

TASK 2 - MOVIE RECOMMENDATION SYSTEM

Table of Contents

- [1 Importing Dataset](#)
- [2 Getting relevant columns](#)
- [3 Data Preprocessing](#)
- [4 Vectorization](#)
- [5 Calculating Cosine distance](#)
- [6 Making a recommendation function](#)
- [7 Using pickle for deploying the model](#)
- [8 This model is deployed using streamlit](#)

```
In [1]: import numpy as np
import pandas as pd
```

```
In [2]: import ast
```

Importing Dataset

```
In [3]: movies = pd.read_csv('tmdb_5000_movies.csv')
credits = pd.read_csv('tmdb_5000_credits.csv')
```

```
In [4]: movies.head()
```

Out[4]:

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popularity	production_companies
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "...]	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id": ...]	en	Avatar	In the 22nd century, a paraplegic Marine is di...	150.437577	[{"name": "Ingr Film Partners", "id": "150437577"}]
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "...]	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "na...]	en	Pirates of the Caribbean: At World's End	Captain Barbosa, long believed to be dead, ha...	139.082615	[{"name": "Walt Disney Pictures", "id": "139082615"}]
2	245000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "...]	http://www.sonypictures.com/movies/spectre/	206647	[{"id": 470, "name": "spy"}, {"id": 818, "name...]	en	Spectre	A cryptic message from Bond's past sends him o...	107.376788	[{"name": "Columbia Pictures", "id": "107376788"}]
3	250000000	[{"id": 28, "name": "Action"}, {"id": 80, "name": "...]	http://www.thedarkknightrises.com/	49026	[{"id": 849, "name": "dc comics"}, {"id": 853, "...]	en	The Dark Knight Rises	Following the death of District Attorney Harve...	112.312950	[{"name": "Legends Entertainment", "id": "112312950"}]
4	260000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "...]	http://movies.disney.com/john-carter	49529	[{"id": 818, "name": "based on novel"}, {"id": ...]	en	John Carter	John Carter is a weary, former military ca...	43.926995	[{"name": "Walt Disney Pictures", "id": "43926995"}]

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

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

Out[6]:

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popularity	production_companies
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "...]	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id": ...]	en	Avatar	In the 22nd century, a paraplegic Marine is di...	150.437577	[{"name": "Ingr Film Partners", "id": "150437577"}]
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "...]	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "na...]	en	Pirates of the Caribbean: At World's End	Captain Barbosa, long believed to be dead, ha...	139.082615	[{"name": "Walt Disney Pictures", "id": "139082615"}]
2	245000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "...]	http://www.sonypictures.com/movies/spectre/	206647	[{"id": 470, "name": "spy"}, {"id": 818, "name...]	en	Spectre	A cryptic message from Bond's past sends him o...	107.376788	[{"name": "Columbia Pictures", "id": "107376788"}]
3	250000000	[{"id": 28, "name": "Action"}, {"id": 80, "name": "...]	http://www.thedarkknightrises.com/	49026	[{"id": 849, "name": "dc comics"}, {"id": 853, "...]	en	The Dark Knight Rises	Following the death of District Attorney Harve...	112.312950	[{"name": "Legends Entertainment", "id": "112312950"}]

4	260000000	{"id": 28, "name": "Action"}, {"id": 12, "name": ...}	http://movies.disney.com/john-carter	49529	["based on novel"], {"id": ...}	en	John Carter	Carter is a war-weary, former military ca...	43.926995	[{"name": "Walt Pictures", "..."}
---	-----------	---	--------------------------------------	-------	---------------------------------	----	-------------	--	-----------	-----------------------------------

5 rows x 23 columns



Getting relevant columns

```
In [7]: movies = movies[['movie_id','title','overview','genres','keywords','cast','crew']]  
movies.head()
```

Out[7]:

	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": ...}	[{"id": 1463, "name": "culture clash"}, {"id": ...]	[{"cast_id": 242, "character": "Jake Sully", ...]	[{"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": ...}	[{"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...	[{"id": 28, "name": "Action"}, {"id": 12, "name": ...}	[{"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...	[{"id": 28, "name": "Action"}, {"id": 80, "name": ...}	[{"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...	[{"id": 28, "name": "Action"}, {"id": 12, "name": ...}	[{"id": 818, "name": "based on novel"}, {"id": ...]	[{"cast_id": 5, "character": "John Carter", "c...]	[{"credit_id": "52fe479ac3a36847f813ea3", "de...

