

Assignment 4

Instructions

This assignment is ungraded, so there is no need to share your solution. Please complete this exercise sheet to practice what you learned in the lecture.

The assignment will be discussed in the tutorial on 23.05.2024.

Exercise 1 - Vector representations

Present 3 different vector representations for the following document d. Discuss the advantages and disadvantages of each approach.

Amateur, no budget films can be surprisingly good ... this however is not one of them.

Exercise 2 - Sentiment classification

Implement a Naïve Bayes classifier for a sentiment classification task. Use the imdb movie review dataset (i.e. from <https://huggingface.co/datasets/stanfordnlp/imdb>) to train and test the model. Draw a confusion matrix of the result and evaluate the performance of the model with appropriate metrics as taught in the lecture. Discuss the advantages and disadvantages of each metric.