ANÁLISE DISCRIMINANTE LINEAR NA CLASSIFICAÇÃO DE VARIEDADES DE VINHO COM BASE NAS CARACTERÍSTICAS QUÍMICAS

Maria Rita dos Reis Muniz [1]
Gabriel Cardoso [2]
Kuang Hongyu [3]
Thiago Duarte da Cruz [4]

RESUMO: A classificação precisa de variedades de vinho com base em suas características químicas tem sido uma preocupação da indústria vinícola e na pesquisa em vinicultura. Este artigo teve como objetivo desenvolver e um modelo de classificação para cultivares de vinho segundo suas características químicas, por meio da Análise Discriminante Linear (LDA). As informações foram obtidas no site UCI Machine Learning Repository em que possui diversos conjuntos de dados para aprendizado de máquina. Os dados das características químicas do vinho estão distribuídos em 13 variáveis e 178 observações, sendo esses dados provenientes de uma região vinícola da Itália. Para a análise dos dados, foi utilizada a técnica de Análise Discriminante Linear (LDA), que classificou a cultivar 1 com 100% da taxa de acerto, para a cultivar 2 a taxa ficou em 97, 18% e para a cultivar 3 foi apresentada uma taxa de 100% de acerto. Os resultados demonstram que o método de LDA foi eficaz para a classificação das cultivares em relação as suas características químicas, destacando seu potencial relevância para ser aplicado em estudos futuros.

Palavras-chave: matriz de confusão; função discriminante; modelo de classificação; método de Jackknife

LINEAR DISCRIMINANT ANALYSIS IN WINE VARIETAL CLASSIFICATION BASED ON CHEMICAL CHARACTERISTICS

ABSTRACT: Accurate classification of wine varieties based on their chemical characteristics has been a concern in the wine industry and viniculture research. This article aimed to develop a classification model for wine cultivars based on their chemical characteristics using Linear Discriminant Analysis (LDA). The data were obtained from the UCI Machine Learning Repository, which provides various datasets for machine learning. The wine's chemical characteristics data are distributed across 13 variables and 178 observations, originating from a wine region in Italy. For data analysis, the Linear Discriminant Analysis (LDA) technique was employed. It achieved a 100% accuracy rate for classifying cultivar 1, a 97.18% accuracy rate for cultivar 2, and a 100% accuracy rate for cultivar 3. The results demonstrate that the LDA method was effective in classifying cultivars based on their chemical characteristics, highlighting its potential relevance for application in future studies.

Keywords: confusion matrix; discriminant function; classification model; Jackknife method.

[1,2,4] Graduandos em Estatística da Universidade Federal de Mato Grosso, Campus Universitário de Cuiabá – mariaritadosreismuniz@gmail.com, gabriel.cardoso.1198@gmail.com e thiagoduarte20151@outlook.com
3 Prof. Dr. Adjunto do Departamento de Estatística. Universidade Federal de Mato Grosso, Av. Fernando Corrêa da Costa, nº 2367, Bairro Boa Esperança. CEP: 78060-900, Cuiabá, MT, Brasil - prof.kuang@gmail.com

INTRODUÇÃO

A busca pela melhora na classificação de variedades de vinho tem sido um tema relevante na indústria vinícola e na comunidade de pesquisa em vinicultura. Como observado por especialistas, a capacidade de distinguir e categorizar diferentes cultivares de vinho com base em suas características químicas é fundamental para a garantia da qualidade e autenticidade dos produtos vinícolas (Magalhães et al., 2021). Nesse contexto, o presente estudo se concentra na aplicação da Análise Discriminante Linear (LDA) como uma ferramenta valiosa para a classificação precisa de vinhos, utilizando um conjunto de dados robusto obtido do UCI Machine Learning Repository, uma fonte bem estabelecida para conjuntos de dados de pesquisa em aprendizado de máquina.

A análise química de vinhos é uma disciplina multifacetada que envolve a avaliação de várias propriedades químicas e físicas das amostras, com o objetivo de identificar padrões distintos entre diferentes tipos de vinhos. De acordo com estudos anteriores, a LDA é uma técnica estatística amplamente reconhecida que tem sido aplicada com sucesso na classificação de vinhos com base em suas características químicas, tornando-se uma escolha promissora para essa tarefa (Sádecká; Jakubíková; Májek, 2018). O conjuntode dados em foco, originado da análise de vinhos cultivados na região italiana, oferece umaoportunidade única para explorar a eficácia da LDA na diferenciação de variedades de vinho com base em treze atributos químicos distintos.

