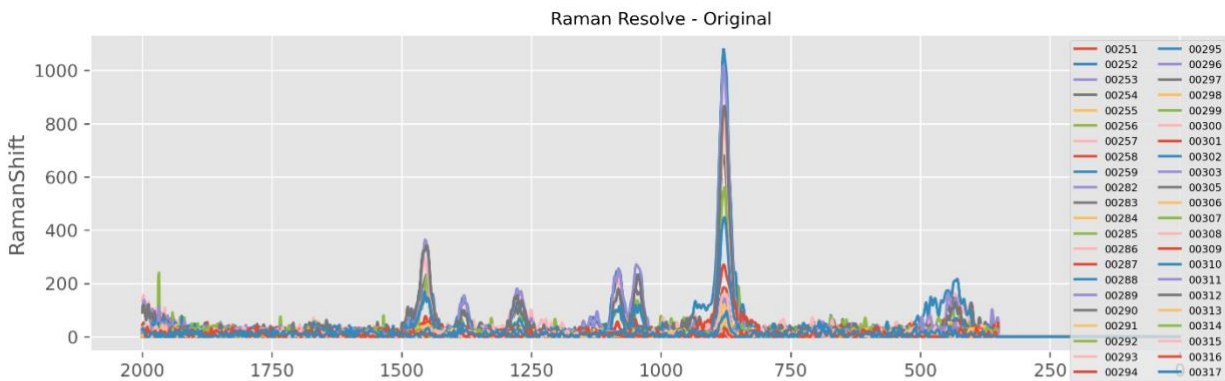
WEBSTER, R. Statistics to support soil research and their presentation. *European Journal of Soil Science*, v. 52, n. 2, p. 331-340, 2001.

