

# Data Mining Exam HWS 2013 (First take)

1. Briefly describe what hierarchical agglomerative clustering is.  
What does the y-axis of an dendogram indicate? Name two advantages and one disadvantage in comparison to K-Means clustering.
2. Explain how the Hunt's Algorithm with the Decision Tree works.  
What impurity measures do you know? Name two of them.
3. Bayesian classification: How does the classifier handle missing values? How is it possible to cope with zero - probabilities?
4. Rule based learning. Briefly explain the characteristics. What is the problem with the Rule Simplification? What can we do against it?
5. Explain the Apriori Principle (From Association Rule)
6. TF-IDF. 3 little sentences, each is a document d. Measure the similarity between the documents. Choose a similarity of your choice (Note: choose a similarity where you don't need to calculate complex terms)