**Wine Quality – Create a model based on physicochemical tests**

# Abstract

Wine is one of most common beverages in every household. Wine (from [Latin](https://en.wikipedia.org/wiki/Latin) vinum) is an [alcoholic beverage](https://en.wikipedia.org/wiki/Alcoholic_beverage) made from [fermented](https://en.wikipedia.org/wiki/Fermentation_(wine)) [grapes](https://en.wikipedia.org/wiki/Grape). These grapes are generally [Vitis vinifera](https://en.wikipedia.org/wiki/Vitis_vinifera" \o "Vitis vinifera), or a hybrid with [Vitis labrusca](https://en.wikipedia.org/wiki/Vitis_labrusca" \o "Vitis labrusca) or [Vitis rupestris](https://en.wikipedia.org/wiki/Vitis_rupestris" \o "Vitis rupestris). Grapes are fermented without the addition of [sugars](https://en.wikipedia.org/wiki/Sugar), [acids](https://en.wikipedia.org/wiki/Acid), [enzymes](https://en.wikipedia.org/wiki/Enzyme), [water](https://en.wikipedia.org/wiki/Water), or other [nutrients](https://en.wikipedia.org/wiki/Nutrient). [Yeast](https://en.wikipedia.org/wiki/Yeast_in_winemaking) consumes the sugar in the grapes and converts it to [ethanol](https://en.wikipedia.org/wiki/Ethanol) and [carbon dioxide](https://en.wikipedia.org/wiki/Carbon_dioxide). Different varieties of grapes and strains of yeasts produce different styles of wine. These variations result from the complex interactions between the biochemical development of the grape, the reactions involved in fermentation, the [terroir](https://en.wikipedia.org/wiki/Terroir), and the production process. Many countries enact legal [appellations](https://en.wikipedia.org/wiki/Appellation) intended to define styles and qualities of wine. These typically restrict the geographical origin and permitted varieties of grapes, as well as other aspects of wine production.

However, as a common consumer we have unanswered question on quality of wine. Every liquor store has variety of wines in various price range – just going thru the ingredients detail we as a consumer not able to determine the quality of the product. We have a common notion that “high price product means better quality”. The goal is to model wine quality based on physicochemical tests.

The objective of this model is not to undermine any specific product per say, rather provide logical conclusion why certain wine products are better than others. This analysis will definitely help Wine manufacturers as well to come up with better product for the consumers.

The goal is to develop a predictive algorithm that can identify the overall quality of the wine based on the following input variables

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

# Dataset

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.).   
  
These datasets can be viewed as classification or regression tasks. The classes are ordered and not balanced (e.g. there are much 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.

The quality of the wine would be determined in the scale of 0 – 10, where 10 means excellent and 0 means worst.

Sample Dataset

