

Movie Recommendation System

MACHINE LEARNING – PROJECT

Naga Nikhila Gonuguntla (U01005492)

Pranavi Adivi (U01000516)

Rachana Krishtipati (U01019754)

PROJECT REPORT

Objective:

The primary objective of this machine learning project is to create a recommendation engine that suggests movies to consumers. A recommendation system makes suggestions to users based on their interests and surfing history. The user's information is used as input. The information is derived from the input, which is in the form of surfing data. This data represents the product's previous usage as well as the awarded ratings.

What is a recommender system?

A recommender system, or a recommendation system (sometimes replacing 'system' with a synonym such as platform or engine), is a subclass of information filtering system that seeks to predict the "rating" or "preference" a user would give to an item. Recommender systems are used in a variety of areas, with commonly recognized examples taking the form of playlist generators for video and music services, product recommenders for online stores, or content recommenders for social media platforms and open web content recommenders.

Recommender systems usually make use of either or both collaborative filtering and content-based filtering (also known as the personality-based approach)

Dataset:

In this project, we are going to use the Movie Lens Dataset. Movie lens also has a website where you can sign up, contribute reviews, and get movie recommendations.

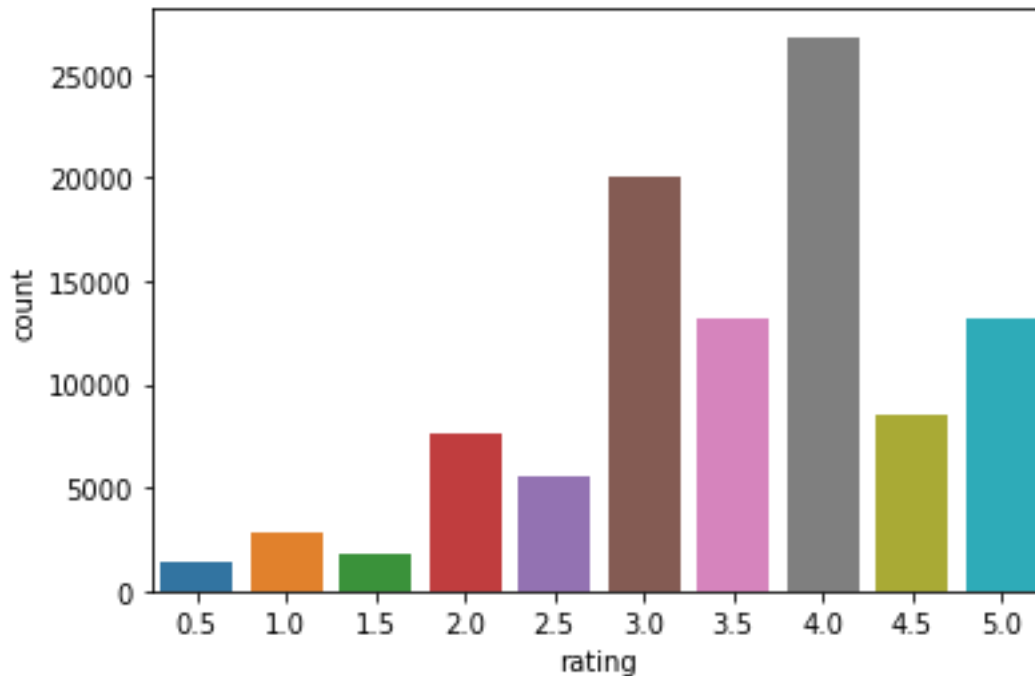
Let's look at what each column represents:

- userId - the ID of the user who rated the movie.
- itemId - the ID of the movie.
- rating - The rating the user gave the movie, between 1 and 5.
- timestamp - The time the movie was rated.
- title - The title of the movie.

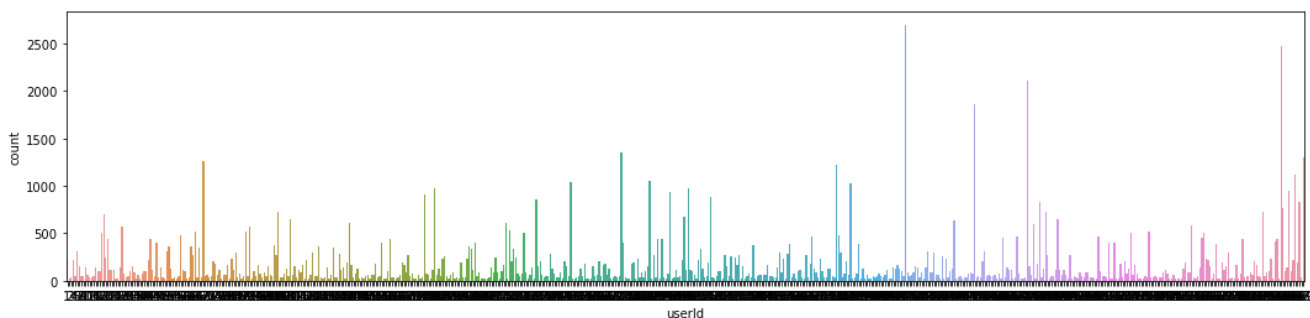
Analysis & Visualization:

The info() and describe() methods of the panda's data frame are used to display the information regarding features, rows, null count, datatypes, and all aggregate functions.

The below count plot shows how the ratings is distributed from rating 0 to rating 5. It shows that most of the ratings are between 2.5 and 5



The below count plot again shows that almost every user rated less than 500 movies except some outliers



We made one more data frame for user ratings to get the count of ratings, so that we can ignore least rated movies when we recommend movies, recommending movies based on data which has one or two ratings is not worth the recommendation.

Next, the highest and lowest rated movies are displayed in the jupyter notebook

Recommender System:

1. Let's now make a matrix with the user ids on one axis and the movie title on the other. Each cell will then include the user's rating for that film. Because most people haven't seen most of the movies, there will be a lot of NaN values
2. get user ratings for the given movie
3. now we correlate the above user ratings in the matrix to obtain similar movies
4. we are making sure the movie has at least 50 ratings to consider for the recommendation
5. sorting the data frame according to the correlation
6. The final output is the top 10 similar movies to the given input movie

Evaluation:

Evaluation of any Machine Learning project requires domain knowledge, here in Movie Recommendation System, we evaluate the project manually with the knowledge of domain, from the output in Movie Recommendation System we can conclude that our recommendation is pretty much working as the output genres are like the input movie genres. There is still a scope to improve the system, or we can go for other ML algorithms for better recommendation.

Resources:

https://en.wikipedia.org/wiki/Recommender_system
<https://grouplens.org/datasets/movielens/>