UNIVERSITY OF SOUTHERN MAINE

Text Mining and Analytics, Spring 2024, Course Project-Part 2

As described in the course project description, in the second phase, students will run one of the existing systems (baseline) as it is on their task. After running the baseline system, you should analyze the results carefully, and provide a detailed analysis of samples for which the system worked well and also failed cases. You should carefully study why the baseline system failed for the cases, and then create a hypothesis for improving the system. In the last part of the course project, you will implement the improvement and test your hypothesis.

For the presentation, each student will have 3 minutes to present work. The presentation should have one slide describing the baseline system. The second slide shows the evaluation results (using evaluation metric) with one positive and one negative example next to each other for comparison. The third slide should explain the proposed approach for the third part of the project. No presentation should go over 3 minutes; failure to do so will result in a penalty and your talk will be stopped.

For the report, each student should have a minimum of two pages, discussing the related work, the reasoning for choosing the baseline system, a description of the baseline system, the evaluation metric used for the experiment, a table showing the results, and an analysis of the results. All the papers should be appropriately cited; if you need to become more familiar with Overleaf, spend some time and learn how to correctly cite papers (this link might help).

For submission, all the slides for individual students should be put together and submitted as one PDF file. You will present your talk from the submitted file. For the report, you should find an appropriate order, combine all the reports, and submit both .zip (source) and .pdf files from Overleaf. The report should be in the Springer template shared before (Springer's LNCS template). You will also provide a link to GitHub repo(s) with your version of code and a comprehensive ReadME file that helps users easily get the data and run your code and regenerate the results. The code should include both running the model and getting the evaluation results.

Presentation Schedule:

04/03 Team 1 – PAN	Evan, Corey, Alex, Jere
04/03 Team 2 – eRisk	Thad, Aiden, Evan
04/03 Team 3 – eRisk	Sarah, Reihaneh
04/03 Team 4 – eRisk	Wyatt, Sarah
04/03 Team 5 – Chat Abuse Identification	Michael
04/03 Team 6 – FormalBERT	Jackson
04/03 Team 7 – SimpleText	Shea, Deiby, Joseph, Jakub
04/03 Team 8 – SimpleText	Nick, Finn, Ben, Gabrielle

Notes:

- The order and the date of talks are fixed and cannot be changed
- All the submissions should be done by the team manager, and other submissions will be ignored
- Submission deadline is April 3rd, before noon. The link for submission will be closed after the deadline
- Submissions not following the requested format will be desk-rejected and receive a 0; three files are expected: .pdf for presentation, .pdf for report, .zip for overleaf source file. Other team members can and should ask the manager for receipt of the submission. The links to the Git repo(s) should be provided in the .pdf file
- Grading is done individually, based on the presentation and report. The rubric for grading is as follows:
    - Presentation: 4%
    - Choice of baseline: 3%
    - Evaluation of baseline: 2%
    - Git Repo and codes: 3%
    - Detailed analysis of the results: 8%
- After the presentation, students might be asked to visit the instructor during office hours to discuss further on their project
- Students should carefully follow the presentations, and find students with similar tasks. This can help you collaborate with other students for your project
- For any questions, please use the project channel on the course Slack group