

# Movie Recommender System

November 22, 2023

## 1 Movie Recommender System

Keerthan Kumar C

Manipal Institute of Technology

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

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

## 2 Datasets

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

```
[3]:      budget      genres \
0  237000000  [{"id": 28, "name": "Action"}, {"id": 12, "nam...
1  300000000  [{"id": 12, "name": "Adventure"}, {"id": 14, "...
2  245000000  [{"id": 28, "name": "Action"}, {"id": 12, "nam...
3  250000000  [{"id": 28, "name": "Action"}, {"id": 80, "nam...
4  260000000  [{"id": 28, "name": "Action"}, {"id": 12, "nam...

      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.thedarkknighttrises.com/  49026
4  http://movies.disney.com/john-carter  49529

      keywords original_language \
0  [{"id": 1463, "name": "culture clash"}, {"id":...  en
1  [{"id": 270, "name": "ocean"}, {"id": 726, "na...  en
2  [{"id": 470, "name": "spy"}, {"id": 818, "name...  en
3  [{"id": 849, "name": "dc comics"}, {"id": 853,...  en
4  [{"id": 818, "name": "based on novel"}, {"id":...  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  \
0  In the 22nd century, a paraplegic Marine is di... 150.437577
1  Captain Barbossa, long believed to be dead, ha... 139.082615
2  A cryptic message from Bond's past sends him o... 107.376788
3  Following the death of District Attorney Harve... 112.312950
4  John Carter is a war-weary, former military ca... 43.926995

```

```

                                production_companies  \
0  [{"name": "Ingenious Film Partners", "id": 289...
1  [{"name": "Walt Disney Pictures", "id": 2}, {"name":
2  [{"name": "Columbia Pictures", "id": 5}, {"name":
3  [{"name": "Legendary Pictures", "id": 923}, {"name":
4  [{"name": "Walt Disney Pictures", "id": 2}]

```

```

                                production_countries  release_date  revenue  \
0  [{"iso_3166_1": "US", "name": "United States o... 2009-12-10 2787965087
1  [{"iso_3166_1": "US", "name": "United States o... 2007-05-19 961000000
2  [{"iso_3166_1": "GB", "name": "United Kingdom"... 2015-10-26 880674609
3  [{"iso_3166_1": "US", "name": "United States o... 2012-07-16 1084939099
4  [{"iso_3166_1": "US", "name": "United States o... 2012-03-07 284139100

```

```

runtime                                spoken_languages  status  \
0  162.0  [{"iso_639_1": "en", "name": "English"}, {"iso... Released
1  169.0  [{"iso_639_1": "en", "name": "English"}] Released
2  148.0  [{"iso_639_1": "fr", "name": "Fran\u00e7ais"},... Released
3  165.0  [{"iso_639_1": "en", "name": "English"}] Released
4  132.0  [{"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

```

```
[4]: credits.head()
```

```
[4]:      movie_id      title \
0      19995      Avatar
1      285  Pirates of the Caribbean: At World's End
2     206647      Spectre
3      49026  The Dark Knight Rises
4      49529      John Carter

      cast \
0 [{"cast_id": 242, "character": "Jake Sully", "...
1 [{"cast_id": 4, "character": "Captain Jack Spa...
2 [{"cast_id": 1, "character": "James Bond", "cr...
3 [{"cast_id": 2, "character": "Bruce Wayne / Ba...
4 [{"cast_id": 5, "character": "John Carter", "c...

      crew
0 [{"credit_id": "52fe48009251416c750aca23", "de...
1 [{"credit_id": "52fe4232c3a36847f800b579", "de...
2 [{"credit_id": "54805967c3a36829b5002c41", "de...
3 [{"credit_id": "52fe4781c3a36847f81398c3", "de...
4 [{"credit_id": "52fe479ac3a36847f813eaa3", "de...
```

