DSBDA MiniProject

April 13, 2023

[4]:
```
print("TEC 06 Shreyas Bhaginath Peherkar")
```

```
TEC 06 Shreyas Bhaginath Peherkar
```

[5]:
```
import pandas as pd
import numpy as np
```

[6]:
```
from sklearn.feature_extraction.text import CountVectorizer
```

[7]:
```
from sklearn.metrics.pairwise import cosine_similarity
```

[8]:
```
df = pd.read_csv("/Users/shreyaspeherkar/Desktop/Dataset/movie_dataset.csv")
```

[9]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4803 entries, 0 to 4802
Data columns (total 24 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   index                 4803 non-null   int64  
 1   budget                4803 non-null   int64  
 2   genres                4775 non-null   object 
 3   homepage              1712 non-null   object 
 4   id                    4803 non-null   int64  
 5   keywords              4391 non-null   object 
 6   original_language     4803 non-null   object 
 7   original_title        4803 non-null   object 
 8   overview              4800 non-null   object 
 9   popularity            4803 non-null   float64
 10  production_companies  4803 non-null   object 
 11  production_countries  4803 non-null   object 
 12  release_date          4802 non-null   object 
 13  revenue               4803 non-null   int64  
 14  runtime               4801 non-null   float64
 15  spoken_languages      4803 non-null   object 
 16  status                4803 non-null   object 
 17  tagline               3959 non-null   object 
```

```
 18  title                 4803 non-null   object
 19  vote_average          4803 non-null   float64
 20  vote_count            4803 non-null   int64
 21  cast                  4760 non-null   object
 22  crew                  4803 non-null   object
 23  director              4773 non-null   object
dtypes: float64(3), int64(5), object(16)
memory usage: 900.7+ KB
```

[10]:
```
df.head(10)
```

[10]:

	index	budget	genres	homepage	id	keywords	original_language
0	0	237000000	Action Adventure Fantasy Science Fiction	http://www.avatarmovie.com/	19995	culture clash future space war space colony so…	en
1	1	300000000	Adventure Fantasy Action	http://disney.go.com/disneypictures/pirates/	285	ocean drug abuse exotic island east india trad…	en
2	2	245000000	Action Adventure Crime	http://www.sonypictures.com/movies/spectre/	206647	spy based on novel secret agent sequel mi6	en
3	3	250000000	Action Crime Drama Thriller	http://www.thedarkknightrises.com/	49026	dc comics crime fighter terrorist secret ident…	en
4	4	260000000	Action Adventure Science Fiction	http://movies.disney.com/john-carter	49529	based on novel mars medallion space travel pri…	en
5	5	258000000	Fantasy Action Adventure	http://www.sonypictures.com/movies/spider-man3/	559	dual identity amnesia sandstorm love of one's …	en
6	6	260000000	Animation Family	http://disney.go.com/disneypictures/tangled/	38757	hostage magic horse fairy tale musical	en
7	7	280000000	Action Adventure Science Fiction	http://marvel.com/movies/movie/193/avengers_ag…	99861	marvel comic sequel superhero based on comic b…	en
8	8	250000000	Adventure Fantasy Family	http://harrypotter.warnerbros.com/harrypottera…	767	witch magic broom school of witchcraft wizardry	en
9	9	250000000	Action Adventure Fantasy	http://www.batmanvsupermandawnofjustice.com/	209112	dc comics vigilante superhero based on comic b…	en

```
                            original_title  \
0                                   Avatar
1  Pirates of the Caribbean: At World's End
2                                  Spectre
3                    The Dark Knight Rises
4                              John Carter
5                             Spider-Man 3
6                                  Tangled
7                  Avengers: Age of Ultron
8   Harry Potter and the Half-Blood Prince
9       Batman v Superman: Dawn of Justice

