

# Anwendungen der KI Challenge

Prof. Dr. York Sure-Vetter, PD Dr. Andreas Harth  
Tobias Weller

24.11.2017

## Aufgabenstellung

Die Aufgabenstellung für die Challenge besteht darin im Team, bestehend aus vier bis fünf Studierenden, Texte hinsichtlich der Stimmung zu klassifizieren. Wir betrachten dabei die binären Klassen **positiv** und **negativ**. Jeder Text lässt sich dazu in genau eine Klasse zuordnen. Die Trainingsdatensätze sind dabei bereits gelabelt (Texte in **traing-neg.txt** sind dabei als negative Kritik anzusehen; Texte in **traing-pos.txt** sind als positive Kritik anzusehen).

Ziel ist es die Testdatensätze (**test.txt**) korrekt zu klassifizieren. Sie reichen ihre Ergebnisse per Email ein (siehe unten). Wenn die Accuracy Ihres Modells auf dem Testdatensatz 75% oder höher ist bekommen Sie einen Bonus von 3 Punkten auf Ihre bestandene Klausur. Der Bonus ist lediglich auf die Prüfungen im WS17/18 als auch im SS18 verfügbar. Die Abgabe erfolgt bis zum 26.01.2018 23:59 Uhr (siehe unten).

Die Ergebnisse der Challenge werden spätestens am 29.01.2018 im ILIAS zur Verfügung stehen. In der Übung von Anwendungen der KI am 30. Januar 2018 werden die Ergebnisse vorgestellt. Die Teams die Ihre Ergebnisse kurz präsentieren sollen (ohne Powerpoint, es beschränkt sich auf die Beschreibung der Vorgehensweise und Diskussion) werden per Email benachrichtigt. Es muss mindestens ein Teammitglied anwesend sein und die Vorgehensweise beschreiben.

Die Besten drei Teams erhalten zusätzlich noch einen nicht prüfungsbezogenen Preis.

## Datensatz

Der Datensatz besteht aus drei Dateien. Die Texte sind allesamt Kritiken von Benutzern auf [imdb](http://imdb.com)<sup>1</sup>. Die Kritiken stammen nicht von einem spezifischen Film, sondern bestehen aus einer gemischten Ansammlung von Filmen. Jede Zeile in den Dateien repräsentiert eine Kritik. Die Kritiken sind alle auf englisch.

- **traing-neg.txt**: Diese Datei enthält alle Kritiken welche sich negativ zu einem Film äußern.
- **traing-pos.txt**: Diese Datei enthält alle Kritiken welche sich positiv zu einem Film äußern.
- **test.txt**: Diese Datei enthält unlabeled positive als auch negative Kritiken zu Filmen. Sie dient zur Evaluation des erstellten Modells.

---

Version von: Freitag 24 November, 2017 13:31

<sup>1</sup><http://imdb.com>

## Hinweise zur Bearbeitung

- Die Teamgröße beträgt vier bis fünf Studierende. Kleinere Teams sind nicht erlaubt. Bitte finden Sie sich rechtzeitig zusammen. Nutzen Sie gegebenenfalls das ILIAS Forum.
- Wenden Sie für die Klassifikation existierende Methoden aus dem Text Mining und Maschinellen Lernen an um die Texte zu klassifizieren.
- Ändern Sie nicht die Reihenfolge der Texte in der Datei `test.txt`. Für die Evaluation erfolgt ein Abgleich Ihrer Vorhersagen mit den korrekten Werten.
- Sie dürfen existierende Tools, Libraries und Programmiersprachen verwenden.
- APIs, welche die Texte klassifizieren oder sonstige Umwandlungen ausführen, dürfen nicht verwendet werden.<sup>2</sup>

## Abgabe

Die Abgabe erfolgt bis zum 26.01.2018 23:59 Uhr per Email an `tobias.weller@kit.edu`. Einreichungen welche später erfolgen sind von der Challenge ausgeschlossen.

Folgende Informationen/Dateien müssen abgegeben werden:

1. Excel Datei mit der Liste an Klassifikationen der Kritiken von `test.txt`.
2. Programmcode und gelerntes Modell / gelernte Modelle.
3. Kurze Beschreibung der Vorgehensweise zur Klassifikation der Texte (4 - 5 Sätze).
4. Accuracy und verwendetes Data Split auf den Trainingsdatensatz.
5. Vollständiger Name, Email Adressen und Matrikelnummer jedes Teammitglieds, als auch ein Name des Teams.

---

<sup>2</sup>APIs wie bspw. WordNet sind erlaubt, da hier keine automatische Umwandlung der Texte ausgeführt wird.