

General Information

Student

- Name: Güneş Özmen Bakan, Alexandru Balan Temocico, Liwia Padowska, Gabrijel Radovčić, Çağan Can Idil, Guilherme Fernando Pereira Sequeira
- Email: c.idil@student.maastrichtuniversity.nl

Advisor

- Name: Francesco Barile
- Academic degree(s): PhD
- Academic title/function: Assistant professor
- Department and Faculty at UM: FSE / DACS
- Email: f.barile@maastrichtuniversity.nl

We, as members of ‘Group 3’, individually confirm that we have thoroughly reviewed and will follow the data protection guidelines.

Description of the Study

- Title of your study
"Evaluating Recommender Algorithms on the MovieLens 20M Dataset: A LensKit Approach"
- At which organisation, if applicable, will the study be carried out?
University of Maastricht.
- In which country (countries) will the study be carried out?
The Netherlands.
- Briefly describe the study. Include very concise information on the background, research questions/aims, research design and methods. Also please explain why this study should take place (in layman's terms) (max 250 words).

This study aims to develop two recommender systems for individual and group users, utilising the MovieLens 20M dataset—a resource with over 20 million movie ratings by 27,278 users. Known for its balanced user selection and extensive movie data, this dataset is invaluable for developing and evaluating recommender systems. Our objective is to assess the performance of individual and group recommenders by obtaining feedback from human participants. We begin with data preprocessing to format the data for analysis. Then using LensKit in Python, initial predictions are made through basic algorithms, with a fallback algorithm ensuring recommendation delivery even if primary algorithms fail. We will explore matrix factorization for its efficiency with large datasets and trial KNN on a subset despite potential computational challenges. Furthermore, we will focus on generating and evaluating group recommendations, forming synthetic groups and implementing social choice aggregation functions to combine individual predictions. Explanations for both individual and group recommendations will be generated and evaluated for effectiveness and understandability, enhancing user experience by providing meaningful recommendations and explanations. The value of this study lies in the increase in customer satisfaction. Imagine Netflix or Amazon recommending shows and products you have no interest in. Not great, right? Using this massive collection of movie ratings, we're trying to build a "movie matchmaker" that gets your tastes right. While some methods are like asking a few friends, others are like asking an entire city.

Ethical Aspects of the Study

The importance of the study

- What justifies in your view of using human research subjects in your study?
Movie recommendations need human interaction to verify accuracy. An algorithm can predict a user's preferences, but only a human can truly verify if the recommendation aligns with their tastes.
- Could comparable research results be attained without using human research subjects?
While many aspects of the research can be addressed using the existing dataset and without direct human interaction, some qualitative and real-time elements of user satisfaction, feedback, and evolving preferences might be best captured with human research subjects.

Possible benefits for the research subjects

- Do the research subjects themselves have any direct benefits from the research results?
In the context of this study there aren't immediate or direct benefits for the subjects however there are some benefits such as getting familiar with recommender systems and their nature.
- Do you expect that in the future, people like these research subjects will benefit from the research results?
Yes, in the context of this study it is highly probable that people who like the research subjects will benefit from the research results in the future. The benefits can be listed as the following: Improved User Experience, Time Savings, Reduced Information Overload, Ethical and Bias Considerations (Continuous research can also ensure that recommendation systems are being built ethically, reducing biases).

Possible harm to the research subjects

- Do you expect that the research subjects may suffer any discomfort or harm from participating in the research (other than time lost on participating in it)?
No.

Respecting the autonomy of the research subjects

- What are you doing to make sure that the research subjects are adequately informed about the nature of your research project?
Before participating, subjects should be given a detailed informed consent document. This document outlines the purpose of the research, what participation entails, potential risks and benefits, and the rights of the participants. A briefing session will also be held, where subjects can learn about the research's purpose and also ask questions.
- Is there any reason to believe that (some) of your research subjects may not be fully free in their decision to participate in your study?
No. Participation is entirely voluntary, no factors will limit the participants' freedom to engage or withdraw from the study at any given time.
- Does your study require any type of deception of research subjects about the true nature of the project? If yes, justify why this is absolutely necessary and whether/when you will reveal to participants the truth about the project.
No.
- How are you securing the consent of your participants?
With a consent form. It will be attached with this.

Respecting the privacy of the subjects

- Are the research subjects remaining completely anonymous?
 - If yes, how are you assuring complete anonymity?
 - If not, why is it necessary to collect identifiable data?
Identifiable data such as preference and rating is essential for this study, which also is arguably not identifiable data.
 - If not, what are you doing to assure that no third person can gain access to the identifiable data or otherwise learn about the identity of the research subjects?
The collected data will not be given away or be shared outside of the research group.
Data will be securely stored in a protected environment with restricted access. Only authorised members of the research team will have access to the data.
 - If not, when will you destroy the original, identifiable data?
Once the study is complete, all the data will be securely and irreversibly destroyed.