                                          overview  popularity  … runtime  \
0  In the 22nd century, a paraplegic Marine is di…  150.437577  …   162.0
1  Captain Barbossa, long believed to be dead, ha…  139.082615  …   169.0
2  A cryptic message from Bond’s past sends him o…  107.376788  …   148.0
3  Following the death of District Attorney Harve…  112.312950  …   165.0
4  John Carter is a war-weary, former military ca…   43.926995  …   132.0
5  The seemingly invincible Spider-Man goes up ag…  115.699814  …   139.0
6  When the kingdom's most wanted-and most charmi…   48.681969  …   100.0
7  When Tony Stark tries to jumpstart a dormant p…  134.279229  …   141.0
8  As Harry begins his sixth year at Hogwarts, he…   98.885637  …   153.0
9  Fearing the actions of a god-like Super Hero l…  155.790452  …   151.0

                                    spoken_languages    status  \
0  [{"iso_639_1": "en", "name": "English"}, {"iso…  Released
1           [{"iso_639_1": "en", "name": "English"}]  Released
2  [{"iso_639_1": "fr", "name": "Fran\u00e7ais"},…  Released
3           [{"iso_639_1": "en", "name": "English"}]  Released
4           [{"iso_639_1": "en", "name": "English"}]  Released
5  [{"iso_639_1": "en", "name": "English"}, {"iso…  Released
6           [{"iso_639_1": "en", "name": "English"}]  Released
7           [{"iso_639_1": "en", "name": "English"}]  Released
8           [{"iso_639_1": "en", "name": "English"}]  Released
9           [{"iso_639_1": "en", "name": "English"}]  Released

                                        tagline  \
0                    Enter the World of Pandora.
1  At the end of the world, the adventure begins.
2                          A Plan No One Escapes
3                               The Legend Ends
4           Lost in our world, found in another.
5                             The battle within.
6       They're taking adventure to new lengths.
7                           A New Age Has Come.
8                          Dark Secrets Revealed
9                             Justice or revenge
```

```
                                      title vote_average vote_count  \
0                                    Avatar          7.2      11800
1  Pirates of the Caribbean: At World's End          6.9       4500
2                                   Spectre          6.3       4466
3                     The Dark Knight Rises          7.6       9106
4                                  John Carter          6.1       2124
5                                Spider-Man 3          5.9       3576
6                                     Tangled          7.4       3330
7                   Avengers: Age of Ultron          7.3       6767
8    Harry Potter and the Half-Blood Prince          7.4       5293
9        Batman v Superman: Dans of Justice          5.7       7004

                                                cast  \
0  Sam Worthington Zoe Saldana Sigourney Weaver S…
1  Johnny Depp Orlando Bloom Keira Knightley Stel…
2  Daniel Craig Christoph Waltz L\u00e9a Seydoux …
3  Christian Bale Michael Caine Gary Oldman Anne …
4  Taylor Kitsch Lynn Collins Samantha Morton Wil…
5  Tobey Maguire Kirsten Dunst James Franco Thoma…
6  Zachary Levi Mandy Moore Donna Murphy Ron Perl…
7  Robert Downey Jr. Chris Hemsworth Mark Ruffalo…
8  Daniel Radcliffe Rupert Grint Emma Watson Tom …
9  Ben Affleck Henry Cavill Gal Gadot Amy Adams J…

                                                crew           director
0  [{'name': 'Stephen E. Rivkin', 'gender': 0, 'd…      James Cameron
1  [{'name': 'Dariusz Wolski', 'gender': 2, 'depa…     Gore Verbinski
2  [{'name': 'Thomas Newman', 'gender': 2, 'depar…         Sam Mendes
3  [{'name': 'Hans Zimmer', 'gender': 2, 'departm…  Christopher Nolan
4  [{'name': 'Andrew Stanton', 'gender': 2, 'depa…     Andrew Stanton
5  [{'name': 'Francine Maisler', 'gender': 1, 'de…          Sam Raimi
6  [{'name': 'John Lasseter', 'gender': 2, 'depar…       Byron Howard
7  [{'name': 'Danny Elfman', 'gender': 2, 'depart…        Joss Whedon
8  [{'name': 'Bruno Delbonnel', 'gender': 0, 'dep…        David Yates
9  [{'name': 'Hans Zimmer', 'gender': 2, 'departm…        Zack Snyder

