Weekly Mentor + Team Meeting: Every Saturday
5:00-6:00 PM Our github repo: [acmucsd-projects/wi23-ai-team-1 (github.com)](#)

Team Meeting Time: 2/25/23 at 5pm in CSE B230
Attendees: Aniket Gupta Jimmy Ying Steven Shi Vincent Tu Vivian Liu Arnav Modi

What have we done so far

- We added another LSTM layer
- We added random seeds so our results are consistent
- Reorganized the github

What is the point of this meeting, what are we going to discuss
- We added regularization layer
  - Dense layer
- Figuring out why our data is overfitting

What will we do going forward
- Attempt to properly implement regularization - Vivian
- Expand model to account for all 6 toxicity labels
- Work on improving our model
- Shuffle the dataset before splitting
  - Stratified shuffle
  - Sklearn stratified shuffle split - Vivian
- Fix the random seed - aniket
- Figure out why training takes so long
  - Maximize the batch size (set to like 64) - steven
  - Cut down the embedding layer: len(encoder.get_vocabulary()) - jimmy
  - GOOGLE HOW TO CUT DOWN TRAINING - jimmy
  - Graph the frequency of words - jimmy
    - Figure out how to narrow down vocabulary
- Use correct metrics (parameter in model.compile function) or fix the class imbalance problem - arnav
- ~~Figure out the workflow (how we gonna collaborate) - done~~
  - A folder for each person
  - Naming conventions
- Double check that the pre-processed format of the dataset is correct when going into the model
- Figure out how to split the dataset

- - Find a good & effective split so that we have a good training set and a good validation dataset

●Someone make a submission to the competition - steven

- Don't use the test dataset, use the validation instead
- Fix these weird problems before we start experimenting
- Standardize our notebooks (implement a format)
- Notebook convention:
  - Setup
  - Cleaning (already done)
  - Preprocessing
  - Model
  - Training