

Wine Quality

This dataset is related to red variants of the Portuguese "Vinho Verde" wine. For more details, consult the reference [Cortez et al., 2009]. Due to privacy and logistic issues, only physicochemical (inputs) and sensory (the output) variables are available (e.g. there is no data about grape types, wine brand, wine selling price, etc.).

The dataset can be viewed as a classification task. The classes are ordered and not balanced (e.g. there are much more normal wines than poor ones).

Input variables (based on physicochemical tests): 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 Output variable (based on sensory data): 12 - quality (score between 0 and 10)

Task:

Use machine learning to determine which physiochemical properties make a wine 'bad' versus 'good'!

Relevant publication

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.

Columns

fixed acidity: most acids involved with wine or fixed or nonvolatile (do not evaporate readily)

volatile acidity: the amount of acetic acid in wine, which at too high of levels can lead to an unpleasant, vinegar taste

citric acid: found in small quantities, citric acid can add 'freshness' and flavor to wines

residual sugar: the amount of sugar remaining after fermentation stops, it's rare to find wines with less than 1 gram/liter and wines with greater than 45 grams/liter are considered sweet

chlorides: the amount of salt in the wine

free sulfur dioxide: the free form of SO₂ exists in equilibrium between molecular SO₂ (as a dissolved gas) and bisulfite ion; it prevents microbial growth and the oxidation of wine

total sulfur dioxide: amount of free and bound forms of S₂; in low concentrations, SO₂ is mostly undetectable in wine, but at free SO₂ concentrations over 50 ppm, SO₂ becomes evident in the nose and taste of wine

density: the density of water is close to that of water depending on the percent alcohol and sugar content

pH: describes how acidic or basic a wine is on a scale from 0 (very acidic) to 14 (very basic); most wines are between 3-4 on the pH scale

sulphates: a wine additive which can contribute to sulfur dioxide gas (SO₂) levels, which acts as an antimicrobial and antioxidant

alcohol: the percent alcohol content of the wine

quality: output variable (based on sensory data, score between 0 and 10)

color: red or white