

Weekly Team Meeting: Every Friday

5:00-6:00 PM Our github repo: [acmucsd-projects/wi23-ai-team-1 \(github.com\)](https://github.com/acmucsd-projects/wi23-ai-team-1)

Team Meeting Time: 2/4/23 at 5pm in CSE B230

Attendees: Aniket Gupta Arnav Modi Jeffrey Lee Jimmy Ying Steven Shi Vincent Tu
Vivian Liu

What have we done so far

- We have brainstormed some ideas and put them on the previous document
- We've gone over how meetings will work
- We went over what kaggle is

What is the point of this meeting, what are we going to discuss

Ideas

- <https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand>
 - Use this dataset to allow user to predict when to book and what hotels to book on vacation trip
- Classifying heart disease:
<https://www.kaggle.com/datasets/abhishek14398/heart-disease-classification>
- <https://www.kaggle.com/datasets?search=leaf+dataset>
 - A lot of datasets about specific categories of leaf to see if they are diseased, we could sort of try to build a front end around it as well to make it actually usable. Probably a dropdown to select crops/plants. We could sort of try to combine multiple datasets each dealing with a specific variety of crops.
- <https://www.kaggle.com/datasets/reihanenamdari/mental-health-corpus>
 - I think we were talking about some sort of sentiment analysis. This dataset is about the same in the context of mental health and depression – some form of NLP I think.
- [YouTube Videos and Channels Metadata | Kaggle](#)
- https://cseweb.ucsd.edu/~jmcauley/datasets.html#amazon_reviews
- <https://www.kaggle.com/datasets/tunguz/200000-jeopardy-questions>
 - Use this dataset, have the user input a question and generate a dollar value
- <https://www.kaggle.com/datasets/zepopo/ukrainian-fake-and-true-news>

NLP tasks:

<https://medium.com/nlplanet/two-minutes-nlp-33-important-nlp-tasks-explained-31e2caad2b1b>

The idea we're going with:

<https://www.kaggle.com/competitions/jigsaw-toxic-comment-classification-challenge/data>

What will we do going forward

Tech tools

- **Pandas**
- **Numpy**
- Matplotlib - visualize data (most common words, etc.)
- Tensorflow
- Pytorch (deep learning)
- Code tab in kaggle competitions

Split data into

- Training (80)
- Validation (10)
- Testing (10)