MAJOR PROJECT 3 SANJAY ANAND V ¶

IMDB MOVIE RECOMMENDATION SYSTEM

1)Importing libraries and datasets

In [1]:

```
import numpy as np
import pandas as pd
```

In [2]:

```
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.metrics.pairwise import cosine_similarity
```

In [3]:

```
df=pd.read_csv("final.csv")
df.head()
```

Out[3]:

	name	year	duration	genre	rating	director	type
0	The Shawshank Redemption	1994	142	Drama	9.3	Frank Darabont	Movie
1	The Godfather	1972	175	Crime,Drama	9.2	Francis Ford Coppola	Movie
2	The Dark Knight	2008	152	Action,Crime,Drama	9.0	Christopher Nolan	Movie
3	The Godfather: Part II	1974	202	Crime,Drama	9.0	Francis Ford Coppola	Movie
4	12 Angry Men	1957	96	Crime,Drama	9.0	Sidney Lumet	Movie

2)User selecting a movie

In [4]:

```
movie=input("Select a movie:")
```

```
Select a movie:The Dark Knight
```

3)Director of movie user selected

In [5]:

```
l = len(df['name'])
directors=[]
for i in range(l):
    if(df.loc[i, 'name'] == movie):
        directors.append(df.loc[i, 'director'])
for n in directors:
    print("The director of the movie is",n)
```

```
The director of the movie is Christopher Nolan
```

4)Top 10 movies based on director of the selected movie

In [6]:

```
l = len(df['name'])
movies=[]
for j in directors:
    for i in range(l):
        if(df['director'][i]==j):
            movies.append(df['name'][i])
print("The movies recommended based on director:")
print("-------------------------------------------------------")
u=0
for k in movies:
    print(u+1,k)
    u=u+1
print("-------------------------------------------------------")
```

```
The movies recommended based on director:
-------------------------------------------------------
1 The Dark Knight
2 Inception
3 Interstellar
4 The Prestige
5 The Dark Knight Rises
6 Memento
7 Batman Begins
8 Dunkirk
-------------------------------------------------------
```

5)Top 10 movies based on name of the movie using one hot encoding

In [7]:

```
count = CountVectorizer()
one_hot = count.fit_transform(df['name'])
similar = cosine_similarity(one_hot)
idx = df[df['name'] == movie].index[0]
similar_movies = list(enumerate(similar[idx]))
sorted_similar_movies = sorted(similar_movies, key=lambda x:x[1], reverse=True)[1:11]
recommendations = [df.iloc[i[0]]['name'] for i in sorted_similar_movies]

print("The movies recommended based on movie similarity:")
print("-------------------------------------------------------")
for i, movie in enumerate(recommendations):
    print(i+1, movie)
print("-------------------------------------------------------")
```

```
The movies recommended based on movie similarity:
-------------------------------------------------------
1 The Dark Knight Rises
2 Dancer in the Dark
3 Dark
4 The Lord of the Rings: The Return of the King
5 The Lord of the Rings: The Fellowship of the Ring
6 The Lord of the Rings: The Two Towers
7 The Silence of the Lambs
8 The Shop Around the Corner
9 The Night of the Hunter
10 The Trial of the Chicago 7
-------------------------------------------------------
```