

J48 decision tree (Weka implementation)

Method description

The OCHEM machine learning method **WEKA-J48** is a Weka[1] implementation[2] of the C4.5 pruned decision tree [3]. Weka 3 is used in OCHEM environment as an external command line tool.

The C4.5 tree tries to recursively partition the data set into subsets by evaluating the normalized information gain (difference in entropy) resulting from choosing a descriptor for splitting the data. The descriptor with the highest information gain is used on every step. The training process stops when the resulting nodes contain instances of single classes or if no descriptor can be found that would result to the information gain. The method is classification-only.

References

1. Weka 3: Data Mining Software in Java, [Weka Website](#)
2. `weka.classifiers.trees.J48`, [J48 class description](#)
3. Ross Quinlan (1993). C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, CA.