Movie Recommendation System using Machine Learning

August 8, 2023

[1]:

```
import numpy as np
import pandas as pd
import difflib
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity
```

[3]:

```
data=pd.read_csv('movies.csv')
```

[4]:

```
data.head()
```

[4]:

```
   index     budget                                    genres  \
0      0  237000000  Action Adventure Fantasy Science Fiction
1      1  300000000                  Adventure Fantasy Action
2      2  245000000                    Action Adventure Crime
3      3  250000000               Action Crime Drama Thriller
4      4  260000000          Action Adventure Science Fiction

                                      homepage      id  \
0                  http://www.avatarmovie.com/   19995
1 http://disney.go.com/disneypictures/pirates/     285
2  http://www.sonypictures.com/movies/spectre/  206647
3           http://www.thedarkknightrises.com/   49026
4         http://movies.disney.com/john-carter   49529

                                            keywords original_language  \
0  culture clash future space war space colony so…                en
1  ocean drug abuse exotic island east india trad…                en
2         spy based on novel secret agent sequel mi6                en
3  dc comics crime fighter terrorist secret ident…                en
4  based on novel mars medallion space travel pri…                en

                             original_title  \
0                                    Avatar
1  Pirates of the Caribbean: At World's End
2                                   Spectre
3                     The Dark Knight Rises
4                               John Carter
```

```
                                            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

                                    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

                                         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.

                                      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

                                              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…

                                              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 rows x 24 columns]
```

[5]:
```
data.shape
```

[5]: (4803, 24)

[6]:
```
#selecting the relevant features for recommendation
```

[12]:
```
selected_features=data[['genres','keywords','tagline','cast','director']]
```

[13]:
```
selected_features.head()
```

[13]:
```
                                      genres  \
0  Action Adventure Fantasy Science Fiction
1                   Adventure Fantasy Action
2                     Action Adventure Crime
3                Action Crime Drama Thriller
4           Action Adventure Science Fiction

                                             keywords  \
0  culture clash future space war space colony so…
1  ocean drug abuse exotic island east india trad…
2         spy based on novel secret agent sequel mi6
3  dc comics crime fighter terrorist secret ident…
4  based on novel mars medallion space travel pri…

                                         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.

                                                cast            director
0  Sam Worthington Zoe Saldana Sigourney Weaver S…       James Cameron
1  Johnny Depp Orlando Bloom Keira Knightley Stel…      Gore Verbinski
2  Daniel Craig Christoph Waltz L\u00e9a Seydoux …          Sam Mendes
3  Christian Bale Michael Caine Gary Oldman Anne …   Christopher Nolan
4  Taylor Kitsch Lynn Collins Samantha Morton Wil…      Andrew Stanton
```

[8]:
```
#replacing the null values with null string
```

[28]:
```
selected_features.isnull().sum()
```

[28]:
```
genres       28
keywords    412
tagline     844
cast         43
director     30
dtype: int64
```

[29]:
```
selected_features=selected_features.fillna(" ")
```

[32]:
```
selected_features.isnull().sum()
```

[32]:
```
genres      0
keywords    0
tagline     0
cast        0
director    0
dtype: int64
```

[33]:
```
#combined all textual features
```

[34]:
```
combined_features=selected_features['genres']+ ' '+␣
↪selected_features['keywords']+ ' ' +selected_features['tagline']+'␣
↪'+selected_features['cast']+' '+selected_features['director']
```

[35]:
```
combined_features
```

[35]:
```
0       Action Adventure Fantasy Science Fiction cultu…
1       Adventure Fantasy Action ocean drug abuse exot…
2       Action Adventure Crime spy based on novel secr…
3       Action Crime Drama Thriller dc comics crime fi…
4       Action Adventure Science Fiction based on nove…
                              …
4798    Action Crime Thriller united states\u2013mexic…
4799    Comedy Romance   A newlywed couple's honeymoon…
4800    Comedy Drama Romance TV Movie date love at fir…
4801        A New Yorker in Shanghai Daniel Henney Eli…
4802    Documentary obsession camcorder crush dream gi…
Length: 4803, dtype: object
```

