Movie Recommender System

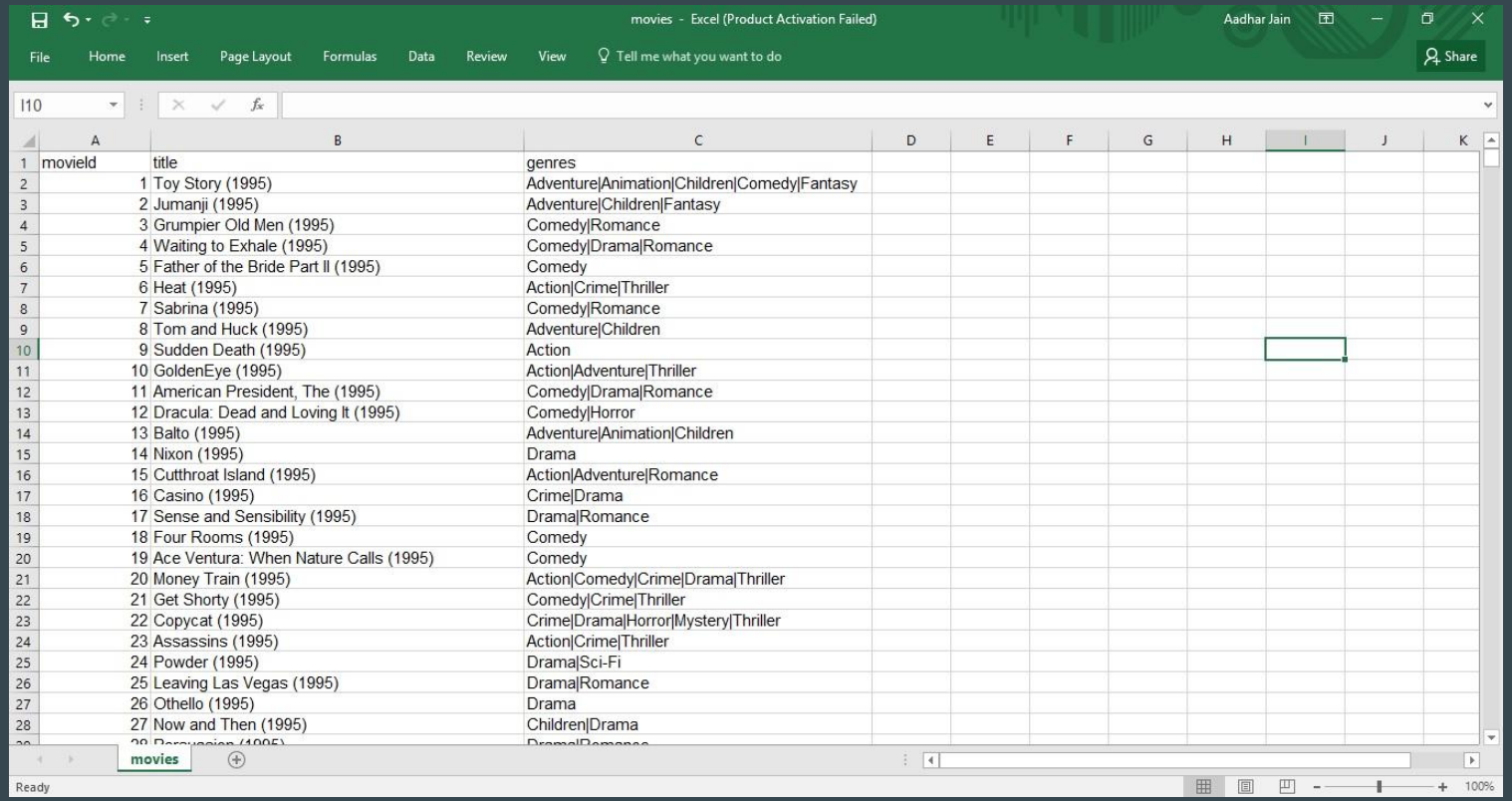
Aadhar Jain
Chirag Agarwal
Pratishtha Patni
Saloni Goyal

What does the System do??