### 3 Data Preprocessing

```
[5]: movies.shape
```

```
[5]: (4803, 20)
```

```
[6]: credits.shape
```

```
[6]: (4803, 4)
```

```
[7]: movies=movies.merge(credits,on="title")
      movies.head(2)
```

```
[7]:      budget      genres \
0  237000000 [{"id": 28, "name": "Action"}, {"id": 12, "nam...
1  300000000 [{"id": 12, "name": "Adventure"}, {"id": 14, "...

      homepage      id \
0  http://www.avatarmovie.com/  19995
1  http://disney.go.com/disneypictures/pirates/  285

      keywords original_language \
0 [{"id": 1463, "name": "culture clash"}, {"id":...      en
1 [{"id": 270, "name": "ocean"}, {"id": 726, "na...      en
```

```

                                original_title \
0                                Avatar
1  Pirates of the Caribbean: At World's End

                                overview popularity \
0  In the 22nd century, a paraplegic Marine is di... 150.437577
1  Captain Barbossa, long believed to be dead, ha... 139.082615

                                production_companies ... runtime \
0  [{"name": "Ingenious Film Partners", "id": 289... ... 162.0
1  [{"name": "Walt Disney Pictures", "id": 2}, {""... ... 169.0

                                spoken_languages status \
0  [{"iso_639_1": "en", "name": "English"}, {"iso... Released
1  [{"iso_639_1": "en", "name": "English"}] Released

                                tagline \
0                                Enter the World of Pandora.
1  At the end of the world, the adventure begins.

                                title vote_average vote_count movie_id \
0                                Avatar                7.2        11800    19995
1  Pirates of the Caribbean: At World's End            6.9         4500     285

                                cast \
0  [{"cast_id": 242, "character": "Jake Sully", "...
1  [{"cast_id": 4, "character": "Captain Jack Spa...

                                crew
0  [{"credit_id": "52fe48009251416c750aca23", "de...
1  [{"credit_id": "52fe4232c3a36847f800b579", "de...

[2 rows x 23 columns]

```

```
[8]: movies=movies[["genres","movie_id","keywords","overview","title","cast","crew"]]
```

```
[9]: movies.head(2)
```

```

[9]:                                genres movie_id \
0  [{"id": 28, "name": "Action"}, {"id": 12, "nam... 19995
1  [{"id": 12, "name": "Adventure"}, {"id": 14, "... 285

                                keywords \
0  [{"id": 1463, "name": "culture clash"}, {"id"::...
1  [{"id": 270, "name": "ocean"}, {"id": 726, "na...

```

```

                                overview \
0  In the 22nd century, a paraplegic Marine is di...
1  Captain Barbossa, long believed to be dead, ha...

                                title \
0                                     Avatar
1  Pirates of the Caribbean: At World's End

                                cast \
0  [{"cast_id": 242, "character": "Jake Sully", "...
1  [{"cast_id": 4, "character": "Captain Jack Spa...

                                crew
0  [{"credit_id": "52fe48009251416c750aca23", "de...
1  [{"credit_id": "52fe4232c3a36847f800b579", "de...

```

```
[10]: movies.isna().sum()
```

```

[10]: genres      0
      movie_id    0
      keywords    0
      overview    3
      title       0
      cast        0
      crew        0
      dtype: int64

```

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

```
[12]: movies.isna().sum()
```

```

[12]: genres      0
      movie_id    0
      keywords    0
      overview    0
      title       0
      cast        0
      crew        0
      dtype: int64

```

```
[13]: movies.duplicated().sum()
```

```
[13]: 0
```

```
[14]: movies.iloc[0].genres
```

```

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

```
[15]: #since list indices must be string literals we use ast
import ast
def convert(obj):
    l=[]
    for i in ast.literal_eval(obj):
        l.append(i["name"])
    return l
```

```
[16]: movies["genres"]=movies["genres"].apply(convert)
```

```
[17]: movies["keywords"]=movies["keywords"].apply(convert)
```

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

