

## **CS 410 Final Project Proposal: Free Topic**

*Jiahao Zhang: jiahao4, Mingqing Sun: ms19, Ruining Tao(Captain) : rtao6, Ziyue Zhou: ziyue5*

**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?**

For our final project, we would like to perform sentiment analysis on the text data from Twitter. We will build a Python application to analyze tweets. In particular, we want to investigate whether a specific tweet is positive, neutral, or negative. We believe that analyzing sentiments can detect possible hate tweets, create a harmonious network environment and protect our community. We also believe that knowing the sentiment of a tweet can be useful for companies who want to do targeted advertisements. To achieve our goal, we will use datasets from Kaggle and utilize Python libraries(e.g. PyTorch, TextBlob) to analyze them. We will train our model using the pre-labelled dataset. Then, we will evaluate our model accuracy using the unlabelled dataset. We want to compare our results with other models. Finally, our expected outcome is to improve our model so that our result has high accuracy and is close to the result of other models.

**Which programming language do you plan to use?**

We would like to use Python for our project.

**Workload Justification(Total :  $4 * 20 = 80h$ ):**

- Collect original dataset (10 hours)
- Tag original data (10 hours)
- Develop sentiment analysis model in Python (40 hours)
- Experiment with other data and improve model performance (20 hours)