



# 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],...

## 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	Demographic	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
quality	Target	Integer		score between 0 and 10		no
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
- 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)

SHOW LESS ^

## Reviews

5  (1 rating)

LOGIN TO WRITE A REVIEW

Issaka Sana



12/8/2023

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

DOWNLOAD



IMPORT IN PYTHON

CITE

” 1 citations

👁 254600 views

## Keywords

Chemistry

## Creators

 Paulo Cortez

 A. Cerdeira

 F. Almeida

 T. Matos

 J. Reis

## DOI

10.24432/C56S3T

## License

This dataset is licensed under a [Creative Commons Attribution 4.0 International](#) (CC BY 4.0) license.

This allows for the sharing and adaptation of the datasets for any purpose, provided that the appropriate credit is given.



**UC Irvine**  
Machine Learning  
Repository

## THE PROJECT

[About Us](#)

[CML](#)

[National Science Foundation](#)

## NAVIGATION

[Home](#)

[View Datasets](#)

[Donate a Dataset](#)

## LOGISTICS

[Contact](#)

[Privacy Notice](#)

[Feature Request or Bug Report](#)