# **Meeting Minutes - Lab 5**

## Team: Guardians of the Algorithm

### Team Members: Pratham Solanki, Jaival Upadhyay, Mayank Patil

### **Wednesday, February 12th, 2025**

**Discussion Topic:** Approach and Methodology

* Reviewed the lab requirements and objectives.
* Decided to use **PRAW API** for data scraping and **MySQL** for storage.
* Discussed data preprocessing steps, including text cleaning and OCR extraction.

### **Thursday, February 13th, 2025**

**Discussion Topic:** Clustering Strategy

* Explored different clustering methods (**K-Means, Hierarchical, DBSCAN**).
* Finalized **K-Means** with **Doc2Vec embeddings**.
* Defined metrics for evaluating clustering performance.
* Discussed API rate limits and error handling.

### **Friday, February 14th, 2025**

**Discussion Topic:** Coding Implementation

* Implemented **web scraping and database storage**.
* Developed **data preprocessing** pipeline (stopword removal, text cleaning, OCR text extraction).
* Ensured data is formatted correctly for clustering.
* Debugged initial errors in API calls and MySQL transactions.

### **Saturday, February 15th, 2025**

**Discussion Topic:** Finalization and Report Writing

* Integrated all code components and ran tests on sample data.
* Generated clustering results and visualizations.
* Compiled findings and methodology into the final report.
* Reviewed and refined the report for submission.