

## Bonusabgabe

### Anwendungen der Künstlichen Intelligenz

Dr. Tobias Käfer, M.Sc. Lukas Kinder,  
B.Sc. Alexander Mittauer

02.12.2024

### 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 hate speech, offensive language oder neither gekennzeichnet sind. Verwenden Sie diese Trainingsmenge, um ein Modell zu entwickeln, das Beiträge in diese drei Klassen möglichst gut klassifizieren kann.



### Daten

Die Datei train.csv umfasst Ihre Trainingsdaten. Jede Zeile enthält den Text eines Beitrags und ein Label, getrennt durch ein Komma. Die Klassenlabels sind „0“ für hate speech, „1“ für offensive language und „2“, für neither.

Die Datei test\_no\_labels.csv enthält die Testdaten, die am Ende zur Bewertung Ihres Modells verwendet werden. Sie enthält keine Klassenlabels.

Hinweis: CSV-Dateien nicht mit Excel öffnen, da die Struktur dadurch teilweise korruptiert werden kann.

### Formalitäten

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

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

(3) Die Aufgabe kann entweder alleine oder zu zweit bearbeitet werden. Wenn Sie im Team arbeiten, geben Sie bitte an, mit wem Sie die Gruppe gebildet haben, und reichen Sie die Aufgabe jeweils individuell ein.

Sollte die Bonusaufgabe in den vergangenen Jahren bereits bearbeitet worden sein, wird der Bonus für Sie **nicht** automatisch auf dieses Jahr übertragen. Sie müssen die Aufgabe erneut einreichen. In diesem Fall ist es nicht erlaubt, mit jemandem zusammenzuarbeiten, der die Bonusaufgabe zum ersten Mal bearbeitet.

<sup>1</sup><https://www.python.org> <sup>2</sup><https://www.r-project.org>

<sup>3</sup><https://pandas.pydata.org> <sup>4</sup><https://scikit-learn.org/stable/>

<sup>5</sup><https://www.nltk.org/> <sup>6</sup><https://keras.io>

## Abgabe

Die Abgabe ist bis zum 31. Januar 2025, 23:59 Uhr über Ilias in dem entsprechenden Ordner möglich. Einreichungen, die nach der Frist eingehen, werden von der Bewertung ausgeschlossen. Die folgenden Unterlagen müssen eingereicht werden:

### 1. Datei mit vorhergesagten Klassenlabels:

- Eine Kopie der Datei `test_no_labels.csv` mit den von Ihnen vorhergesagten Klassenlabels.
- Die Reihenfolge der Zeilen muss exakt mit der ursprünglichen Datei übereinstimmen, und die verwendeten Klassenlabels müssen den im Training genutzten entsprechen.
- Benennen Sie die Datei bitte **test\_with\_label.csv**.

### 2. Programmcode:

- Der vollständige von Ihnen verwendete Code, inklusive einer **README-Datei**, die beschreibt, wie der Code ausgeführt wird.
- Die Ausführung des Codes sollte die Datei **test\_with\_label.csv** (siehe Punkt 1) erzeugen.
- Stellen Sie sicher, dass Ihr Programm konsistent arbeitet und immer dieselbe .csv-Datei erzeugt (verwenden Sie gegebenenfalls Random-Seed-Werte).

### 3. PDF-Dokument:

- Vollständigem Namen, u-Kürzel und Matrikelnummer
- Ggf. Teampartner und dessen u-Kürzel
- Kurze Beschreibung des Ansatzes für die Klassifizierung in 4-5 Sätzen (z.B. Preprocessing, Trainingsplit, Modell, wichtige Parameter, ..)

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 Testdatei werden mit Hilfe eines gewichteten F1 Scores bewertet. Wenn Ihre Vorhersagen einen Weighted-F1-Score von 0,80 erreichen, erhalten Sie einen Bonuspunkt, bei 0,85 zwei Bonuspunkte und bei 0,90 drei Bonuspunkte auf Ihre bestandene Prüfung. Die beste Abgabe von allen erhält zusätzlich noch einmal drei Bonuspunkte.

### Testen

Vor der Abgabe können Zwischenlösungen getestet werden. Hierfür steht eine eigene Website zur Verfügung, welche unter folgendem Link erreicht werden kann: <http://193.196.36.35:5000/>. Nach Upload der Datei `test_with_label.csv` bekommt man Feedback bzgl. der erreichten Bonuspunkte (1-3).

Happy policing.