[36]:
```
#converting the text data to feature vectors
```

[37]:
```
vectorizer=TfidfVectorizer()
```

[38]:
```
vectorized_features=vectorizer.fit_transform(combined_features)
```

[95]:
```
print(vectorized_features)
```

```
  (0, 2432)     0.17272411194153
  (0, 7755)     0.1128035714854756
  (0, 13024)    0.1942362060108871
  (0, 10229)    0.16058685400095302
  (0, 8756)     0.22709015857011816
  (0, 14608)    0.15150672398763912
  (0, 16668)    0.19843263965100372
  (0, 14064)    0.20596090415084142
  (0, 13319)    0.2177470539412484
  (0, 17290)    0.20197912553916567
  (0, 17007)    0.23643326319898797
  (0, 13349)    0.15021264094167086
```

```
(0, 11503)    0.27211310056983656
(0, 11192)    0.09049319826481456
(0, 16998)    0.1282126322850579
(0, 15261)    0.07095833561276566
(0, 4945)     0.24025852494110758
(0, 14271)    0.21392179219912877
(0, 3225)     0.24960162956997736
(0, 16587)    0.12549432354918996
(0, 14378)    0.33962752210959823
(0, 5836)     0.1646750903586285
(0, 3065)     0.22208377802661425
(0, 3678)     0.21392179219912877
(0, 5437)     0.1036413987316636
:      :
(4801, 17266) 0.2886098184932947
(4801, 4835)  0.24713765026963996
(4801, 403)   0.17727585190343226
(4801, 6935)  0.2886098184932947
(4801, 11663) 0.21557500762727902
(4801, 1672)  0.1564793427630879
(4801, 10929) 0.13504166990041588
(4801, 7474)  0.11307961713172225
(4801, 3796)  0.3342808988877418
(4802, 6996)  0.5700048226105303
(4802, 5367)  0.22969114490410403
(4802, 3654)  0.262512960498006
(4802, 2425)  0.24002350969074696
(4802, 4608)  0.24002350969074696
(4802, 6417)  0.21753405888348784
(4802, 4371)  0.1538239182675544
(4802, 12989) 0.1696476532191718
(4802, 1316)  0.1960747079005741
(4802, 4528)  0.19504460807622875
(4802, 3436)  0.21753405888348784
(4802, 6155)  0.18056463596934083
(4802, 4980)  0.16078053641367315
(4802, 2129)  0.3099656128577656
(4802, 4518)  0.16784466610624255
(4802, 11161) 0.17867407682173203
```

1 Cosine Similarity

[41]:
```
#getting the similarity scores
```

[43]:
```
cosine_similarity=cosine_similarity(vectorized_features)
```

[44]:
```
cosine_similarity
```

[44]:
```
array([[1.        , 0.07219487, 0.037733  , …, 0.        , 0.        ,
        0.        ],
       [0.07219487, 1.        , 0.03281499, …, 0.03575545, 0.        ,
        0.        ],
       [0.037733  , 0.03281499, 1.        , …, 0.        , 0.05389661,
        0.        ],
       …,
       [0.        , 0.03575545, 0.        , …, 1.        , 0.        ,
        0.02651502],
       [0.        , 0.        , 0.05389661, …, 0.        , 1.        ,
        0.        ],
       [0.        , 0.        , 0.        , …, 0.02651502, 0.        ,
        1.        ]])
```

[45]:
```
cosine_similarity.shape
```

[45]:
```
(4803, 4803)
```

2 Getting the movie name from the user

[51]:
```
movie_name=input('Enter your favorite movie  :')
```

```
Enter your favorite movie  :bat man
```

