

Progress Report

Link to proposal: <https://github.com/nykznykz/CourseProject/blob/main/Proposal.pdf>

Completed Tasks

We have completed the first task:

1. Scraping
 - a. Scraping reviews & ratings
 - b. Scraping recipes

Challenges

1. Dataset size
 - a. While we managed to perform scraping on the website smoothly, getting a large dataset of users ratings is challenging. We expect to overcome this eventually as we will continue scraping and growing the dataset along with the completion of subsequent tasks.
2. Sparsity
 - a. One concern is that the dataset might be sparse as each recipe is only reviewed by a small subset of users. We plan to address this by framing the recommendations as a classification problem and in doing so, we will be able to only train on observed ratings.

Pending Tasks

1. Construction of dataset
 - a. Basic dataset of user_id, recipe_id, rating
 - b. Construction of features
 - i. Exploratory Data Analysis
 - ii. Feature engineering
2. Setting up baseline approach
3. Model training
 - a. Creating general model architecture
 - b. Tuning of hyperparameters
 - c. Others (e.g. dropout, l2 normalization, embedding layers etc.)
4. Model Evaluation

- a. Evaluation of baseline & actual model
- 5. Post analysis
- 6. Compiling results and writing of report