

O USO DE MACHINE LEARNING NA INDÚSTRIA VINÍCOLA

A influência das características físico-químicas e sensoriais na classificação do vinho

GARRONI, Leonardo Caixeta¹

CRUZ, Lucas Pereira da¹

SANTOS, Flávia Aparecida Oliveira²

¹Discente – Ciência da Computação - Universidade José do Rosário Vellano

²Docente – Ciência da Computação - Universidade José do Rosário Vellano

RESUMO

A partir de uma base de dados contendo duas variantes de vinhos, sendo eles, tinto e branco, foi utilizada uma técnica na área de Inteligência Artificial, chamada Machine Learning. Nessa técnica, foi explorado o modelo de Classificação com base nas características físico-químicas dos vinhos. O objetivo é fazer previsões da qualidade do vinho, analisando doze atributos existentes na base de dados. Para executar os experimentos, foi usado o software Weka 3.8.5. Os resultados obtidos foram comparados com a classificação feita previamente por especialistas.

Palavras-chave

Machine Learning, Classificação, Aprendizado Supervisionado

ABSTRACT

From a database containing two wine variants, red and white, an Artificial Intelligence technique, called Machine Learning, was used. In this technique, the classification model based on the physical-chemical characteristics of the

wines was explored. The objective is to make the quality of the wine, analyze twelve attributes existing in the database. To run the experiments, it was used in the Weka 3.8.5 software. The results obtained were compared with the classification previously made by experts.

Keywords

Machine Learning, Classification, Supervised Learning

1 INTRODUÇÃO

Durante o advento da pandemia da Covid-19, no ano de 2020, muitas mudanças aconteceram no cenário mundial. Com a crise da situação econômica e, consequentemente, com a desvalorização da moeda, as atividades de lazer de menor custo e com menor quantidade de

pessoas aglomeradas têm se destacado. Uma dessas atividades é a apreciação de vinhos. Conforme informações do Ministério do Turismo (2020), a tecnologia e a industrialização tiveram uma grande importância para o aumento da quantidade de vinho produzido.

Um dos grandes desafios para os apreciadores de vinho é selecionar qual produto adquirir, considerando suas preferências de durabilidade e tempo para consumo. Segundo CARVALHO (2019), atualmente os vinhos tintos elaborados no Vale do São Francisco têm sido indicados para o rápido consumo, pois o mesmo não é apto ao envelhecimento. Tal problemática deve-se às elevadas concentrações de potássio nos solos, que é transferido à bebida. Isso resulta em alto valor de pH e acidez titulável inferior nos vinhos, promovendo maior suscetibilidade do vinho a alterações químicas e microbiológicas, além de sedimentação na garrafa e modificação na coloração do vinho (CARVALHO et al, 2019).

Para adquirir o melhor vinho de acordo com suas preferências pessoais, os apreciadores devem conhecer e analisar suas características físico-químicas, de forma a dar suporte à sua decisão. Dessa forma, um ponto a se destacar é a aplicação da técnica de Inteligência Artificial, chamada Machine Learning, que pode ajudar a classificar a qualidade do vinho a partir da análise inteligente de suas características físico-químicas (COSTA, 2016).

Neste trabalho, foi aplicada a técnica de regressão para prever a qualidade do vinho tinto e vinho branco (FORTI, 2018). Foi utilizada uma base de dados que contém as características físico-químicas do vinho, por exemplo, valor de PH, acidez volátil e acidez fixa, além da sua classificação de qualidade.

O algoritmo de classificação por árvore de decisão foi utilizado para processamento dos dados. Algoritmos que induzem árvores de decisão pertencem a família de algoritmos *Top Down Induction of Decision Table - TDIDT* (MONARD; BARANAUSKAS, 2003, p. 60).

Uma árvore de decisão é uma estrutura de dados definida recursivamente como:

- um nó folha que corresponde a uma classe ou;
- um nó de decisão que contém um teste sobre algum atributo. Para cada resultado do teste existe uma aresta para uma subárvore. Cada subárvore tem a mesma estrutura que a mesma.