A Análise Discriminante Linear (LDA) é uma técnica poderosa na estatística multivariada que se concentra na diferenciação de grupos ou classes com base em uma combinação linear de variáveis independentes (Varella, 2008). De acordo com Johnson e Wichern (2007), a LDA tem aplicações em uma variedade de campos, como biologia, medicina, finanças, reconhecimento de padrões e engenharia. Ela é amplamente utilizada quando se deseja maximizar a separação entre grupos distintos, sendo um método valioso para a classificação de observações em múltiplas categorias.

Um dos fundamentos essenciais da LDA é a identificação de funções discriminantes lineares. Como destacado por Tabachnick, Fidell e Ullman (2013), essas funções discriminantes são derivadas a partir das médias e das matrizes de covariância dos grupos e fornecem uma maneira eficaz de classificar novas observações em grupos predefinidos. A contribuição de Fisher em seu artigo de 1936, intitulado "The Use of Multiple Measurements in Taxonomic Problems," continua a ser uma referência fundamental na literatura da LDA. Fisher estabeleceu as bases teóricas e matemáticas da técnica, demonstrando como as funções discriminantes lineares podem ser obtidas para maximizar a separação entre grupos. Ele também destacou a utilidade da LDA em problemas de classificação multivariada, abrindo portas para uma ampla gama de aplicações em pesquisa e prática.

Este artigo tem como objetivo apresentar uma análise abrangente dos resultados obtidos com aaplicação da LDA, incluindo a avaliação da precisão do modelo, critérios de previsão e uma análise da matriz de confusão. A contribuição deste estudo reside na demonstração da utilidade da LDA como uma ferramenta eficaz para a classificação precisa de vinhos com base em suas características químicas. Além disso, abre portas para investigações adicionais no campo da classificação de cultivares de vinho.

MATERIAIS E MÉTODO

Dados

Os dados utilizados neste estudo foram adquiridos a partir do UCI Machine Learning Repository, um repositório de pesquisa em aprendizado de máquina que abriga uma variedade de conjuntos de dados. Os criadores originais destes dados foram Aeberhard, Stefan e Forina (1991). Estes dados representam os resultados de uma análise química de vinhos cultivados na região da Itália, originados de três variedades distintas. A análise abordou as concentrações de 13 componentes químicos presentes em cada um dos três tipos de vinhos, a saber: teor alcoólico, ácido málico, cinzas, alcalinidade das cinzas, magnésio, fenóis totais, flavonóides, fenóis não flavonóides, proantocianinas, intensidade de cor, matiz, OD280/OD315 de vinhos diluídos e prolina (Tabela 1). Neste contexto de classificação, o conjunto de dados se destaca como um problema bem definido, com classes que se comportam de maneira coesa. Para o estudo foi utilizada a Análise Discriminante Linear (LDA).

Tabela 1. Descrição das variáveis utilizadas no estudo.

Variáveis	Tipo dos dados	Descrição
X1	Categórico	Tipo de cultivar
X2	Contínuo	Teor alcoólico do vinho
X3	Contínuo	Quantidade de ácido málico presente no vinho
X4	Contínuo	Conteúdo de cinzas no vinho
X5	Contínuo	Nível de alcalinidade das cinzas no vinho
X6	Inteiro	Quantidade de magnésio no vinho
X7	Contínuo	Quantidade total de fenóis no vinho
X8	Contínuo	Quantidade de flavonoides no vinho
X9	Contínuo	Quantidade de fenóis que não são flavonoides no vinho
X10	Contínuo	Quantidade de pro antocianinas no vinho
X11	Contínuo	Intensidade da cor do vinho
X12	Contínuo	Matiz da cor do vinho
X13	Contínuo	Relação óptica no vinho diluído
X14	Inteiro	Quantidade de prolina no vinho

Análise Discriminante Linear de Fisher

A LDA teve sua origem na proposta inicial de Fisher em 1936 para discriminação e classificação entre dois ou mais grupos. A ideia principal era transformar observações multivariadas em observações univariadas por meio de combinações lineares das variáveis originais, a fim de minimizar classificações incorretas de indivíduos em diferentes populações (Regazzi, 2000). Assim, a classificação é realizada por meio da construção de uma função linear que envolve as características observadas das variáveis e a determinação dessa função é então baseada na otimização da classificação correta dos casos nos grupos apropriados. A função discriminante linear é uma combinação linear das características originais, projetada de forma a maximizar a separação entre duas populações (Varella, 2008).

