**Project Timeline:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Duration** | **Expect End Date** | **Status** |
| Study the Background of CR | 3 months | 17/11 | Completed |
| ~~Study and Select the Dataset~~ | ~~2 months~~ | ~~17/11~~ | Adjusted |
| Research on System Setup | 1 month | 17/11 | Completed |
| ~~Data Pre-Processing~~ | ~~2 weeks~~ | ~~31/11~~ | Removed |
| 1st Stage BE (Retrieval) | 1 month | 31/12 | Completed |
| 1st Stage Testing & Evaluation Design | 2 weeks | 9/2 | / |
| 2nd Stage BE (Retrieval + Ranking) | *2 months* | *28/2* | / |
| 1st Stage Models Optimization | 2 months | 28/2 | / |
| 1st Stage FE | 1 month | 16/3 | / |
| 2nd Stage Testing & Evaluation Design | 2 weeks | 16/3 | / |
| 3rd Stage BE (Ranking) | 1 month | 31/3 | / |
| Final Report (1st half) | 2 weeks | 31/3 | / |
| 2nd Stage Model Optimization | 1 month | 13/4 | / |
| 2nd Stage FE | 1 month | 13/4 | / |
| Features Add-on (Optional) | 1 month | 20/4 | / |
| Final Report (2nd half) | 2 weeks | 20/4 | / |
| Video Production | 1 month | 30/4 (TBC) | / |

Check [Trello](https://trello.com/b/5JG6Hmrf/milestones-tasks) for task detail and distribution.

**Task Update:**

1. **Study the Background of CR (Completed)**

* Completed an initial review of retrieval models, including User/Item Collaborative Filtering, Deep Structured Semantic Model (DSSM), and Deep Retrieval according to a [tutorial](https://github.com/wangshusen/RecommenderSystem) of Red Recommendation System.
* Will continue exploring models addressing specific challenges and documenting pain points as part of further research.

1. **Study and Select the Dataset (Adjusted)**

* Currently using the [Yelp Dataset](https://www.yelp.com/dataset) for learning and testing.
* Open to exploring additional datasets that might align better with project goals.

1. **Research on System Setup (Completed)**

* Modelling: *Python* (using *TensorFlow* & *Scikit-learn*; may expand).
* Backend: *Python* (*Flask* for API and real-time processing).
* Frontend: *React* (default *Flask* pages used temporarily).
* Database: *SQLite* (Structure DB), *Faiss* (Vector DB), *npy/pkl* (Simple Key/Index).
* Hosting: Linux machine provided by the department (exploring *AWS*/*Azure* as alternatives).

1. **Data Pre-Processing (Removed)**

Pre-processing will depend on the selected models. This task has been removed to allow flexibility.

1. **1st Stage BE (Retrieval) (Adjusted)**

* Designed and trained the minimum performance models for Item Collaborative Filtering (Item CF) and Deep Structured Semantic Model (DSSM).
* Integrated the two models into the CR system for qualitative assessment using Flask.