

Machine Learning

Assignment 1

CLO3

Ida Bagus Dwi Satria Kusuma
1301140297

September 16, 2017

1. In this problem we will consider similarity measures for movies on the Movielens dataset. Download the Movielens data that we sent to you through email. In addition to the data, the file also contains some functions for easily loading the data into Matlab/Octave/R and some example code that you can use if you wish. See the README files for details.
 - (a) We will now construct a similarity measure over the movies. For simplicity, let us first consider a simple measure that does not use the explicit (numerical) ratings given by the users, nor the time stamps of the ratings, but only whether or not a given movie was rated by a given user.
 - i. **(20 points)** Create a function that, given two different movie IDs as input, outputs the Jaccard coefficient: the number of users who rated both movies divided by the number of users who rated at least one of the movies. For example, for the movies 'Toy Story' and 'GoldenEye' the coefficient should be 0.217.

Answer Gunakan jaccard_coeff.m untuk mendapatkan Jaccard Coefficient. Parameternya adalah dua id film yang ingin diukur.

```
>> jaccard_coeff(1,2)
Jumlah user yang merate kedua film :      104
Jumlah user yang merate film (1) Toy Story (1995) :      348
Jumlah user yang merate film (2) GoldenEye (1995) :      27
Jaccard coefficient : 0.217
```

- ii. (5 points) What is the Jaccard coefficient between 'Three Colors: Red' and 'Three Colors: Blue'?

Answer Use searchMovie.m to get the movie's id and use jaccard_coeff.m to get the Jaccard Coefficient.

```
>> searchMovie('Three Colors: Red')
Indeks of Three Colors: Red is : 59
>> searchMovie('Three Colors: Blue')
Indeks of Three Colors: Blue is : 60
>> jaccard_coeff(59,60)
Jumlah user yang merate kedua film :      55
Jumlah user yang merate film (59) Three Colors: Red (1994) :      28
Jumlah user yang merate film (60) Three Colors: Blue (1993) :      9
Jaccard coefficient : 0.598
```

- iii. (10 points) What are the 5 movies with highest Jaccard coefficient to 'Taxi Driver'?

Answer Gunakan searchMovie.m untuk mendapatkan id film 'Taxi Driver' dan gunakan jaccard_coeff_n.m untuk mendapatkan film-film dengan Jaccard Coefficient sebanyak n tertinggi. dalam kasus ini, $n = 5$.

```
Showing movies that related to Taxi Driver (1976)
ID : 182      | GoodFellas (1990)      | Jaccard Coeff : 0.417
ID : 187      | Godfather: Part II, The (1974) | Jaccard Coeff : 0.417
ID : 179      | Clockwork Orange, A (1971) | Jaccard Coeff : 0.404
ID : 134      | Citizen Kane (1941)      | Jaccard Coeff : 0.397
ID : 654      | Chinatown (1974)      | Jaccard Coeff : 0.394
```

- iv. (10 points) Select a movie of your own choosing (which you are familiar with), what are the 5 movies with highest Jaccard coefficient to that movie? Do they make sense?

Answer Gunakan searchMovie.m untuk mendapatkan id film 'Star Wars' dan gunakan jaccard_coeff_n.m untuk mendapatkan film-film dengan Jaccard Coefficient sebanyak n tertinggi. dalam kasus ini, $n = 5$.

```
Showing movies that related to Star Wars (1977)
ID : 181      | Return of the Jedi (1983)      | Jaccard Coeff : 0.787
ID : 174      | Raiders of the Lost Ark (1981) | Jaccard Coeff : 0.610
ID : 1      | Toy Story (1995)      | Jaccard Coeff : 0.583
ID : 172      | Empire Strikes Back, The (1980) | Jaccard Coeff : 0.570
ID : 100      | Fargo (1996)      | Jaccard Coeff : 0.565
```