

DSCI 560 - Data Science Professional Practicum
Lab 4 Part 1

Team name: Cleanup Crew

Team members: Jonathan Kusuma (ID:4897950942), James Susanto (ID:3466104463)

Minutes:

February 13: Investigated Reddit API rate limits for bulk data extraction and optimized SQL queries for efficiency. Assigned tasks for enhancing code modularity and reducing technical debt.

February 14: Debugged and resolved issues in the post-fetching loop; identified and commenced enhancements on vectorization integration errors.

February 15: Finalized the implementation of a KMeans clustering algorithm on the vectorized dataset; strategized the setup for the automation script to periodically trigger data processing.

February 16: Fixed threading issues to ensure smooth automation; fine-tuned the Doc2Vec model's hyperparameters for better content representation.

February 17: Improved the automation script's error handling mechanisms; established context management for robust database connections; conducted a comprehensive review of the system in preparation for the demonstration and submission.