

MACHINE LEARNING 2 - PROJECT

1. GROUP MEMBERS

- A. Paul Wecker
- B. Robert Wienröder
- C. Vipin Singh

2. DATASET: RED WINE QUALITY

A. Source:

- Kaggle:
<https://www.kaggle.com/datasets/uciml/red-wine-quality-cortez-et-al-2009>
- UCI machine learning repository:
<https://archive.ics.uci.edu/dataset/186/wine+quality>

B. Synopsis:

The dataset contains physiochemical properties of red wine alongside with a quality score (target variable) based on sensory data. The dataset scores are not distributed equally, since there are a lot more “normal” quality wines than “poor” or “excellent” quality wines.

C. Number of observations: 1599

D. Variables: (11 predictor variables and 1 output variable)

Variable name	Variable description	Variable type
Fixed acidity	most acids involved with wine or fixed or nonvolatile	Continuous
Volatile acidity	amount of acetic acid in wine, which at too high of levels can lead to an unpleasant, vinegar taste	Continuous
Citric acid	found in small quantities, citric acid can add 'freshness' and flavor to wines	Continuous
Residual sugar	amount of sugar remaining after fermentation stops, it's rare to find wines with less than 1 gram/liter and wines with greater than 45 grams/liter are considered sweet	Continuous
Chlorides	amount of salt in the wine	Continuous
Free sulfur dioxide	free form of SO ₂ exists in equilibrium between molecular SO ₂ (as a dissolved gas) and bisulfite ion; it prevents microbial growth and the oxidation of wine	Continuous
Total sulfur dioxide	amount of free and bound forms of SO ₂ ; in low concentrations, SO ₂ is mostly undetectable in wine, but at free SO ₂ concentrations over 50	Continuous

	ppm, SO ₂ becomes evident in the nose and taste of wine	
Density	density of water is close to that of water depending on the percent alcohol and sugar content	Continuous
pH	describes how acidic or basic a wine is on a scale from 0 (very acidic) to 14 (very basic); most wines are between 3-4 on the pH scale, (Paul: pH is logarithmic as far as I know)	Continuous
Sulphates	a wine additive which can contribute to sulfur dioxide gas (SO ₂) levels, which acts as an antimicrobial and antioxidant	Continuous
Alcohol	ethanol	Continuous
Quality	target variable	Discrete

3. ML-METHODS THAT SHOULD BE USED FOR THE PROJECT

- A. Non-linear models (Spline smoothing)
- B. Pursuit projection regression