

Wine Type and Quality Classification



Introduction

Wine is an alcoholic beverage made from fermented grapes. Yeast consumes the sugar in the grapes and converts it to ethanol, carbon dioxide, and heat. It is a pleasant tasting alcoholic beverage, loved and celebrated. It will definitely be interesting to analyze the physicochemical attributes of wine and understand their relationships and significance with wine quality and types classifications. To do this, We will proceed according to the standard Machine Learning and data mining workflow models like the CRISP-DM model, mainly for:

- Predict if each wine sample is a red or white wine.
- Predict the quality of each wine sample, which can be low, medium, or high.
-

The dataset is related to red and white variants of the "Vinho Verde" wine. Vinho verde is a unique product from the Minho (northwest) region of Portugal. Medium in alcohol, is it particularly appreciated due to its freshness (specially in the summer). This dataset is publicly available for research purposes only, for more information, read 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.).

Attribute Information

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 - sulfates 11 - alcohol Output variable (based on sensory data): 12 - quality (score between 0 and 10)

UCI Notes About the Dataset

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.