

## Künstliche Intelligenz (SS 2021)

### Aufgabenblatt 2

zu bearbeiten bis: 14.06.2021

---

#### Aufgabe 2.1 (Naive Bayes: Implementierung (Teil 1))

Implementieren Sie das Training eines Naive-Bayes-Klassifikators für Songtexte: Der Klassifikator soll automatisch Lyrics von Songtexten ihrem Genre (wie *Metal*, *Rock*, *Pop*, ...) zuordnen.

- Verwenden Sie die Dateien `train*.csv` (siehe `read.mi`), vorwiegend `train_small.csv`. Kombinieren
- In `classifier.py` (siehe `Read.MI`) finden Sie ein Programmgerüst für Ihren Klassifikator. Modifizieren Sie die `main`-Routine, so dass die gefilterten Dateien eingelesen werden. Erstellen Sie mit Hilfe Ihres Vokabulars (letzte Woche) Bag-of-Words-Features für Ihren Klassifikator.
- Sie starten das Skript mit folgendem Befehl:

```
python classifier.py --train
```
- Lesen Sie sich den Code in `classifier.py` durch. Ihre Aufgabe ist es, die Klasse `NaiveBayesLyricsClassifier` zu implementieren:
  - Verwenden Sie – wie in der Vorlesung – boolesche Bag-of-Words Features: Kommt ein Wort in einem Dokument vor, ist das entsprechende Merkmal 1, ansonsten 0. Die *Häufigkeit* des Wortes ist irrelevant.
  - Implementieren Sie diese Woche nur den Konstruktor `__init__()` und die Methode `train()`. In `train()` sollten Sie die Wahrscheinlichkeiten  $P(x_i = 1|c)$  für jeden Term  $x_i$  und jede Klasse  $c$  lernen, sowie den Prior  $P(c)$ .
  - Schreiben Sie außerdem etwas Code, um die gelernten Wahrscheinlichkeiten zu speichern und später wieder zu verwenden. Hier bietet sich z.B. das Python-Modul `pickle` an.