

Base de dados “Wine Quality”

8598861 - Bernardo Simões Lage G. Duarte
8122585 - Eder Rosati Ribeiro
8936993 - Gabriel Luiz Ferraz Souto
8936648 - Giovanni Ortolani Barbosa
8531887 - Giovanni Robira
8937271 - Rafael Bueno da Silva

A base de dados¹ escolhida apresenta é dividida em duas bases menores, uma delas é referente à análise sensorial de vinho branco e a outra de vinho tinto. Existem 4898 instâncias na base com 12 atributos no total (11 de entrada e 1 de saída). Os dados podem ser utilizados tanto para problemas de classificação quanto de regressão.

O objetivo da aplicação de aprendizado de máquina nesta base é prever a qualidade do vinho através dos 11 atributos psicoquímicos de entrada. Também podem ser extraídos conhecimentos se existe uma relação forte entre um determinado atributo de entrada e a respectiva classificação

Os 11 atributos de entrada presentes na base foram obtidos através de testes psicoquímicos, já o atributo de saída foi classificado de acordo com uma análise sensorial de especialistas.

Atributos

Entrada

1. Acidez fixada
2. Acidez volátil
3. Ácido cítrico
4. Açúcar residual
5. Cloreto
6. Dióxido de enxofre livre
7. Dióxido de enxofre total
8. Densidade
9. pH
10. Sulfatos
11. Álcool

Todos os atributos são valores numéricos.

Saída

1. Qualidade

É um valor entre 0 (ruim) e 10 (bom) que indica a qualidade do vinho.

As classes não estão ordenadas e estão desbalanceadas, havendo um maior número de vinhos intermediários do que vinhos ruins ou bons.

¹ P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. Modeling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, 47(4):547-553, 2009.