Lab5 Meeting Notes

Day 1 (2/11/2025)
- Discussed key points of the project.
- Split and divide project tasks.
- Searched for the properties of BeautifulSoup and Praw API. And decided to use Praw API to fetch Reddit data.

Day 2 (2/12/2025)
- Shared progress of each member's tasks.
- Scraped Reddit posts and comments. Observed the characteristics of the acquired data and discussed how to clean the data.
- Finished the data collection and storage part.

Day 3 (2/13/2025)
- Shared progress of each member's tasks.
- Finished the data preprocessing part.
- Discussed how to use the doc2vec and issue of embedding.

Day 4 (2/14/2025)
- Shared progress of each member's tasks.
- Finished the clustering part.

Day 5 (2/15/2025)
- Shared progress of each member's tasks.
- Combine the 3 scripts and run them together, then debug them. Discussed and wrote the main.py to perform the automation process.
- Completed the writing of various documents and the uploading of codes.