Data Preprocessing

```
In [8]: movies.isnull().sum()
```

```
Out[8]: movie_id    0  
title      0  
overview    3  
genres     0  
keywords    0  
cast       0  
crew       0  
dtype: int64
```

```
In [9]: movies.dropna(inplace=True)
```

```
In [10]: movies.duplicated().sum()
```

Out[10]: 0

```
In [11]: def convert(object):  
    l = []  
    for i in ast.literal_eval(object):  
        l.append(i['name'])  
    return l
```

```
In [12]: movies['genres'] = movies['genres'].apply(convert)
```

```
In [13]: movies['keywords'] = movies['keywords'].apply(convert)
```

```
In [14]: def convert(object):  
    l = []  
    counter = 0  
    for i in ast.literal_eval(object):  
        if counter != 3:  
            l.append(i['name'])  
            counter+=1  
        else:  
            break  
    return l
```

```
In [15]: movies['cast'] = movies['cast'].apply(convert)
```

```
In [16]: def fetch_director(object):  
    l = []  
    for i in ast.literal_eval(object):  
        if i['job'] == 'Director':  
            l.append(i['name'])  
        break  
    return l
```

```
In [17]: movies['crew'] = movies['crew'].apply(fetch_director)
```

```
In [18]: movies['overview'] = movies['overview'].apply(lambda x:x.split())
```

```
In [19]: movies.head()
```

Out[19]:

	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,	[Action, Crime, Drama]	[dc comics, crime fighter,	[Christian Bale, Michael Caine,]	[Christopher Nolan]

	movie_id	title	overview	genres	keywords	cast	crew	tags
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]	[Andrew Stanton]

```
In [20]: movies['genres'] = movies['genres'].apply(lambda x:[i.replace(" ","") for i in x])
movies['keywords'] = movies['keywords'].apply(lambda x:[i.replace(" ","") for i in x])
movies['cast'] = movies['cast'].apply(lambda x:[i.replace(" ","") for i in x])
movies['crew'] = movies['crew'].apply(lambda x:[i.replace(" ","") for i in x])
movies['overview'] = movies['overview'].apply(lambda x:[i.replace(" ","") for i in x])
```

```
In [21]: movies['tags'] = movies['genres'] + movies['keywords'] + movies['cast'] + movies['crew'] + movies['overview']
```

```
In [22]: movies.head()
```

```
Out[22]:
```

	movie_id	title	overview	genres	keywords	cast	crew	tags
0	19995	Avatar	[In, the, 22nd, century, a, paraplegic, Marin...]	[Action, Adventure, Fantasy, Sciencefiction]	[cultureclash, future, spacewar, spacecolony, ...]	[SamWorthington, ZoeSaldana, SigourneyWeaver]	[JamesCameron]	[Action, Adventure, Fantasy, ScienceFiction, c...
1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa, long, believed, to, be, d...]	[Adventure, Fantasy, Action]	[ocean, drugabuse, exoticisland, eastindiatrad...	[JohnnyDepp, OrlandoBloom, KeiraKnightley]	[GoreVerbinski]	[Adventure, Fantasy, Action, ocean, drugabuse,...]
2	206647	Spectre	[A, cryptic, message, from, Buru's, past, send...]	[Action, Adventure, Crime]	[spy, basedonnovel, secretagent, sequel, mi6, ...]	[DanielCraig, ChristophWaltz, LéaSeydoux]	[SamMendes]	[Action, Adventure, Crime, spy, basedonnovel, ...]
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...]	[Action, Crime, Drama, Thriller]	[dccomics, crimefighter, terrorist, secretiden...]	[ChristianBale, MichaelCaine, GaryOldman]	[ChristopherNolan]	[Action, Crime, Drama, Thriller, dccomics, cri...
4	49529	John Carter	[John, Carter, is, a, war-weary., former, mili...	[Action, Adventure, Sciencefiction]	[basedonnovel, mars, medallion, spacetravel, p...]	[TaylorKitsch, LynnCollins, SamanthaMorton]	[AndrewStanton]	[Action, Adventure, ScienceFiction, basedonnov...

```
In [23]: new_df = movies[['movie_id','title','tags']]
```

```
In [24]: new_df.head()
```

```
Out[24]:
```

	movie_id	title	tags
0	19995	Avatar	[Action, Adventure, Fantasy, ScienceFiction, c...
1	285	Pirates of the Caribbean: At World's End	[Adventure, Fantasy, Action, ocean, drugabuse, ...]
2	206647	Spectre	[Action, Adventure, Crime, spy, basedonnovel, ...]
3	49026	The Dark Knight Rises	[Action, Crime, Drama, Thriller, dccomics, cri...
4	49529	John Carter	[Action, Adventure, ScienceFiction, basedonnov...

```
In [25]: new_df['tags'] = new_df['tags'].apply(lambda x:" ".join(x))
```

```
C:\Users\HP\AppData\Local\Temp\ipykernel_19220\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 [26]: new_df.head()
```

```
Out[26]:
```

	movie_id	title	tags
0	19995	Avatar	Action Adventure Fantasy ScienceFiction cultur...
1	285	Pirates of the Caribbean: At World's End	adventure fantasy action ocean drugabuse exoti...
2	206647	Spectre	Action Adventure Crime spy basedonnovel secret...
3	49026	The Dark Knight Rises	Action Crime Drama Thriller dccomics crimefigh...
4	49529	John Carter	Action Adventure ScienceFiction basedonnovel m...

```
In [27]: new_df['tags'] = new_df['tags'].apply(lambda x:x.lower())
```

```
C:\Users\HP\AppData\Local\Temp\ipykernel_19220\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 [28]: new_df.head()
```

```
Out[28]:
```

