

# MiniProject 2 (20%)

CS3448: Recommender Systems /

CSX4207/ITX4207: Decision Support and Recommender Systems

IT4454: Selected Topic in Decision Support and Recommender Systems

# Mini Project 2

- **Part I: A simple recommendation algorithm (8%)**

1. Download the datasets from the subfolder 'miniproject/p2' in Google shared drive:

1. rating\_trainset.csv
2. rating\_testset.csv

Note: Each file has 3 attributes: userid, placeid, and rating in the range of 1-10 (separated by Tab).

2. (5%) Write a program to

- Read a training set from the 'rating\_trainset.csv' file.
- Read a test set from the 'rating10user91testset.csv' file.
- Implement **User-based Nearest Neighbor algorithm** that
  - Use **Cosine Similarity** to calculate the similarity between all possible pairs of user in the training set. << Store the result in the file 'Group1\_Part1\_COSINE\_11.csv'.
  - Based on the computed similarity matrix, predict the rating of all unseen books for all users in 'rating\_testset.csv', where  $k = 7$ .
    - **Note:** Ignore the 'true' ratings given in the 'rating10user91\_testset.csv' when making a prediction. They will be used when evaluating the results.

3. (3%) Display **the top 5 not-yet-visited restaurants (with the following details: TargetUserID, 1stNNUserID, 2ndNNUserID, 3rdNNUserID, 4thNNUserID, 5thNNUserID, 6thNNUserID, 7thNNUserID, PlaceID, predicted rating)** with respect to the similarity results for each user in 'rating\_testset.csv'. << Store the result in the file 'Group1\_Part1\_RECOMMEND\_12.csv'.

# Mini Project 2 -- Cont.

- **Part II: A simple collaborative based filtering algorithm and evaluation (8%)**
  1. Use the same dataset as given in Part I.
  2. Select and implement a collaborative based approach discussed in the classes ***that is different from Part I*** to recommend the top 10 not-yet-visited restaurants to the users in 'rating\_testset.csv':
    - (1%\*) Store the created profiles in the file 'Group1\_Part2\_PROFILE\_21.csv' with column headers.
    - (4%) Store the similarity matrix and/or model used in the file 'Group1\_Part2\_SIMILARITY\_22.csv' with row and column header.
    - (1%) Display the **10 not-yet-visited restaurants** (UserID, PlaceID, model's calculated value, e.g., predicted rating) for each target user. << Store the result in the file 'Group1\_Part2\_RECOMMEND\_23.csv'.
    - (2%) Calculate **RMSE, Precision, Recall** of all unseen items for all target users in the testset. << Store the result in the file 'Group1\_Part2\_EVAL\_24.csv'.
- *Remark \*: The score will be given only if the algorithm is applied on the dataset as well.*

# Submission and Presentation (4%)

- Submit the code (a zip file), **the following 6 files** and the presentation one day before the deadline (Feb. 26, 2023 before midnight) – 2% Score deduction will be applied for any late submission:
  1. Group1\_Part1\_COSINE\_11.csv
  2. Group1\_Part1\_RECOMMEND\_12.csv
  3. Group1\_Part2\_PROFILE\_21.csv
  4. Group1\_Part2\_SIMILARITY\_22.csv
  5. Group1\_Part2\_RECOMMEND\_23.csv
  6. Group1\_Part2\_EVAL\_24.csv
- (4%) Every team member must present your individual contribution (***individual scoring may be applied***) in class on Feb. 27, 2023. Otherwise, there is no score given.
  - Details of the presentation slides:
    - The detail of all tasks in Part I and II.
  - Prepare to answer any implemented codes for verification.
  - Also explain the tasks of each member in the presentation slides