[MOVIES RECOMMENDATION SYSTEM]

[26]

Team Members:

1. Ahmed Ayman Aabed
2. Ahmed Emad Ahmed
3. Amany Salah Zahran
4. Moumen Hamada Nagah
5. Mohamed Mostafa Abd-elhakim

Project Title

Introduction:

Recommend the N closest movies to a certain movie based on users ratings.

input -> system -> output

movie -> Recommendation system -> similar movies based on users ratings similarities

Objective:

Develop a recommendation system using ML techniques to recommend movies similar to a particular movie,

based on users ratings.

Methodology:

The feature set is **movies names** and all the available **users rating** for each movie.

Preprocessing:

- Remove noisy data

- remove the movie if a movie has less than 10 users ratings

- remove the user if a user voted on less than 50 movies

- Removing sparsity

creating a table with movies VS users ratings is a very sparse table with a lot of empty cells because users don't rate all the movies.

- Convert the matrix into a compressed sparse row matrix format which stores [(i, j), value] only

Model:

- kNN (k-Nearest-Neighbor)

The problem can be easily solved with a kNN model which will find the k nearest movies to all the movies.

Data Set Summary:

1- What is the data set used?

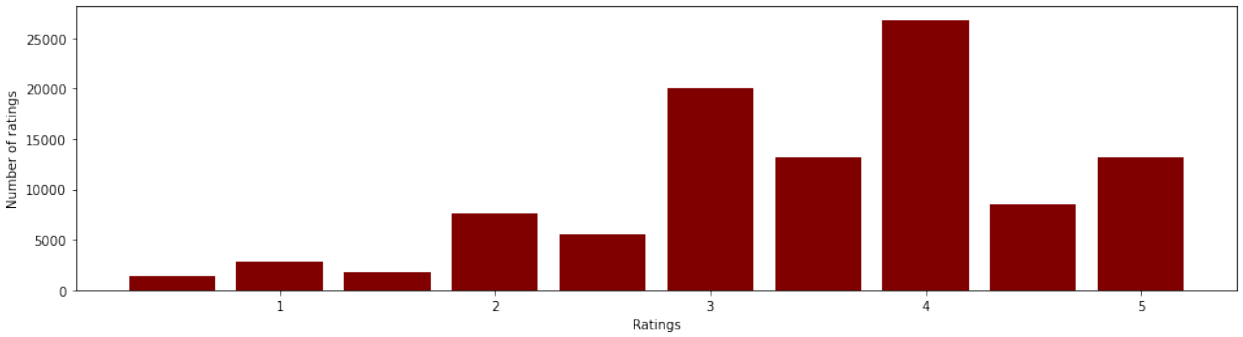
MovieLens-small dataset

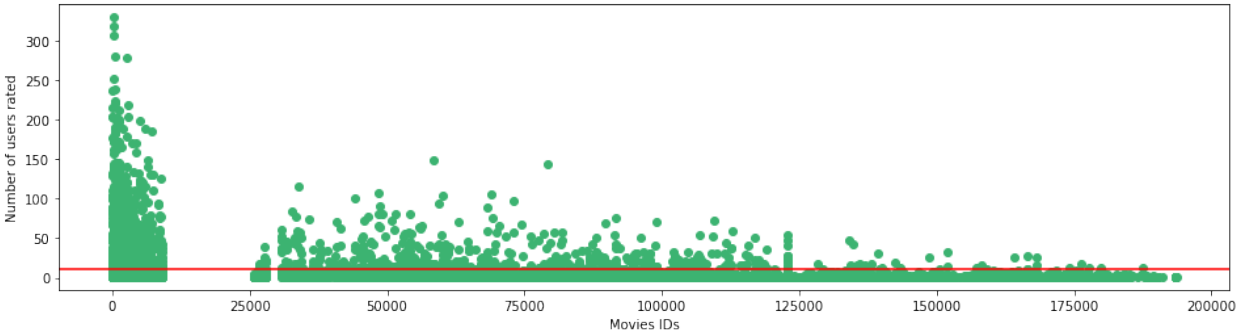
2- What is the summary of the dataset columns?

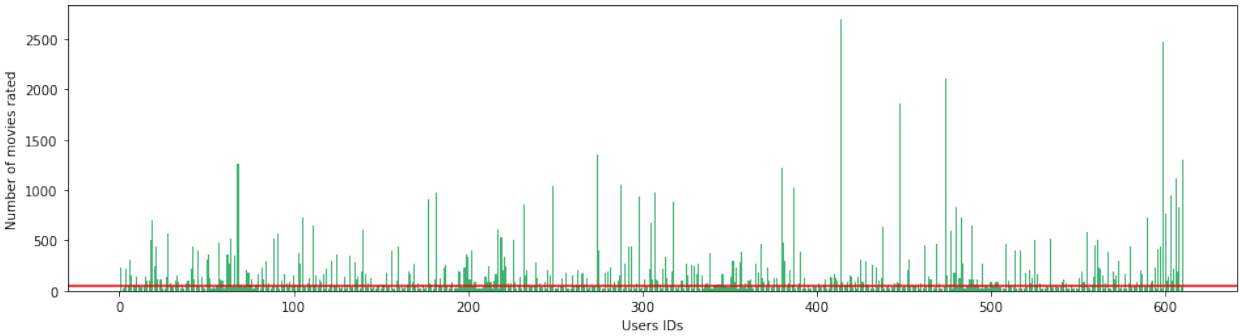
	movieId
count	9742.000000
mean	42200.353623
std	52160.494854
min	1.000000
25%	3248.250000
50%	7300.000000
75%	76232.000000
max	193609.000000

	userId	movieId	rating	timestamp
count	100836.000000	100836.000000	100836.000000	1.008360e+05
mean	326.127564	19435.295718	3.501557	1.205946e+09
std	182.618491	35530.987199	1.042529	2.162610e+08
min	1.000000	1.000000	0.500000	8.281246e+08
25%	177.000000	1199.000000	3.000000	1.019124e+09
50%	325.000000	2991.000000	3.500000	1.186087e+09
75%	477.000000	8122.000000	4.000000	1.435994e+09
max	610.000000	193609.000000	5.000000	1.537799e+09

3- Visualize the dataset statistics







Results:

Results of Input: “Batman”