Week 5 - Practice

Topic Modeling

For a set of 15 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. Use Latent Semantic Analysis with SVD for
    a. the bag-of-words encoding and
    b. the TF-IDF encoding
3. Use Non-negative matrix factorization
4. Use LDA
5. Document on the evaluation metrics from the gensim library and apply them on ypu results

Use **sklearn** library in python. Check the tutorial on LDA:
https://towardsdatascience.com/end-to-end-topic-modeling-in-python-latent-dirichlet-allocation-lda-35ce4ed6b3e0