

CS 410 Project Proposal

- 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.

Name: Melissa Cho NetID: mccho3

Name: Iris Kan NetID: iriskan2

Captain: Melissa Cho

- 2. What topic have you chosen? Why is it a problem? How does it relate to the theme and to the class?**

We have chosen the track of Free Topics and the topic of using sentiment analysis on Google reviews of apartments in Urbana-Champaign, because this will make it easier for students to determine the sentiment of the reviews by other students for each apartment complex, which will aid them in their search for an apartment. This relates to the class in general, because we will be implementing a program that uses sentiment analysis to analyze Google reviews of apartments in Urbana-Champaign.

- 3. Briefly describe any datasets, algorithms or techniques you plan to use.**

We plan to use sentiment analysis in our program, as well as web scraping to collect the data of the reviews for the analysis.

- 4. How will you demonstrate that your approach will work as expected?**

It will be clear that our approach is working as expected if the sentiment analysis returns a sentiment that matches the overall number of stars of the reviews.

- 5. Which programming language do you plan to use?**

We plan to use Python.

- 6. Please justify that the workload of your topic is at least $20 \cdot N$ 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.**

N : 2 People

Week 10: (8 hrs/person)

- Web scrape google reviews of apartments in Urbana-Champaign
- Building dataframe of google reviews of apartments in Urbana-Champaign
- Classify/give scores to word with positive/negative connotations
- Assigning a sentiment score to each google review

Week 11: (8 hrs/person)

- Build sentiment analysis model
 - data cleaning

- split data frame
- create bag of words
- import logistic regression
- split target and independent variables
- fit model on data
- make predictions

Week 12: (2 hrs/person)

- test model's accuracy
- classification report

Week 13: (2 hrs/person)

- Prepare for presentation