In your proposal, please answer the following questions:

1. What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.
2. What is your free topic? Please give a detailed description. What is the task? Why is it important or interesting? What is your planned approach? What tools, systems or datasets are involved? What is the expected outcome? How are you going to evaluate your work?
3. Which programming language do you plan to use?
4. Please justify that the workload of your topic is at least 20\*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.

1.

Name: Mingchong Zhai; NetID: mzhai4; Captain

Only one student is in this team.

2.

The topic is about sentiment analysis on tweets from Twitter.

The task is to pass tweets to the model and predict the sentiment in the tweets.

It is interesting because we can identify whether people are being positive or negative in their posts.

The approach is to train models on a certain dataset containing tweets.

Tools used will include jupyter notebook, pickle (to save trained models as a binary file), NTLK (for data pre-processing and cleaning), Tweepy (the twitter API) and some other python library.

The expected outcome will be consisting of tweets and the predicted sentiment with a probability.

We can manually check some tweets as well as the predicted sentiment and use different parameter in training to improve the accuracy.

3.

Python.

4.

Find the initial dataset used for training. (2-3h)

Do data pre-processing and cleaning. (4-5h)

Fetch new data from the twitter API. (4-6h)

Build model(s) for analyzing the data. (6-10h)

Analyze and print the data with some visualization. (2-3h)