Para comparação dos resultados, foi utilizada a classificação feita por profissionais da área, que atribuíram notas entre 0 e 10, sendo 0 muito ruim e 10 excelente (CORTEZ et al., 2009). Com este trabalho, espera-se aplicar os algoritmos funcionais de *machine learning* na seleção dos melhores vinhos, além de contribuir com a indústria na melhoria da qualidade dos vinhos produzidos.

2 METODOLOGIA

A técnica de *machine learning* para classificação foi utilizada na análise da base de dados. Foram utilizados todos os atributos físico-químicos da base de dados para treinamento do modelo de classificação. Além da previsão da qualidade do vinho, também foram obtidas análises gráficas dos dados.

Os atributos do vinho presentes na base de dados são: acidez fixa, acidez volátil, ácido cítrico, açúcar residual, cloretos, dióxido de enxofre livre, dióxido de enxofre total, densidade, pH, sulfatos, álcool e a classificação da qualidade.

De acordo com as questões de privacidade e logística, apenas as variáveis físico-químicas (entradas) e sensoriais (saída) estão disponíveis. Dessa forma, não há dados sobre os tipos de uva, marca de vinho, preço de venda do vinho, entre outras características.

As classes utilizadas são ordenadas e não balanceadas, conforme descrições a seguir (COLI et al, 2015):

- **Acidez fixa.** É a soma dos ácidos fixos, que são os mais importantes, chamados de Tartárico e Málico. Quanto mais elevada for a acidez fixa, mais baixa será a acidez volátil, que também é levada em conta pela base de dados. Quanto maior a quantidade dos ácidos, maior será a dificuldade para as bactérias acéticas se desenvolverem.
- **Acidez volátil.** É a soma dos ácidos voláteis, que se libertam por forma de ebulição ou destilação do vinho e traduz o nível de ataque aceto bacteriano ao vinho. O valor máximo permitido por lei é de 1,2 gramas de ácido acético por litro de vinho.
- **Ácido cítrico.** É um ácido orgânico forte, normalmente presente em fracas quantidades nos mostos de uva e geralmente ausente nos vinhos. A sua concentração aumenta ligeiramente a fermentação do álcool no vinho. A adição do ácido cítrico é degradável pela grande maioria das bactérias lácticas do vinho.
- **Açúcar residual.** A concentração residual do açúcar no vinho refere-se à quantidade de sólidos de açúcar em um determinado volume de vinho após o final da fermentação, além de qualquer adição de açúcar. Sua principal função é equilibrar o sabor ácido do vinho tendo um impacto muito importante na sua qualidade.
- **Cloretos.** Geram no vinho um gosto salgado que pode causar uma reação negativa dos consumidores. Se a concentração ultrapassar certos limites,

o vinho não terá permissão para ser comercializado.

- **Dióxido de enxofre.** É o resultado da fermentação do vinho. A maioria dos produtores utilizam o dióxido de enxofre como um conservante do vinho. Uma grande característica do dióxido de enxofre é trazer condições melhores para os processos de vinificação do vinho, elimina bactérias e leveduras frágeis e indesejáveis, o que permite que apenas as melhores prossigam com o processo de fermentação.
- **Densidade.** A densidade do vinho está relacionada ao seu teor alcoólico e de açúcares residuais do mesmo.
- **pH.** O pH, em termos simples, mede a força da acidez. Sua escala pode variar de 0 a 14, com 0 sendo muito ácido e 14 alcalino. O valor do pH sendo 7 é uma solução neutra, vinhos normalmente variam de 2,8 a 4,0.
- **Sulfatos.** Eles formam-se naturalmente em comidas e bebidas, como resultado da fermentação do vinho. A maioria dos produtores utilizam o sulfitos com o dióxido de enxofre para conservar o vinho.
- **Álcool.** O álcool é resultado da fermentação sem oxigênio do vinho, chamada de fermentação alcoólica, em que o açúcar contido nas frutas é

transformado em etanol. Essa conversão é feita através de uma complexa série de reações químicas.