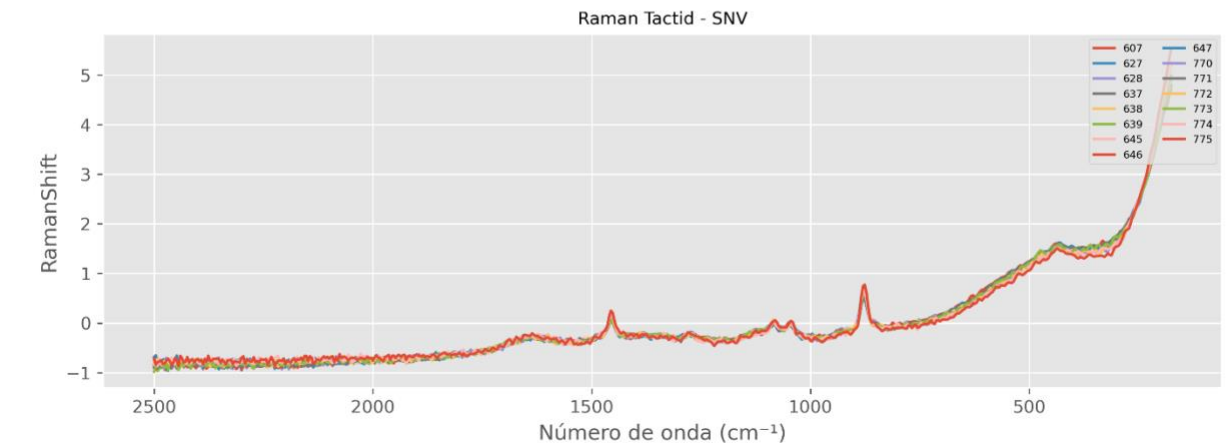
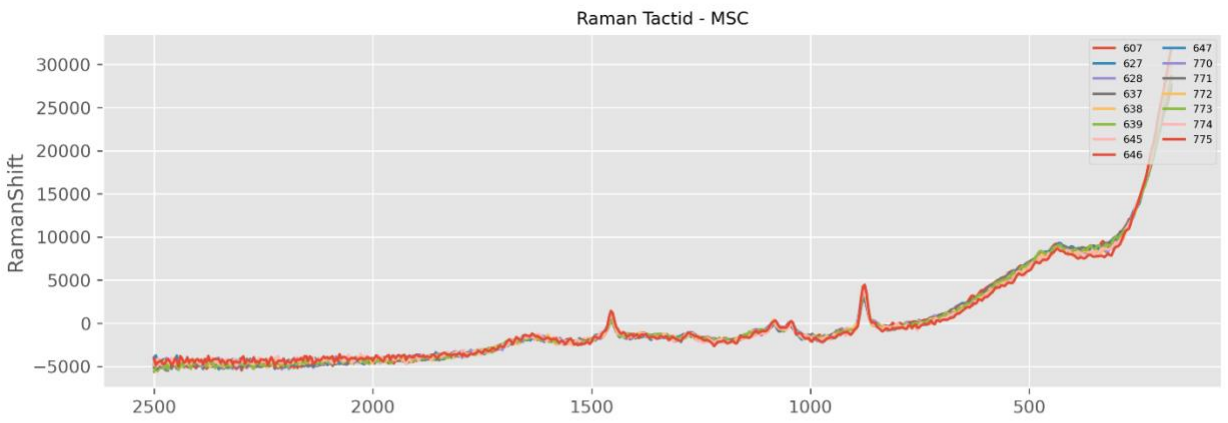
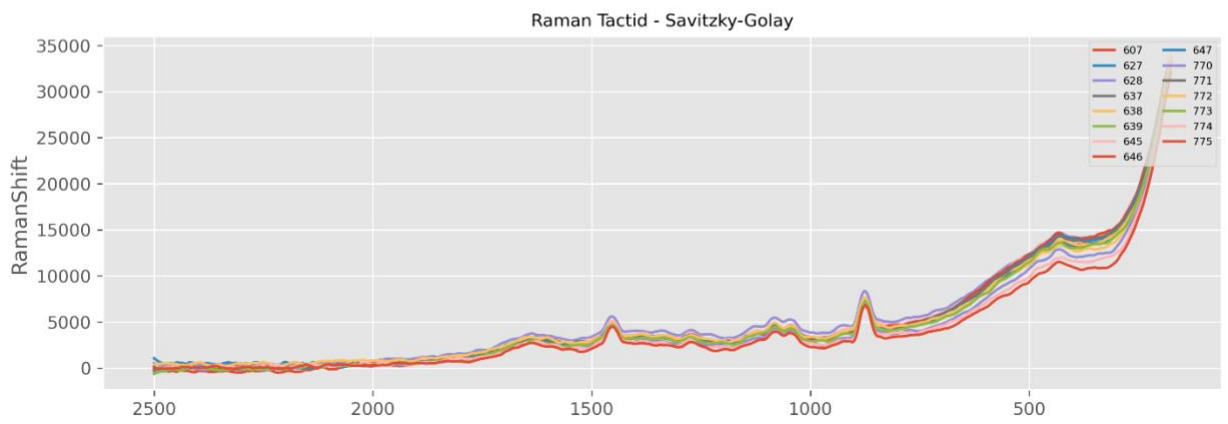
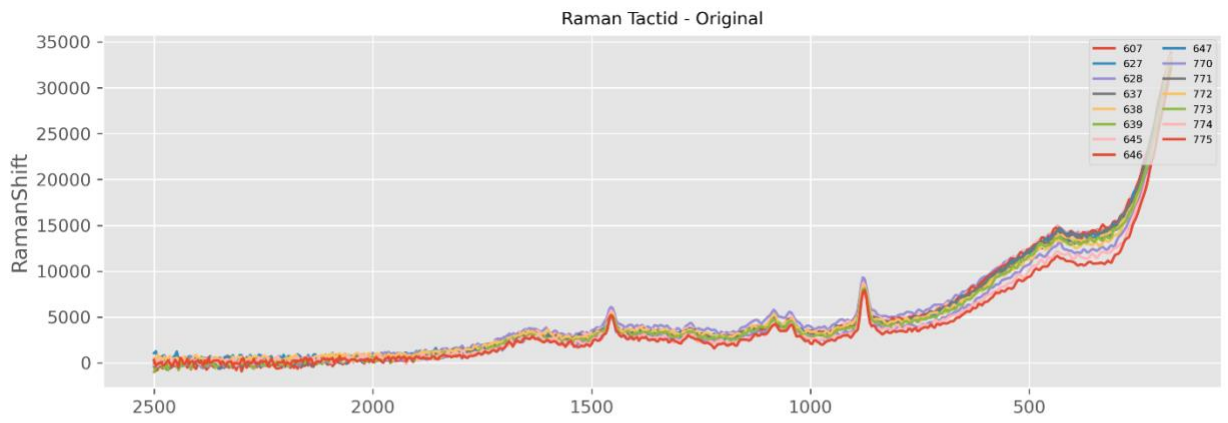
ZADBUKE, N. et al. Recent trends and future of pharmaceutical packaging technology. *Journal of Pharmacy and Bioallied Sciences*, v. 5, n. 2, p. 98-110, 2013.

