3.1. Should we play golf?

The *Golf data* set is one of the examples that are delivered together with RapidMiner. The data set models different aspects of the weather (outlook, temperature, humidity, forecast) that are relevant for deciding whether one should play golf or not.

1. Learn a Naïve Bayes Model from the Golf data set (Operator: *Naïve Bayes*). Use the model to classify the examples in the Golf-Testset, which is also delivered together with RapidMiner (Operator: *Apply Model*).

2. Think about ways how you can evaluate the performance of your model. What measures can be calculated from the resulting dataset?

3. Does a k-nearest-neighbor classifier work better for this task? Replace the *Naïve Bayes* operator with a *K-NN* operator and check how the accuracy of your classifier changes to find out. Do different values of k improve the performance?

3.2. Learning a classifier for the Iris Data Set

You want to learn and evaluate a classifier for recognizing different types of Iris flowers. In this case you have only one single dataset.

1. Let’s try Naïve Bayes first. Build a process that (1) loads the Iris data into your process. (2) Afterwards, the process should use the *Split Validation* operator (split ratio=0.7, stratified sampling) to generate a training and test data set. (3) As inner operator of the split validation, the process should use the *Naïve Bayes* operator to learn a model and the *Performance (Classification)* operator to evaluate the accuracy of the learned model.

2. Now, place a *Sort* operator, sorting by label before your Split-Validation and select linear sampling. Observe the changes in the result and try to explain the behavior.

3. Try a k-nearest-neighbor classifier on the problem. Does it perform better?

4. Also try normalizing your data and play around with different similarity measures. What is the best accuracy you can reach?