

732A81/732A92/TDDE16 Text Mining (2022)

# Text classification

Marco Kuhlmann

Department of Computer and Information Science

# Announcements

# This session

- Questions and answers
- Assessment criteria for the project
- Introduction to the lab

Questions and answers

# Overview of text classification

1. Introduction to text classification
2. Evaluation of text classifiers
3. The Naive Bayes classifier
4. The Logistic regression classifier
5. Support vector machines

# Quiz: Evaluation of text classifiers

<https://forms.office.com/r/rBEH9D27yc>



## Beehives (10 min)

- Name one concept that you were already familiar with
- Name one concept that you found hard to understand
- Select one of the 'hard' concepts and write it on the whiteboard.

Assessment criteria for the project

# Project structure

- |                           |                    |
|---------------------------|--------------------|
| 1. Identify your problem  | 8 hours (w44–w48)  |
| 2. Design your approach   | 32 hours (w49–w50) |
| 3. Evaluate your approach | 32 hours (w51–w01) |
| 4. Produce your report    | 16 hours (w02)     |

# Phase 1: Identify your problem

- A good way to identify a problem is to start with a data set that you find interesting. Have a look at sites such as Kaggle!
- You could also have a look at the project abstracts from previous years. It's fine to do a project that is inspired by an old project!
- In previous years, several students have used their Text Mining project to test an idea for a thesis project.

# Assessment criteria – Method

Are the methods applied in the project suitable to solve the stated problem? Are the experimental results analysed with appropriate evaluation methods? Are the findings from these analyses correctly interpreted?

- F: The work should have been done or evaluated differently. The findings from the analyses of the experimental results are misinterpreted, e.g. because there is no proper point of comparison (baseline, related work).
- E: The methods applied in the project are suitable to solve the stated problem. The experimental results are analysed with appropriate evaluation methods. The findings from the analyses are interpreted correctly.
- A: The technical approach is well-suited for the stated problem. The experimental results are analysed in detail (e.g., ablation studies) or from different perspectives. The findings from the analyses are compared with related work.

## Beehives (10 min)

Read the assessment criteria for Aspect 2, Method.

Suppose that you have done a project on sentiment analysis. Taking the assessment criteria as given, what would be examples

- that show that your report meets a certain criterion?
- that show that your report *does not* meet a certain criterion?

Discuss in the group!

# Introduction to the lab