

```
In [1]: #Recommender system
#1.Content based(tags)
#2.Collaborative filtering(User based system)
#3.hybrid
import numpy as np
import pandas as pd
```

C:\Users\ASUS\AppData\Roaming\Python\Python311\site-packages\pandas\core\arrays\maskd.py:60: UserWarning: Pandas requires version '1.3.6' or newer of 'bottleneck' (version '1.3.5' currently installed).
from pandas.core import (

```
In [2]: movies = pd.read_csv('tmdb_5000_movies.csv')
credits = pd.read_csv('tmdb_5000_credits.csv')
```

```
In [3]: movies.head(1)
```

	budget	genres	homepage	id	keywords	original_language	original_title
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "name..."}]	http://www.avatarmovie.com/	19995	{"id": 1463, "name": "culture clash"}, {"id": ...}	en	Avatar

```
In [4]: credits.head(1) #[ 'crew'].values
```

	movie_id	title	cast	crew
0	19995	Avatar	[{"cast_id": 242, "character": "Jake Sully", "..."}]	[{"credit_id": "52fe48009251416c750aca23", "de..."}]

```
In [5]: #merged two dataset
movies = movies.merge(credits, on='title')
```

```
In [6]: movies.head(1)
```

	budget	genres	homepage	id	keywords	original_language	original_title
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "name..."}]	http://www.avatarmovie.com/	19995	{"id": 1463, "name": "culture clash"}, {"id": ...}	en	Avatar

1 rows × 23 columns

Include part

```
In [7]: # genres  
# id  
# keywords  
# title  
# overview  
# cast  
#crew
```

```
movies = movies[['movie_id','title','overview','genres','keywords','cast','crew']]
```

```
In [8]: #movies['original_Language'].value_counts()  
movies.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 4809 entries, 0 to 4808  
Data columns (total 7 columns):  
 #   Column      Non-Null Count  Dtype     
---  --          -----            
 0   movie_id    4809 non-null   int64    
 1   title       4809 non-null   object    
 2   overview    4806 non-null   object    
 3   genres      4809 non-null   object    
 4   keywords    4809 non-null   object    
 5   cast        4809 non-null   object    
 6   crew        4809 non-null   object    
dtypes: int64(1), object(6)  
memory usage: 263.1+ KB
```

```
In [9]: movies.head()
```

Out[9]:	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[{"id": 28, "name": "Action"}, {"id": 12, "name": "culture clash"}]	[{"id": 1463, "name": "culture clash"}, {"id": 726, "name": "na..."}]	[{"cast_id": 242, "character": "Jake Sully", "name": "Sully", "id": 1463}]	{"credit_id": "52fe48009251416c750aca23", "de..."}]
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[{"id": 12, "name": "Adventure"}, {"id": 14, "name": "pirates"}]	[{"id": 270, "name": "ocean"}, {"id": 726, "name": "na..."}]	[{"cast_id": 4, "character": "Captain Jack Sparrow", "name": "Jack Sparrow", "id": 270}]]	{"credit_id": "52fe4232c3a36847f800b579", "de..."}]
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[{"id": 28, "name": "Action"}, {"id": 12, "name": "spy"}]	[{"id": 470, "name": "spy"}, {"id": 818, "name": "name..."}]	[{"cast_id": 1, "character": "James Bond", "name": "James Bond", "id": 470}]]	{"credit_id": "54805967c3a36829b5002c41", "de..."}]
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[{"id": 28, "name": "Action"}, {"id": 80, "name": "dc comics"}]	[{"id": 849, "name": "dc comics"}, {"id": 853, "name": "na..."}]	[{"cast_id": 2, "character": "Bruce Wayne / Batman", "name": "Bruce Wayne / Batman", "id": 849}]]	{"credit_id": "52fe4781c3a36847f81398c3", "de..."}]
4	49529	John Carter	John Carter is a war-weary, former military ca...	[{"id": 28, "name": "Action"}, {"id": 12, "name": "based on novel"}]	[{"id": 818, "name": "based on novel"}, {"id": 726, "name": "na..."}]	[{"cast_id": 5, "character": "John Carter", "name": "John Carter", "id": 818}]]	{"credit_id": "52fe479ac3a36847f813eaa3", "de..."}]

