

Exercise Week 07

Sprachverarbeitung (VL + Ü)

Nils Reiter, nils.reiter@uni-koeln.de

May 9, 2023 (Summer term 2023)

1. Create an ARFF file from `titanic.csv`. For this, you need to specify the header with the data types, while the actual data set can remain as it is. You need to make a copy of the file into your own directory first.

You can test your file by asking the class `weka.core.converters.ArffLoader` to load it:

```
java -cp /tools/weka/weka.jar weka.core.converters.ArffLoader FILE
```

2. Train and evaluate a machine learning model with Weka. The simplest command to do that is

```
java -cp /tools/weka/weka.jar weka.classifiers.Evaluation weka.classifiers.bayes.NaiveBayes -t DATASET.
```

Please replace `DATASET` with the filename of your ARFF file. Inspect the output and verify that this is indeed predicting the survival of the passengers. If you leave out the dataset argument, available options are printed on the command line. Alternatively, you can also read about them in the API documentation: <https://weka.sourceforge.io/doc/stable-3-8/>

3. If you're certain Weka does what it's supposed to do, inspect the output of the evaluation and identify parts you don't understand.