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 apply the machine learning models covered in the course to suitable problems, evaluate their performance with appropriate validation methods, correctly interpret the results, and critically assess your work with well-developed judgements.

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'). *Well-developed judgements* are ones that (among others) are based on a substantial amount of relevant material (such as experiments or theoretical analysis), contain nuanced arguments (rather than unsubstantiated opinion), make use of concrete and relevant examples, and are clearly written.[1]

Suggested structure

The suggested structure for the report is the following:

1. Introduction. Introduce the task or question that you have addressed in your project. What were you trying to do? Why did you choose this project?

2. Theory. Present relevant theoretical background, and in particular the models that you have used. Where appropriate, use mathematical formulas.

3. Data. Present your data. What information does it contain? Where did you get it from? What preprocessing did you do, if any?

[1] More information on assessment criteria for critical judgments is available in a separate document.

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. What do you make of it? Discuss your results and present your analysis. Relate your results to your original task or question.

7. Conclusion. 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 requirements. Failure to comply with these may lead to your report being rejected without assessment.

Title The first page of your report must contain the project title as well as your full name and LiU-ID. 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 Every report must include an abstract. This abstract must be one paragraph, and may not exceed 250 words. The abstract should provide a concise summary of the project purpose, method, and results. 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, abstract, figures, tables, mathematical formulas, and references.

References When using ideas, code, or text from others, you must appropriately credit and acknowledge your sources. This also applies to materials obtained from the internet. When formatting references, choose whatever style you are comfortable with, but be consistent.[2]

Proofreading I attach a lot of importance to proofreading. Before submitting your report, you should carefully look for typos as well as errors in grammar and style.

[2] For an example style, see Section V of the IEEE Editorial Style Manual.

Assessment criteria

To grade your report, I will assess it with respect to the six criteria defined below.[3] For each criterion, I will assign a score between 1 (below standard) and 5 (excellent). Descriptors are provided for scores 1, 3, and 5. Based on the criteria-specific scores, I will compute an *overall score* according to the following algorithm: If all of the individual scores are at least 3, then the overall score is the average of the individual scores, rounded upwards to the next integer if necessary. Otherwise, the overall score is 3 minus the number of individual scores that are lower than 3. The overall score determines your grade as specified in the following table:

Overall score	≤ 0	1	2	3	4	5
Grade 732A92	Fx	E	D	C	B	A
Grade TDDE16	U	U	U	3	4	5

Clarity For the reasonably well-prepared reader, is it clear what was done in this project, and why? Is the report well-written and well-structured?

1 There are some important questions about the method, results, or analysis that I (the examiner) was not able to resolve even with effort.

3 The report was mostly understandable to me (the examiner).

5 I (the examiner) believe that any student who has successfully completed the course will understand what was done in this project, and why.

Innovativeness How exciting and innovative is the project? Does the project use any methods that were not covered in the course?

1 There are no original or innovative elements in this project. The project is essentially a repetition of one of the lab assignments.

3 The project applies methods that have been covered in the course to a new and suitable problem.

5 The choice of the problem and/or the approach are innovative. The project goes significantly beyond what has been covered in the course.

[3] The assessment criteria and their definitions are based on the review form (as of August 2018) for submissions to the *Transactions of the Association for Computational Linguistics* in the section *Empirical and Data-Driven Methods*.

Soundness and correctness Is the technical approach sound and well-chosen? Are the claims made in the report supported by proper experiments or theoretical analysis, and are the results correctly interpreted?

1 Troublesome. There may be some ideas worth salvaging here, but the work should really have been done or evaluated differently.

3 Fairly reasonable work. The approach is not bad, the validation methods are appropriate, and at least the main claims are probably correct.

5 The approach is very apt, and the claims are convincingly supported.

Related work Does the report show awareness and understanding of related work, and is it clear how the work done in the project compares to that work?

1 The report shows very little awareness and understanding of related work, or it contains a flawed comparison with such work.

3 The report shows some awareness and understanding of related work. Related work is adequately referenced.

5 The report features a precise and enlightening comparison with related work, including peer-reviewed research articles.

Substance Based on the report, does this project have enough substance (in terms of amount of work), or would there have been room for more ideas, results, or analysis?

1 Seems thin. I (the examiner) would have expected significantly more ideas, results, or analysis for a project with this timeframe.

3 Represents an appropriate amount of work for a project in this course.

5 Contains more ideas, results, and analysis than most projects in this course.

Language and formal aspects How well does the report conform to conventions of academic writing? How well has it been proofread?

1 The report contains many errors in grammar and punctuation. The language or notation used is not up to academic standards.

3 The report shows awareness of the formal conventions of academic writing. It makes correct use of mathematical notation.

5 The report is well-polished. In regard to language and formal aspects, it would have no problem to get accepted in an academic journal.