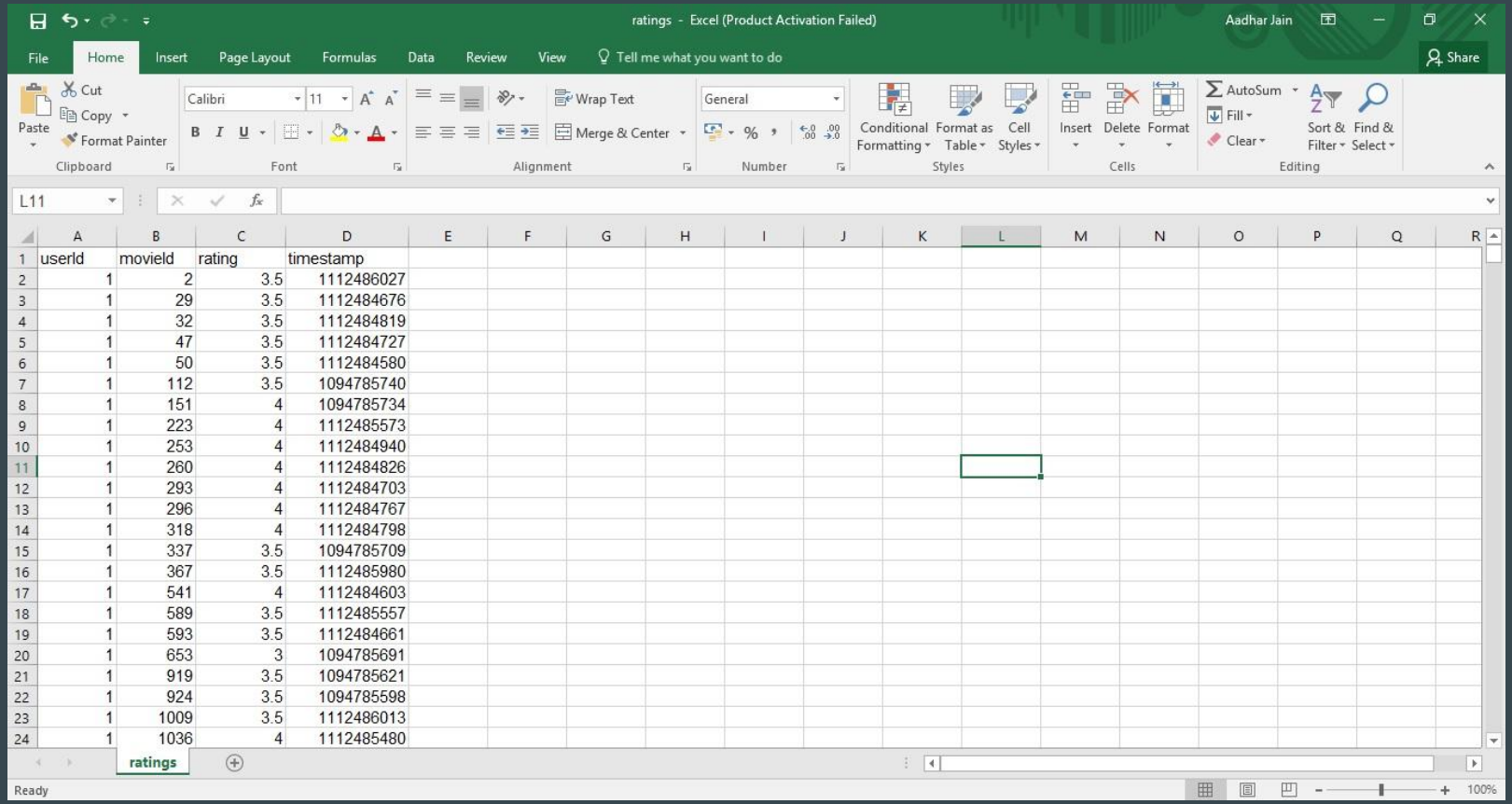
- Predict Task - The user's preference for a movie.
- Recommend Task - Produce best ranked list of n-items for user's need.
- Gives you recommendations based on your present mood using gnome score.
- Gives you recommendations based on your previous choices.
- The end result is an application where a user is recommended top 10 movies.

How does the System work??

