Lab Exercise 3

Decision Trees Comparison

Task: Based on the introduction to data classification and decision tree algorithms from Lecture 3, 4 and Lab 2, perform an experiment to compare various Decision Tree algorithms and their performance. Prepare an experiment report (max. 4 A4 pages) presenting your results according to the instructions below.

1. Download dataset **weather** and **iris** (both attached to the WEKA package).
2. Prepare data for import to WEKA.
3. Start WEKA and generate the following four classification models:
    a. Decision tree – **RandomTree,**
    b. Decision tree – **j48**,
    c. Decision tree – **REPTree,**
    d. and **Decission table**.
4. Changing the classifiers' parameters and analysing the Confusion Matrix for each algorithm, create a satisfactory decision tree based classification model and calculate and compare their performance parameters such as:
    a. Accuracy
    b. Sensitivity
    c. Specificity – How WEKA names this parameter?
    d. False-positive rate
5. By analyzing the above parameters and the ROC curve (Receiver Operating Characteristic), compare the created algorithms and rank them in the order from best to worst.
6. Write the experiment report containing:
    a. Part describing principles of operation for each algorithm (use WEKA documentation and online materials)
    b. Describe data sets (how many / what are the attributes, what does the data collection describe, comment on data diversity and distribution as well as add any other comments/descriptions about the data set that you consider important).
    c. Present results from the classification process presenting each decision tree / decision table for each data set.
    d. Discuss the results as well as present the conclusion and the summary of the experiment.

The report should be sent by email in **ONE** pdf file. When naming the file please use the following naming convention: **DM_LAB3_name_surname.pdf**. An email with the file should be sent to the email address of the lecturer and titled: **DM_LAB3_name_surname**.