	movie_id	title	tags
0	19995	Avatar	action adventure fantasy sciencefiction cultur...
1	285	Pirates of the Caribbean: At World's End	adventure fantasy action ocean drugabuse exoti...
2	206647	Spectre	action adventure crime spy basedonnovel secret...
3	49026	The Dark Knight Rises	action crime drama thriller dccomics crimefigh...
4	49529	John Carter	action adventure sciencefiction basedonnovel m...

```
In [29]: pip install nltk
```

```
Requirement already satisfied: nltk in c:\users\hp\anaconda3\lib\site-packages (3.7)
Requirement already satisfied: joblib in c:\users\hp\anaconda3\lib\site-packages (from nltk) (1.1.0)
Requirement already satisfied: regex>=2018.3 in c:\users\hp\anaconda3\lib\site-packages (from nltk) (2022.7.9)
Requirement already satisfied: tqdm in c:\users\hp\anaconda3\lib\site-packages (from nltk) (4.64.1)
Requirement already satisfied: click in c:\users\hp\anaconda3\lib\site-packages (from nltk) (8.0.4)
Requirement already satisfied: colorama in c:\users\hp\appdata\roaming\python\python39\site-packages (from click->nltk) (0.3.9)
Note: you may need to restart the kernel to use updated packages.
```

```
WARNING: Ignoring invalid distribution -ensorflow-intel (c:\users\hp\appdata\roaming\python\python39\site-packages)
```

```
WARNING: Ignoring invalid distribution -ensorflow-intel (c:\users\hp\appdata\roaming\python\python39\site-packages)
```

```
In [30]: import nltk
from nltk.stem.porter import PorterStemmer
ps = PorterStemmer()
```

```
In [31]: def stem(text):
    y = []
    for i in text.split():
        y.append(ps.stem(i))
    return " ".join(y)
```

```
In [32]: new_df['tags'] = new_df['tags'].apply(stem)

C:\Users\HP\AppData\Local\Temp\ipykernel_19220\3213734980.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 [33]: new_df.head()
```

```
Out[33]:
```

	movie_id	title	tags
0	19995	Avatar	action adventur fantasi sciencefict culturecla...
1	285	Pirates of the Caribbean: At World's End	adventur fantasi action ocean drugabus exotic...
2	206647	Spectre	action adventur crime spi basedonnovel secreta...
3	49026	The Dark Knight Rises	action crime drama thriller dccomic crimefight...
4	49529	John Carter	action adventur sciencefict basedonnovel mar m...

Vectorization

```
In [34]: from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer(max_features=5000,stop_words='english')
```

```
In [35]: vectors = cv.fit_transform(new_df['tags']).toarray()
```

```
In [36]: cv.get_feature_names()
```

```
C:\Users\HP\anaconda3\lib\site-packages\sklearn\utils\deprecation.py:87: FutureWarning: Function get_feature_names is deprecated; get_feature_names is deprecated in 1.0 and will be removed in 1.2. Please use get_feature_names_out instead.
warnings.warn(msg, category=FutureWarning)
```

```
Out[36]: ['000',
'007',
'10',
'100',
'11',
'12',
'13',
'14',
'15',
'16',
'17',
'17th',
'18',
'18th',
'18thcenturi',
'19']
```

Calculating Cosine distance

```
In [37]: from sklearn.metrics.pairwise import cosine_similarity
```

```
In [38]: similarity = cosine_similarity(vectors)
```

```
In [39]: sorted(list(enumerate(similarity[0])),reverse=True, key = lambda x:x[1])[1:6]
```

```
Out[39]: [(1216, 0.28676966733820225),
(2409, 0.26901379342448517),
(3730, 0.2605130246476754),
(507, 0.255608593705383),
(539, 0.25038669783359574)]
```

Making a recommendation function

```
In [40]: def recommend(movie):
    movie_index = new_df[new_df['title'] == movie].index[0]
    distances = similarity[movie_index]
    movies_list = sorted(list(enumerate(distances)),reverse=True, key = lambda x:x[1])[1:6]

    for i in movies_list:
        print(new_df.iloc[i[0]].title)
```

```
In [41]: recommend('Batman Begins')
```

```
The Dark Knight
Batman
Batman
The Dark Knight Rises
10th & Wolf
```

```
In [42]: recommend('Avatar')
Aliens vs Predator: Requiem
Aliens
Falcon Rising
Independence Day
Titan A.E.
```

```
In [43]: recommend('Men in Black 3')
Time Changer
I Origins
Timecop
About Time
The R.M.
```

Using pickle for deploying the model

```
In [44]: import pickle
```

```
In [45]: pickle.dump(new_df,open('movies.pkl','wb'))
```

```
In [46]: pickle.dump(new_df.to_dict(),open('movie_dict.pkl','wb'))
```

```
In [47]: pickle.dump(similarity,open('similarity.pkl','wb'))
```

This model is deployed using streamlit

END