

Instructions for the project report

Marco Kuhlmann

Introduction

The project module of the course 732A92 / TDDE16 Text Mining is examined based on a written report. The knowledge requirements for the module are as follows:

You identify and formulate a *substantial* text mining problem *with some help from a teacher*. In working on your problem, you implement and apply *suitable* text mining methods, analyse experimental results *with appropriate evaluation methods*, and summarise them *with simple judgements*. You clearly present and discuss the conclusions of your work.

A *problem* can either be a specific task ('build a system for sentiment classification of movie reviews') or the answering of a limited-scale research question ('investigate which text segmentation strategy yields the most coherent topic models').

Suggested structure

The suggested structure for the report is the following:

1. Introduction. Introduce the problem that you have addressed in your project. What did you do? How did you do it? What are your main results?
2. Theory. Present relevant theoretical background, and in particular those concepts and methods that were not covered in the course.
3. Data. Present your data. What information does it contain? Where did you get it from? What preprocessing did you do, if any?
4. Method. Explain how you carried out your study. Aim to be detailed enough for others to reproduce your results.

5. Results. Present your results in an objective way. Use tables and charts, but do not forget to also include a summary in text form. Do not interpret your results.
6. Discussion. Analyse your results. Compare your study to related work, such as scientific articles addressing a similar problem.
7. Conclusion. Summarise your results. Based on your results and their analysis, what new knowledge do you take away from your project?

Formal requirements

Your report must meet the following formal requirements.

Title page The first page of your report must contain the title of your project as well as your full name, LiU-ID, and course code (732A92 or TDDE16). Do not use a generic title such as ‘Text Mining Project Report’; instead, find a title that expresses what you did in your project, such as ‘Predicting Personality Types from Written Text’.

Abstract Together with your report, you must submit an abstract (via an online form). This abstract must be one paragraph, and may not exceed 200 words. The abstract should provide a concise summary of your project’s purpose, method, and results. Note that the abstracts of all projects will be published on the course website.

Length Your report should contain between 2,000 and 4,000 words, corresponding to approximately 4–8 pages of single-spaced text. This guidance excludes the title and abstract, as well as any non-textual elements such as figures, tables, mathematical formulas. Even references are excluded from this guidance.

References When relating your work to the work of others or when using ideas, code, or text from others, you must appropriately cite your sources. This also applies to materials obtained from the internet. When formatting references, choose a style that you are comfortable with.

Code Together with your report, you must submit (via an online form) a link to a repository containing the code that you wrote for the project. For example, this could be a public repository on GitHub or a private repository on LiU’s GitLab server. In case of the latter, you must grant us (aliba43, marku61) read access to the repository.

Proofreading Before submitting your report, make sure to carefully proofread your text and check it for errors in spelling and grammar. Consider using a writing assistant service such as *Grammarly* or *Ginger*. If you need help with your writing, consider seeking guidance from Academic English Support (*Språkverkstaden*).

Failure to comply with these requirements If you fail to comply with any of the formal requirements listed here, we will return your report to you, ask you to make the necessary changes, and submit the report for one of the additional examination opportunities. (Dates can be found on the course website.)

Assessment

When grading your report, we will assess it with respect to three aspects: problem, method, and conclusions. For each aspect, we will assign a component grade from the scale A, C, E (all passing grades) and F (fail). Descriptors for A, E and F are provided below. If your report meets all of the criteria in the descriptor for E and at least one of the criteria in the descriptor for A, your grade for the corresponding component will be C. To pass, you must have a passing grade in each of the components. Your final grade is computed as follows:

Final grade 732A92 Your final grade is your lowest component grade, or the next highest ECTS grade in case two of your component grades are higher than the lowest component grade. (Example: If your component grades are E, C, E, your final grade is E; if they are C, E, C, your final grade is D.)

Final grade TDDE16 As for 732A92, but the final ECTS grade is converted into a numerical grade on the scale 3, 4, 5. In this context, A–B correspond to 5; C corresponds to 4, and D–E correspond to 3. (Example: If your component grades are C, E, C, your final grade is 3; if they are A, C, A, your final grade is 5.)

Aspect 1: Problem

Is it clear what was done in this project, why it was done, and how it was done? Does the project go beyond what has been covered in the course? Does the project have enough substance, or would there have been room for more experiments or analysis?

- F The report does not contain a clear problem statement. The project is essentially a repetition of one of the lab assignments. For a project with this timeframe, I (the examiner) would have expected significantly more experiments or analysis.

- E The problem is clearly stated. The project goes significantly beyond the lab assignments, e.g. by using a method not covered there (and explaining it in the report). The project represents an appropriate amount of work.
- A The problem is well-motivated and placed in a broader scientific or societal context (including references). There are several creative elements. The project contains significantly more experiments and analysis than expected.

Aspect 2: 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.

Aspect 3: Conclusions

Are the conclusions of the work clearly presented and discussed? Does the report show awareness of the limitations of the work? Does it show awareness and understanding of related work documented in external sources (e.g., research articles)?

- F It is not clear what conclusions the author draws from this work. There is no discussion of the limitations of the work. There is no account of how the work done in the project compares to related work.
- E The conclusions are clearly presented and convincingly supported. The report shows awareness of the limitations of the work. The relation between the project and the work documented in the external sources is clear.
- A The report contains a detailed discussion of the limitations of the project. The report features a precise and enlightening comparison with related work. The majority of the external sources are peer-reviewed research articles.

Grading table

The following table shows all combinations of component grades together with their corresponding final grades for the two courses 732A92 and TDDE16. To avoid unnecessary repetitions, combinations are sorted in lexicographic order. Combinations that do not yield a passing grade (i.e., where at least one of the component grades is F) are excluded.

Combination	Final grade 732A92	Final grade TDDE16
A A A	A	5
A A C	B	5
A A E	D	3
A C C	C	4
A C E	D	3
A E E	E	3
C C C	C	4
C C E	D	3
C E E	E	3
E E E	E	3