

# Bonusabgabe

## Anwendungen der Künstlichen Intelligenz

Dr. Michael Färber, Dr. Tobias Käfer, Anna Nguyen

10.12.2021

### Aufgabenstellung

Gerade in den letzten Jahren wurde immer mehr deutlich, was für eine Reichweite Kommentare und Beiträge im Internet haben können. Der Großteil der Meinungsäußerung findet online statt. Dadurch steigt jedoch auch die Anzahl an unangemessen und schädlichen Inhalten, die von Mobbing bis hin zu Shit-Storms ausarten können.

Ihre Aufgabe bei dieser Herausforderung wird es sein, die Rolle einer „Content-Polizei“ zu übernehmen. Um ein zivilisiertes und ordnungsgemäßes Verhalten auf einer imaginären Social Media-Plattform zu gewährleisten, müssen Sie alle unerwünschten Inhalte aussortieren. Zu diesem Zweck wird Ihnen ein Trainingsset mit anonymisierten Benutzerbeiträgen zur Verfügung gestellt, die jeweils als angemessen oder nicht angemessen gekennzeichnet sind. Verwenden Sie diese Trainingsmenge, um ein Modell zu entwickeln, das Beiträge in angemessene und unangemessene klassifizieren kann.



### Data

Die Datei train.tsv umfasst Ihre Trainingsdaten. Jede Zeile enthält den Text eines Beitrags und ein Label, getrennt durch ein TAB-Zeichen. Erwähnungen von Benutzernamen innerhalb des Textes wurden durch „@USER“ und URLs durch „{{URL}}“ ersetzt. Die Klassenlabels sind „OK“ für geeignete Inhalte und „NG“ („not good“) für unangemessene Inhalte.

Die Datei test.tsv.dist enthält die Testdaten, die am Ende zur Bewertung Ihres Modells verwendet werden. Sie enthält keine Klassenlabels.

### Formalitäten

(1) Verwenden Sie bestehende Methoden des maschinellen Lernens für die Klassifizierung.

(2) Sie können existierende Tools (e.g. Weka<sup>1</sup>), Libraries (e.g. Sklearn<sup>2</sup>, NLTK<sup>3</sup>, Keras<sup>4</sup>) und Programmiersprachen (vorrangig Python<sup>5</sup>; R<sup>6</sup> auch möglich, falls bereits erlernt) verwenden.

---

<sup>1</sup><https://www.cs.waikato.ac.nz/ml/weka/>

<sup>2</sup><https://scikit-learn.org/stable/>

<sup>3</sup><https://www.nltk.org/>

<sup>4</sup><https://keras.io>

<sup>5</sup><https://www.python.org>

<sup>6</sup><https://www.r-project.org>

## Abgabe

Die Abgabe ist bis zum 30. Januar 2022, 23:59 Uhr über Ilias in dem Ordner Bonusabgabe möglich. Einreichungen, die nach der Frist eingehen, werden von der Bewertung ausgeschlossen. Die folgenden Unterlagen müssen eingereicht werden:

1. Eine Kopie der Datei test.tsv.dist mit den von Ihnen vorhergesagten Klassenlabels (die Reihenfolge der Zeilen muss der ursprünglichen test.tsv.dist entsprechen; verwenden Sie \OK" und \NG" als Klassenlabels).
2. Programmcode, falls verwendet. Wenn Weka oder ein anderes Tool verwendet wurde, senden Sie bitte Screenshots, die die verwendeten Funktionen/Parameter/etc. veranschaulichen.
3. PDF-Datei mit:
  - Vollständigem Namen, E-Mail-Adresse und Matrikelnummer.
  - Kurzer Beschreibung des Ansatzes für die Klassifizierung (4-5 Sätze)
  - Informationen zu den Bewertungskriterien und dem Training/Test-Split, den Sie bei der Erarbeitung verwendet haben.

Bitte laden Sie alle oben genannten Informationen/Daten in einem ZIP-Archiv von angemessener Größe (maximal einige MB) hoch und benennen Sie dieses wie folgt: u-Kürzel.zip. Abgaben, die formal nicht korrekt sind (z.B. inkorrekte Dateinamensgebung, falsches Dateiformat oder fehlende Dateien in der Abgabe), werden nicht bewertet.

## Ergebnisse

### Auswertung

Die von Ihnen vorhergesagten Labels für die Einträge in test.tsv.dist werden mit Hilfe eines F1 Score bewertet.

### Bonuspunkte

Wenn Ihre Vorhersagen, die anhand der Testdaten ausgewertet werden, einen Makro-F1-Wert von 0,85 erreichen, erhalten Sie einen Bonuspunkt, bei 0,90 zwei Bonuspunkte und bei 0,95 drei Bonuspunkte auf Ihre bestandene Prüfung. Die beste Abgabe von allen erhält zusätzlich noch einmal drei Bonuspunkte.

Happy policing.