```
[18]:
```

	genres	movie_id	\
0	[Action, Adventure, Fantasy, Science Fiction]	19995	
1	[Adventure, Fantasy, Action]	285	
2	[Action, Adventure, Crime]	206647	
3	[Action, Crime, Drama, Thriller]	49026	
4	[Action, Adventure, Science Fiction]	49529	

	keywords	\
0	[culture clash, future, space war, space colon...	
1	[ocean, drug abuse, exotic island, east india ...	
2	[spy, based on novel, secret agent, sequel, mi...	
3	[dc comics, crime fighter, terrorist, secret i...	
4	[based on novel, mars, medallion, space travel...	

	overview	\
0	In the 22nd century, a paraplegic Marine is di...	
1	Captain Barbossa, long believed to be dead, ha...	
2	A cryptic message from Bond's past sends him o...	
3	Following the death of District Attorney Harve...	
4	John Carter is a war-weary, former military ca...	

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

	cast	\
0	[{"cast_id": 242, "character": "Jake Sully", "...	
1	[{"cast_id": 4, "character": "Captain Jack Spa...	
2	[{"cast_id": 1, "character": "James Bond", "cr...	
3	[{"cast_id": 2, "character": "Bruce Wayne / Ba...	

```

4 [{"cast_id": 5, "character": "John Carter", "c...

                                crew
0 [{"credit_id": "52fe48009251416c750aca23", "de...
1 [{"credit_id": "52fe4232c3a36847f800b579", "de...
2 [{"credit_id": "54805967c3a36829b5002c41", "de...
3 [{"credit_id": "52fe4781c3a36847f81398c3", "de...
4 [{"credit_id": "52fe479ac3a36847f813eaa3", "de...

```

```

[19]: #we need for cast only top 3 casts
def convertfor3(obj):
    l=[]
    counter=0;
    for i in ast.literal_eval(obj):
        if counter<3:
            l.append(i["name"])
            counter=counter +1;
        else:
            break;
    return l

```

```

[20]: movies["cast"]=movies["cast"].apply(convertfor3)

```

```

[21]: movies

```

```

[21]:
                                genres  movie_id  \
0      [Action, Adventure, Fantasy, Science Fiction]    19995
1                      [Adventure, Fantasy, Action]       285
2                      [Action, Adventure, Crime]    206647
3      [Action, Crime, Drama, Thriller]    49026
4      [Action, Adventure, Science Fiction]    49529
...
4804                      [Action, Crime, Thriller]    9367
4805                      [Comedy, Romance]    72766
4806      [Comedy, Drama, Romance, TV Movie]    231617
4807                      []    126186
4808                      [Documentary]    25975

                                keywords  \
0      [culture clash, future, space war, space colon...
1      [ocean, drug abuse, exotic island, east india ...
2      [spy, based on novel, secret agent, sequel, mi...
3      [dc comics, crime fighter, terrorist, secret i...
4      [based on novel, mars, medallion, space travel...
...
4804      [united states-mexico barrier, legs, arms, pap...
4805

```

4806 [date, love at first sight, narration, investi...  
 4807 []  
 4808 [obsession, camcorder, crush, dream girl]

#### overview \

0 In the 22nd century, a paraplegic Marine is di...  
 1 Captain Barbossa, long believed to be dead, ha...  
 2 A cryptic message from Bond's past sends him o...  
 3 Following the death of District Attorney Harve...  
 4 John Carter is a war-weary, former military ca...  
 ...  
 4804 El Mariachi just wants to play his guitar and ...  
 4805 A newlywed couple's honeymoon is upended by th...  
 4806 "Signed, Sealed, Delivered" introduces a dedic...  
 4807 When ambitious New York attorney Sam is sent t...  
 4808 Ever since the second grade when he first saw ...

#### title \

0 Avatar  
 1 Pirates of the Caribbean: At World's End  
 2 Spectre  
 3 The Dark Knight Rises  
 4 John Carter  
 ...  
 4804 El Mariachi  
 4805 Newlyweds  
 4806 Signed, Sealed, Delivered  
 4807 Shanghai Calling  
 4808 My Date with Drew

#### cast \

0 [Sam Worthington, Zoe Saldana, Sigourney Weaver]  
 1 [Johnny Depp, Orlando Bloom, Keira Knightley]  
 2 [Daniel Craig, Christoph Waltz, Léa Seydoux]  
 3 [Christian Bale, Michael Caine, Gary Oldman]  
 4 [Taylor Kitsch, Lynn Collins, Samantha Morton]  
 ...  
 4804 [Carlos Gallardo, Jaime de Hoyos, Peter Marqua...  
 4805 [Edward Burns, Kerry Bishé, Marsha Dietlein]  
 4806 [Eric Mabius, Kristin Booth, Crystal Lowe]  
 4807 [Daniel Henney, Eliza Coupe, Bill Paxton]  
 4808 [Drew Barrymore, Brian Herzlinger, Corey Feldman]

#### crew

0 [{"credit\_id": "52fe48009251416c750aca23", "de...  
 1 [{"credit\_id": "52fe4232c3a36847f800b579", "de...  
 2 [{"credit\_id": "54805967c3a36829b5002c41", "de...



