MiniProject 1 (20%)

CSX4207/ITX4207: Decision Support and Recommender Systems

Mini Project 1

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

1. Download the datasets from the subfolder 'miniproject1' in MS Teams' :
    1. HOTEL_DATASET.csv
    2. UserData.csv
2. (2%) Preprocess and/or select a subset of the attributes of hotels used for generating recommendations.
3. (2%) Create the **profile of 5 users** whose ids are in the file UserData.csv. << Store the created profiles in the file '**Group1**_Part1_Profile11.csv' (change **Group1** to your group number) with column headers. *Hint: you may use the user profiling technique discussed in class.*
4. (2%) Calculate **Jaccard similarity** of each user profiles and not-yet-visited-hotel, and then display the results (user/hotel similarity matrix). << Store the similarity matrix in the file '**Group1**_Part1_SimMatrix12.csv' with row and column header.
5. (2%) Display 1) ***the first 5 not-yet-visited-hotel*** (user ID, Item ID, hotel ID, hotel's name, similarity values) with respect to step 4) the similarity results ***for each user***. << Store the result in the file '**Group1**_Part1_Recommendation13.csv'.

Mini Project 1 -- *Cont.*

- **Part II: A simple content based filtering algorithm (7%)**

1. Use the same dataset as given in Part I.
2. Select a content based approach discussed in the classes ***that is different from Part I*** to create the content based recommendation to recommend top-10 not-yet-visited hotel to the users:
   - (1%) Store the created profiles in the file '**Group1**_Part2_Profile21.csv' with column headers.
   - (5%) Implement the algorithm used and store the similarity matrix and/or model obtained in the file '**Group1**_Part2_Model22.csv' with row and column header.
   - (1%) Display 1) the **top 10 not-yet-visited hotel** (user ID, Item ID, hotel ID, hotel's name, model's calculated value, e.g., similarity result) **for each user**. << Store the result in the file '**Group1**_Part2_Recommendation23.csv'.

Submission and Presentation (5%)

- Submit the code (a zip file), the following 6 files and the presentation one day before the deadline (Aug. 25, 2025 before midnight) – 2% Score deduction will be applied for any late submission:
  1. **Group1**_Part1_Profile11.csv
  2. **Group1**_Part1_SimMatrix12.txtac
  3. **Group1**_Part1_Recommendation13.csv
  4. **Group1**_Part2_Profile21.csv
  5. **Group1**_Part2_Model22 .txt
  6. **Group1**_Part2_Recommendation23.csv

- (5%) Every team member must present your individual contribution ***(individual scoring may be applied)*** in class on Aug. 26, 2025. 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