- **Qualidade.** A qualidade do vinho é avaliada neste banco de dados por especialistas que dão uma nota ao saborear os vinhos analisados. Eles analisam conforme o sabor, textura e aspectos físicos do vinho.

Para treino e avaliação do modelo, foi utilizado o software Weka, na versão 3.8.5. A base de dados foi dividida, aleatoriamente, em dados para treino e dados para testes.

3 RESULTADOS E DISCUSSÃO

Após o treinamento do modelo utilizando o Weka 3.8.5, foram obtidas as classificações dos vinhos de acordo com a base de dados de treino e testes utilizada. Na Figura 1, são apresentados os resultados obtidos.

O resultado apresentado na Figura 1 foi obtido utilizando o algoritmo de árvore de decisão. Os resultados apresentados na coluna *Quality* são as classificações feitas pelos especialistas a partir dos atributos informados. Por outro lado, na coluna *Classification*, são apresentados os resultados obtidos com o algoritmo de árvore de decisão.

De acordo com os resultados obtidos, é possível observar que a maioria dos vinhos foram

No	1: fixed acidity	2: volatile acidity	3: citric acid	4: residual sugar	5: chlorides	6: free sulfur dioxide	7: total sulfur dioxide	8: density	9: pH	10: sulphates	11: alcohol	12: quality	13: classification
	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric
1	7.4	0.7	0.0	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5.0	5.1951219512...
2	7.8	0.88	0.0	2.6	0.098	25.0	67.0	0.9968	3.2	0.68	9.8	5.0	4.6666666666...
3	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.997	3.26	0.65	9.8	5.0	5.0
4	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.998	3.16	0.58	9.8	6.0	5.6
5	7.4	0.7	0.0	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5.0	5.1951219512...
6	7.4	0.66	0.0	1.8	0.075	13.0	40.0	0.9978	3.51	0.56	9.4	5.0	5.1951219512...
7	7.9	0.6	0.06	1.6	0.069	15.0	59.0	0.9964	3.3	0.46	9.4	5.0	5.0
8	7.3	0.65	0.0	1.2	0.065	15.0	21.0	0.9946	3.39	0.47	10.0	7.0	5.3333333333...
9	7.8	0.58	0.02	2.0	0.073	9.0	18.0	0.9968	3.36	0.57	9.5	7.0	5.1951219512...
10	7.5	0.5	0.36	6.1	0.071	17.0	102.0	0.9978	3.35	0.8	10.5	5.0	5.5675675675...