```

3      [{"credit_id": "52fe4781c3a36847f81398c3", "de...
4      [{"credit_id": "52fe479ac3a36847f813eaa3", "de...
...
4804 [{"credit_id": "52fe44eec3a36847f80b280b", "de...
4805 [{"credit_id": "52fe487dc3a368484e0fb013", "de...
4806 [{"credit_id": "52fe4df3c3a36847f8275ecf", "de...
4807 [{"credit_id": "52fe4ad9c3a368484e16a36b", "de...
4808 [{"credit_id": "58ce021b9251415a390165d9", "de...

```

[4806 rows x 7 columns]

```

[23]: #we need only director name from crew
def getdirector(obj):
    l=[];
    for i in ast.literal_eval(obj):
        if(i['job']=='Director'):
            l.append(i["name"])
    return l

```

```

[24]: movies["crew"]=movies["crew"].apply(getdirector)

```

```

[26]: movies.head()

```

```

[26]:
          genres  movie_id  \
0  [Action, Adventure, Fantasy, Science Fiction]      19995
1                [Adventure, Fantasy, Action]         285
2                [Action, Adventure, Crime]      206647
3                [Action, Crime, Drama, Thriller]    49026
4                [Action, Adventure, Science Fiction]  49529

```

```

          keywords  \
0  [culture clash, future, space war, space colon...
1  [ocean, drug abuse, exotic island, east india ...
2  [spy, based on novel, secret agent, sequel, mi...
3  [dc comics, crime fighter, terrorist, secret i...
4  [based on novel, mars, medallion, space travel...

```

```

          overview  \
0  In the 22nd century, a paraplegic Marine is di...
1  Captain Barbossa, long believed to be dead, ha...
2  A cryptic message from Bond's past sends him o...
3  Following the death of District Attorney Harve...
4  John Carter is a war-weary, former military ca...

```

```

          title  \
0                Avatar
1  Pirates of the Caribbean: At World's End

```

```

2             Spectre
3         The Dark Knight Rises
4             John Carter

```

```

                                cast                crew
0  [Sam Worthington, Zoe Saldana, Sigourney Weaver]  [James Cameron]
1      [Johnny Depp, Orlando Bloom, Keira Knightley]  [Gore Verbinski]
2      [Daniel Craig, Christoph Waltz, Léa Seydoux]  [Sam Mendes]
3      [Christian Bale, Michael Caine, Gary Oldman]  [Christopher Nolan]
4      [Taylor Kitsch, Lynn Collins, Samantha Morton]  [Andrew Stanton]

```

```

[27]: #overview column splitting to list of words
      movies["overview"]=movies["overview"].apply(lambda x:x.split())

```