Essa técnica aborda a avaliação da capacidade de distinguir entre dois ou mais conjuntos de indivíduos, considerando medidas em várias variáveis para esses indivíduos (Khattree e Naik, 2000). Uma regra de classificação eficaz deve reduzir ao mínimo os erros de

classificação. Portanto, a análise discriminante deve estabelecer uma função para a definição das regiões R_1 e R_2, de modo a minimizar as probabilidades de classificação incorreta. Para distinguir entre duas ou mais populações, é necessário criar uma combinação linear das características observadas que demonstre a maior capacidade de discriminação. Essa combinação é conhecida como Função Discriminante Linear de Fisher (Johnson e Wichern, 2007).

Portanto, considerando μ_i como os vetores de médias e $\mathbf{\Sigma}$ como a matriz de covariâncias comuns das populações π_i, a Função Discriminante Linear de Fisher de um vetor aleatório $\mathbf{X}$, que busca a máxima separação entre as duas populações, é expressa da seguinte forma:

$$D(\mathbf{X}) = a'.\mathbf{X} = [\mu^1 - \mu^2]'\mathbf{\Sigma}^{-1}\mathbf{X}$$

Em que, D ($\mathbf{X}$) é a função discriminante linear de Fisher, $\mathbf{a}'$é um vetor discriminante contendo a diferença entre as médias das duas populações $[\mu_1 - \mu_2]'$, $\mathbf{X}$ é o vetor de características a ser classificado e $\mathbf{\Sigma^{-1}}$ é a matriz inversa da matriz de covariância comum ($\mathbf{\Sigma}$) das populações (Johnson e Wichern, 2007).

Assim, a expressão para o valor da função discriminante de Fisher para uma observação específica x_0 é a seguinte:

$$D(x_0) = [\mu^1 - \mu^2]'\boldsymbol{\Sigma}^{-1}x_0$$

Portanto, o ponto médio entre as duas médias populacionais univariadas μ_1 e μ_2 é o seguinte:

$$m = \frac{1}{2}[D(\mu_1) + D(\mu_2)]$$

Desse modo, o princípio de classificação das características com base na função discriminante linear de Fisher é designar x_0 em:

$$\mu_1, \text{ se } D(x_0) = [\mu_1 - \mu_2]' \cdot \boldsymbol{\Sigma}^{-1} \cdot x_0 \geq m$$
$$\text{e}$$
$$\mu_2, seD(x_0) = [\mu_1 - \mu_2] \cdot \boldsymbol{\Sigma}^{-1} \cdot x_0 < m$$

A função discriminante linear amostral de Fisher é calculada substituindo os parâmetros μ_1, μ_2 e $\mathbf{\Sigma}$ pelas quantidades amostrais correspondentes: $\bar{x}_1$, $\bar{x}_2$ e S_c (Johnson e Wichern, 2007). A expressão para a função discriminante linear amostral de Fisher é a seguinte:

$$D(x) = \hat{\boldsymbol{a}}'\mathbf{X} = [\bar{x}_1 - \bar{x}_2]'\boldsymbol{S}_c^{-1}\boldsymbol{X}]$$

Aqui, $D(x)$ representa a função discriminante linear amostral de Fisher, $\hat{\boldsymbol{a}}'$é a estimativa do vetor discriminante, $\bar{x}_1$é a média amostral da população π_1 e $\bar{x}_2$ é a média amostral da população π_2 (Johnson e Wichern, 2007). É importante destacar que ao utilizar a função discriminante de Fisher, que pressupõe duas populações com matrizes de covariância iguais, o ponto de separação entre essas duas populações corresponde ao ponto médio entre os valores que representam as funções discriminantes para as médias das amostras das duas populações (Soares, 1997).

A estatística de teste utilizada nesse estudo para comparar vetores de médias amostrais foi o Traço de Pillai, que pode ser descrita como autovalores de λ_1 a λ_p dessa seguinte maneira:

$$V = \sum_{i=1}^{p} \frac{\lambda_i}{(1 + \lambda_i)}$$

Em que, λ_1 a λ_p são os autovalores e p é o número de variáveis. Conforme destacado por Onson (1974), quando essa estatística produz valores significativamente elevados, isso sugere fortemente que as amostras sob exame possuem médias vetoriais distintas em relação às suas populações de origem.