In [10]: <code>movies.isnull().sum()</code>
Out[10]:
<pre>movie_id 0 title 0 overview 3 genres 0 keywords 0 cast 0 crew 0 dtype: int64</pre>
In [11]: <code>movies.dropna(inplace=True)</code>
In [12]: <code>movies.duplicated().sum()</code>
Out[12]: 0
In [13]: <code>movies.iloc[0].genres</code>

```
Out[13]: '[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]'
```

```
In [14]: # '[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]'  
# ['Action', 'Adventure', 'Fantasy', 'scifi']
```

```
In [15]: import ast  
def convert(obj):  
    L = []  
    for i in ast.literal_eval(obj):  
        L.append(i['name'])  
    return L
```

```
In [16]: #convert(L)  
import ast  
ast.literal_eval('[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]')
```

```
Out[16]: [{'id': 28, 'name': 'Action'},  
          {'id': 12, 'name': 'Adventure'},  
          {'id': 14, 'name': 'Fantasy'},  
          {'id': 878, 'name': 'Science Fiction'}]
```

```
In [17]: movies['genres'] = movies['genres'].apply(convert)
```

```
In [18]: movies.head()
```

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[{"id": 1463, "name": "culture clash"}, {"id": ...]	[{"cast_id": 242, "character": "Jake Sully", "id": ...}	[{"credit_id": "52fe48009251416c750aca23", "de...]
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[{"id": 270, "name": "ocean"}, {"id": 726, "na...]	[{"cast_id": 4, "character": "Captain Jack Spa...}	[{"credit_id": "52fe4232c3a36847f800b579", "de...]
2	206647	Spectre	A cryptic message from Bond's past sends him O...	[Action, Adventure, Crime]	[{"id": 470, "name": "spy"}, {"id": 818, "name...]	[{"cast_id": 1, "character": "James Bond", "cr...}	[{"credit_id": "54805967c3a36829b5002c41", "de...]
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action, Crime, Drama, Thriller]	[{"id": 849, "name": "dc comics"}, {"id": 853,...]	[{"cast_id": 2, "character": "Bruce Wayne / Ba...}	[{"credit_id": "52fe4781c3a36847f81398c3", "de...]
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action, Adventure, Science Fiction]	[{"id": 818, "name": "based on novel"}, {"id": ...]	[{"cast_id": 5, "character": "John Carter", "c...}	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...]

In [19]: `movies['keywords'] = movies['keywords'].apply(convert)`

In [20]: `#movies['cast'][0]`

In [21]: `def convert3(obj):
 L = []
 counter = 0
 for i in ast.literal_eval(obj):
 if counter != 3:
 L.append(i['name'])
 counter+=1
 else:
 break
 return L`

In [22]: `movies['cast'] = movies['cast'].apply(convert3)`

In [23]: `movies.head(5)`

	<code>movie_id</code>	<code>title</code>	<code>overview</code>	<code>genres</code>	<code>keywords</code>	<code>cast</code>	<code>crew</code>
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...]	[Sam Worthington, Zoe Saldana, Sigourney Weaver]	{"credit_id": "52fe48009251416c750aca23", "de...
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...]	[Johnny Depp, Orlando Bloom, Keira Knightley]	{"credit_id": "52fe4232c3a36847f800b579", "de...
2	206647	Spectre	A cryptic message from Bond's past sends him O...	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...]	[Daniel Craig, Christoph Waltz, Léa Seydoux]	{"credit_id": "54805967c3a36829b5002c41", "de...
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action, Crime, Drama, Thriller]	[dc comics, crime fighter, terrorist, secret i...]	[Christian Bale, Michael Caine, Gary Oldman]	{"credit_id": "52fe4781c3a36847f81398c3", "de...
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel...]	[Taylor Kitsch, Lynn Collins, Samantha Morton]	{"credit_id": "52fe479ac3a36847f813eaa3", "de...

In [24]: `def fetch_director(obj):`
 `L = []`
 `for i in ast.literal_eval(obj):`
 `if i['job'] == 'Director':`
 `L.append(i['name'])`
 `break`
 `return L`

In [25]: `movies['crew'] = movies['crew'].apply(fetch_director)`

In [26]: `#movies['overview'][0]`
`movies['overview'] = movies['overview'].apply(lambda x:x.split())`

