

# Artificial Intelligence 2022/2023

## Exercise Sheet 5b: Supervised Learning

### 5.3 Iris flower extended data set –Classification using different Algorithms.

Continuing with the Iris dataset, suppose that we have Iris already identified in the 3 classes but now we also have the Iris packed in different types of packages: “Simple – 0”, “Gift – 1” and “Luxury – 2”. We also have a new variable “price” with three possibilities: “Low”, “Medium”, “High”.



Iris setosa



Iris versicolor



Iris virginica

We now have a different classification problem in which we want to predict the “price” classification based on the remaining characteristics: sepal\_length\_cm, sepal\_width\_cm, petal\_length\_cm, petal\_width\_cm, iris\_type, and package.

- a) Create a new notebook and start by importing the needed libraries.
- b) Read the data from the CSV file and check the data using the head(), describe(), and other Pandas commands.
- c) Using only the attribute sepal\_length\_cm, sepal\_width\_cm, petal\_length\_cm, petal\_width\_cm, fit a simple decision tree model to the data, using holdout, with 75% for training.
- d) Analyze the accuracy, precision, recall and f-measure achieved.
- e) Create and analyze a confusion matrix of the results.
- f) Using only the attribute sepal\_length\_cm, sepal\_width\_cm, petal\_length\_cm, petal\_width\_cm, fit a simple nearest neighbor model to the data using holdout with 75% for training.
- g) Analyze the accuracy, precision, recall and f-measure achieved and the confusion matrix.
- h) Use two different methods for balancing the dataset and repeat the previous analyses.
- i) Using all the attributes available, and the balanced dataset, fit distinct models such as Nearest Neighbor, Decision Trees, SVMs and Neural Networks to the data and try different configuration parameters, using holdout with 75% for training.
- j) Analyze the accuracy, precision, recall and f-measure achieved and the confusion matrix.
- k) Repeat the previous analysis but experimenting with all models using 5-fold Cross-Validation.
- l) Use Grid Search to define the best parameters for the best two algorithms.
- m) Analyze the accuracy, precision, recall, f-measure and the confusion matrix for the best model.