Report Datasets

# Datasets:

Mushroom: <http://archive.ics.uci.edu/ml/datasets/Mushroom>

Wine-Quality: <http://archive.ics.uci.edu/ml/datasets/Wine+Quality>

# Motivation:

## Why did those datasets catch our eye?

We were looking for datasets that we could relate to. Therefor food and drinks seemed suitable. Also the classification is something one could use in the real world.

## Characteristics:

Dataset: Wine Quality

The wine quality dataset consists of 1599 red wine and 4898 white wine samples of the Portuguese “Vinho Verde” wine, as described in: http://www.vinhoverde.pt/en/ or the reference [Cortez et al., 2009]. Those samples were classified according to their quality. This classification is purely based upon sensory data obtained by the median of 3 evaluations made by wine experts.

## Attributes:

The dataset does not contain any missing values and every attribute is a numeric value. Each sample comes with 12 attributes, including the quality (classification):

Those are:

- fixed acidity (g(tartaric acid)/dm3)

- volatile acidity (g(tartaric acid)/dm3)

- citric acid (g/dm3)

- residual sugar (g/dm3)

- chlorides (g(sodium chloride/dm3)

- free sulfur dioxide (mg/dm3)

- total sulfur dioxide (mg/dm3)

- density (g/cm3)

- pH (0-14)

- sulphates (g(tassium sulphate)/dm3)

- alcohol (vol.%)

- output variable: quality (score between 0 and 10)