CS410 Course Project Proposal

Team name: Noobs

Team members: Jiarui Zhang (jiaruiz2), **Xueyang Li (xl112)(C)**, Ziyang Zhan(ziyangz6)

Topic: (Free topic) Sentiment analysis for Chatbot. We will implement sentiment analysis and train a simple Chatbot based on the result. This Chatbot targets daily conversation.

Example scenario: “I’m so upset” -- “I’m sorry to hear that, do you want some food?”.

Task: The purpose of this project is to integrate what we’ve learned in this course and develop a simple Chatbot featuring the function of sentiment analysis, which has the ability to identify the user’s sentiment based on user inputs and give a proper response that simulates real life chatting.

Important/Interesting thing: Given the current COVID situation and the nature of Computer Science, it’s not surprising that we sometimes feel lonelier and maybe more depressed than ever. We need friends we can talk to, someone we can relate to, someone that knows our emotions, interests and feelings. That someone isn’t always available. Therefore, we hope to use the knowledge and techniques we’ve learned in this course, such as word identification and sentiment analysis, to build a chatbot that’s able to provide a similar experience.

Planned approach: We will first implement sentiment analysis based on daily chat conversations records from social media, such as twitter or facebook. After this, we will train multiple models by which we can predict the sentiment of a sentence. Then, we will use the model with the best performance to implement a simple chatbot which can detect the sentiment of a sentence and reply based on sentiment.

Involved tools: We will use Python Notebook/Pycharm as our coding platform. The main libraries we will use are sklearn, NLTK, Metapy, operator, and numpy. The dataset we will use is “Sentiment140 dataset with 1.6 million tweets” from Kaggle.

Expected outcome: A Chatbot with the ability to catch users’ emotion and communicate with users (daily chatting).

Evaluation principle: The rate of Chatbot correctly understanding users’ emotion and giving proper corresponding responses.

Programming language: Python

Main tasks and time commitment:

Environment setup	3
Data cleaning and processing	6
Multiple Model training	15
Model selection(based on F1, recall, precision....)	6

Classification and prediction with selected model	10
Classification and prediction test	6
Chatbot - front-end UI	18
Chatbot - backend interaction with trained model	20
Chatbot test	6
Report and documentation	6
Total	96*

*Some advanced features may be removed if we do not have enough time, to better meet the 20*N hours, as we only have 3 team members.