```

[29]: movies.head()

```

```

[29]:                                     genres  movie_id  \
0  [Action, Adventure, Fantasy, Science Fiction]      19995
1                        [Adventure, Fantasy, Action]      285
2                        [Action, Adventure, Crime]    206647
3      [Action, Crime, Drama, Thriller]      49026
4      [Action, Adventure, Science Fiction]      49529

```

```

                                keywords  \
0  [culture clash, future, space war, space colon...
1  [ocean, drug abuse, exotic island, east india ...
2  [spy, based on novel, secret agent, sequel, mi...
3  [dc comics, crime fighter, terrorist, secret i...
4  [based on novel, mars, medallion, space travel...

```

```

                                overview  \
0  [In, the, 22nd, century,, a, paraplegic, Marin...
1  [Captain, Barbossa,, long, believed, to, be, d...
2  [A, cryptic, message, from, Bond's, past, send...
3  [Following, the, death, of, District, Attorney...
4  [John, Carter, is, a, war-weary,, former, mili...

```

```

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

```

```

                                cast                crew
0  [Sam Worthington, Zoe Saldana, Sigourney Weaver]  [James Cameron]
1      [Johnny Depp, Orlando Bloom, Keira Knightley]  [Gore Verbinski]

```

```

2      [Daniel Craig, Christoph Waltz, Léa Seydoux]      [Sam Mendes]
3      [Christian Bale, Michael Caine, Gary Oldman]      [Christopher Nolan]
4      [Taylor Kitsch, Lynn Collins, Samantha Morton]      [Andrew Stanton]

```

```

[30]: #since we are doing this for tags a space would make it difficult so removing
      ↪spaces
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["overview"]=movies["overview"].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])

```

```

[31]: movies.head()

```

```

[31]:
      genres  movie_id  \
0  [Action, Adventure, Fantasy, ScienceFiction]      19995
1              [Adventure, Fantasy, Action]          285
2              [Action, Adventure, Crime]      206647
3      [Action, Crime, Drama, Thriller]      49026
4      [Action, Adventure, ScienceFiction]      49529

      keywords  \
0  [cultureclash, future, spacewar, spacecolony, ...
1  [ocean, drugabuse, exoticisland, eastindiatrad...
2  [spy, basedonnovel, secretagent, sequel, mi6, ...
3  [dccomics, crimefighter, terrorist, secretiden...
4  [basedonnovel, mars, medallion, spacetravel, p...

      overview  \
0  [In, the, 22nd, century,, a, paraplegic, Marin...
1  [Captain, Barbossa,, long, believed, to, be, d...
2  [A, cryptic, message, from, Bond's, past, send...
3  [Following, the, death, of, District, Attorney...
4  [John, Carter, is, a, war-weary,, former, mili...

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

      cast      crew
0  [SamWorthington, ZoeSaldana, SigourneyWeaver]  [JamesCameron]

```

```

1      [JohnnyDepp, OrlandoBloom, KeiraKnightley]      [GoreVerbinski]
2      [DanielCraig, ChristophWaltz, LéaSeydoux]      [SamMendes]
3      [ChristianBale, MichaelCaine, GaryOldman]      [ChristopherNolan]
4      [TaylorKitsch, LynnCollins, SamanthaMorton]      [AndrewStanton]

```

```

[32]: #combining all columns to a single column to get tags
movies["tags"]=movies["genres"]+movies["keywords"]+movies["overview"]+movies["cast"]+movies["crew"]

```

```

[33]: movies.head()

```

```

[33]:
      genres  movie_id \
0  [Action, Adventure, Fantasy, ScienceFiction]      19995
1                [Adventure, Fantasy, Action]         285
2                [Action, Adventure, Crime]      206647
3      [Action, Crime, Drama, Thriller]      49026
4      [Action, Adventure, ScienceFiction]      49529

      keywords \
0  [cultureclash, future, spacewar, spacecolony, ...
1  [ocean, drugabuse, exoticisland, eastindiatrad...
2  [spy, basedonnovel, secretagent, sequel, mi6, ...
3  [dccomics, crimefighter, terrorist, secretiden...
4  [basedonnovel, mars, medallion, spacetravel, p...

      overview \
0  [In, the, 22nd, century,, a, paraplegic, Marin...
1  [Captain, Barbossa,, long, believed, to, be, d...
2  [A, cryptic, message, from, Bond's, past, send...
3  [Following, the, death, of, District, Attorney...
4  [John, Carter, is, a, war-weary,, former, mili...

