MiniProject 2 (20%)

CSX4207/ITX4207: Decision Support and Recommender Systems

Mini Project 2

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

1. Download the datasets from the subfolder 'mini-project2' in MS Teams' class channel:
   1. rating10user91_trainset.csv
   2. rating10user91_testset.csv

   Note: Each file has 3 attributes: userid, isbn, and rating in the range of 1-10 (separated by Tab).
2. Write a program to
   - Read a training set from the 'rating10user91_trainset.csv' file.
   - Read a test set from the 'rating10user91_testset.csv' file.
   - Implement User-based Nearest Neighbor algorithm (as shown in the lecture slides) that
     - Use Pearson Correlation Coefficient to calculate the similarity between all possible pairs of user in the training set. << Store the result in the file 'P2Part1_1PCC_Group [group_no].csv'.
     - Based on the computed similarity matrix, ~~predict the rating of all unseen books for all users in 'rating10user91_testset.csv',~~ where k = 5.
       - **Note**: Ignore the 'true' ratings given in the 'rating10user91_testset.csv' when making a prediction. They will be used when evaluating the results.
3. Display the first 5 not-yet-read books (TargetUserID, 1stNNUserID, 2NNUserID, 3NNUserID, 4NNUserID, 5NNUserID, Book's ISBN, predicted rating) with respect to the similarity results for each user in 'rating10user91_testset.csv'. << Store the result in the file 'P2Part1_2Recommendation_Group [group_no].csv'.

Mini Project 2 -- *Cont.*

- **Part II: A simple collaborative based filtering algorithm (7%)**
  1. Use the same dataset as given in Part I.
  2. Select a collaborative based approach discussed in the classes ***that is different from Part I*** to recommend top-10 not-yet-read books to the users in 'rating10user91_testset.csv':
     - Store the created profiles in the file 'P2Part2_1Profile_Group[group_no].csv' with column headers.
     - Store the similarity matrix and/or model used in the file 'P2Part2_2Model_Group[group_no].csv' with row and column header.
     - Display 1) the **top 10 not-yet-read books** (User ID, Book's ISBN, model's calculated value, e.g., predicted rating) for each target user. << Store the result in the file 'P2Part2_3Recommendation_Group [group_no].csv'.
     - Calculate RMSE of all unseen items for all target users in the testset. << Store the result in the file 'P2Part2_4RMSE_Group[group_no].csv'.

Submission and Presentation (5%)

- Submit the code (a zip file), six resulted files and the presentation (.pdf) one day before the deadline (Sep. 22, 2025 before midnight).
  1. P2Part1_1PCC_Group [group_no].csv
  2. P2Part1_2Recommendation_Group [group_no].csv
  3. P2Part2_1Profile_Group[group_no].csv
  4. P2Part2_2Model_Group[group_no].csv
  5. P2Part2_3Recommendation_Group [group_no].csv
  6. P2Part2_4RMSE_Group[group_no].csv
- Every team member must present your individual contribution in class on Sep. 23, 2025. Otherwise, there is no score given for the absent student.