11	6.7	0.58	0.08	1.8	0.097	15.0	65.0	0.9959	3.28	0.54	9.2	5.0	5.1951219512...
12	7.5	0.5	0.36	6.1	0.071	17.0	102.0	0.9978	3.35	0.8	10.5	5.0	5.5675675675...
13	5.6	0.615	0.0	1.6	0.089	16.0	59.0	0.9943	3.58	0.52	9.9	5.0	5.3488372093...
14	7.8	0.61	0.29	1.6	0.114	9.0	29.0	0.9974	3.26	1.56	9.1	5.0	5.5
15	8.9	0.62	0.18	3.8	0.176	52.0	145.0	0.9986	3.16	0.88	9.2	5.0	5.0
16	8.9	0.62	0.19	3.9	0.17	51.0	148.0	0.9986	3.17	0.93	9.2	5.0	5.0
17	8.5	0.28	0.56	1.8	0.092	35.0	103.0	0.9969	3.3	0.75	10.5	7.0	6.1590909090...
18	8.1	0.56	0.28	1.7	0.368	16.0	56.0	0.9968	3.11	1.28	9.3	5.0	5.0
19	7.4	0.59	0.08	4.4	0.086	6.0	29.0	0.9974	3.38	0.5	9.0	4.0	5.0
20	7.9	0.32	0.51	1.8	0.341	17.0	56.0	0.9969	3.04	1.08	9.2	6.0	5.3333333333...
21	8.9	0.22	0.48	1.8	0.077	29.0	60.0	0.9968	3.39	0.53	9.4	6.0	5.5
22	7.6	0.39	0.31	2.3	0.082	23.0	71.0	0.9982	3.52	0.65	9.7	5.0	5.3541666666...
23	7.9	0.43	0.21	1.6	0.106	10.0	37.0	0.9966	3.17	0.91	9.5	5.0	5.3333333333...
24	8.5	0.49	0.11	2.3	0.084	9.0	67.0	0.9968	3.17	0.53	9.4	5.0	5.2826086950...
25	6.9	0.4	0.14	2.4	0.085	21.0	40.0	0.9968	3.43	0.63	9.7	6.0	5.3541666666...
26	6.3	0.39	0.16	1.4	0.08	11.0	23.0	0.9955	3.34	0.56	9.3	6.0	5.3541666666...
27	7.6	0.41	0.24	1.8	0.08	4.0	11.0	0.9962	3.28	0.59	9.5	5.0	5.3541666666...
28	7.9	0.43	0.21	1.6	0.106	10.0	37.0	0.9966	3.17	0.91	9.5	5.0	5.3333333333...
29	7.1	0.71	0.0	1.9	0.08	14.0	35.0	0.9972	3.47	0.55	9.4	5.0	5.1212121212...
30	7.8	0.645	0.0	2.0	0.082	8.0	16.0	0.9964	3.38	0.59	9.8	6.0	5.3488372093...
31	6.7	0.675	0.07	2.4	0.089	17.0	82.0	0.9958	3.35	0.54	10.1	5.0	5.3488372093...
32	6.9	0.685	0.0	2.5	0.105	22.0	37.0	0.9966	3.46	0.57	10.6	6.0	5.6829268292...
33	8.3	0.655	0.12	2.3	0.083	15.0	113.0	0.9966	3.17	0.66	9.8	5.0	5.3488372093...
34	6.9	0.605	0.12	10.7	0.073	40.0	83.0	0.9993	3.45	0.52	9.4	6.0	5.1951219512...
35	5.2	0.32	0.25	1.8	0.103	13.0	50.0	0.9957	3.38	0.55	9.2	5.0	5.3541666666...
36	7.8	0.645	0.0	5.5	0.086	5.0	18.0	0.9986	3.4	0.55	9.6	6.0	5.1951219512...
37	7.8	0.6	0.14	2.4	0.086	3.0	15.0	0.9975	3.42	0.6	10.8	6.0	5.6829268292...
38	8.1	0.38	0.28	2.1	0.066	13.0	30.0	0.9968	3.23	0.73	9.7	7.0	5.5315789473...