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

      cast      crew \
0  [SamWorthington, ZoeSaldana, SigourneyWeaver]      [JamesCameron]
1      [JohnnyDepp, OrlandoBloom, KeiraKnightley]      [GoreVerbinski]
2      [DanielCraig, ChristophWaltz, LéaSeydoux]      [SamMendes]
3      [ChristianBale, MichaelCaine, GaryOldman]      [ChristopherNolan]
4      [TaylorKitsch, LynnCollins, SamanthaMorton]      [AndrewStanton]

      tags
0  [Action, Adventure, Fantasy, ScienceFiction, c...

```

```

1 [Adventure, Fantasy, Action, ocean, drugabuse,...
2 [Action, Adventure, Crime, spy, basedonnovel, ...
3 [Action, Crime, Drama, Thriller, dcomics, cri...
4 [Action, Adventure, ScienceFiction, basedonnov...

```

```
[34]: new_df=movies[["title","movie_id","tags"]]
```

```
[35]: new_df.head()
```

```
[35]:
```

	title	movie_id	\
0	Avatar	19995	
1	Pirates of the Caribbean: At World's End	285	
2	Spectre	206647	
3	The Dark Knight Rises	49026	
4	John Carter	49529	

  

```

tags
0 [Action, Adventure, Fantasy, ScienceFiction, c...
1 [Adventure, Fantasy, Action, ocean, drugabuse,...
2 [Action, Adventure, Crime, spy, basedonnovel, ...
3 [Action, Crime, Drama, Thriller, dcomics, cri...
4 [Action, Adventure, ScienceFiction, basedonnov...

```

```
[36]: #converting tags column to a string
new_df["tags"]=new_df["tags"].apply(lambda x : ' '.join(x))
```

C:\Users\KEERTHAN\AppData\Local\Temp\ipykernel\_11988\2105490204.py:2:

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](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))
```

```
[37]: new_df.head()
```

```
[37]:
```

	title	movie_id	\
0	Avatar	19995	
1	Pirates of the Caribbean: At World's End	285	
2	Spectre	206647	
3	The Dark Knight Rises	49026	
4	John Carter	49529	

  

```

tags
0 Action Adventure Fantasy ScienceFiction cultur...
1 Adventure Fantasy Action ocean drugabuse exoti...
2 Action Adventure Crime spy basedonnovel secret...

```

```

3 Action Crime Drama Thriller dccomics crimefigh...
4 Action Adventure ScienceFiction basedonnovel m...

```

```

[38]: #converting tags column to all lower case
new_df["tags"]=new_df["tags"].apply(lambda x : x.lower())

```

```

C:\Users\KEERTHAN\AppData\Local\Temp\ipykernel_11988\2760261440.py:2:
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())

```

```

[39]: new_df

```

```

[39]:
           title  movie_id \
0           Avatar      19995
1  Pirates of the Caribbean: At World's End       285
2           Spectre     206647
3  The Dark Knight Rises     49026
4       John Carter     49529
...           ...           ...
4804       El Mariachi       9367
4805       Newlyweds      72766
4806  Signed, Sealed, Delivered     231617
4807  Shanghai Calling     126186
4808  My Date with Drew     25975

```

```

           tags
0  action adventure fantasy sciencefiction cultur...
1  adventure fantasy action ocean drugabuse exoti...
2  action adventure crime spy basedonnovel secret...
3  action crime drama thriller dccomics crimefigh...
4  action adventure sciencefiction basedonnovel m...
...           ...
4804  action crime thriller unitedstates-mexicobarri...
4805  comedy romance a newlywed couple's honeymoon i...
4806  comedy drama romance tvmovie date loveatfirsts...
4807  when ambitious new york attorney sam is sent t...
4808  documentary obsession camcorder crush dreamgir...

