

**Information mining exercise, winter term 2022/23**

Dr.-Ing. Benedikt Loepp (benedikt.loep@uni-due.de)

**Exercise sheet 8**Presentation on **22.12.2022****Exercise 1: Precision and recall**

- (a) In classification tasks, the following four outcomes can occur. Please explain these outcomes using an illustrative example.
- true positive (TP)
  - true negative (TN)
  - false negative (FN)
  - false positive (FP)
- (b) Given the four outcomes, also describe how precision and recall are defined. It may be useful to draw a so-called Venn diagram for illustration purposes.

**Exercise 2: More metrics**

- (a) Explain the following metrics and provide the definition.
- Accuracy
  - Mean-squared error
  - F-measure
- (b) Which of the metrics mentioned in this and the task before can be used to determine whether you succeeded i) in the CPU speed prediction task on exercise sheet 4 (task 3) and ii) when using the Naïve Bayes classifier on exercise sheet 5 (task 3)? Shortly justify your answer.

**Exercise 3: Cross validation in RapidMiner**

- (a) Describe the general idea of cross validation, and, more specifically, of stratified 10-fold cross validation.
- (b) Given the process you had to create for the last exercise sheet to measure the performance in the CPU prediction task (sheet 7, task 4). Adapt the process so that it uses the **Cross Validation** operator to measure the performance of your linear classifier more appropriately.

**Exercise 4: Significance testing in RapidMiner**

- (a) Please explain why it is recommended to use the Student's t-test to compare two machine learning schemes.
- (b) On exercise sheet 3, we created a decision tree classifier (task 3). Now, we want to find out if it performs better with i) gain ratio or ii) information gain. For this, please apply cross validation as in the task above. Then, use the **T-Test** operator to compare the results of the two **DecisionTrees** that you need to create for this purpose. Hint: The **Multiply** operator can be used to provide the input for the **Cross Validation** operators.