Para o indicador de eficácia, foi utilizada a matriz de confusão, que representa uma métrica eficaz na avaliação da classificação, pois ela oferece informações sobre o número de classificações corretas para cada grupo específico (MATOS et al., 2009). Uma matriz de confusão é uma representação matricial quadrada na qual as linhas e colunas contêm números que indicam a quantidade de amostras atribuídas a categorias específicas em relação à categoria real (Suarez e Candeias, 2012). Assim, a diagonal principal dessa matriz revela as classificações precisas.

Posteriormente, para a avaliação do modelo foi utilizado o Método de Jackknife, que que se realiza a amostragem do conjunto de dados, essa técnica é essencial na análise de dados e na construção e validação de modelos de classificação. Ela envolve a criação de várias amostras ou subconjuntos a partir do conjunto de dados original. Costa, (1988) descreve o método de Jackknife como um método de reamostragem que envolve a criação de várias versões da amostra original, omitindo um único ponto de dados de cada vez. O método pode ser descrito de tal maneira:

Considerando um cenário em que está sendo estimado um parâmetro θ com base em uma amostra de n valores, como segue:

$$\hat{\theta} = f(x_1, x_2, \dots, x_n)$$

Nesse contexto, a i-ésima replicação Jackknife pode ser definida como o valor estimado $\hat{\theta}$ porém, excluindo a i-ésima observação da amostra, conforme:

$$\hat{\theta}_{(i)} = f(x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$$

Além disso, pode-se introduzir o conceito de i-ésimo pseudovalor, que é definido da seguinte forma:

$$x_i^* = N\hat{\theta} - (N-1)\hat{\theta}_i$$

Com base nos pseudovalores, pode-se calcular a seguinte expressão:

$$\hat{\theta}_{jk} = \frac{1}{n}\sum_{i=1}^{n} x_i^* = N\hat{\theta} - (N-1)\hat{\theta}_{(.)}$$

em que:

$$VAR_{jk}(\hat{\theta}) \cong \frac{n-1}{n}\sum_{i=1}^{n} x_i^*(\hat{\theta}_{(i)} - \hat{\theta}_{(.)})$$

RESULTADOS E DISCUSSÕES

Primeiramente, foram analisadas as médias simples de cada grupo de cultivar (Tabela 2), o primeiro grupo, grupo 1, apresentou uma média de teor alcoólico de 13,745, a mais alta entre os grupos. Em contraste, o grupo 2 registrou a menor média de teor alcoólico, com 12,279. O grupo 3, por sua vez, se apresentou em um ponto intermediário, com uma média de 13,154 de teor alcoólico. Outra característica relevante observada foi a quantidade de ácido málico presente nos vinhos. O grupo 3 liderou com uma média notavelmente alta de 3,334, enquanto o grupo 2 mostrou a menor média de ácido málico, com 1,933. O grupo 1, mais uma vez, apresentou valores intermediários, com uma média de 2,011.

Além disso, o nível de alcalinidade das cinzas no vinho também variou entre os grupos. O grupo 3 exibiu a maior média, 21,417, indicando um nível significativo de alcalinidade. O grupo 2, em contraste, teve a menor média de alcalinidade das cinzas, registrando 20,238. O grupo 1, mais uma vez, se mostrou no meio do caminho, com uma média de 17,037.

Tabela 2. Médias simples de cada variável na análise química para cada tipo de cultivar de vinho.

Variáveis	Média por grupo		
	1	2	3
X2	13,745	12,279	13,154
X3	2,011	1,933	3,334
X4	2,456	2,245	2,437
X5	17,037	20,238	21,417
X6	106,339	94,549	99,313
X7	2,840	2,259	1,679
X8	2,982	2,081	0,781
X9	0,290	0,364	0,448
X10	1,899	1,630	1,154
X11	5,528	3,087	7,396
X12	1,062	1,056	0,683
X13	3,158	2,785	1,684
X14	1.115,712	519,507	629,896

Posteriormente, empregou-se a Análise Discriminante Linear (LDA) para investigar a capacidade de diferentes variáveis independentes em diferenciar grupos de vinhos representados por grupo 1, grupo 2 e grupo 3, com base em uma variável dependente (X1). Primeiramente foi aplicado a estatística de Pillai, que retornou p-valor de 0,90, o que demonstra uma forte evidência de diferenças significativas entre os grupos independentes na variável dependente X1. Esta estatística é indicativa da proporção substancial da variabilidade total que pode ser atribuída às diferenças entre grupos, com um alto grau de significância estatística ($p < 2 \times 10^{-16}$).

