

Assignment 7: Decision Trees

In this assignment we use a data set of animals and their attributes. Using a decision tree classifier the computer learns to classify animals into different categories (mammals, fish, reptiles etc). Use the **zoo.xls** data set

1. Study the animals in the Excel document ([zoo.xls](#)). Without using a data mining tool, draw a decision tree of three to five levels deep that classifies animals into a mammal, bird, reptile, fish, amphibian, insect or invertebrate.
2. Open the zoo.arff data set in WEKA and Find out:
 - 2.1 How many animals are there in the data set?
 - 2.2 How many attributes are known of each animal?
3. Let us build some classifiers. Go to the classifier tab. We will use 66% of the animals to build the models, and the remaining 34% to evaluate the quality of the model, so select **percentage split – 66%**. First we will build a ‘naïve’ model that just predicts the most occurring class in the data set for each animal. This corresponds to a decision tree of depth 0. Click start to build a model.
 - 3.1 What % of animals is correctly classified?
 - 3.2 Into what category are all these animals classified and why?
4. Now build a **decision tree of depth 1** (a.k.a. a decision stump - select choose – trees – **decision stump**).
 - 4.1 Draw the discovered decision tree.
 - 4.2 What % of animals is correctly classified?
 - 4.3 Give an example of an animal that would not be classified correctly by this model.
5. Now build a decision tree of **any depth** (a.k.a. a **J48 tree**). Go to the classifier tab and select the decision tree classifier j48. Click on the line behind the choose button. This shows you the parameters you can set and a button called 'More'.
 - 5.1 Which algorithm is implemented by j48?
 - 5.2 Draw the discovered decision tree.
 - 5.3 What percentage of instances is correctly classified by j48?
 - 5.4 Which families are mistaken for each other?
6. Again go to the parameter settings by clicking on the box after the 'Choose' button. Now change binarySplit to true and build a new decision tree.
 - 6.1 Draw the discovered decision tree.
 - 6.2 What is the difference?
 - 6.3 What % of animals is correctly classified?
 - 6.4 Give an example of an animal that would not be classified correctly by this model.