

## NLP Assignment 2

### 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 der Datei Messages\_train.csv finden Sie Texte von weiblichen und männlichen Personen aus drei Altersklassen:

10s (13 bis 17 Jahre)	20s (23 bis 27 Jahre)	30s (33 bis 37 Jahre)
-----------------------	-----------------------	-----------------------

Die Texte sind wie folgt strukturiert:

„id“ , „text“ , „person\_gender“ , „person\_age“

Entwickeln Sie drei verschiedene *Classifier* für

- die Identifikation von weiblichen und männlichen Personen,
- die Identifikation der Altersklasse und
- die kombinierte Zuordnung sowohl zur Altersklasse als auch zum Geschlecht.

Das Ziel soll sein, für jeden der drei verschiedenen *Classifier* eine möglichst **grosse accuracy** zu erreichen.

Berechnen Sie für jeden *Classifier* die *confusion matrix* und die *accuracy* sowie die *precision*, *recall* und *F1-score* sowohl für jede Klasse als auch die gewichteten Mittelwerte.

Teilen Sie die Daten allenfalls in ein Trainings- und ein Dev-Set auf.

Hinweis:

- Die Qualität der *Classifier* werden wir mit einer anderen Test-Datei messen, für welche stellvertretend vorerst die Datei dummy\_test.csv eingelesen wird.

Abgabe : In Form eines Jupyter Notebooks  
per Email an [dominik.frefel@fhnw.ch](mailto:dominik.frefel@fhnw.ch) bis 14. April 2019