Institute for Advancing Intelligence, TCG CREST
(TCG Centres for Research and Education in Science and Technology)

Design and Analysis of Algorithms
PhD Coursework, Semester-II, Session: 2023-24

Project-I

Maximum Marks: 5 Submission Deadline: **2023-Mar-10**

Title: **Analyzing Movie Ratings using Hash Tables and Divide and Conquer**

Problem Statement

You are given a dataset containing movie ratings provided by users. Each entry in the dataset consists of the following information:

- User ID
- Movie ID
- Rating (on a scale of 1 to 5)
- Timestamp of the rating

There are 1000 users and 5000 movies. Your task is to perform the following analyses using hash tables and divide and conquer algorithms:

Part 1: Hash Table Analysis

1. Implement a hash table (Create, Insert and Delete) to efficiently store the movie ratings data.
2. If you are using python, then use dict to store the ratings. Finally, run 10000 insertions followed by 10000 search operations and compare the time taken by your implemented one and the dict library

Part 2: Movie rating

1. Compute and store the average rating for each movie in a file `avg_ratings.csv`.
2. Use Previous hash table to efficiently store user-wise ratings.
3. Compute the average rating provided by each user.

Part 3: Divide and Conquer Analysis

1. Implement a divide and conquer algorithm to efficiently sort the dataset based on movie ID.
2. Implement a binary search algorithm to find the average rating of a specific movie given its ID.

Dataset

You are provided with a large dataset containing movie ratings in the following format:

```
User ID,Movie ID,Rating,Timestamp
```

The dataset is stored in a CSV file named `ratings.csv`.

Submission Guidelines

1. Submit your source code files along with a README file containing instructions to compile and execute the code.
2. Submit the analysis report in a file PDF file.
3. Convert the codes in a PDF file.
4. Submit the soft copies within due date.
5. Submit the hard copies on the next day to the due date.

Note

- You are allowed to use any open source codes and tools.
- Collaboration is not permitted.