Dataset (MovieLens Data)



	A	B	C
1	movieId	title	genres
2	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy
3	2	Jumanji (1995)	Adventure|Children|Fantasy
4	3	Grumpier Old Men (1995)	Comedy|Romance
5	4	Waiting to Exhale (1995)	Comedy|Drama|Romance
6	5	Father of the Bride Part II (1995)	Comedy
7	6	Heat (1995)	Action|Crime|Thriller
8	7	Sabrina (1995)	Comedy|Romance
9	8	Tom and Huck (1995)	Adventure|Children
10	9	Sudden Death (1995)	Action
11	10	GoldenEye (1995)	Action|Adventure|Thriller
12	11	American President, The (1995)	Comedy|Drama|Romance
13	12	Dracula: Dead and Loving It (1995)	Comedy|Horror
14	13	Balto (1995)	Adventure|Animation|Children
15	14	Nixon (1995)	Drama
16	15	Cutthroat Island (1995)	Action|Adventure|Romance
17	16	Casino (1995)	Crime|Drama
18	17	Sense and Sensibility (1995)	Drama|Romance
19	18	Four Rooms (1995)	Comedy
20	19	Ace Ventura: When Nature Calls (1995)	Comedy
21	20	Money Train (1995)	Action|Comedy|Crime|Drama|Thriller
22	21	Get Shorty (1995)	Comedy|Crime|Thriller
23	22	Copycat (1995)	Crime|Drama|Horror|Mystery|Thriller
24	23	Assassins (1995)	Action|Crime|Thriller
25	24	Powder (1995)	Drama|Sci-Fi
26	25	Leaving Las Vegas (1995)	Drama|Romance
27	26	Othello (1995)	Drama
28	27	Now and Then (1995)	Children|Drama


ratings - Excel (Product Activation Failed)
Aadhar Jain
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Tell me what you want to do
Share
Paste
Cut
Copy
Format Painter
Clipboard
Calibri
11
Font
Wrap Text
Merge & Center
Alignment
General
Number
Conditional Formatting
Format as Table
Cell Styles
Styles
Insert
Delete
Format
Cells
AutoSum
Fill
Clear
Sort & Filter
Find & Select
Editing
L11
userId movieId rating timestamp
1 2 3.5 1112486027
1 29 3.5 1112484676
1 32 3.5 1112484819
1 47 3.5 1112484727
1 50 3.5 1112484580
1 112 3.5 1094785740
1 151 4 1094785734
1 223 4 1112485573
1 253 4 1112484940
1 260 4 1112484826
1 293 4 1112484703
1 296 4 1112484767
1 318 4 1112484798
1 337 3.5 1094785709
1 367 3.5 1112485980
1 541 4 1112484603
1 589 3.5 1112485557
1 593 3.5 1112484661
1 653 3 1094785691
1 919 3.5 1094785621
1 924 3.5 1094785598
1 1009 3.5 1112486013
1 1036 4 1112485480
ratings
Ready
100%