A LDA considera que a distribuição de probabilidade das amostras é previamente conhecida e pode ser representada por meio dos parâmetros de média e dispersão das amostras (XAVIER et al., 2011). As probabilidades prévias dos grupos indicaram a distribuição relativa das observações em cada grupo. indicando que os grupos 1, 2 e 3 têm tamanhos amostrais distintos. Consequentemente, a probabilidade prévia para o grupo 1 foi de 33,14% , para o grupo 2 foi de 39,89% e para o grupo 3 atingiu 26,96%. Posteriormente, os coeficientes dos discriminantes lineares das equações LD1 e LD2, foram calculados indicando como as variáveis independentes contribuem para a discriminação entre grupos.

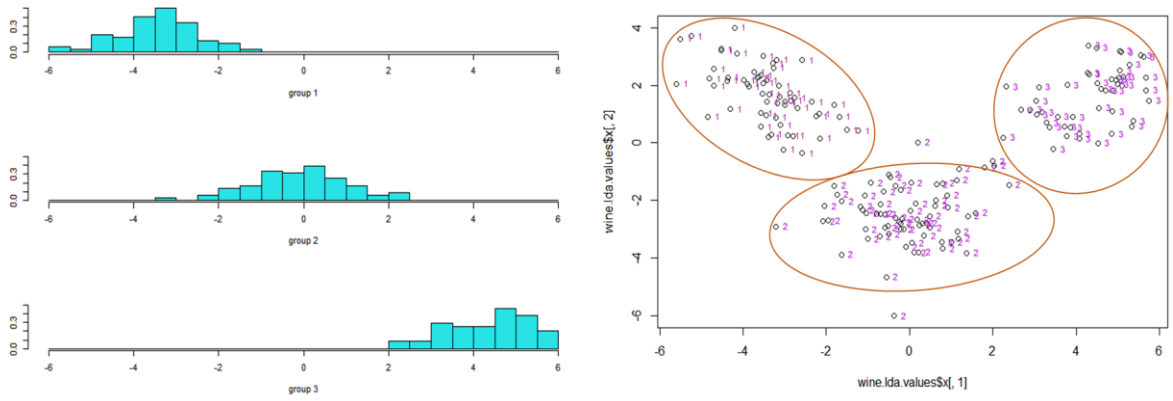
Dessa forma, com o objetivo de encontrar um sistema de coordenadas ideal, a LDA utiliza uma transformação linear que busca maximizar a separabilidade entre as classes presentes no conjunto de dados, conforme destacado por Santos em 2005. Como, LD1 explicou 68,75% da variância total, enquanto LD2 explicou apenas 31,25% da variância total, LD1 foi utilizada, pois ela captura uma proporção maior da variância total explicada, o que reforça sua importância na diferenciação dos grupos. Dessa forma a função discriminante LD1 que será utilizada no modelo tem seus coeficientes expressos na Tabela 3.

Tabela 3. Coeficientes da função discriminante LD1 para as variáveis X2 a X13.

Variáveis	Coeficientes da função LD1
X_2	-0,403
X_3	0,165
X_4	-0,369
X_5	0,155
X_6	-0,002
X_7	0,618
X_8	-1,661
X_9	-1,496
X_{10}	0,134
X_{11}	0,355
X_{12}	-0,818
X_{13}	-1,158
X_{14}	-0,003

Posteriormente, é possível predizer a posteriori de cada observação de pertencer a uma das três classes de vinho. Como pode ser observado na Figura 1 (a) e (b), a discriminação entre o grupo 1 e o grupo 3 demonstrou ser altamente satisfatória, uma vez que nenhuma observação entre os dois se cruzou. Da mesma forma, a discriminação entre o grupo 1 e o grupo 2 também se mostrou eficiente, com apenas algumas observações ultrapassando o limite de corte. O mesmo padrão foi observado ao avaliar a discriminação entre o grupo 2 e o grupo 3.