[52]:
```
#creating a list with all name of movies in the dataset
```

[55]:
```
movie_list=data['title'].tolist()
```

[102]:
```
movie_list[0:50]
```

[102]:
```
['Avatar',
 "Pirates of the Caribbean: At World's End",
 'Spectre',
 'The Dark Knight Rises',
 'John Carter',
 'Spider-Man 3',
 'Tangled',
 'Avengers: Age of Ultron',
 'Harry Potter and the Half-Blood Prince',
 'Batman v Superman: Dawn of Justice',
 'Superman Returns',
 'Quantum of Solace',
 "Pirates of the Caribbean: Dead Man's Chest",
 'The Lone Ranger',
 'Man of Steel',
 'The Chronicles of Narnia: Prince Caspian',
 'The Avengers',
```

```
 'Pirates of the Caribbean: On Stranger Tides',
 'Men in Black 3',
 'The Hobbit: The Battle of the Five Armies',
 'The Amazing Spider-Man',
 'Robin Hood',
 'The Hobbit: The Desolation of Smaug',
 'The Golden Compass',
 'King Kong',
 'Titanic',
 'Captain America: Civil War',
 'Battleship',
 'Jurassic World',
 'Skyfall',
 'Spider-Man 2',
 'Iron Man 3',
 'Alice in Wonderland',
 'X-Men: The Last Stand',
 'Monsters University',
 'Transformers: Revenge of the Fallen',
 'Transformers: Age of Extinction',
 'Oz: The Great and Powerful',
 'The Amazing Spider-Man 2',
 'TRON: Legacy',
 'Cars 2',
 'Green Lantern',
 'Toy Story 3',
 'Terminator Salvation',
 'Furious 7',
 'World War Z',
 'X-Men: Days of Future Past',
 'Star Trek Into Darkness',
 'Jack the Giant Slayer',
 'The Great Gatsby']
```

3 finding the close match for the movie name given by user

[59]:
```
find_close_match=difflib.get_close_matches(movie_name,movie_list)
```

[60]:
```
print(find_close_match)
```

```
['Batman', 'Batman', 'Catwoman']
```

[62]:
```
close_match=find_close_match[0]
```

[63]:
```
close_match
```

[63]:
```
'Batman'
```

[64]:
```
# finding the index of the movie with title
```

[65]:
```
index_of_the_movie=data[data.title == close_match]['index'].values[0]
```

[66]:
```
index_of_the_movie
```

[66]: 1359

[67]:
```
#getting list of similar movies
```

[75]:
```
similarity_score=list(enumerate(cosine_similarity[index_of_the_movie]))
```

[101]:
```
similarity_score[0:50]
```

[101]:
```
[(0, 0.06605757327786128),
 (1, 0.05494299286841542),
 (2, 0.022474655275199845),
 (3, 0.18229768118495995),
 (4, 0.04826852550842614),
 (5, 0.031153020937531885),
 (6, 0.007816453213423849),
 (7, 0.03428924698332946),
 (8, 0.020070942434549842),
 (9, 0.10380582477831074),
 (10, 0.20133128757695565),
 (11, 0.01951398074561013),
 (12, 0.025827307032424),
 (13, 0.016928223234223393),
 (14, 0.3188610536407353),
 (15, 0.018553049692010756),
 (16, 0.03251349500357992),
 (17, 0.02470681749845619),
 (18, 0.029643169623505952),
 (19, 0.04417369541584225),
 (20, 0.029947941232879607),
 (21, 0.011401180198063948),
 (22, 0.03639562527544031),
 (23, 0.025066323882186783),
 (24, 0.03654732000312345),
 (25, 0.0),
 (26, 0.03538948489610589),
 (27, 0.03529828509945782),
 (28, 0.038753351499167796),
 (29, 0.0142136220030666),
 (30, 0.025493376585254864),
 (31, 0.04277585721181915),
 (32, 0.03875437083506119),
```

```
 (33, 0.03407619761223045),
 (34, 0.0),
 (35, 0.035771481945846714),
 (36, 0.03223295986231689),
 (37, 0.020432362116956287),
 (38, 0.025815233384302975),
 (39, 0.036492247430502514),
 (40, 0.013244821920171627),
 (41, 0.10975905674983476),
 (42, 0.10149944282849166),
 (43, 0.08576604259167658),
 (44, 0.006319182166201161),
 (45, 0.06462598213154251),
 (46, 0.05309679823241051),
 (47, 0.03761158537170053),
 (48, 0.025917949138402414),
 (49, 0.003998237151396815)]
```

[77]:
```
#sorting the similarity_score based on their similarity score
```

[78]:
```
sorted_similarity_score=sorted(similarity_score,key=lambda x:x[1],reverse=True)
```

