

▷ Exercise 2: RapidMiner - Basic Workflow in Data Science

This exercise

- covers basic data exploration and modeling in RapidMiner Studio

1. Task: Data Exploration

Explore the iris data set:

- What is the number of observations and attributes?
- Are there special attributes?
- Use Histograms and Boxplots to inspect data scales.
- Use the operator Statistics. What does it yield?

2. Task: Modeling

Objective: *Determine the type of flowers depending on 4 numerical properties.*

- Make sure to see: This is multiclass classification.
- Prepare a workflow for k NN on the iris data set.
- What are variables, paramter and hyperparameter in this case?
- What is the accuracy of k NN ($k=5$) on the iris data set?
- What is the accuracy of k NN ($k=1$) on the iris data set?
- Which one performs better, $k=1$ or $k=5$? Why?
- What is the meaning of the confusion matrix?

3. Task: Inspection

Continue the 2. task.

- Mark misclassified examples using the operator Generate Attributes.
- Show classification result in (scatter) plot.
- Use Filter Examples for showing a table of only misclassified examples.