Instructions: Provide the link to the shared task that you would like to participate in, and give a 1-paragraph description of the approach you hope to apply.

You are recommended to select a task from SemEval 2019, but you may select another shared task as long as it is a natural language processing task, and it has an official online leaderboard that you are able to submit to. The shared task timeline does not need to match the semester time frame, as long as the task still allows you to submit unofficial results after the official evaluation is over. (This is the case for most SemEval 2019 tasks.)

OffensEval (Task 6)
CodaLab (Submission): https://competitions.codalab.org/competitions/20011
Leaderboard (Published in ArXiV Preprint): https://arxiv.org/pdf/1903.08983.pdf

Logistics:
Emmi kindly e-mailed the task organizers on behalf of several students, and the task organizers have opened up the submission page. Rahul successfully submitted a test run for **Sub-Task A** to verify that the CodaLab site was generating an output log with (1) accuracy, (2) F1-micro, (3) F1-macro, and (4) F1-weighted scores.

Project Description:
I will build a classifier for sub-task A in the OffensEval shared task in SemEval 2019, which will detect the presence of offensive language in tweets. I will use the pre-trained BERT-Base model to acquire word embeddings that will be used as inputs for a classifier. I will build the classifier by stacking bidirectional LSTMs/GRUs. If time allows, I will fine-tune BERT's pre-trained model on the HPC using the GPU nodes on Ocelote or on Google Cloud TPUs via Google Colab for the provided training dataset. Additionally, I will investigate ensembling techniques to improve the overall performance of this classifier.