

Progress Project 2

CSIS 4495-050 Applied Research Project

300372531 Hyunhee Kim

1. Work Logs

| Date | Number of Hours | Description of work done |
|------------|-----------------|--|
| 2025-09-30 | 2 | Merged two “job & skills” datasets; created a Meeting Summary folder to share meeting notes. |
| 2025-10-02 | 2 | Started Coursera course & skills dataset preprocessing: removed unnecessary columns; handled null values. |
| 2025-10-06 | 2 | Researched course datasets on Kaggle and gathered references. |
| 2025-10-07 | 1.5 | Finished Coursera dataset preprocessing. |
| 2025-10-08 | 3.5 | Began merging skill_jobs datasets with Coursera dataset for modeling ; held a virtual meeting with teammate to discuss the data-modeling plan. |
| 2025-10-09 | 4 | Finalized the merged & preprocessed dataset (3h); drafted the data-modeling plan (1h). |

2. Description of Work Done

During this period, I completed preprocessing of the Coursera course/skills dataset (column pruning and null-value handling) and initiated—then completed—the merge with the job-skills data. I standardized schema and skill naming conventions, and produced a “final merged dataset” suitable for modeling. While incorporating the final Coursera and job-skills sources, the dataset became very large, which extended the time required for cleaning and alignment. To handle this, I ultimately selected a representative sample, applied targeted filters, and used a clear methodology to isolate only the skills essential for modeling. I also met virtually with a teammate to align on the modeling approach (feature set, target definition, validation strategy).

3. Repo Check-ins (Completed)

- Folder structure: Created a Meeting Summary folder and uploaded weekly notes.
- Data preprocessing artifacts: Committed Coursera preprocessing scripts and logs.
- Merged dataset artifacts: Uploaded draft and final merged datasets, plus a draft data dictionary.
- Planning doc: Added a modeling plan outlining objectives, metrics, validation, and a baseline approach.