WhiteWineClassification Challenge

In this Github repository you find `winequality-white.csv`. This dataset has 11 features which are the results of chemical tests on wine samples. The target variable is the wine quality, which is an integer between 0 (very bad) and 10 (excellent). You can find the description of the dataset in `wineQualityInfo.txt`.

Task

Your task consists of the following steps: 1. Data Exploration 2. Feature Selection 3. Train one Clustering **or** Regression **or** Classification model of your choice (if you choose regression or classification, split the data into training and test set) 4. Interpret the results and explain your workflow

This task should take you about 4-6 hours. The goal is first and foremost not having a perfect model score, but presenting and explaining a consistent workflow.

Data Exploration

Preliminary analysis of the dataset for a better

Feature Selection

The dataset has 11 features. This might be too much, regarding the small dataset. Also, not all of them might influence the target variable. Reduce the number of features/dimensionality of the dataset while losing as little information as possible.

Model Training

Train one machine learning model. You can choose between a clustering, classification or regression algorithm.

Results

Present, interpret and explain the results of your data analysis.

Jupyter Notebook

You are free in your choice of programming language/toolkit. We recommend using the Jupyter notebook we already started for you: `WhiteWineAnalysis.ipynb`.

Questions

If there are any questions left, feel free to contact me under julia.kraus@rohde-schwarz.com.

Have fun with the data! Looking forward to seeing you again.