# 04. Classification – Homework

## Overview

* [Large Movie Review Dataset](http://ai.stanford.edu/~amaas/data/sentiment/)
* Target – pos/neg folders

## Evaluation Criteria

Scoring (8 max):

* Text preprocessing *– 3*
  + Removing stop words
  + Stemming / Lemmatization
  + Bag of words / TF-IDF
  + N-grams
* Words importance - 2
* Hyperparameters tuning – 1
* Compare performance of models: SGDClassifier; SVM; Naive Bayes - 2

Metric: AUC-ROC

Libraries: scikit‑learn + NLTK