

Homework #6: Predicting Wine Quality and Type

In this assignment, your job is to develop a classification workflow to predict wine **quality** and **type** based on the provided **wine-quality** datasets. To clarify, wine **quality** is a subjective rating between 1 and 10 that is *provided* as a column in the training datasets. The **type**, either “*red*” or “*white*”, is NOT a column in the provided datasets...

You have **two** tasks to complete:

- 1) Use your ML / DM tool of choice (e.g., KNIME) to experiment with the two provided training datasets (i.e., **wine-quality-red-train.csv** and **wine-quality-white-train.csv**).
 - Your goal is to find the “best” features, classification models, and algorithm parameters to predict wine **quality** and **type**.
 - i. You must build **two separate** classifiers: one to predict the **quality** and another to classify the wine’s **type**.
 - ii. The test dataset for Task #2 includes both **red** and **white** wines; you should combine both the “red” and “white” wine datasets for training.
 - Submit a description (with screenshots) explaining your experimentation. E.g.,
 - What *features* did you select? Why?
 - What *algorithms* and *parameters* did you try?
 - What were the results (e.g., recall, precision, accuracy)?
- 2) Once you determine the “best” algorithm and parameters, implement your classification workflow in **Python** to predict both the **quality** and **type** of each wine (row) in the provided *test* dataset (i.e., **wine-quality-test.csv**). Your solution must write the predictions to a file named **<lastname>.csv** as detailed below. **Submit your Python code.**
 - **wine-quality-test.csv** file:
 - i. includes an ID attribute
 - ii. includes both *red* and *white* wine data objects (rows)
 - iii. does not include the **quality** column (that’s your job!)
 - Your output file **<lastname>.csv** must have the following format:
 - i. **id ; quality ; type**
 - ii. Where:
 1. **id** matches the id column in the test dataset
 2. **quality** is your quality prediction (e.g., 3.0)
 3. **type** is your type prediction (e.g., 0 for red, 1 for white)
 - You may *hard-code* the input file names, i.e.,
 - i. wine-quality-red-train.csv
 - ii. wine-quality-white-train.csv
 - iii. wine-quality-test.csv