The data can be used to test (ordinal) **regression** or **classification** (in effect, this is a **multi-class** task, where the classes are **ordered**) methods. Other research issues are **feature selection** and **outlier detection**. The data includes two datasets:

 winequality-red.csv - red wine preference samples;

 winequality-white.csv - white wine preference samples;

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data Set Characteristics:** | Multivariate | **Number of Instances:** | 4898 | **Area:** | Business |
| **Attribute Characteristics:** | Real | **Number of Attributes:** | 12 | **Date Donated** | 2009-10-07 |
| **Associated Tasks:** | Classification, Regression | **Missing Values?** | N/A | **Number of Web Hits:** | 1120801 |

**Vinho verde** is a unique product from the Minho (northwest) region of **Portugal**. Medium in alcohol is it particularly appreciated due to its freshness (especially in the summer). More details can be found at: <http://www.vinhoverde.pt/en/>

**Data Set Information:**

The two datasets are related to red and white variants of the Portuguese "Vinho Verde" wine. For more details, consult: [[Web Link]](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.

**Attribute 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)