

NLP: Sprachbestimmung

Christian Wartena

8. April 2021

1 Aufgabe

1. Erstellen Sie ein Programm, das die Sprache eines Textes bestimmt und dabei aus zwei Sprachen wählt, zum Beispiel Deutsch und Englisch, oder Spanisch und Portugiesisch
2. Trainieren Sie hierfür zwei Sprachmodelle auf einem Korpus mit mindestens 100 000 Wörtern
3. Testen Sie das Programm mit möglichst kurzen Sätzen oder einzelne Wörter.

2 Lernziele

Am Ende dieser Lerneinheit sollen Sie in der Lage sein:

1. n-Gramme aus Texten zu extrahieren;
2. Einfache Sprachmodelle auf der Grundlage von n-Gramm-Statistiken zu erstellen;
3. ein Classifier zu erstellen um die Sprache eines Textes zu bestimmen

3 Materialien

Um die Aufgabe zu lösen gibt es folgende Materialien:

1. Ein Jupyter Notebook