# Swe584 - Term Project Wine Quality

Progress Report

Zeynep Deniz Yılmaz

# Tool 1 - Decision Trees

- I used decision trees on the red wine dataset. I tried different setups as in:
  - Training set size (10% / 20%) - default is %20
  - Selection measure (gini / entropy) - the default is gini
  - Split strategy (best/random) - the default is best
- For each run, the accuracy and its details is provided.

# Sample result

1. Accuracy value
2. How close the predictions are to the real quality scores

Example:

Accuracy: 0.64375
Size of the test set: 160
accurate prediction: 103
predictions that were one off: 39
predictions that were two off: 14
predictions that failed: 4

# Training set size

%80 training, %20 test (default)

Accuracy: 0.628125
Size of the test set: 320
accurate prediction: 201
predictions that were one off: 98
predictions that were two off: 20
predictions that failed: 1

%90 training, %10 test

Accuracy: 0.71875
Size of the test set: 160
accurate prediction: 115
predictions that were one off: 38
predictions that were two off: 6
predictions that failed: 1

# Selection measure

Criterion = “gini” (default)

Accuracy: 0.628125
Size of the test set: 320
accurate prediction: 201
predictions that were one off: 98
predictions that were two off: 20
predictions that failed: 1

Criterion = “entropy”

Accuracy: 0.609375
Size of the test set: 320
accurate prediction: 195
predictions that were one off: 99
predictions that were two off: 24
predictions that failed: 2

# Split strategy

Splitter = “best” (default)

Accuracy: 0.628125
Size of the test set: 320
accurate prediction: 201
predictions that were one off: 98
predictions that were two off: 20
predictions that failed: 1

Splitter = “random”

Accuracy: 0.646875
Size of the test set: 320
accurate prediction: 207
predictions that were one off: 97
predictions that were two off: 13
predictions that failed: 3