

MOVIE RECOMMENDATION SYSTEM

By:
Himanshu Baheti - 33106

INTRODUCTION

A movie recommendation system's fundamental idea is pretty straightforward. Every recommender system primarily consists of two components: users and items. Users receive movie predictions from the system, and the actual movies are the products.

Filtering and predicting only the movies that a matching user is most likely to wish to see is the main objective of movie recommendation systems. The user information from the system's database is used by the ML algorithms for these recommendation systems. Based on information from the past, this data is used to forecast the user in question's behavior in the future.

To assist users in finding the most pertinent films, movie recommendation systems employ a variety of filtration techniques and algorithms. The content-based filtering and collaborative filtering systems subcategories of the ML algorithms used for movie recommendations are the most well-liked ones:

- **Content Based Filtering:** content based filtering is a method of filtering movies in movie recommendation systems that makes advantage of the items' data (movies). This information, which is taken from just one user, is quite important in this case. This technique uses an ML algorithm to suggest movies that are comparable to the user's past choices.
- **Collaborative Filtering:** As the name implies, this filtering technique is based on the interactions between the relevant person and other users.

AGENDA

Agenda of Movie Recommendation System is to develop a efficient system that will look for stuff that may be of interest to a particular person. It will basically give results on string matching where the string will be the Movie Name another attributes for recommending the movie will be based on Actors involved in the movies and on the basis of genres like if the movie is action movie then the movie recommendation will be of action movies and related to that genre.

LITERATURE SURVEY

Content based filtering and collaborative filtering are the two main methods used.

- Content based filtering mechanism is that the recommendation engine examines the user's prior viewing habits before using this data to look for movies that are comparable. Database access is provided for this data (e.g., lead actors, director, genre, etc.). After that, the system offers the user movie suggestions. Having said that, the fundamental component of content-based filtering is the use of only one user's data to generate predictions.
- Collaborative filtering based on the history of every user in the database serves as the foundation of both this movie recommendation system and the machine learning algorithm. In essence, collaborative filtering relies on system users interacting with the objects (movies). This means that the results of this ML-based recommendation system are influenced by all users, as opposed to content-based filtering, which exclusively models using input from a single user.
- We use TF-IDF in our project for recommendation. TF-IDF is an abbreviation for Term Frequency Inverse Document Frequency. This is a very common algorithm to transform text into a meaningful representation of numbers which is used to fit a machine algorithm for prediction.
- We use cosine similarity for recommendation of top similar movies.

DATASET

- Dataset consists of 7,52,014 rows and 20 columns.
- The columns consists of basic features of movies like genre, description, actors, etc.
- There are missing values in some columns so as a part of data cleaning process I have replaced the values.

ML MODEL USED

- SnowballStemmer
- TfidfVectorizer
- cosine_similarity

```
In [22]: pd.set_option('display.max_columns', None) # code to display all columns

# Visualisation libraries
import matplotlib.pyplot as plt
import seaborn as sns

# Libraries for text processing
import nltk
from nltk.stem.snowball import SnowballStemmer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity

# to display images
from skimage import io

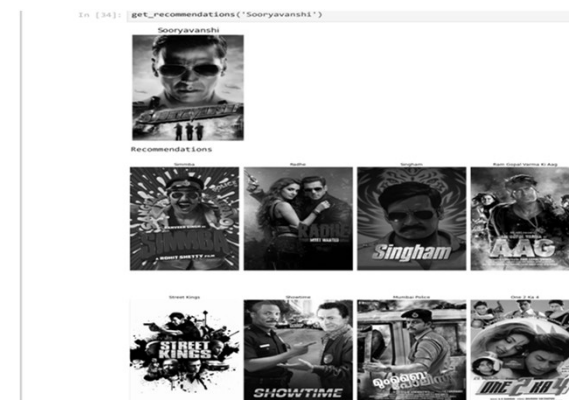
# to save the required files
import pickle

import warnings
warnings.filterwarnings('ignore') # To prevent kernel from showing any warning
```

WORKING

- Recommendation system consists of three types of filtering techniques in which the content based filtering and collaborative techniques are favored . In collaborative filtering there are two types of model based filtering and memory based filtering in which we will use memory based filtering consisting of user based and item based for recommending movies based on the input received.
- We use TF-IDF in our project for recommendation. TF-IDF is an abbreviation for Term Frequency Inverse Document Frequency. This is a very common algorithm to transform text into a meaningful representation of numbers which is used to fit a machine algorithm for prediction.
- We have used cosine similarity for recommendation of top similar movies.

SAMPLE REPRESENTATIVE SCREENSHOTS



CONCLUSION

In this project we have seen two different methods of recommender systems that is content based filtering technique and collaborative filtering technique for recommending movies to users. Currently users can get a maximum of eight movies based on the user's input movie. There is a set of seven lakh different movies with different genres so the user gets a wide variety of options.

Thank You