Figura 1. Histograma empilhados dos valores da LDA (a) e Scatterplots da função discriminante (b).



Para definir a regra de alocação, primeiramente, foram encontrados os valores médios da função discriminante para cada grupo de cultivar (Tabela 4). Para a cultivar 1, o valor médio da primeira função discriminante foi encontrado como −3,422, enquanto para a cultivar 2 foi de −0,079, e para a cultivar 3, foi de 4,325.

Tabela 4. Valores médios da função discriminante obtidos por meio da LDA.

Grupo de cultivar	Média	Desvio-padrão
1	-3,422	0,939
2	-0,079	1,084
3	4,325	0,940

Assim, foram obtidos os pontos intermediários, que se são importantes para a criação de critérios de previsão. Portanto, se a primeira função discriminante de uma amostra for menor ou igual a −1,751, é apropriado prever que essa amostra pertence a cultivar 1. Se estiver no intervalo entre −1,751 e 2,123, a previsão indicará que a amostra pertence a cultivar 2. Por fim, se a primeira função discriminante for maior que 2,123, a amostra será prevista como pertencente a cultivar 3. Esses critérios de previsão são fundamentais para a classificação precisa das amostras com base em suas características na primeira função discriminante, contribuindo para uma análise mais detalhada das cultivares em questão. Posteriormente, foi realizada a matriz de confusão (Tabela 5) calculada por meio dos pontos intermediários obtidos anteriormente. Essa matriz é uma ferramenta fundamental para avaliar o desempenho de algoritmos de classificação.

Tabela 5. Matriz de confusão calculada por meio dos pontos intermediários.

Grupo de cultivar	Classificação como grupo 1	Classificação como grupo 2	Classificação como grupo 3
1	56	3	0
2	5	65	1
3	0	0	48

No grupo de cultivar, o modelo avaliou corretamente 56 amostras como pertencentes ao grupo 1. No entanto, identificou incorretamente 5 delas como pertencentes ao grupo 1, quando na realidade elas eram do grupo 2. Quanto ao grupo 2, o modelo apresentou um desempenho melhor, classificando corretamente 65 amostras como pertencentes a esse grupo, com apenas 3 amostras erroneamente classificadas como grupo 2 quando, na verdade, faziam parte do grupo 1. Por fim, o grupo 3 também foi avaliado pelo modelo, e conseguiu acertar todas as classificações corretamente, identificando 48 amostras como pertencentes ao grupo 3.

Portanto, a matriz de confusão revelou que o modelo classificou incorretamente 9 das 178 amostras de vinho. Isso resulta em uma taxa de erro de classificação de aproximadamente 5,06%. Vale ressaltar que a taxa de classificação errada é relativamente baixa, o que indica um desempenho geralmente sólido do modelo. Além disso, a porcentagem de amostras corretamente classificadas, alcançou 94,94%. Entretando, ainda é preciso fazer a validação do modelo.

Para a validação do modelo, inicialmente, foi empregada uma técnica de reamostragem no conjunto de dados utilizando o método de Jackknife. Esse método consiste em 11 retirar uma única observação do conjunto total de amostras, recalculando o estimador com base nos valores remanescentes. Em seguida, para a avaliação da precisão da previsão do modelo foi utilizada a matriz de confusão utilizando a amostragem realizada com o método de Jackknife (Tabela 6).

Tabela 6. Matriz de confusão a partir dos dados reamostrados com o método Jackknife.

Grupo de Cultivar	Classificação como grupo1	Classificação como grupo 2	Classificação como grupo 3
1	59	0	0
2	1	69	1
3	0	0	48

Portando, a análise da acurácia das previsões do modelo revelou resultados promissores. Para a cultivar 1, observa-se uma taxa de classificação correta de 100%, indicando que todas as previsões para essa categoria foram precisas. O desempenho para a cultivar 2 teve uma taxa de classificação correta de 97,18%, indicando uma precisão substancialmente elevada nas previsões. De maneira similar, a cultivar 3, teve uma taxa de classificação correta de 100%, indicando que todas as previsões para esse grupo foram corretas. Nesse contexto, o modelo demonstra uma taxa total de classificação correta 98,88%. Esse resultado, significativamente alto, reitera a capacidade geral do modelo em realizar previsões precisas em relação à variável dos grupos (X1) no conjunto de dados Wine. Essa taxa de acerto robusta ressalta a eficácia do modelo e sua utilidade em tarefas de classificação relacionadas a classificação de cultivares de vinhos.