Figura 1 - Resultados obtidos com o Weka 3.8.5

classificados como normais. Utilizando arredondamento, o algoritmo de Inteligência Artificial acertou, aproximadamente, 63% das classificações, comparando os resultados obtidos com as classificações feitas pelos profissionais.

Nas Figuras 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 e 12 são apresentados os resultados individuais para cada atributo da base de dados.

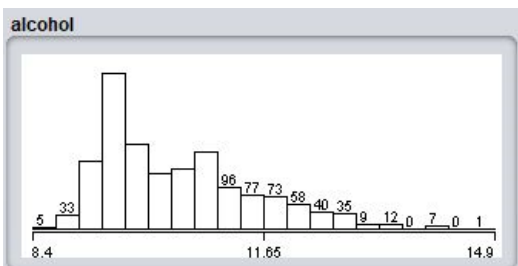


Figura 2 - Representa o atributo álcool

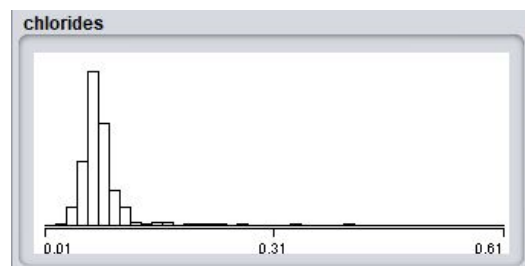


Figura 3 - Representa o atributo cloretos

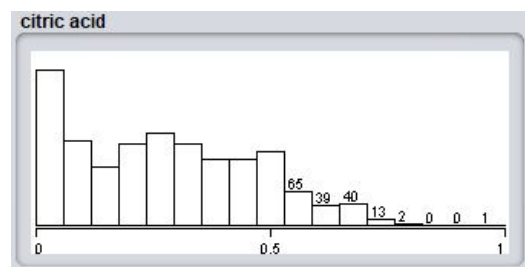


Figura 4 - Representa o atributo ácido cítrico

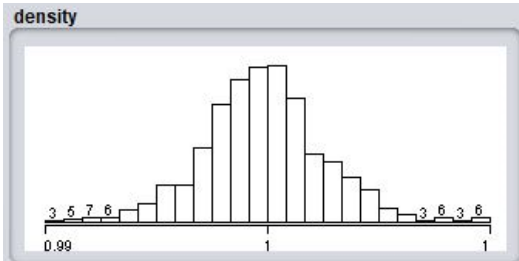


Figura 5 - Representa o atributo densidade

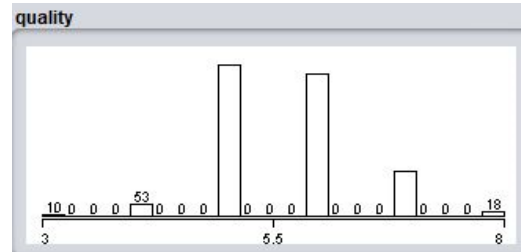


Figura 9 - Representa o atributo qualidade

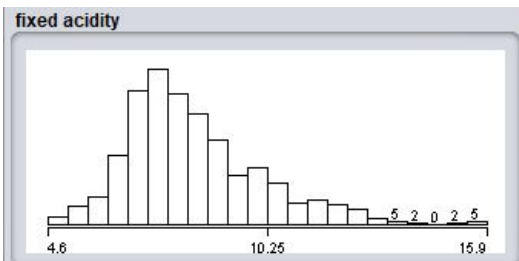


Figura 6 - Representa o atributo acidez fixa

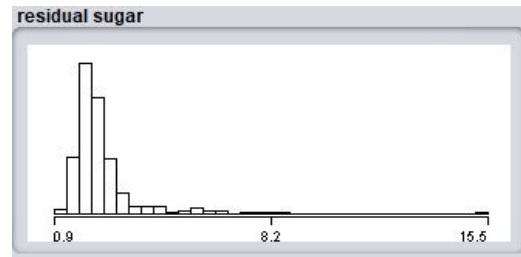


Figura 10 - Representa o atributo açúcar residual

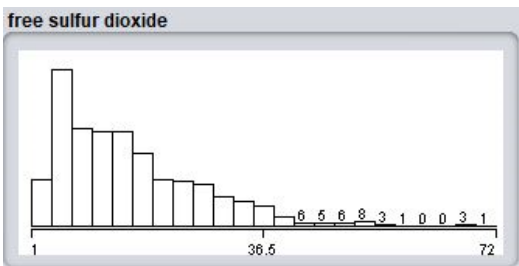


Figura 7 - Representa o atributo dióxido de enxofre

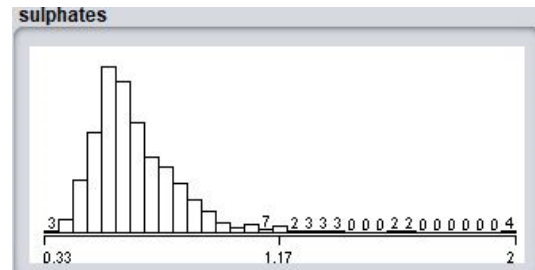


Figura 11 - Representa o atributo sulfatos

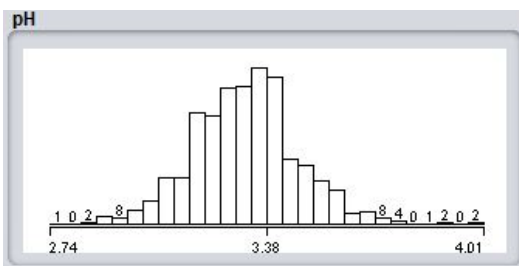


Figura 8 - Representa o atributo pH

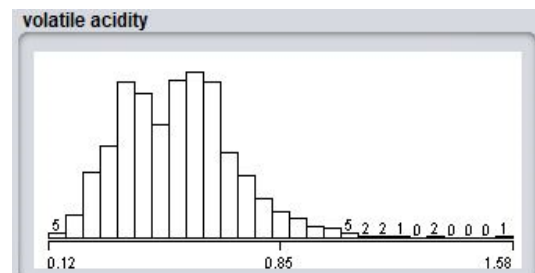


Figura 12 - Representa o atributo acidez volátil

4 CONCLUSÃO

Após a análise dos resultados, observa-se que há mais vinhos com a classificação normal do que vinhos excelentes ou pobres. Isso ocorre pelo fato de que as classes estudadas são ordenadas e não balanceadas. Analisando a classificação feita pela máquina e a classificação obtida por especialistas, pode-se concluir que a máquina fez previsões assertivas em relação aos especialistas, sendo que a menor nota dada por especialistas foi entre 3 e 8. Após a utilização do algoritmo, pode-se observar que, tanto na menor nota, quanto na maior nota, obteve-se a mesma faixa de pontuação, que varia entre 3 e 8. A porcentagem de erros do algoritmo em relação a nota de qualidade, segundo especialistas, foi de apenas 37%.

