Final Project Instructions: Introducing to NLP

Yair Lakretz, Dauphine University 2024.

Overview:
This final project is designed to give you practical experience with natural language processing (NLP) and machine learning techniques. You will work in groups of three, with each member taking responsibility for a crucial component of the project. This will involve preparing data, creating and optimizing a machine learning model, and establishing a baseline for comparison.

Groups:
Each group must consist of three students (exceptions must be confirmed with me directly). You are encouraged to collaborate closely, though each member will have a specific role:

1. Data Preparation: Responsible for collecting, cleaning, and preparing the data set that will be used for training and evaluating the model.
2. Modeling: In charge of designing, training, and evaluating the NLP model. This includes selecting algorithms, implementing the model, and optimizing its performance.
3. Baseline Comparison: Tasked with setting up a baseline model to compare against the primary model, providing an essential benchmark to gauge improvements.

Project Phases:

1. Proposal Submission:
   Due Date: **19/04/2024, Midnight**
   Submit a brief document (half a page) outlining your chosen problem, data sources, anticipated challenges, and initial roles of each group member.

2. Final Submission:
   - Due Date: **16/05/2024, Midnight**

     Your final submission should include:

     - Code: All scripts used in the project, well-documented and organized, preferably in an online repository (e.g., **GitHub**).
     - Report: A detailed report documenting your methodologies, analyses, and findings. The report should include:
       - Introduction to the problem and justification of its importance.
       - Description of the data and its preparation.
       - Detailed explanation of the model architecture, training process, and evaluation metrics.
       - Explanation and results of the baseline model.
       - Comparative analysis of your model and the baseline.

    - Conclusion discussing the implications of your findings, limitations, and possible future work.
- Presentation: Prepare a 15-minute presentation summarizing your work and findings, to be delivered during the final project session.

Deliverables:

1. Code: All source code used in the project, provided via a shared repository (e.g., GitHub).
2. Report: A comprehensive written report (PDF format), submitted to yair.lakretz@gmail.com. **Max length 5 pages (not including references), Arial font size 11, reasonable page margin.**
3. Presentation: Slides and any other visual aids used during your final presentation.

Evaluation Criteria:

- Innovation and complexity of the solution
- Effectiveness and efficiency of the data preparation
- Accuracy and robustness of the model
- Relevance and clarity of the baseline comparison
- Quality and clarity of the written report and presentation

Additional Guidelines:

- Ensure that all data used complies with privacy and usage regulations.
- Regularly communicate with your team members and manage your project time effectively to meet the deadlines.
- Seek feedback from peers and to refine your approach and/or contact me at the above email address.

This project is your opportunity to demonstrate your ability to apply advanced NLP and machine learning concepts in a real-world scenario.

Good luck!