CONCLUSÃO

Este estudo empregou a LDA para avaliar a capacidade das variáveis independentes em diferenciar grupos de vinhos com base na variável dependente (X1). A LDA revelou um desempenho notavelmente sólido, com uma taxa de classificação correta de aproximadamente 98,88% e uma taxa de erro de classificação de cerca de 5,06%.

Os resultados indicam que a primeira função discriminante (LD1) desempenhou um papel crucial na diferenciação dos grupos de vinhos, explicando 68,75% da variância total. A matriz de confusão gerada com o método Jackknife também destacou a eficácia do modelo, com taxas de classificação correta de 97,18% para as diferentes cultivares de vinho.

Esses resultados sólidos demonstram a eficácia da Análise Discriminante Linear como uma ferramenta confiável para classificar com precisão as amostras de vinho com base em suas características químicas. Este estudo contribui significativamente para a compreensão e aprimoramento da classificação de cultivares de vinhos e fornece uma base sólida para futuras investigações nessa área.

REFERÊNCIAS BIBLIOGRÁFICAS

AEBERHARD, S.; FORINA, M. Wine. 1991. **UCI Machine Learning Repository**. DOI: https://doi.org/10.24432/C5PC7J. 2

COSTA, F. d. S. **Aplicação de técnicas estatísticas de reamostragem em hidrologia.** Universidade Federal do Rio de Janeiro, 1988. 8

FISHER, Ronald A. **The use of multiple measurements in taxonomic problems**. Annals of eugenics, v. 7, n. 2, p. 179-188, 1936.

HAIR, J. F. *et al.* **Multivariate data analysis**. 6. ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2006.

HAIR, J. Jf; babin, b.; money, ah; samouel, p. **Fundamentos de métodos de pesquisa em administração**. Porto Alegre: Bookman, 2005. 2, 4.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 6. ed. Upper Saddle River, N.J: Pearson Prentice Hall, 2007.

MAGALHÃES, V. M. et al. **Uma classificação de vinhos baseada em regras fuzzy utilizando o algoritmo farc-hd**. In: SBC. Anais do VI Workshop-Escola de Informática Teórica. [S.l.], 2021. p. 148–155. 2

MAROCO, J. **Análise estatística com utilização do SPSS**. [S.l.: s.n.], 2007. 6.

MATOS, P. F. et al. **Relatório técnico "métricas de avaliaçao**. Universidade Federal de São Carlos, 2009. 7.

OLSON, Chester L. Comparative robustness of six tests in multivariate analysis of variance. **Journal of the American Statistical Association**, v. 69, n. 348, p. 894-908, 1974.

R Core Team (2023). _R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Áustria. Disponível em: https://. R - project.org/.

SÁDECKÁ, J.; JAKUBÍKOVÁ, M.; MÁJEK, P. **Fluorescence spectroscopy for discrimination of botrytized wines**. Food control, Elsevier, v. 88, p. 75–84, 2018. 2.

SHARMA, S. **Applied multivariate techniques**. [S.l.]: John Wiley & Sons, Inc., 1995. 4.

SOARES, Regina Maria Agostinho. **Métodos não paramétricos em Análise Discriminante**. 1997. Tese de Doutorado. Instituto Superior de Economia e Gestão.

SUAREZ, A. F.; CANDEIAS, A. L. B. **Avaliação de acurácia da classificação de dados de sensoriamento remoto para o município de Maragogipe**. IV Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação. Recife, Brasil, 2012. 7.

TABACHNICK, Barbara G.; FIDELL, Linda S.; ULLMAN, Jodie B. **Using multivariate statistics**. Boston, MA: Pearson, 2013.

TEAM, R. C. R: A. **Language and Environment for Statistical Computing**. Vienna, Austria, 2023. Disponível em: <https://www.R-project.org/>. 4.

VARELLA, C. A. A**. Análise Discriminante**. Análise Multivariada Aplicada às Ciências Agrárias, 2008.

XAVIER, A. C. et al. **Análise discriminante e classificação de imagens 2d de ultrassonografia mamária.** In: VII Workshop de Visão Computacional. [S.l.: s.n.], 2011. p. 67–72. 9..