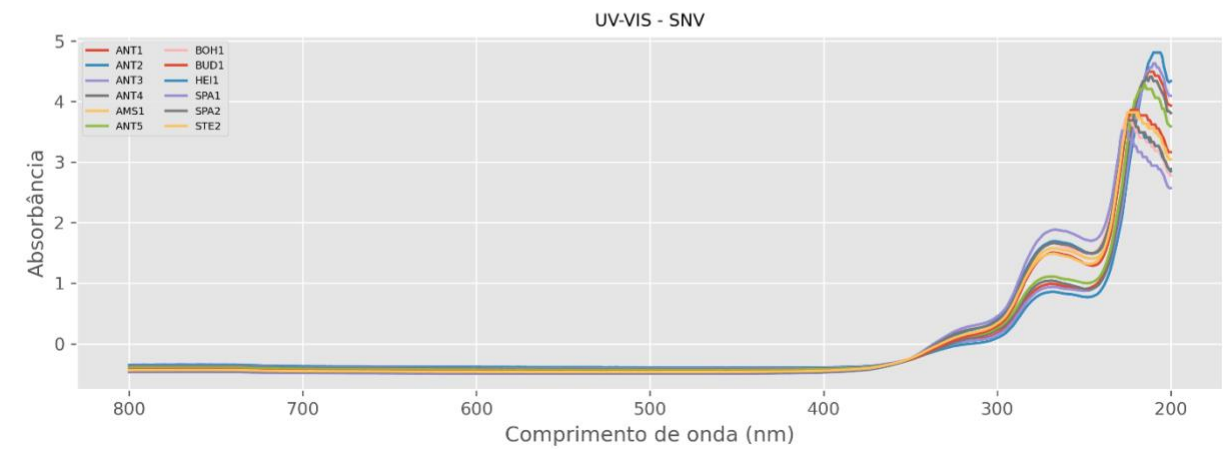
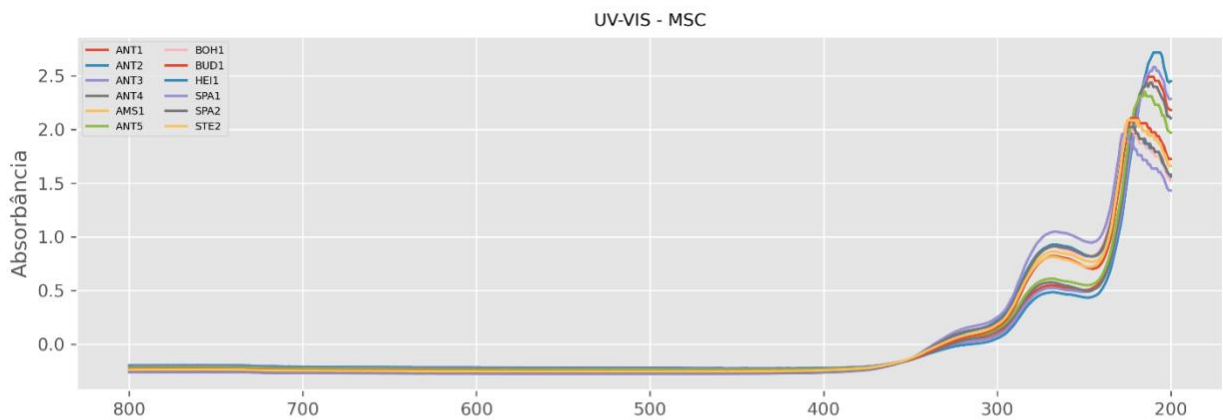
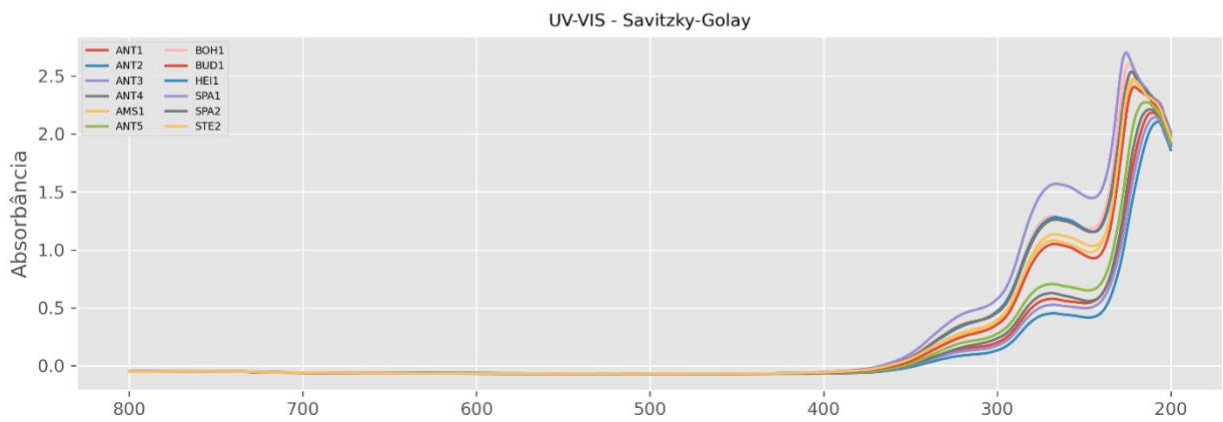
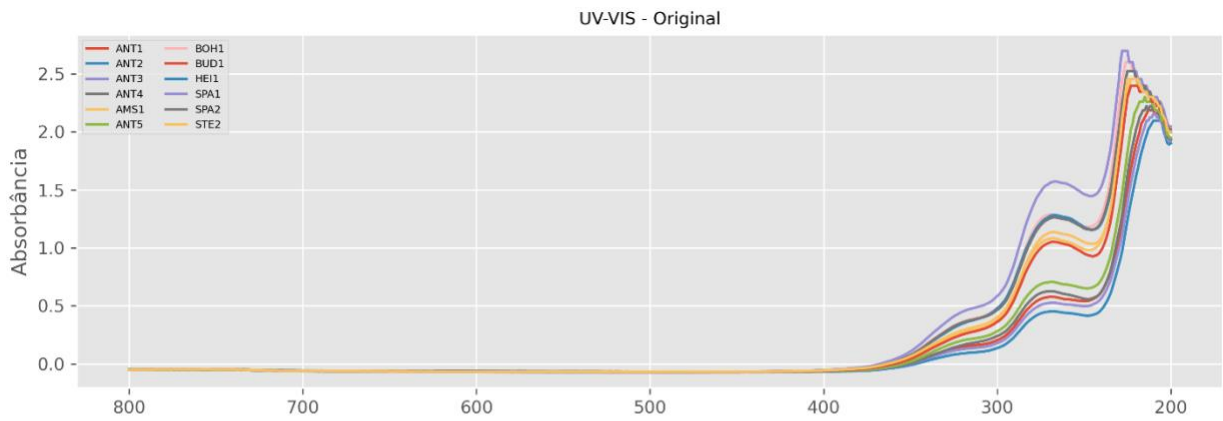
ZITKO, V. Principal component analysis in the evaluation of environmental data. *Marine Pollution Bulletin*, v. 28, n. 10, p. 597-600, 1994.

MATERIAL COMPLEMENTAR









**DECLARAÇÃO DE ORIGINALIDADE DE DISSERTAÇÃO DE MESTRADO OU TESE DE
DOUTORADO**

Declaro que a presente dissertação/tese é original, elaborada especialmente para este fim, não tendo sido apresentada para obtenção de qualquer título e que identifico e cito devidamente todas as autoras e todos os autores que contribuíram para o trabalho, bem como as contribuições oriundas de outras publicações de minha autoria.

Declaro estar ciente de que a cópia ou o plágio podem gerar responsabilidade civil, criminal e disciplinar, consistindo em grave violação à ética acadêmica.

Brasília, (dia) de (mês) de (ano).

Assinatura do/a discente: Mariana Moura Moutinho

Programa: Programa de Pós- Graduação em química - Instituto de Química - UnB

Nome completo: Mariana Moura Moutinho

Título do Trabalho: Análises Instrumentais e Quimiometria na Discriminação de
Cervejas Industriais Nacionais: Um Combate à Falsificação.

Nível: Mestrado () Doutorado

Orientador/a: Grace Ferreira Ghesti