

## **Project Deliverable 2**

### **Movie Recommendation System**

#### **1. Problem statement:**

- The aim of this project is to build a python machine learning model to suggest movies for users. The users will simply input the movies that they like and the system will output suggesting movies.

#### **2. Data Preprocessing:**

- The dataset is the small movielens dataset from kaggle.com, and two main data tables are movies.csv and ratings.csv. The movies.csv contains more than 9000 movies from 1995 to 2017, including their titles and genres. The ratings.csv contains more than 100k ratings from 671 unique users.
- Since I am only interested in userId, movieId, rating from the rating data frame, i only extract those features. For the movies dataset, I first remove films that does not have any ratings, then merge this data frame with the rating data frame.
- Now the data is ready for machine learning model.

#### **3. Machine Learning Model:**

- The approach here is using collaborative filtering to suggest new movies. The idea is that two users with same ratings on a movie may have the same preference and the system can suggest movies that one user have watched to another.
- For this approach, I simply created a scr\_matrix from the python scipy module since note every user rated every movies.
- The matrix created has rows and columns corresponding to userId and movieId, and each cell is rating of a user to a movie.
- I then used different distance measurement algorithms, from Euclidean to Cosine to measure how different between two movies based on their ratings.
- After the model completed calculating, I simply made it print out 10 movies with smallest differences from the origin movie. The code can be found on Data\_Exploration.ipynb file.

#### **3. Preliminary Results:**

- At this point, the model is simply take a movie as an input and suggest new movies based on collaborative filtering method.

## **5. Next Steps:**

- At this point, I only know about this simple method to build a recommendation system.

The next approach maybe building it using content-based algorithm or do a supervised learning algorithm to suggest new movies and ratings that a user will rate for those new movies.