```

```

[4806 rows x 3 columns]

```

## 4 Stemming

```
[74]: #to remove repeated words in tags like loving,loved,lovable and all as it_
      ↪ affects in vectorization stemming is being used
import nltk
```

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

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

```
[78]: new_df["tags"]=new_df["tags"].apply(stem)
```

```
C:\Users\KEERTHAN\AppData\Local\Temp\ipykernel_11988\1027202188.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)
```

```
[79]: new_df.head()
```

```
[79]:
```

	title	movie_id	\
0	Avatar	19995	
1	Pirates of the Caribbean: At World's End	285	
2	Spectre	206647	
3	The Dark Knight Rises	49026	
4	John Carter	49529	

  

	tags
0	action adventur fantasi sciencefict culturecla...
1	adventur fantasi action ocean drugabus exoticici...
2	action adventur crime spi basedonnovel secreta...
3	action crime drama thriller dccomic crimefight...
4	action adventur sciencefict basedonnovel mar m...

we are ready with the dataset to work with now for our requirement

## 5 Vectorization

Why Vectorization?

converting all tags to a vector in dimensional space so that based on the distances between the vectors similarity can be calculated hence movie can be recommended

so we combine all the tags and count the max say 5000 most frequent words and then check each frequent word count for eqch movie(i.e in tag of that movie) which actually represent a vector

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

```
[81]: cv=CountVectorizer(max_features=5000,stop_words='english')#no stop words (like_
      ↪are in a etc..)
```

```
[82]: vector=cv.fit_transform(new_df["tags"]).toarray()
      vector
```

```
[82]: 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)
```

```
[83]: vector.shape
```

```
[83]: (4806, 5000)
```

we can see it has been converted to a (4806, 5000) approx 5k dimensional space so each movie is a vector

means that 4806 movies ka each column of 5k repeated words whether that is present in the movie or not and plotted based on it

Starting Similarity between vectors

we are using cosine distance(angles) not euclidian as dimension increases euclidian is less significant

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

```
[85]: similarity=cosine_similarity(vector)
```

```
[86]: similarity
```

```
[86]: array([[1.          , 0.08346223, 0.0860309 , ..., 0.04499213, 0.
            0.          ],
            [0.08346223, 1.          , 0.06063391, ..., 0.02378257, 0.
            0.02615329],
            [0.0860309 , 0.06063391, 1.          , ..., 0.02451452, 0.
            0.          ],
            ...,
            [0.04499213, 0.02378257, 0.02451452, ..., 1.          , 0.03962144,
            0.04229549],
```



```
[0.          , 0.          , 0.          , ..., 0.03962144, 1.          ,
 0.08714204],
[0.          , 0.02615329, 0.          , ..., 0.04229549, 0.08714204,
 1.          ]])
```

```
[87]: similarity.shape
```

```
[87]: (4806, 4806)
```

so each movie of 4806 movies ka distance is calculated with the all 4806 movies

```
[90]: similarity[0] #represent movie 1 ka distance with all 4806 movies
```

```
[90]: array([1.          , 0.08346223, 0.0860309 , ..., 0.04499213, 0.          ,
 0.          ])
```

## 6 Recommending

```
[97]: def recommend(movie):
    movie_indices = new_df[new_df["title"] == movie].index
    if len(movie_indices) == 0:
        print(f"Sorry, '{movie}' is not found in the dataset.")
        return

    movie_index = movie_indices[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)
```

```
[101]: #asking input and recommending movies
mov=str(input("Please enter movie name \n"))
print("Recommended movies are \n")
recommend(mov)
```

```
Please enter movie name
Batman Begins
Recommended movies are
```

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

The Name of the movie entered should be proper and according to the name in the dataset  
DONE!!!