Progress Project 3

<div align="right">

CSIS 4495-050 Applied Research Project

300372531 Hyunhee Kim

</div>

1. Work Logs

Date	Number of Hours	Description of work done
2025-10-18	2	Creating normalized columns, and adding extra necessary columns for consistent schema structure
2025-10-19	2	Planning the data modeling process and constructing a logical data modeling diagram for inclusion in the midterm report
2025-10-20	3	Developing normalized views for jobs, courses, and edges to improve model quality; creating a modeling workflow diagram.
2025-10-21	2	Beginning embedding training using TF-IDF and LSA to convert text into numerical form, reduce dimensionality, and identify meaningful patterns.
2025-10-22	4	Implement recommendation function codes and drafting midterm report,
2025-10-23	2	Finalization and editing the midterm report for submission

2. Description of Work Done

During this period, I focused on refining the data structure and advancing the modeling workflow for the project. I created normalized columns and additional necessary fields to maintain data consistency across multiple datasets. Following that, I planned the data modeling process and constructed a detailed logical diagram to visually explain the model's workflow for the midterm report. I developed normalized views for the job, course, and edge tables to improve data quality and enhance the model's performance. I then began embedding training using TF-IDF and Latent Semantic Analysis (LSA) to transform text-based features into numerical vectors, reduce dimensionality, and uncover meaningful relationships between job and course data.

Later in the week, I implemented the recommendation function, integrating the trained embeddings and similarity metrics to produce course recommendations aligned with job skills. Finally, I focused on writing and finalizing the midterm report, ensuring all modeling processes, diagrams, and implementation details were well-documented.

3. Repo Check-ins (Completed)

- Data preprocessing: Misc / Dataset preprocessing/ create_columns.ipynb: Script for generating normalized views for jobs, courses, and edges datasets. Normalization columns commit - Added consistent schema alignment and additional metadata columns(check-ins were pushed to the **main** branch.)
- Misc/Dataset preprocessing/create_columns.ipynb: create additional columns and create job view, course view, edge view to handle duplication (check-ins were pushed to the **kim** branch.)
- Misc/Data Modeling/Modeling_Diagram.drawio.png -Visual diagram illustrating the end-to-end modeling workflow and logical data relationships. (check-ins were pushed to the **kim** branch.)
- Misc/Data Modeling/Career_Recommendation_modeling1.ipynb- Initial model development notebook containing TF-IDF and LSA embedding setup. (check-ins were pushed to the **kim** branch.)
- Misc/ Data Modeling/Career_Recommendation_modeling2.ipynb- Extended version including refined similarity functions and evaluation logic. (check-ins were pushed to the **kim** branch.)