

NLP Assignment 3

Dieses Assignment wird bewertet !

Die Zusammenarbeit unter Studierenden ist erwünscht, so lange sich diese auf die Diskussion von Konzepten oder Problemen mit python konzentrieren. Das Kopieren von Code ist nicht erlaubt, in diesem Falle werden alle Lösungen der beteiligten Parteien mit 0 Punkten bewertet; dazu werden alle Lösungen (manuell und automatisiert mit Plagiat-Checkern) auf Gruppenarbeit und Kopien untersucht.

In diesem Assignment geht es um die Analyse von *Sentiment* und *Emotionality* in Reviews von Amazon-Produkten (die meisten aus dem Baby-Sortiment). Die Reviews mit einem Rating (von 1 bis 5) finden Sie in der Datei `Amazon_Baby_train.txt`.

- 1.) Entwickeln Sie für diese Amazon-Reviews einen *Sentiment-Analyzer*, welcher den Reviews einen *Sentiment-Score* $\in \{1, 2, 3, 4, 5\}$ zuordnet. Das Ziel dabei ist, eine möglichst grosse **accuracy** zu erzielen. Stellen Sie das Resultat mit einer *confusion matrix* dar.
- 2.) Entwickeln Sie einen zweiten *Sentiment-Analyzer* mit dem Ziel, für die Klassen 4 und 5 einen möglichst grossen (durchschnittlichen) *F1*-Wert zu erzielen.
- 3.) Entwickeln Sie einen dritten *Sentiment-Analyzer*, welcher für die Klasse 1 einen möglichst grossen *F1*-Wert erzielt.
- 4.) Bestimmen Sie mit einer Methode (oder einer Kombination von Methoden) die *Emotionality* der Reviews und geben Sie die 10 Reviews mit der höchsten *Emotionality* an.

Hinweise:

- Die Arbeit von „*Adding Sentiment Analysis support to the NLTK Python Platform*“ von Pierpaolo Pantone (liegt auf AD) könnte Ihnen allenfalls hilfreich sein.

Abgabe : In Form eines Jupyter Notebooks
per Email an dominik.frefel@fhnw.ch bis 20. Mai 2018