Algorithms Used

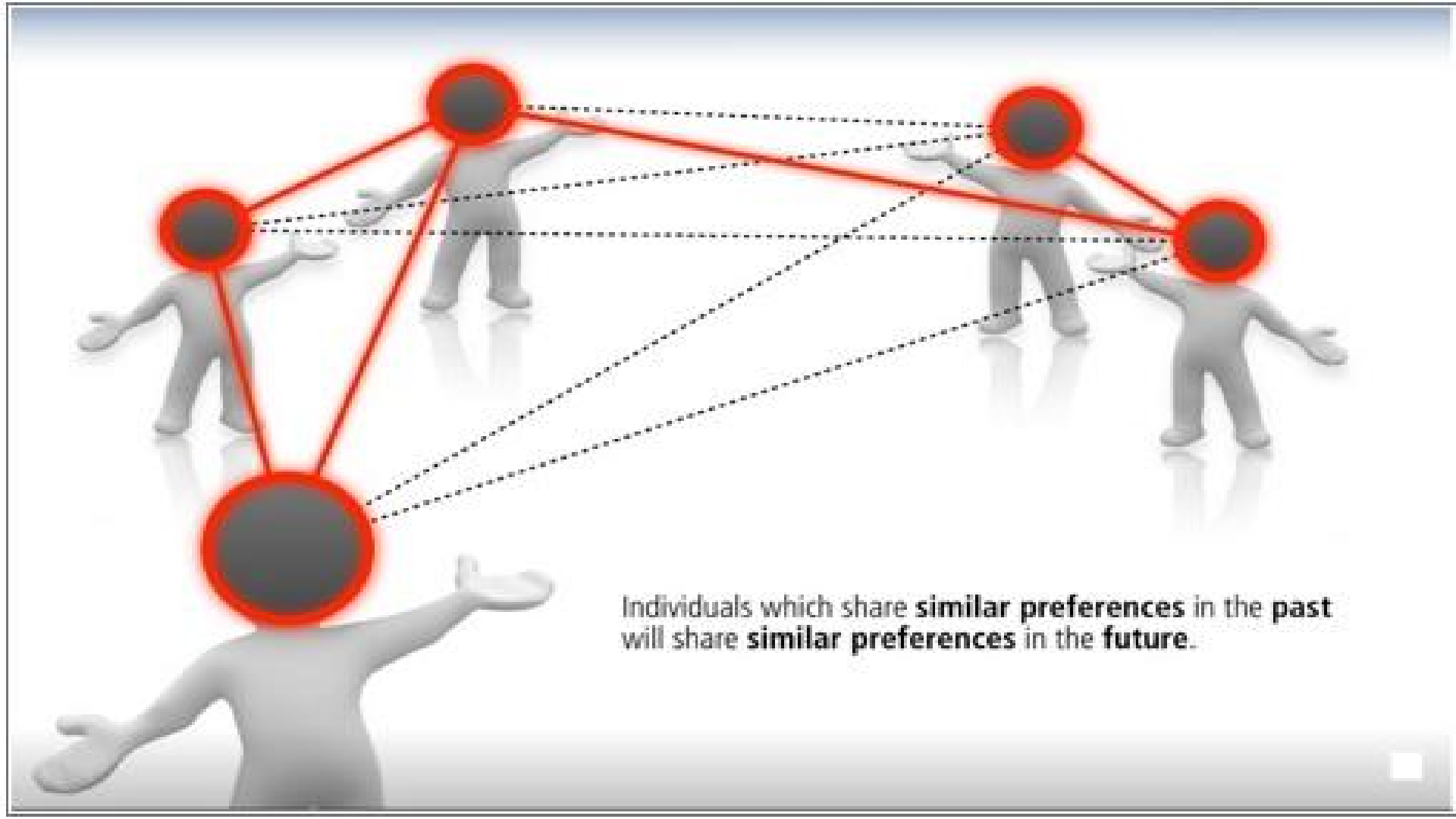
- Collaborative Filtering
- K-means Clustering
- Jacckard Index (Similarity)

Preprocessing of Data

- Parsed the data from .csv files
- Used hashing for users and movies.
- <Key,Value> -- <User/Movie Id, User/Movie Object>
- Hashing done on the tags where key is the name of the Tag and value is Tag Id.

