## 1. Introduction

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

2. Data Collection and Description

* **Method Used and Dataset Overview:** 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. 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.

3. Data Labeling

* **AI model:** Explain the AI model used to label the data, including the original sample used to train the model and how their definition of the label aligns with your research question.

4. Discussion on Ethical and Privacy challenges

* **Ethical principles:** Discuss the alignment of your research project with ethical principles
* **Privacy regulations:** Explain the steps taken to ensure GDPR and other regulations are fullfill.

5. Discussion on Errors and Biases

* **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.
* **Fairness:** Discuss what errors may be present in the data that may create differences between subpopulations.
* **Correcting for errors:** Explain what steps could be done to correct for errors and biases

6. Conclusion

* Provide a final reflection on the answer to the research question and how different errors may affect the results.

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

**Rubric**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **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 and the participation in the feedback moments)** | The amount of work done was less than would have been expected on the basis of the amount of ECTS. The student did not participate in the feedback moments. | The amount of work done was what would have been expected on the basis of the amount of ECTS. The student participated actively. | The amount of work done was more than would have been expected on the basis of the amount of ECTS. The student participated actively, allow others to express themselves, and showed a deep understanding of the project. | 0.3 |  |
| Part B: Quality of the report or presentation | | | | | |
| **Problem definition and data collection**  (assessed from the presentation) | The purpose or the research question is still not clearly articulated. | Effectively formulates a clear description of the research question and clearly argues its relevance. New data is not collected. | Effectively formulates a clear description of the research question and clearly argues its relevance. New data is collected if it was mentioned in the feedback. | 0.10 |  |
| **Data labeling**  (assessed from the presentation) | The data was not labeled, or labeled in a way that did not match the RQ. | The data is labeled correctly, matching the RQ. The students show understanding of biases created in the data labeling. | The data is labeled correctly and effectively contributes to answering the RQ. The students show understanding of biases created in the data labeling. The students produce at least one output showing the results of the labeling (e.g. a graph) and connecting it to the research question. | 0.15 |  |
| **Discussion of ethical and privacy challenges**  (assessed from the presentation) | No discussion of the alignment of the project with ethical and privacy principles, or the discussion is clearly incorrect. | The discussion of the alignment of the project with ethical and legal principles is mostly correct. No or few steps have been done to ensure ethical and privacy principles are respected. | The discussion of the alignment of the project with ethical principles is correct and steps have been done to ensure ethical and privacy principles are respected. At least three principles are discussed. | 0.15 |  |
| **Discussion of errors**  (assessed from the presentation) | 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, potential consequences for the analysis and recommendations on how to expand the study that are logical and consistent with the analysis and evidence. | The discussion of errors is generally correct. Conclusions, potential consequences of errors for the analysis and recommendations on how to expand the study are insightful, coherent, well supported, logically consistent. | 0.20 |  |
| **Q&A** | Questions are not answered correctly | Questions are answered mostly correctly | The answers to the questions show a deep understanding of the weaknesses and strengths of their methods. | 0.10 |  |
| **General comments** |  | | | | |
| **Overall grade** |  | | | | |