Overview

The main objective of the project is to build an end-to-end recommender system with standardized development & evaluation framework.
The team should consist of 4 or 5 students.
Each team will be scored based on a written report and project source code.
Teams should be formed before May 26th.
Each team should submit a document with team name, team member and link to the github repository.

Datasets

Standard Datasets

- [MovieLens 1M Dataset](#) - dataset with movie ratings
- [Book-Crossing Dataset](#) - dataset with books ratings
- [Last.fm Dataset](#) - dataset with artist listening records

Custom Datasets

Teams can propose and develop a system using a different dataset. If opting for this, the dataset must be submitted for review before June 1st.

Components

Final Report [10 points]

- Exploratory Data Analysis
- Development & Experiment Journal
- Comparison of models and approaches
- Summary & Conclusions

Evaluation Framework [30 points]

Offline Implementation [20 points]

- Standardized pipeline for model evaluation
- Implementation of multiple metrics for algorithm performance scoring using the most suitable for specific recommendation task
- Train-test split strategy adapted to the dataset

- Description of online evaluation techniques relevant for specific task
- Clearly defined methodology for performing online testing on the dataset

Modeling [35 points]

- Baseline Recommender [3 points]
- Content-Based Filtering Algorithm [7 points]
- Collaborative Filtering Algorithm [10 points]
- Matrix Factorization Algorithm [15 points]

For Significant Model Quality Improvement [25 points]

This category aims to award teams for significant improvements in model performance and incentivise high quality experiments.
Achieving high quality performance may be difficult using the list of standard algorithms defined in the Modeling Part.

For Engineering Excellence [25 points]

This category aims to award teams for excellence in code organization and infrastructure that supports experiments.
The following aspects will be taken into account: code organization, code readability, ease of extending the functionality and reusability.
The optimal solution should have a common model interface, ways to serialize the model, unified training pipeline, robust process for models inferences, versioning strategy, etc.

Deliverables

Report file in pdf or html format
GitHub repository with README file that contains clear setup instructions and module description

Deadlines

May 26th - team formation deadline
June 1st - custom dataset submission deadline
June 4th - dataset selection deadline
July 17th 01:00 - project submission deadline
July 24th 01:00 - project submission deadline with penalty of 30 points