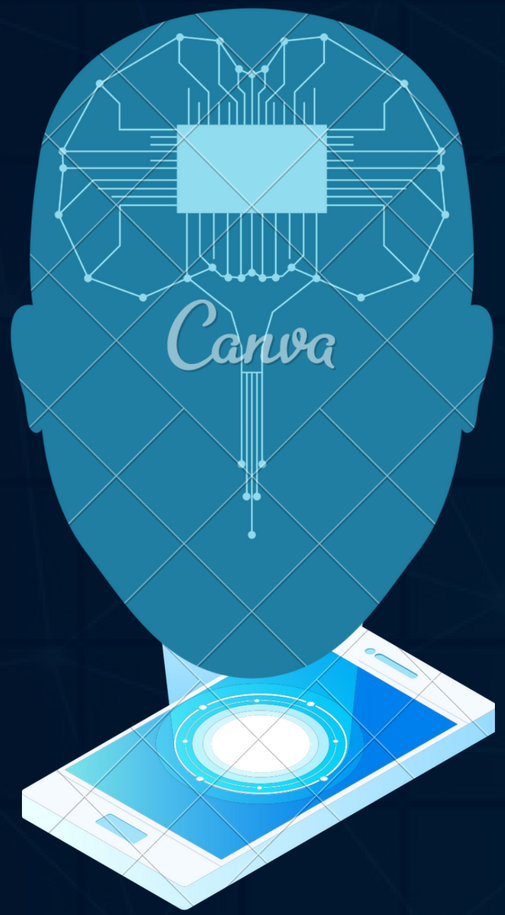


ARTIFICIAL INTELLIGENCE MINI-PROJECT

TOPIC :

CONTENT BASED MOVIE RECOMMENDATION SYSTEM

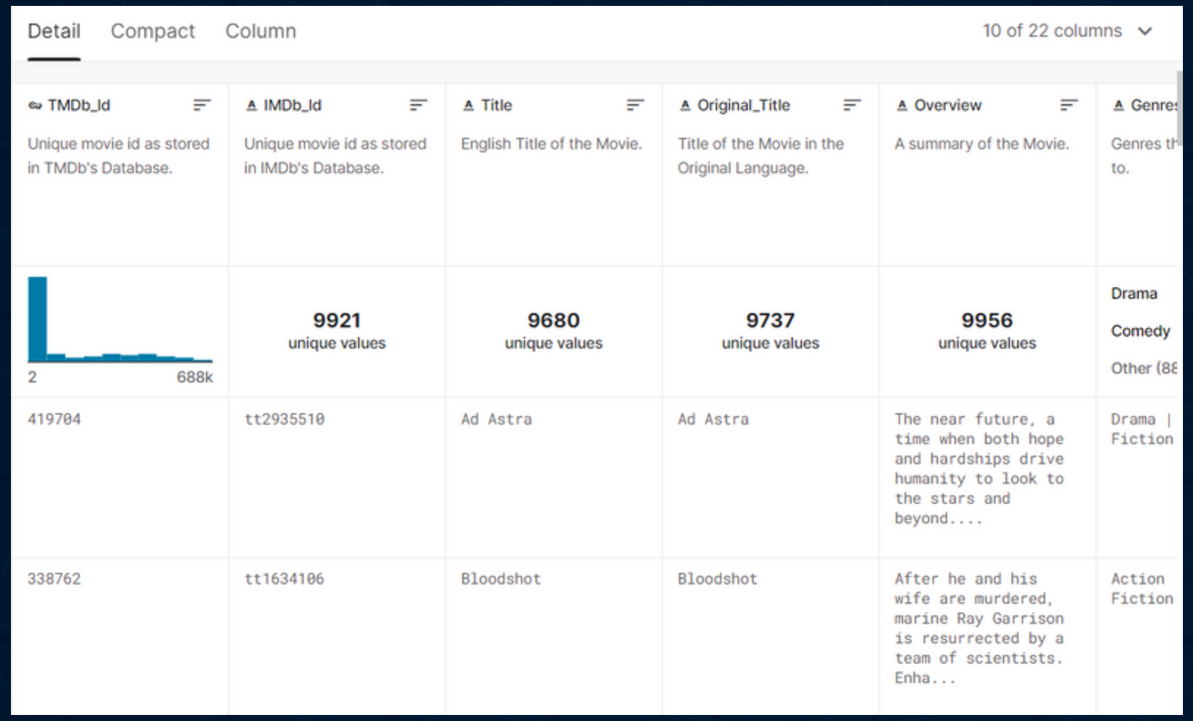
INTRODUCTION

Movie recommendation system is a collaborative filtering or content based recommendation technique where movie titles are recommended to users based on their choices.

