Title of Report

**Student names and numbers**

## 1. Introduction

* **Problem Statement:** Briefly describe the societal or research issue you aim to address with this project. Why is this important to study?
* **Research Question:** Clearly state the research question you have chosen. Ensure that the question is relevant for later analysis using text classification models.

2. Data Collection Approach

* **Method Used:** Explain the data collection method selected (e.g., plug-ins, data donation, scraping, APIs). Describe the rationale for choosing this method based on your research question and the intended analysis. Mention any ethical considerations.

3. Data Description

* **Dataset Overview:** Provide a description of the dataset you acquired, including the number of data points, the type of data (e.g., text, metadata), and any key characteristics. You may use figures or tables to illustrate this if necessary.

4. Reflection on Errors

* **Measurement Errors:** Explore any issues related to how the data was measured (e.g., ambiguous text, incorrect data labels). How might these impact your ability to answer the research question?
* **Representation Errors:** Discuss representation errors, such as biases in the dataset that might make it less representative of the broader population or phenomenon you are studying.

5. Conclusion

* Provide a final reflection on the overall quality of the dataset in relation to your research question. Briefly mention how you will use the data in the second assignment (e.g., applying text classification models) and how the errors identified will inform your future work.

References

* Include any sources referenced in your report (e.g., literature, tools, methods).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Grade:** | **Less than 5.5****Unsatisfactory:** | **Between 6 and 7.5****Satisfactory to good** | **Between 8 and 10****Good to Excellent** | **Weight** | **Grade** |
| Part A: Process |
| **Amount of work done (assessed by the list of tasks done by the student)** | The amount of work done was less than would have been expected on the basis of the amount of ECTS. | The amount of work done was what would have been expected on the basis of the amount of ECTS. | The amount of work done was more than would have been expected on the basis of the amount of ECTS. | 0.2 |  |
| **Active participation in the feedback moments** | The student did not participate in the feedback moments. | The student participated actively. | The student participated actively, allow others to express themselves, and showed a deep understanding of the project. | 0.1 |  |
| Part B: Quality of the report or presentation |
| **Problem definition and research question** | The purpose or the research question is not clearly articulated. | Adequately identifies and describes (or sketches out) the research question and its importance. | Effectively formulates a clear description of the research question and clearly argues its relevance. | 0.10 |  |
| **Data collection and description** | No or little data was collected (e.g. only used the data from the labs). The type or size of the data collected is not clear. | The data collected for the project expanded significatively the data collected in the labs. The reader can easily understand the type and size of the data collected. | The data collected for the project expanded significatively the data collected in the labs, and is highly relevant to answer the research question. The type and size of the data collected is detailed clearly, using figures or tables when necessary. | 0.20 |  |
| **Ethical form** | The ethical form was not submitted or submitted without putting much effort. | The ethical form was submitted and generally correct. | The ethical form was submitted and most fields were well argued. | 0.05 |  |
| **Reflection on errors and conclusion** | The discussion of errors is often inaccurate or incomplete; presents little if any analysis or interpretation; conclusions or recommendations are often not well supported, inaccurate, and/or inconsistent, and are presented in a vague or rudimentary manner; reflection/discussion is missing or lacks depth. | The discussion of errors is generally correct, outlines conclusions and recommendations that are logical and consistent with the analysis and evidence. | The discussion of errors is generally correct, external datasets are used to test at least one of the errors (e.g. aggregated statistics are used to test representation bias). Conclusions and recommendations are insightful, coherent, well supported, logically consistent. | 0.30 |  |
| **Report style** | Report is poorly organized, few references are present. | Report is well organized and formatted, the use of English is correct, clear and fair citations are presented in most case. | Report is well organized and formatted, extra media is presented (e.g. figures or illustrations) the use of English is correct, clear and fair citations are presented in most cases. | 0.05 |  |
| **General comments** |  |
| **Overall grade** |  |