**Practice assignment № 2 «Naïve Bayes Classification for spam filtering»**

Modules of scikit: sklearn.naive\_bayes, cross\_validation, metrics, probably pandas

1. Introduction to the subject can be found here https://web.stanford.edu/~jurafsky/slp3/6.pdf

1. Choose an arbitrary dataset of spam/ham messages (i.e. it can be found here <http://www.esp.uem.es/jmgomez/smsspamcorpus/>).

2. If necessary, apply basic text preprocessing methods (transform to lower case, remove punctuation symbols, remove stop-words…)

3. Create matrix with objects and features using bag of words model.

4. Train different classifiers and compare using cross-validation: а) Gaussian Naïve Bayes; b) Multinomial Naïve Bayes; в) Bernoulli Naïve Bayes.