Assignment 1

If you look at the examples made with the user-based collaborative filtering method, you'll see that most users have watched and rated many of the movies and anime in the whole dataset. User-based systems try to define the dataset using one big table, which is sometimes called a pivot table. You can see a sample of this table in Figure 1. In this table, rows show individual users, and columns show the movies or anime they rated.

User_id	Anime_id_1	Anime_id_2	Anime_id_3	Anime_id_4	Anime_id_5
10001	5 (rating)	8	9	6	10
10002	8	4	-	8	7
10003	-	6	7	8	3
10004	5	6	5	9	10

Figure 1: Pivot matrix for user-based recommendation systems

However, because the dataset in this assignment is quite large, with many users and anime involved, creating a table like the one in Figure 1 can be very costly. In such situations, it's important to avoid making a big table and instead use the ratings given by users by combining them from different rows. In this assignment, it's necessary to represent the dataset as illustrated in Figure 2.

Also, because each user has watched between 10 and 25 anime, using the pivot table method means that even anime that no one has watched will still take up space in the table.

User_id	Anime_id	Genre_Advanture	Genre_Mystery	...	...	...	...	Duration(min)	Rating
10001	1	0	1						5
10001	34	0	0						4
10001	5	1	0						8
10001	67	1	1						10
10001	88	0	1						3
10001	1123	0	1						6
10002	5	1	0						8
10002	6	0	0						2
10002	88	1	1						9
10002	77	1	0						5

Figure 2: Example dataset for training.