[10 rows x 24 columns]
```

[11]:
```
#now here these dataset features are to much for only selected features we are␣
↪usign -
```

[12]:
```
features = ['keywords', 'cast', 'genres', 'director']
for feature in features:
    df[feature] = df[feature].fillna('')
```

[13]:
```
def combined_features(row):
    return row['keywords']+" "+row['cast']+" "+row['genres']+" "+row['director']
df["combined_features"] = df.apply(combined_features, axis =1)
```

[14]:
```
cv = CountVectorizer()
count_matrix = cv.fit_transform(df["combined_features"])
print("Count Matrix:", count_matrix.toarray())
```

```
Count Matrix: [[0 0 0 … 0 0 0]
 [0 0 0 … 0 0 0]
 [0 0 0 … 0 0 0]
 …
 [0 0 0 … 0 0 0]
 [0 0 0 … 0 0 0]
 [0 0 0 … 0 0 0]]
```

[15]:
```
cosine_sim = cosine_similarity(count_matrix)
```

[16]:
```
movie_user_likes = "Dead Poets Society"
def get_index_from_title(title):
    return df[df.title == title]["index"].values[0]
movie_index = get_index_from_title(movie_user_likes)
```

[17]:
```
similar_movies = list(enumerate(cosine_sim[movie_index]))
```

[18]:
```
features = ['keywords','cast','genres','director']
```

[19]:
```
def combine_features(row):return row['keywords'] +" "+row['cast']+"␣
 ↪"+row["genres"]+" "+row["director"]
```

[20]:
```
for feature in features:
    df[feature] = df[feature].fillna('')
    df["combined_features"] = df.apply(combine_features,axis=1)
```

[21]:
```
cv = CountVectorizer()
```

[22]:
```
count_matrix = cv.fit_transform(df["combined_features"])
cosine_sim = cosine_similarity(count_matrix)
```

[23]:
```
def get_title_from_index(index):
    return df[df.index == index]["title"].values[0]
```

[24]:
```
def get_index_from_title(title):
    return df[df.title == title]["index"].values[0]
```

[25]:
```
movie_user_likes = "Avatar"
movie_index = get_index_from_title(movie_user_likes)
```

```
similar_movies =  list(enumerate(cosine_sim[movie_index]))
```

[26]:
```
sorted_similar_movies = sorted(similar_movies,key=lambda x:x[1],reverse=True)[1:
↪]
```

[28]:
```
i=0
print("Top 5 similar movies to "+movie_user_likes+" are:\n")
for element in sorted_similar_movies:
    print(get_title_from_index(element[0]))
    i=i+1
    if i>=5:
        break
```

```
Top 5 similar movies to Avatar are:

Guardians of the Galaxy
Aliens
Star Wars: Clone Wars: Volume 1
Star Trek Into Darkness
Star Trek Beyond
```

[29]:
```
#another try for another movie name like
```

[30]:
```
movie_user_likes = "Spider-Man"
movie_index = get_index_from_title(movie_user_likes)
similar_movies =  list(enumerate(cosine_sim[movie_index]))
```

[31]:
```
sorted_similar_movies = sorted(similar_movies,key=lambda x:x[1],reverse=True)[1:
↪]
```

[33]:
```
i=0
print("Top 5 Similar movies to"+movie_user_likes+" are:\n")
for element in sorted_similar_movies:
    print(get_title_from_index(element[0]))
    i=i+1
    if i>=5:
        break
```

```
Top 5 Similar movies toSpider-Man are:

Spider-Man 3
Spider-Man 2
Highlander: Endgame
The Count of Monte Cristo
Cold Mountain
```

[38]:
```
#another movie name for checking in the datset is it avaible or not.
```

[35]:
```
movie_user_likes = "Avengers: Age of Ultron"
movie_index = get_index_from_title(movie_user_likes)
similar_movies =  list(enumerate(cosine_sim[movie_index]))
```

[36]:
```
sorted_similar_movies = sorted(similar_movies,key=lambda x:x[1],reverse=True)[1:
↪]
```

[37]:
```
i=0
print("Top 5 Similar movies to"+movie_user_likes+" are:\n")
for element in sorted_similar_movies:
    print(get_title_from_index(element[0]))
    i=i+1
    if i>=5:
        break
```

```
Top 5 Similar movies toAvengers: Age of Ultron are:

The Avengers
Iron Man 2
Captain America: Civil War
Captain America: The Winter Soldier
Thor: The Dark World
```