In [27]: `movies.head()`

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...]	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...]	[Sam Worthington, Zoe Saldana, Sigourney Weaver]	[James Cameron]
1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d...]	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...]	[Johnny Depp, Orlando Bloom, Keira Knightley]	[Gore Verbinski]
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...]	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...]	[Daniel Craig, Christoph Waltz, Léa Seydoux]	[Sam Mendes]
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...]	[Action, Crime, Drama, Thriller]	[dc comics, crime fighter, terrorist, secret i...]	[Christian Bale, Michael Caine, Gary Oldman]	[Christopher Nolan]
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...]	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel...]	[Taylor Kitsch, Lynn Collins, Samantha Morton]	[Andrew Stanton]

```
In [28]: def collapse(L):
    L1 = []
    for i in L:
        L1.append(i.replace(" ", ""))
    return L1
```

```
In [29]: movies['cast'] = movies['cast'].apply(collapse)
movies['crew'] = movies['crew'].apply(collapse)
movies['genres'] = movies['genres'].apply(collapse)
movies['keywords'] = movies['keywords'].apply(collapse)
```

```
In [30]: movies['tags'] = movies['overview'] + movies['genres'] + movies['keywords'] + movies['
```

```
In [31]: new_df = movies[['movie_id','title','tags']]
```

```
In [32]: new_df['tags'] = new_df['tags'].apply(lambda x: " ".join(x))
```

```
C:\Users\ASUS\AppData\Local\Temp\ipykernel_10800\3089450492.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
    new_df['tags'] = new_df['tags'].apply(lambda x: " ".join(x))
```

```
In [33]: new_df['tags'] = new_df['tags'].apply(lambda x:x.lower())
```

```
C:\Users\ASUS\AppData\Local\Temp\ipykernel_10800\3214958533.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
    new_df['tags'] = new_df['tags'].apply(lambda x:x.lower())
```

In [34]: `new_df.head()`

	movie_id	title	tags
0	19995	Avatar	in the 22nd century, a paraplegic marine is di...
1	285	Pirates of the Caribbean: At World's End	captain barbossa, long believed to be dead, ha...
2	206647	Spectre	a cryptic message from bond's past sends him o...
3	49026	The Dark Knight Rises	following the death of district attorney harve...
4	49529	John Carter	john carter is a war-weary, former military ca...

In [35]: `from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer(max_features=5000, stop_words='english')`

In [36]: `vector = cv.fit_transform(new_df['tags']).toarray()
vector`

Out[36]: `array([[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]], dtype=int64)`

In [37]: `vector.shape`

Out[37]: `(4806, 5000)`

In [38]: `# cv.get_feature_names()`

In [39]: `import nltk`

In [40]: `from nltk.stem.porter import PorterStemmer
ps = PorterStemmer()`

In [41]: `def stem(text):
 y = []

 for i in text.split():
 y.append(ps.stem(i))
 return " ".join(y)`

In [42]: `new_df['tags']=new_df['tags'].apply(stem)`

```
C:\Users\ASUS\AppData\Local\Temp\ipykernel_10800\3514595201.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
    new_df['tags']=new_df['tags'].apply(stem)
```

```
In [43]: # p = ['Love', 'Loving', 'Loves']
```

```
In [44]: # ps.stem("Loves")
```

```
In [45]: from sklearn.metrics.pairwise import cosine_similarity
```

```
In [46]: similarity = cosine_similarity(vector)
```

```
In [47]: similarity[0]
```

```
Out[47]: array([1.          , 0.08964215, 0.05976143, ..., 0.02519763, 0.02817181,
   0.          ])
```

```
In [48]: # new[new['title'] == 'The Lego Movie'].index[0]
```

```
In [49]: def recommend(movie):
    movies_index = new_df[new_df['title'] == movie].index[0]
    distances = similarity[movies_index]
    movies_list = sorted(list(enumerate(distances)), reverse=True, key = lambda x: x[1])
    for i in movies_list:
        print(new_df.iloc[i[0]].title)
```

```
In [50]: recommend('Batman')
```

```
Batman
Batman & Robin
The Dark Knight Rises
Batman Begins
Batman Returns
```

```
In [51]: import pickle
```

```
In [55]: pickle.dump(new_df.to_dict(),open('movie_dict.pkl','wb'))
# pickle.dump(similarity,open('similarity.pkl','wb'))
```

```
In [ ]: pickle.dump(similarity,open('similarity.pkl','wb'))
```

```
In [ ]: # new['title'].values
```

```
In [ ]: # pickle.dump(new.to_dict(),open('movie_dict.pkl','wb'))
```

```
In [ ]: # pickle.dump(similarity,open('similarity.pkl','wb'))
```