Team name: IDK

Team members: Ming Feng Chua, Rena Pei Qi Chong, Yingxuan Wu, Kevin Li
Date: 03/10/23

This journal has to be uploaded to 1) a **dropbox folder (https://www.dropbox.com/request/oMzqzRNvauGRuF97kjyh). 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, Yingxuan Wu

Recorder(s): Yingxuan Wu, Zijun Li

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

Planner(s): Rena Pei Qi Chong

Team Contact: Chua Ming Feng

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.

Items:

1. We held a Zoom meeting over the weekend and determined the work for the two week phase. Each member selected some potentially relevant datasets from the internet, briefly introduced their datasets, and discussed them through the meeting.
2. In the first week, we explored the movie datasets that each of us found and conducted data explorations on three different potential datasets among them. We ultimately compared and decided on the dataset that we will use.
3. Over the weekend, Zijun and Yingxuan conducted exploratory data analysis for their chosen datasets on python and used the files from Kaggle as references. The datasets were obtained from Kaggle which contained information of movies and tv shows from Netflix, and movies from TMDB. They reported the graphs and links to the rest of the team on Monday and expected to get feedback from them by Wednesday.
4. In the second week, Rena and Ming Feng conducted further detailed data analysis on the selected dataset from MovieLens, specifically, we are using the MovieLens 25M dataset. At the same time, we broke down the data files, determined the project's objectives, and clarified our work direction. In addition, we discussed the analysis methods and algorithms that will be used in the initial stages of the model, and completed the writing of the project proposal. We will be building our recommender system using the item-based collaborative filtering approach for the initial stages. We have also updated our website with our project proposal.

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).

In these two weeks of work, all group members encountered special situations such as mid-term exams for other courses or other conditions that caused time conflicts. This led to our communication being less timely, so there is still room for improvement in the rationality of time allocation.

Since none of us have any knowledge in building a recommender system, a lot of time and effort was required in researching the different approaches to recommendation systems and the techniques, such as item-based collaborative filtering, content-based filtering, matrix factorization etc., that we can employ to build the recommender system.

Due to time constraints and short timeline of the project, there was limited time to research current techniques available for recommender systems and to perform EDA for potential datasets. As such, we may have missed out on other insights that we could have explored in our EDA. Following that, we will continue to work on EDA for our chosen dataset to draw some preliminary insights.

Lastly, since our group was busy, we were unable to schedule more meetings to discuss the flow and manpower allocation of the project moving forward.

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 are aiming to start working on the recommender system from spring break / after spring break. We will be working on building the recommender system using the item-based collaborative filtering approach.

What each team member will be working on:

Rena: I will research into the methods to build the recommender system using the item-based collaborative filtering approach. I will also look more into the data and explore how we can better make use of the data, such as the tags data as we currently have little idea how those data can be used together with essential information such as movie genre for the recommendation system..

Ming Feng: During spring break, I will research more into potential improvements beyond the item-based collaborative filtering approach that we will be adopting. I will read up on current industry approaches and standards for recommender systems through research papers to better understand the math and statistics behind the algorithms. After spring break, I will begin writing the code for the recommender system and perform model diagnosis if possible.

Yingxuan: I have no travel plans during the spring break, so I will use this time to further familiarize myself with the programming environment. At the same time, I will gain some background knowledge on the algorithms and item-based collaborative filtering approach that we will be using.

Zijun: I will start to learn about how to program on Visual Studio Code, have a closer look at datasets, and research on the collaborative filtering approach for more details. Also, I’ll think about an alternative way to connect with my teammates to make sure they see and respond timely.

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.