Paul Wecker, Robert Wienröder, Vipin Singh

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 a lot more “normal” quality wines than “poor” or “excellent” quality wines.

C. Number of observations: 1599

D. Variables:

Fixed acidity		Continuous
Volatile acidity		Continuous
Citric acid		Continuous
Residual sugar		Continuous
Chlorides		Continuous
Free sulfur dioxide		Continuous
Total sulfur dioxide		Continuous
Density		Continuous
pH		Continuous
Sulphates		Numeric
Alcohol		Numeric
Quality		Ordinal

3. CHOSEN ML-METHODS

A. Non-linear models (Spline smoothing)

B. Pursuit projection regression