A Inteligência Artificial é de suma importância para a melhoria dos processos industriais e, por consequência, auxilia na melhoria dos produtos que chegam até o consumidor. Neste trabalho, foi apresentado apenas uma, de tantas outras aplicações, onde a técnica de *machine learning* pode ser aplicada para trazer algum benefício à sociedade.

REFERÊNCIAS

- [1] EMBRAPA. Disponível em: <https://www.infoteca.cnptia.embrapa.br/bitstream/doc/871709/1/pub140.pdf> Acesso em: 28 jan. 2021.
- [2] CINGOLANI, Mauro. POTENCIALIDADES DA ENOGASTRONOMIA NO CONTEXTO DA ESTRATÉGIA EM DUAS VINÍCOLAS DA SERRA GAÚCHA. 2021. 57f. Dissertação - UNIVERSIDADE DE CAXIAS DO SUL, Caxias do Sul, 2021.
- [3] COSTA, Nattane Luíza da. Mineração de dados para a classificação de alguns vinhos *Vitis Vinifera* da América do Sul. 2016. 99f. Dissertação - UNIVERSIDADE FEDERAL DE GOIÁS INSTITUTO DE INFORMÁTICA, Goiânia, 2016.
- [4] CORTEZ, Paulo et al. Modeling wine preferences by data mining from physicochemical properties. In *Decision Support Systems*. Disponível em: <<http://dx.doi.org/10.1016/j.dss.2009.05.016>>. Acesso em: 24 set. 2021.
- [5] CARVALHO, Erika Samantha Santos de et al. Influência de correção do pH do Mosto do Vinho Utilizando Diferentes Ácidos Orgânicos de Vinhos Tintos do Vale do Submédio São Francisco. *Anais da II Jornada de Integração da Pós-graduação da Embrapa Seminário*. Salvador, v.2, n.1, p. 203-210, Jan/Jun. 2019.
- [6] MONARD, Maria Carolina; BARANAUSKAS, José Augusto. *Indução de Regras e Árvores de Decisão*. 5 ed. São Paulo: Usp, 2003.
- [7] FORTI, Melissa. *Técnicas de Machine Learning Aplicadas na Recuperação de Crédito do Mercado Brasileiro*. 2018. 75f. Dissertação - FUNDAÇÃO GETÚLIO VARGAS ESCOLA DE ECONOMIA DE SÃO PAULO, São Paulo, 2018.
- [8] COLI, Marina Sonegheti et al. Conteúdo de cloretos em vinhos brancos de diferentes países. *Revista de Ciências Farmacêuticas Básica e Aplicada*. Vitória, v.36, n.4, p. 503-507, Jan/Jun. 2015.
- [9] LIMA, Isafas; PINHEIRO, Carlos AM; SANTOS, Flávia A. Oliveira. *Inteligência artificial*. Elsevier Brasil, 2016.