[103]:
```
sorted_similarity_score[0:50]
```

[103]:
```
[(813, 1.0),
 (870, 0.4582408553739051),
 (2433, 0.3864273245926594),
 (14, 0.3188610536407353),
 (1296, 0.2529397869699989),
 (823, 0.21578853909502327),
 (10, 0.20133128757695565),
 (428, 0.18596641221431343),
 (72, 0.18512906308616195),
 (2793, 0.18272642077822948),
 (3, 0.18229768118495995),
 (1420, 0.1727129308212663),
 (119, 0.17180650719081145),
 (65, 0.16160582880593957),
 (3337, 0.1515876731444143),
 (4401, 0.14678890114385917),
 (4267, 0.14520436352015406),
 (1359, 0.14414128303962467),
 (210, 0.14402535832262298),
 (1238, 0.1306893940867623),
 (1282, 0.12714887083401077),
 (587, 0.1257927430818875),
 (2492, 0.1253438322333649),
```

```
 (164, 0.12497636481940003),
 (1024, 0.12419846557095111),
 (4432, 0.12366856042396873),
 (1510, 0.12144757299512393),
 (1959, 0.11861641964942257),
 (1890, 0.11744412178112948),
 (955, 0.11597960277260794),
 (1183, 0.1138043381601171),
 (1740, 0.11327489168242302),
 (634, 0.11273647980632893),
 (613, 0.11245293131098388),
 (473, 0.11152736228262713),
 (3552, 0.11081484589361365),
 (41, 0.10975905674983476),
 (505, 0.10966848766340348),
 (1188, 0.10936090976553697),
 (299, 0.10881877855976649),
 (1477, 0.1080855817629384),
 (149, 0.10618963102178569),
 (1525, 0.10397516619953215),
 (9, 0.10380582477831074),
 (1358, 0.10378325738223843),
 (969, 0.10369857164487108),
 (178, 0.10161128779663497),
 (42, 0.10149944282849166),
 (1035, 0.09925170538152248),
 (93, 0.09836617663461664)]
```

[80]:

```
#print the name of similar movie
```

[83]:

```
i=1
for movie in sorted_similarity_score:
    index=movie[0]
    tittle_of_the_movie=data[data.index == index]['title'].values[0]
    if (i<20):
        print(i,'.',tittle_of_the_movie)
        i=i+1
```

```
1 . Batman
2 . Batman Returns
3 . Batman & Robin
4 . The Dark Knight Rises
5 . Batman Begins
6 . The Dark Knight
7 . A History of Violence
8 . Superman
9 . Beetlejuice
```

```
10 . Bedazzled
11 . Mars Attacks!
12 . The Sentinel
13 . Planet of the Apes
14 . Man of Steel
15 . Suicide Squad
16 . The Mask
17 . Salton Sea
18 . Spider-Man 3
19 . The Postman Always Rings Twice
```

4 Movie Recommendation System

[105]:
```python
movie_name=input('Enter your favorite movie  :')
find_close_match=difflib.get_close_matches(movie_name,movie_list)
close_match=find_close_match[0]
index_of_the_movie=data[data.title == close_match]['index'].values[0]
similarity_score=list(enumerate(cosine_similarity[index_of_the_movie]))
sorted_similarity_score=sorted(similarity_score,key=lambda x:x[1],reverse=True)
i=1
for movie in sorted_similarity_score:
    index=movie[0]
    tittle_of_the_movie=data[data.index == index]['title'].values[0]
    if (i<20):
        print(i,'.',tittle_of_the_movie)
        i=i+1
```

```
Enter your favorite movie  :spider man
1 . Spider-Man
2 . Spider-Man 3
3 . Spider-Man 2
4 . The Notebook
5 . Seabiscuit
6 . Clerks II
7 . The Ice Storm
8 . Oz: The Great and Powerful
9 . Horrible Bosses
10 . The Count of Monte Cristo
11 . In Good Company
12 . Finding Nemo
13 . Clear and Present Danger
14 . Brothers
15 . The Good German
16 . Drag Me to Hell
17 . Bambi
18 . The Queen
19 . Charly
```

[ ]:

```

```