

Introduction to Natural Language Processing

Lecturer: Prof. Dr. Lucie Flek

Course Coordinator: Vahid Sadiri Javadi

Exercise Instructor: David Kaczér

Project Report (PR) – Guideline

The final submission of your project should consist of **your code (Problem Solving)** and **written final report in PDF format (Project Report)**. You should submit all the files in a ZIP file with the following name:

Team_< Team Number>.zip

The final report should provide a comprehensive overview of your project, its objectives, methodology, results, and conclusions. It should be structured in a clear and coherent manner to help readers understand the project's scope and outcomes. Below are the key components that should be included in the final report of your project:

1. Introduction:

- Briefly introduce the project and its purpose.
- State the project's objectives and the problem being addressed.
- Provide a clear research question or hypothesis.

2. Related Work:

- Discuss relevant background research and existing literature related to the project.
- Summarize key methods and techniques used in similar projects.

3. Data Exploration and Preprocessing:

- Describe the data sources and explain how the data were obtained.
- Provide a brief introduction to the dataset, including its size, and any relevant background information.

- Calculate and present basic statistics about the text data by performing an in-depth analysis of the data such as the average length of documents, the number of words per document, or the most common words to identify patterns, trends, and potential issues.
- Visualize the data including the distribution of labels or categories and any other relevant metrics that characterize the dataset using plots such as histograms, bar charts, or pie charts.
- Detail the data preprocessing steps, such as tokenization, stemming, lemmatization, cleaning, etc. (if applicable)

4. Methodology:

- Explain the techniques, algorithms, and models used in the project.
- Describe the architecture of any custom models or systems developed for the project.
- If you used pre-trained models, specify the architectures and models used.
- Discuss the choice of libraries, frameworks, or APIs used.

5. Experimental Setup:

- Explain the training process for your model, including hyperparameters and optimization techniques.
- If applicable, describe any fine-tuning or customization done on pre-trained models.
- Explain the train-test split or cross-validation approach used to evaluate the model.

6. Evaluation and Results:

- Outline the evaluation metrics used to assess the model's performance.
- Present the quantitative and qualitative results of the model's performance.
- Compare your results with existing benchmarks or other models if available.
- Include tables, charts, or visualizations to help readers understand the findings.

7. Analysis and Discussion:

- Interpret the results and findings from the previous section.
- Analyze the strengths and weaknesses of your approach and the project overall.
- Discuss any unexpected or interesting observations.
- Address any challenges or limitations encountered during the project.
- Discuss how the outcomes of the project can be utilized in practical scenarios.

8. Conclusion:

- Summarize the key findings and achievements of your project.
- Revisit the research question or hypothesis and state whether it was achieved.

9. Future Work:

- Suggest potential areas of improvement or extension for the project.
- Identify any unresolved issues or limitations that could be addressed in future work.

10. References:

- Cite all the sources, papers, datasets, and tools used throughout the project.