Content Based Recommendation System: It uses attributes such as genre, director, description, actors,for movies, to make suggestions for the users. The intuition behind this sort of recommendation system is that if a user liked a particular movie or show, he/she might like a movie similar to it.

DATASET

For training this model we have TMDB 5000-movie dataset. This dataset contains the data of 5000 most Popular Movies in TMDbs's Database. It contains many columns such as movie_Id, crew details, budget, description, tags etc. Processing the dataset into a required manner we can make it easier for the model to recommend the movies, given a movie we give as input.



Detail | Compact | Column | 10 of 22 columns

TMDb_Id	IMDb_Id	Title	Original_Title	Overview	Genres
Unique movie id as stored in TMDb's Database.	Unique movie id as stored in IMDb's Database.	English Title of the Movie.	Title of the Movie in the Original Language.	A summary of the Movie.	Genres th to.
2 – 688k	9921 unique values	9680 unique values	9737 unique values	9956 unique values	Drama, Comedy, Other (88
419704	tt2935510	Ad Astra	Ad Astra	The near future, a time when both hope and hardships drive humanity to look to the stars and beyond....	Drama | Fiction
338762	tt1634106	Bloodshot	Bloodshot	After he and his wife are murdered, marine Ray Garrison is resurrected by a team of scientists. Enha...	Action Fiction

IMPLEMENTATION

PRE-PROCESSING THE DATA

- We have many attributes that aren't required for the recommendation of the movie
- So, we eliminate all the columns and keep only the required ones and trim the dataset
- Then we combine all the existing columns into tags and convert them into a single string
- We stem the string for duplicates and similar words (like go, goes, going) using the natural language processing library in python
- We eliminate primitive words (like is, are, they) using the library and now we have the string of required words

IMPLEMENTATION

MORE INTO IMPLEMENTATION

- In this model we combine some particular attributes such as genre,director,description, actors of all the movies in the dataset and make a separate attribute called tag.
- Each movie with its tag value is represented as a vector in the vector space and the Euclidean Distance is calculated among all theses vectors
- After that we create a matrix of n*n dimensions where n=number of movies in the dataset and it represents the similarity of one movies with another movie. Similarity is calculated as the inverse of the Euclidean Distance
- Now if the user searches for a particular movie the machine will first find the index of that particular movie in the dataset and then it will sort that particular row in descending order so that the movies with best similarity value is at front and then finally it will display a certain number of movies

WORKING AND RUNNING OF THE CODE

- We define a function recommend( ) which will take a movie as an input and recommend the list of similar movies using the avaiable vector of euclidian distances to each of the movies in the dataset

The recommend function

```
def recommend(movie):
    movie_index=new_dataset[new_dataset['title']== movie].index[0]
    node_distances=distances[movie_index]
    recommended_list=sorted(list(enumerate(node_distances)),key=lambda x:x[1])[0:10]
    temp=[];
    for i in recommended_list:
        temp.append(new_dataset.iloc[i[0]].title)
    return temp
```

TEST CASES

```
In [187]: recommend("Avatar")

Out[187]: ['Avatar',
           'The Helix... Loaded',
           'Khiladi 786',
           'Harrison Montgomery',
           "Amidst the Devil's Wings",
           'Sardaarji',
           'Falcon Rising',
           'Pink Narcissus',
           'Krull',
           'The Ten']

In [188]: recommend("The Dark Knight Rises")

Out[188]: ['The Dark Knight Rises',
           'Nighthawks',
           "Amidst the Devil's Wings",
           'Batman',
           'Exiled',
           'Mi America',
           'Nine Queens',
           'The Lords of Salem',
           'The Helix... Loaded',
           'Carlos']

In [189]: recommend("John Carter")

Out[189]: ['John Carter',
           'The Helix... Loaded',
           'Khiladi 786',
           'Harrison Montgomery',
           "Amidst the Devil's Wings",
           'Devil',
           'Krull',
           'The Loss of Sexual Innocence',
           'Sardaarji',
           'Son of God']
```

CONCLUSION

Content-based recommendation systems can be highly accurate and explainable ways to recommend items. They require less information from other users, and can utilize a wide range of input (genre, director name, keywords, tags, etc.).

Hence, we developed a model that will recommend us based on the similarity with other movies. This model can also used for other similar purposes, changing the dataset because we didn't use anything that specific to the movies, rather a generic model is developed.

THANKS!