

## Week 4 - Practice

### Topic Modeling

For a set of documents extracted from Wikipedia for three different topics at your choice, perform a semantic analysis as follows:

1. Use text preprocessing techniques (stemming/lematization, stop words removal) and create the bag-of-words and TF-IDF vectorizations
2. Using Latent Semantic Analysis with SVD for a) the bag-of-words encoding and b) the TF-IDF encoding
3. Using Non-negative matrix factorization
4. Using LDA

Use **sklearn** library in python. Check the tutorial here: <https://nlpforhackers.io/topic-modeling/>