

Project Proposal Feedback Form

For cells in this form, state whether the project proposal met expectations or whether additional work is needed. If additional work is needed, or there are other concerns, leave comments with your explanation.

Problem Definition
<p>Clarity. Is the problem definition clear?</p> <p>Meets expectations. Problem definition is clear.</p>
<p>Creativity & Novelty. Is the project nontrivial?</p> <p>Additional work is needed. There are not enough novel components in the project (only embeddings are used as features). The task is the same task as in the dataset paper. Early on, you mention other features (LIWC and LDA), how are you planning to use these in your project?</p>
Datasets and Features
<p>For this section, start comments with: “novel” / “satisfactory” / “needs work”</p>
<p>Dataset. Is the dataset clearly specified? Is it clear how the dataset is going to be used? Evaluated? If the dataset is not listed in instructions, is there a link to it?</p> <p>Satisfactory. Dataset is clearly stated and readily available.</p>
<p>Features. Is it clear what features are planned to be used in the project? Is it clear if feature selection will be part of the project, and if so, is it explained clearly?</p> <p>Needs Work. It is not completely clear what features you are planning to use and if you are planning on doing any feature selection as part of the project. <u>Please elaborate.</u></p>
Methods
<p>Source. Does the project build off existing code, or use pre-existing libraries?</p> <p>Needs Work. It is not clear whether pre-existing code or libraries are going to be used. Models similar to the ones used in the dataset paper are planned to be used, but it is not clear if code released by the dataset paper authors will be used. <u>Please clarify.</u></p>
<p>Planned Methods. Is it clear what methods they are planning to use and implement?</p>

Satisfactory. The methods are clearly stated.

Contribution. Is there a clear distinction between the planned contributions of the students and what is already implemented before?

Needs Work. As it is right now, the only contribution would be the new feature (embeddings). I would like to see a contribution in feature selection, and possibly using another model that was not used in the dataset paper to increase performance. Please clarify your plans.

Baselines. Are there planned baselines for experiments? Do the baselines seem appropriate for the task/problem?

Needs Work. No baselines were specified. Please consider an appropriate baseline (maybe models implemented by dataset paper).

Evaluation. Are there planned evaluations for experiments? Are evaluations appropriate for the task/problem?

Satisfactory. Evaluation plan was clearly stated.

For the cells below, please comment on the following:

Deliverables

For this section, start the comment with: "too much" / "just right" / "not enough"

Are deliverables scoped correctly?

Not enough. As of right now, your contributions fall a little too short.

1. Feature selection is not clear
2. Only adding one more feature is not enough, please consider adding other features (LDA or LIWC).
3. Models planned are the same as the dataset paper, please consider using an additional model to improve upon.

Ethical Concerns

Are there any ethical concerns with the dataset or task?

Studying hate speech is important for society, but there are ethical considerations to take into account. See [4].

General Comments

Any related work, etc.

My final recommendation is to resubmit your proposal with the changes I've suggested. Please let me know if you have any questions. The plan is to give you another round of feedback on your proposal and give you full points once we've converged with a project we all agree upon.

There is a lot of related work and publicly available datasets in this area. Some are listed below if you need more ideas to add to your deliverables.

- [1] Z. Zhang, D. Robinson, and J. Tepper, "Detecting Hate Speech on Twitter Using a Convolution-GRU Based Deep Neural Network," in *The Semantic Web*, Cham, 2018, pp. 745–760, doi: 10.1007/978-3-319-93417-4_48.
- [2] A.-M. Founta *et al.*, "Large Scale Crowdsourcing and Characterization of Twitter Abusive Behavior," *arXiv:1802.00393 [cs]*, Apr. 2018, Accessed: Nov. 11, 2020. [Online]. Available: <http://arxiv.org/abs/1802.00393>.
- [3] M. Zampieri, S. Malmasi, P. Nakov, S. Rosenthal, N. Farra, and R. Kumar, "Predicting the Type and Target of Offensive Posts in Social Media," *arXiv:1902.09666 [cs]*, Apr. 2019, Accessed: Nov. 11, 2020. [Online]. Available: <http://arxiv.org/abs/1902.09666>.
- [4] M. Sap, D. Card, S. Gabriel, Y. Choi, and N. A. Smith, "The Risk of Racial Bias in Hate Speech Detection," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, Jul. 2019, pp. 1668–1678, doi: 10.18653/v1/P19-1163.

Final Recommendation: Resubmit Full Proposal

Either: "fine as is" / "answer questions" / "resubmit full proposal"