Recommender System for Entertainment

When you want to watch your favorite movies or listen to songs or perhaps shop online, have you noticed that the items suggested to you are perfectly aligned with your interests? This is the beauty of AI. These smart recommendation systems learn your behavior and interests from your online activities and offer you similar content. The personalized experience is made possible by continuous training. The data is collected at the frontend (from the user), stored as big data and analyzed through machine learning and deep learning. It is then able to predict your preferences by recommendations that keep you entertained without having to search any further. **User profiles** play an important role in the success of the **recommendation** process since the **profiles** represent the **users'** information needs.

The objective of your project is to:

Implement a **collaborative recommendation engine** that exploits the user's likes and dislikes to recommend various items. Collaborative filtering is a technique that can filter out items that a user might like on the basis of reactions by similar users. Most websites like Amazon, YouTube, and Netflix use collaborative filtering as a part of their sophisticated recommendation systems.

Data Set
To experiment with recommendation algorithms, you'll need data that contains a **set of items** and a **set of users** who have reacted to some of the items. The reaction can be **explicit** (rating on a scale of 1 to 5, likes or dislikes) or **implicit** (viewing an item, adding it to a wish list, the time spent on an article). While working with such data, you'll mostly see it in the form of a **matrix** consisting of the reactions given by a set of users to some items from a set of items. Each row would contain the ratings given by a user, and each column would contain the ratings received by an item. A matrix with five users and five items could look like this:

	i_1	i_2	i_3	i_4	i_5
u_1	5		4	1	
u_2		3		3	
u_3		2	4	4	1
u_4	4	4	5		
u_5	2	4		5	2

Fig1: Rating Matrix

There are a lot of datasets that have been collected and made available to the public for research and benchmarking. Here's a list of high-quality data sources that you can choose from. The best one to get started would be the MovieLens data set (https://grouplens.org/datasets/movielens/) collected by GroupLens Research.

This dataset consists of many files that contain information about the movies, the users, and the ratings given by users to the movies they have watched. For example,

- **u.item:** contains the list of movies
- **u.data:** contains the list of ratings given by users

The file u.data that contains the ratings is a tab separated list of user ID, item ID, rating, and timestamp. The first few lines of the file look like this:

user_id	item_id	rating	timestamp
196	242	3	881250949
186	302	3	891717742
22	377	1	878887116
244	51	2	880606923
166	346	1	886397596

Fig2 : First 5 Rows of MovieLens 100k Data

As shown above, the file tells what rating a user gave to a particular movie. This file contains 100,000 such ratings, which will be used to predict the ratings of the movies not seen by the users.

Requirements:

To develop a **collaborative recommendation engine**, there are **two** approaches:

- **User-based Collaborative filtering**-Find similar users and then recommend movies
- **Item based Collaborative Filtering** –Find similar movies and then recommend movies

a) Implement a recommendation engine in any high level programming language of your choice by selecting one of the above approaches.(**10 marks**)
b) Enhance the data set with some of the latest movies and users of your choice and show the movie recommendations by the developed engine. (**5marks**)
c) Prepare a presentation of not more than **15 slides** that explains your adopted technique of developing a recommendation engine and demo of the project. (**10 marks**)

Team Members:
The project team can contain at most **two** members with equal delegation of tasks as discussed with the course instructor.

Evaluation Criteria:
The institute has **zero** tolerance for plagiarism. Anything copied shall result into **0 marks.**

Submission Deadline: 20th of June' 2021 (No late submissions would be accepted.)