

Dataset: Wine Quality

Attributes: 11 numeric, 1 nominal,

Class: Numeric

Instances: 6497

Missing Values: None

In the original form of this dataset, two datasets were created, using red and white wine samples. Here, these two datasets have been combined into one dataset. The inputs include objective tests (e.g. PH values) and the output is based on sensory data (median of at least 3 evaluations made by wine experts). Each expert graded the wine quality between 0 (very bad) and 10 (very excellent).

The two datasets are related to red and white variants of the Portuguese "Vinho Verde" wine. 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.).

Attributes:

- 1 - fixed acidity, numeric
- 2 - volatile acidity, numeric
- 3 - citric acid, numeric
- 4 - residual sugar, numeric
- 5 - chlorides, numeric
- 6 - free sulfur dioxide, numeric
- 7 - total sulfur dioxide, numeric
- 8 - density, numeric
- 9 - pH, numeric
- 10 - sulphates, numeric
- 11 - alcohol, numeric
- 12 - R/W, nominal - R= red, W = white

Class:

- quality (score between 0 and 10)

Reference:

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.