

## Project 3: Recommender System

### General Introduction:

In this project, you are asked to learn about algorithms in recommender system. Each team should submit codes and a report via Canvas.

### Dataset Description: MovieLens 100K

Data link and description: <https://grouplens.org/datasets/movielens/100k/>

This dataset contains 100000 ratings by 943 users on 1682 items. Each user has rated at least 20 movies.

### Complete the following tasks:

Suppose you are working for a recommender system, given training data of collected user ratings, you need to train a model that can predict the user rating in test data. Try to improve the performance as much as you can.

- **Data preprocessing:** Preprocess the raw data (with any technique you find useful, e.g., dimension reduction, normalization, in-advance similarity computation, ...).
- **Algorithm:** train the prediction model, you can use any method you like (e.g., content-based, collaborative filtering, hybrid...). Try different parameters and techniques to improve the performance.
- **Data usage:** Depend on your algorithm. The dataset includes both rating scores and user/item attributes. Many algorithms only need the rating data to train the model, but if you want, you can also use user/item attributes.
- **Train/Test data:** The dataset `uk.base` and `uk.test` ( $k=1, \dots, 5$ ) are 5-fold 80%/20% train/test split of the user rating data.
- **Metrics:** accuracy and RMSE. Please report your final performance in the above **5-fold cross validation** in the form of **mean  $\pm$  standard deviation**.

Note: Existing packages of data preprocessing, math, and algorithms (such as collaborative filtering) can be directly used.

### Project Submission:

• Prepare your submission. **One team only needs to provide one submission.** Make a zipped folder named "**CaseID\_CaseID\_Proj3.zip**", where "CaseID" refers to your group members' Case IDs. In the folder, you should include:

1. **Report:** A pdf file named **Recommendation\_report.pdf**. Describe the whole method that you used, including the data preprocessing and the prediction model. Try to tune your model with different parameters or techniques and give comparison for your tuning process, report your experimental results with the above comparison and your own analysis. The report is suggested (but not required) to **not exceed** 4 pages.
2. **Code:** A zipped folder named **code.zip**, which contains all codes used in this part. Inside the folder, please also provide a **README** file which describes how to run your code.

Note that copying code/results/report from another group or source is not allowed and may result in an F in the grades of all the team members.

**Grades will be given based on the following criterion:**

- Report (65 pts):
  - (10 pts) Data processing: description of the data processing process
  - (25 pts) Algorithm: describe the design of your prediction model
  - (10+10+5=25 pts) Result evaluation, parameter/technique tuning, analysis
  - (5 pts) Overall coherence and clarity
- Code (35 pts)
  - (30 pts) Correctness and Reproducibility of the results in the report
  - (5) Readme: Readability and clarity