**Title – Wine Quality**

**License**:

Citation Request:

This dataset is public available for research. The details are described in [Cortez et al., 2009].

Please include this citation if you plan to use this database:

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. ISSN: 0167-9236.

Available at: [@Elsevier] http://dx.doi.org/10.1016/j.dss.2009.05.016

[Pre-press (pdf)] http://www3.dsi.uminho.pt/pcortez/winequality09.pdf

[bib] <http://www3.dsi.uminho.pt/pcortez/dss09.bib>

**Questions that can be answered:**

1. Wine companies can use this data to understand what all ingredients combine to form a good quality wine
2. People at health sector can use this data to understand the components of a wine and the percentage in which these contents are present, if they are harmful or not and accordingly advice patients.
3. Researcher can use this data to experiment and replace with the harmful elements in a wine and find a replacement for them.
4. **URL**: <https://archive.ics.uci.edu/ml/datasets/Wine+Quality>

1. **Sources**

Created by: Paulo Cortez (Univ. Minho), Antonio Cerdeira, Fernando Almeida, Telmo Matos and Jose Reis (CVRVV) @ 2009

1. **Past Usage:**

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. ISSN: 0167-9236.

In the above reference, two datasets were created, using red and white wine samples.

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). Several data mining methods were applied to model

these datasets under a regression approach. The support vector machine model achieved the

best results. Several metrics were computed: MAD, confusion matrix for a fixed error tolerance (T),

etc. Also, we plot the relative importances of the input variables (as measured by a sensitivity

analysis procedure).

1. **Relevant Information:**

The two The two datasets containing survey results of both red and white wine

1. **Number of Instances:** red wine - 1599; white wine - 4898.

6. **Attribute information:**

* volatile acidity
* citric acid
* residual sugar
* chlorides
* free sulfur dioxide
* total sulfur dioxide
* density
* pH
* sulphates
* alcohol

Output variable (based on sensory data):

* quality (score between 0 and 10)

7. **Missing Attribute Values:** None

 