

## **Assignment 1: User-based Collaborative Filtering Recommendations**

Due: **November 8, 2023**

Points: 20/100

Submit: A file upload at Moodle

The goal of the first assignment is to implement a user-based collaborative filtering approach.

The assignment may be completed in pairs. Each pair submits one only assignment, and both students are expected to understand, be able to explain, and be able to modify the implementation.

(a) Download the **MovieLens 100K** rating dataset from <https://grouplens.org/datasets/movielens/> (the *small dataset recommended for education and development*). Read the dataset, display the first few rows to understand it, and display the count of ratings (rows) in the dataset to be sure that you download it correctly. **Score: 10%**

(b) Implement the user-based collaborative filtering approach, using the Pearson correlation function for computing similarities between users (**Score: 30%**), and (c) the prediction function presented in class for predicting movies scores (**Score: 30%**).

(d) Select a user from the dataset, and for this user, show the 10 most similar users and the 10 most relevant movies that the recommender suggests. **Score: 5%**

(e) Design and implement a new similarity function for computing similarities between users. Explain why this similarity function is useful for the collaborative filtering approach. Hint: Exploiting ideas from related works are highly encouraged. **Score: 25%**

Any programming language for your assignment is acceptable.

Please explain any assumptions you made.

Submit your files at Moodle before **NOVEMBER 8, 2023** (before 11.00pm). Some instructions on how to run your codes are necessary.

Later this month, we may have face-to-face meetings for examining the students implementations of Assignment 1.

Please send your groups to [konstantinos.stefanidis@tuni.fi](mailto:konstantinos.stefanidis@tuni.fi) before 5.00pm, **November 6**.

**Note:** For each 1-week delay in an assignment submission, you lose 10% of your assignment score.