

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.
 - a. Captain: Havish Rani: hrani2
 - b. Acchindra Thev: Athev2
 - c. Nishit Samarth samarth9
 - d. Harsha Gajja: hgajja2
 - e. Vinayak Manepalli: vmanep2
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?

Our free topic is restaurants and we are going to be looking at restaurants in a college town and a large city to see if the sentiment analysis is different. The task involves using Python to scrape Reddit subreddits related to restaurants in the Champaign and Chicago areas. The goal is to collect and compare sentiments expressed in posts and comments about restaurants in these two regions. This project seeks to understand the public's sentiments, both positive and negative, towards restaurants in Champaign and Chicago and analyze any regional variations. This project is significant as it can provide valuable insights for restaurant owners, tourists, and residents in Champaign and Chicago. It can help identify popular restaurants, customer experiences, and areas for improvement. It's an interesting project as it combines web scraping, natural language processing, and data analysis to draw meaningful conclusions about two different geographic regions.

The planned approach includes the following tasks, which would have to be part of the project in one way or another, but there are a multitude of ways that they could be done. We found some ways online and will look into which suits our purpose the best as we go through the project. Some tasks that we need to do include using the Reddit API and Python (PRAW) to scrape posts and comments from Champaign and Chicago restaurant-related subreddits. After that, we would have to utilize natural language processing libraries like NLTK, TextBlob, or VADER to perform sentiment analysis on the collected text data to determine sentiment scores. The next stage would include, storing the sentiment data, along with relevant post details, in a structured format. Then we move on to analyze and compare sentiment scores between Champaign and Chicago to identify trends and differences. The final part is to create visualizations to present the sentiment comparisons.

Tools, Systems, or Datasets:

- Python with libraries like PRAW for Reddit data, NLTK, TextBlob, or VADER for sentiment analysis. Possibly also PyTorch, and APIs such as OpenAI and LangChain.
- Reddit API for data scraping.
- Datasets will be generated from Reddit posts and comments.

Expected Outcome:

The expected outcome is a comprehensive analysis that highlights the sentiments expressed in Champaign and Chicago restaurant-related subreddits. The project should provide insights into the public's perception of restaurants in these areas and potentially show regional differences.

Project Evaluation Metrics:

- Comparing sentiment scores between the two regions.
 - Generating visualizations that make the differences and trends clear.
 - Assessing the project's impact on decision-making (e.g., restaurant owners' responses to feedback).
 - Ensuring that the project adheres to ethical guidelines and respects privacy.
1. Which programming language do you plan to use?
 - a. Python
 2. 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.
 - a. Project Planning: 5 hours
 - b. Learn how to Scrape: 10 hours
 - c. Data Collection and Scraping: 20 hours
 - d. Data Preprocessing: 15 hours
 - e. Sentiment Analysis: 20 hours
 - f. Data Storage and Management: 10 hours
 - g. Data Analysis and Comparison: 10 hours
 - h. Data Visualization: 10 hours