Wine Quality
Donated on 10/6/2009

Two datasets are included, related to red and white vinho verde wine samples, from the north of Portugal. The goal is to model wine quality based on physicochemical tests (see [Cortez et al., 2009], http://www3.dsi.uminho.pt/pcortez/wine/).

Dataset Characteristics
Multivariate

Subject Area
Business

Associated Tasks
Classification, Regression

Feature Type
Real

 # Instances
4898

 # Features
11

Dataset Information

Additional Information

The two datasets are related to red and white variants of the Portuguese "Vinho Verde" wine. For more details, consult: http://www.vinhoverde.pt/en/ or 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.).

These datasets can be viewed as classification or regression tasks. The classes are ordered and not balanced (e.g. there are many more normal wines than excellent or poor ones). Outlier detection algorithms could be used to detect the few excellent or poor wines. Also, we are not sure if all input variables are relevant. So it could be interesting to test feature selection methods.

SHOW LESS ⌃

Has Missing Values?
No

Introductory Paper

[Modeling wine preferences by data mining from physicochemical properties](#)
By P. Cortez, A. Cerdeira, Fernando Almeida, Telmo Matos, J. Reis. 2009
Published in Decision Support Systems

Variables Table

Variable Name	Role	Type	Description	Units	Missing Values
fixed_acidity	Feature	Continuous			no
volatile_acidity	Feature	Continuous			no
citric_acid	Feature	Continuous			no
residual_sugar	Feature	Continuous			no
chlorides	Feature	Continuous			no
free_sulfur_dioxide	Feature	Continuous			no
total_sulfur_dioxide	Feature	Continuous			no
density	Feature	Continuous			no
pH	Feature	Continuous			no
sulphates	Feature	Continuous			no
alcohol	Feature	Continuous			no

DOWNLOAD

IMPORT IN PYTHON

CITE

Keywords

Chemistry

Creators

👤 Paulo Cortez

👤 A. Cerdeira

👤 F. Almeida

👤 T. Matos

👤 J. Reis

DOI

10.24432/C56S3T

License

Variable Name	Role	Type	Description	Units	Missing Values
color	Other	Categorical	red or white		no

Rows per page 25 0 to 13 of 13 ‹ ›

Additional Variable Information ⌃

For more information, read [Cortez et al., 2009].
Input variables (based on physicochemical tests):
  1 - fixed acidity...
SHOW MORE ⌄

Dataset Files ⌃

File	Size
winequality-white.csv	258.2 KB
winequality-red.csv	82.2 KB
winequality.names	3.2 KB

Reviews ⌃

5.0 ★★★★★ (3 ratings)

LOGIN TO WRITE A REVIEW

Mars Nunez
★★★★★   Excelent
1/21/2025

Huynh Thanh Phong
B2207555
★★★★★   _
9/11/2024

Issaka Sana
★★★★★   régression par réseaux de neurones des vins blancs. On créera une nouvelle variable d'intérêt (bon vin"grade<5"/mauvais vin"grade>=5").
12/8/2023

THE PROJECT
About Us
CML
National Science Foundation

NAVIGATION
Home
View Datasets
Donate a Dataset

LOGISTICS
Contact
Privacy Notice
Feature Request or Bug Report