Collaborative Filtering

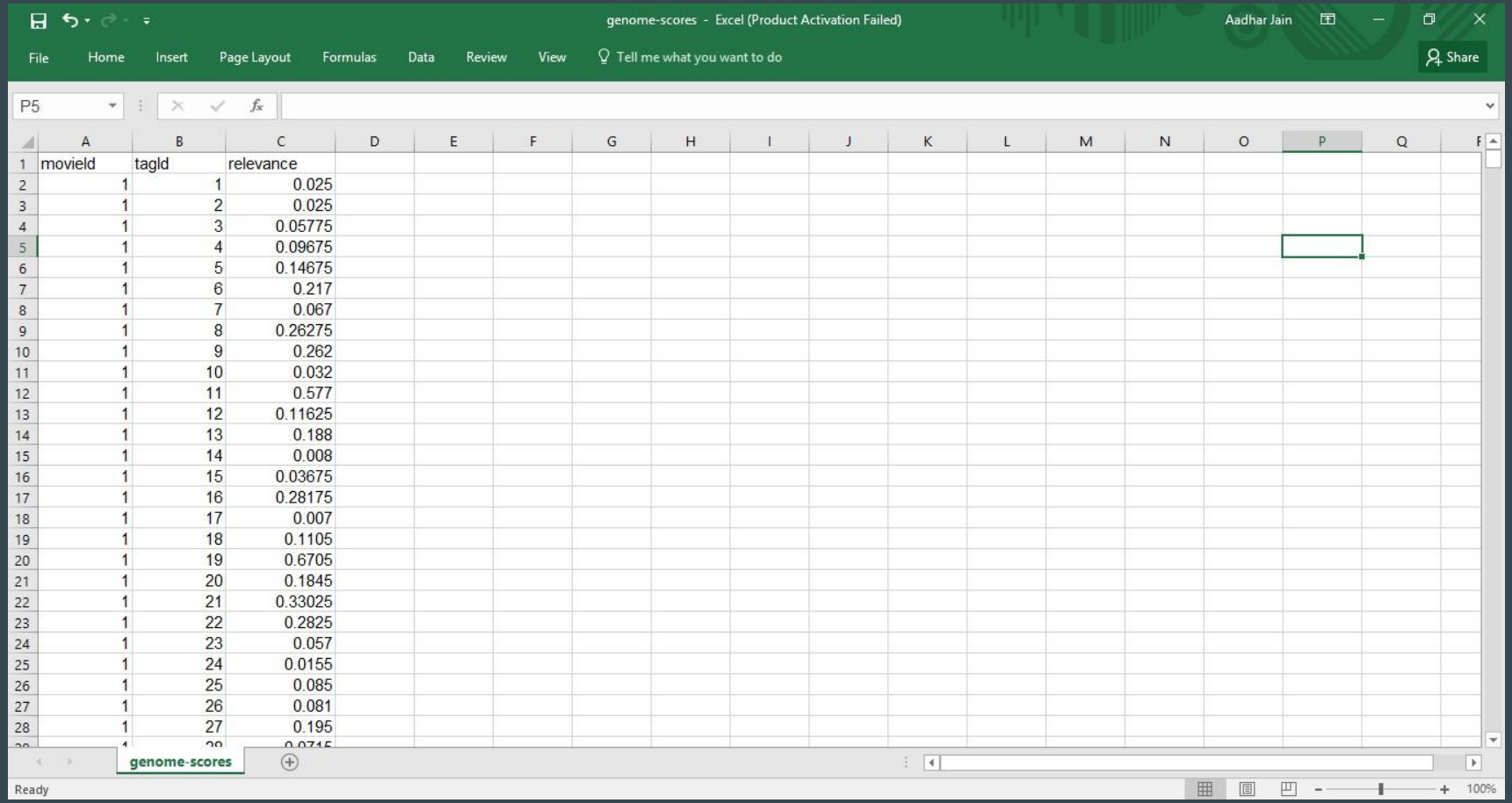
- People relied on the recommendations from their peers. This method doesn't take the personal preference of the user into account.
- Similar users to the active user (user that recommendations are prepared for) are found.
- By weighting the users, a recommendation list is prepared from other user data.


Individuals which share **similar preferences** in the **past**
will share **similar preferences** in the **future**.

- Used Jacckard Index to find Similarity among users
- Recommended movies using several heuristics (age, movie genre, gender, occupation etc)
- This method even though gave accurate results, its performance deteriorated when user-movie ratings matrix became sparse.

Mood Based Recommendation

- Recommendation is given based on user's current mood
- Asking user the genre of the he wants to see
- Implemented using gnome score of movies



	movieId	tagId	relevance
1	1	1	0.025
2	1	2	0.025
3	1	3	0.05775
4	1	4	0.09675
5	1	5	0.14675
6	1	6	0.217
7	1	7	0.067
8	1	8	0.26275
9	1	9	0.262
10	1	10	0.032
11	1	11	0.577
12	1	12	0.11625
13	1	13	0.188
14	1	14	0.008
15	1	15	0.03675
16	1	16	0.28175
17	1	17	0.007
18	1	18	0.1105
19	1	19	0.6705
20	1	20	0.1845
21	1	21	0.33025
22	1	22	0.2825
23	1	23	0.057
24	1	24	0.0155
25	1	25	0.085
26	1	26	0.081
27	1	27	0.195

Future Goals

- Incorporate Content Based Filtering by asking questions from user and recommending movies on that basis.
- Cluster Visualization
- Bringing innovation in the existing system
- Comparative analysis of various algorithms using Recall, Precision and ROC curve