

Team name: IDK

Team members: Ming Feng Chua, Rena Pei Qi Chong, Yingxuan Wu, Kevin Li

Date: 03/31/23

This journal has to be uploaded to 1) a **dropbox folder** (<https://www.dropbox.com/request/oMzqzRNvauGRuF97kjh>). **AND** 2) posted to your project website every two weeks by 11:59 pm on the second Friday (except the spring break). When uploading to the dropbox folder, change the file name to "Team_name_MM-DD-2023" where MM = month and DD= day of upload).

Team roles for this report (write down name):

Facilitator(s): Zijun Li, Ming Feng Chua, Rena Pei Qi Chong

Recorder(s): Zijun Li, Ming Feng Chua, Rena Pei Qi Chong

Deliverer(s): Ming Feng Chua, Rena Pei Qi Chong

Planner(s): Ming Feng Chua, Rena Pei Qi Chong

Team Contact: Yingxuan Wu

See last page for description of roles. Obviously one person can take more than one role or there can be more than one person per role or make your own roles!

- 0. Describe briefly what the main goal of your team is (so the peer reviewer has some context). E.g. we are working on image classification for blah de blah. Our goal is blah de blah etc. In the initial part of the semester before your proposal it is ok to put down "we are still coming up with ideas on team project".**

Our chosen topic for the project is a movie recommender system. We will be working towards a movie recommender system that uses the item-based collaborative filtering approach.

- I. **What was done during the report period regarding the project:** If you want to include code include this in the Appendix. Describe what the group did (including contributions of individual team members) with regards to the group project during this report period. Give enough details so I understand what you folks have been doing over the week. Include dates of your meeting(s) and who met on these days.
1. We discuss the potential framework for the movie recommender system based on existing projects and research articles. Zijun found several websites with similar datasets / methods and discussed the possibility of using similar logic or function in our project.
 2. Ming Feng and Rena have narrowed down our approach for the recommender system through various research articles. We have discussed a rough pipeline for our project to build the recommendation system, which is to be finalized during our next meeting.
 3. Ming Feng has started on feature engineering to transform the data into the desired format for the subsequent task of model building.
 4. Rena has started looking into the different algorithms and models that can be used to identify similar movies by their features and user ratings.
- II. **What were obstacles faced if any in working on the project?** This could be technical (like not being able to implement or understand particular techniques) or time issues (midterms for other courses etc).

Since our data is large (eg. 25 million rows in rating.csv), difficulties were faced during feature engineering to transform the data into the required format for our subsequent tasks, such as doing one-hot encoding for categorical data (eg. movie genres). Moving forward, we would export the transformed data out into csv files to reduce the need for any large recomputation to speed up our coding.

The theoretical aspect of the project is challenging to group members who don't have prior experience. Also, it is hard to figure out the distribution of work when everyone has a different level of background to the method and programming language.

III. What is the plan for the next reporting period including what each team member is planning to work on. Describe goals and potential timelines (“ I plan to finish understanding x to see if it can be implemented for our project by Wednesday etc”.)

We plan to have a meeting to finalize the project pipeline and properly delegate the work to each of the members. To ensure that everyone is clear about our finalized approach, we will read up on research articles regarding our item-based collaborative approach by the end of this week to ensure that each member is caught up on the theory and algorithms behind the item-based collaborative approach by the end of this week to promote more productive discussions, and for the ease of delegation of work once our machine learning pipeline has been designed.

What each team member will be working on:

Rena: I have started exploring how to use different parts of our data for our chosen approach and read up on the possible ways to build the model. Moving forward, I will be working on the feature engineering and coding part of the machine learning model for our finalized approach, since our chosen approach will involve many parts to build the final recommender system.

Ming Feng: I will continue to perform feature engineering to prepare the data for model building. Moreover, I will continue exploring research articles on item-based collaborative filtering, such as the prediction of movie rating scores using KNN, matrix factorization and deep neural networks, to gain inspiration on the most ideal and effective model for our dataset. I will also begin designing the backend machine learning pipeline such that work regarding each section of the pipeline can be easily delegated amongst the group members.

Yingxuan: I will do more academic research on the topic of recommendation systems, reading articles, papers and existing projects shared by other group members and try to give more insights by the end of the weekend. Also trying to learn more about potential predicting models that could be useful.

Zijun: I will try to finish reading the introduction and method of the research article sent by my group member and discuss the implementation and question by the end of this weekend. I am going to learn about coding in the platform we're trying to use and join the programming part of the project next week.

While in the biweekly document above you will describe what your team did with regards to the team project (with proper attributions of who did what in the week) there are 4 pre-defined roles. I urge you to have different people do these jobs every week so that you gain experience in each of the jobs. There can also be more than one person per job for example 2 people recording the weekly journal.

Facilitator: Manages the group for this week including setting up times for group members to meet, making sure everyone has a say in the meetings etc.

Recorder: Person in charge of recording the meetings as well as the happenings of the past two weeks and describing what was accomplished in the meeting and writing up this report.

Deliverer: Person in charge of checking the entire report and uploading the file to dropbox folder and project website, as well as the representative of the group getting in touch with the instructor.

Planner: Person in charge of what will be happening next two weeks as well as thinking about longer term goals (what more needs to be done for the project).

Team contact: Person I can email if I see any issues in the biweekly report instead of mass spamming everyone in the team.