

## Künstliche Intelligenz (SoSe 2024)

### Aufgabenblatt 2

zu bearbeiten bis: 15.05.2024

---

#### Aufgabe 2.1 (Naive Bayes: Implementierung 2 – Anwendung)

Finalisieren Sie Ihre Naive-Bayes-Implementierung:

- (a) **Anwendung:** Implementieren Sie die Methode `apply()`, die den Eingabedokumenten jeweils eine Klasse zuweist. *Beachten Sie folgende Hinweise, die wir auch im Praktikum noch einmal besprechen:*
- **Glätten** Sie die Einzelwahrscheinlichkeiten auf eine Mindest-Wahrscheinlichkeit  $\epsilon$ .
  - Beachten Sie, dass in Ihr Produkt **alle Worte des Vokabulars einfließen** sollten. Kommt ein Wort nicht im Zieldokument vor, gilt die Gegenwahrscheinlichkeit.
  - Da viele einzelne Wahrscheinlichkeiten aufmultipliziert werden, erhalten wir sehr kleine (im Allgemeinen zu kleine) Zahlen. Rechnen Sie deshalb statt mit den  $W$ 'keiten mit ihren **Logarithmen**.
- (b) **Testen:** Klassifizieren Sie die Dokumente der Testmenge mit folgendem Befehl:
- ```
python classifier.py --apply
```
- [Deliverable] Geben Sie für alle Dokumente den Titel, die korrekte Klasse, und die von Ihrem Modell geschätzte Klasse aus. Welche Klassen werden scheinbar häufig miteinander verwechselt?
- (c) **Messen:** Erweitern Sie Ihr Programm, so dass die Genauigkeit – d.h. der Prozentsatz korrekt klassifizierter Testdokumente – gemessen wird.
- [Deliverable] Geben Sie die Genauigkeit Ihres Modells an.
- (d) **Optimieren:** Testen Sie in einer Grid Search verschiedene geeignete Werte für die Minimalwahrscheinlichkeit  $\epsilon$ . Welches  $\epsilon$  funktioniert am besten?
- [Deliverable] Fassen Sie in einer Tabelle zusammen, welche Werte  $\epsilon$  Sie getestet haben und welche Genauigkeiten Sie jeweils erzielt haben.