MOVIE RECOMMENDER SYSTEM

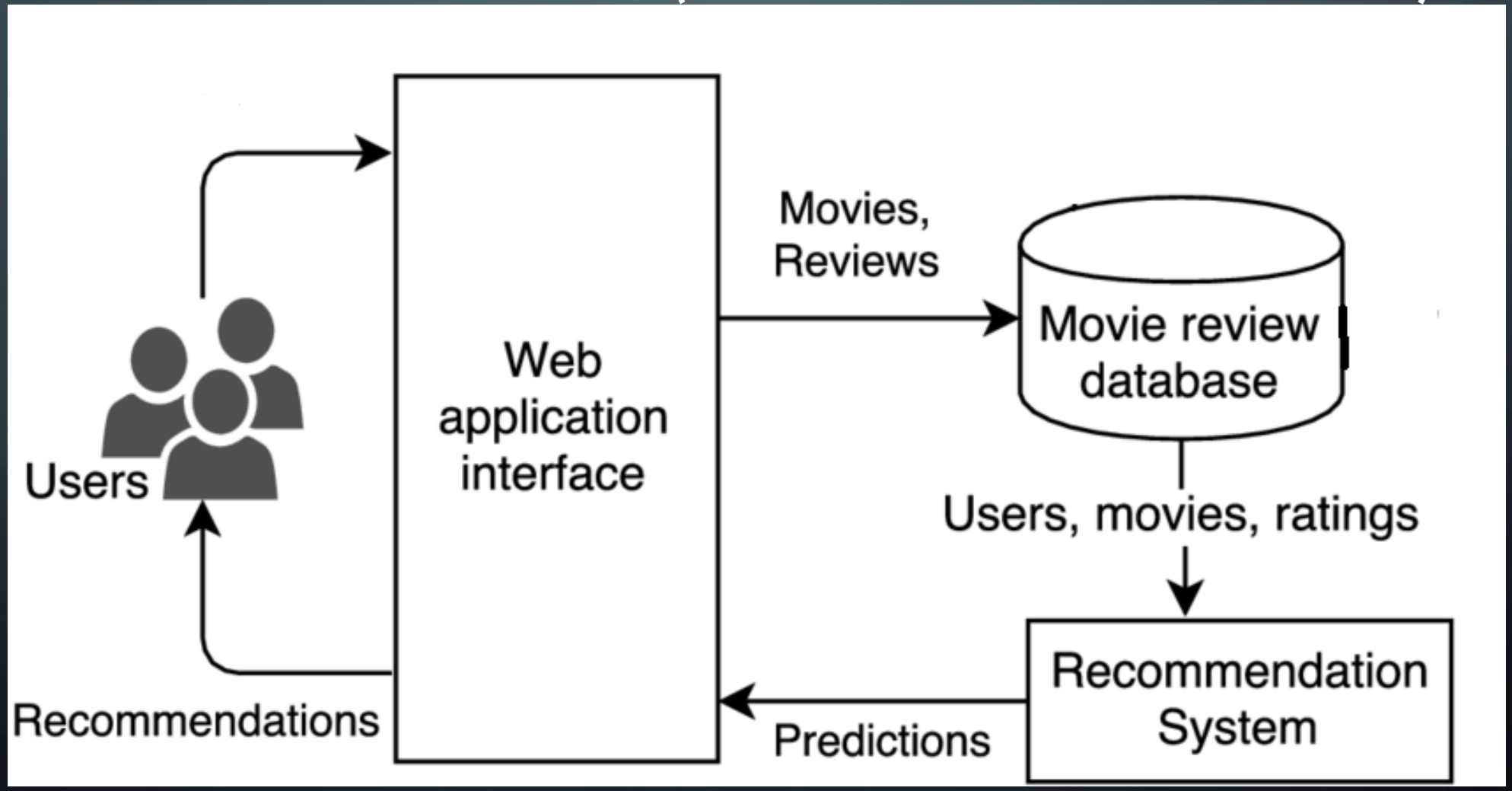
SHAURYA YADAV 21BCE8437

VANSH SINGH 21BCE8889

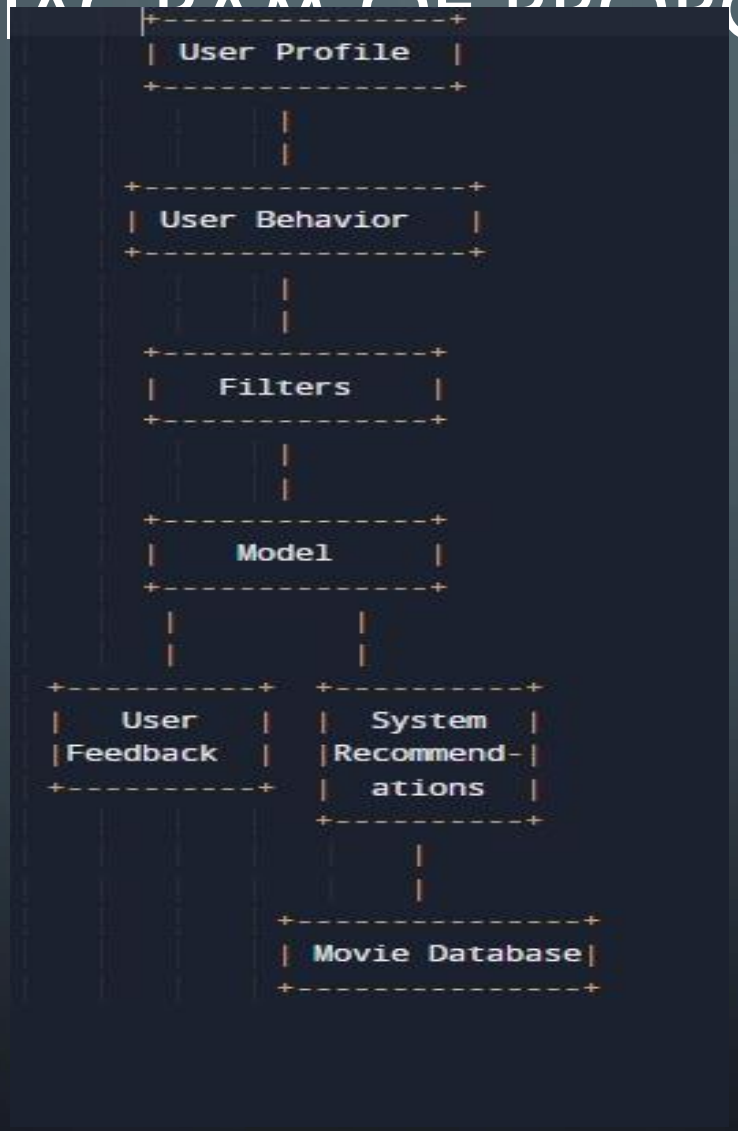
ABSTRACT

- The recommendation system plays an essential role in the modern era and is used by many prestigious applications. A movie recommender system provides a set of of movies to the user based on his/her preferences , and his watching style.
- We acquire movie data from TMDB which includes the poster , the ratings and the summary of it. Usage of panda, numpy, picle, ast, nltk, lemmatization and stemming would be done.
- The back end will be supported by python and front end will use stream lit which will be locally hosted on the college.

SYSTEM ARCHITECTURE(CONCEPTUAL DIAGRAM)



DATA FLOW DIAGRAM OF PROPOSED WORK



MODULE DETAILS

1. Data Collection Module: This module would be responsible for collecting data about movies, users, ratings, and other relevant information. It could include web scraping tools or APIs to gather data from movie databases like IMDb, Rotten Tomatoes, or The Movie Database.
2. Data Preprocessing Module: This module would be responsible for cleaning and preparing the collected data for analysis. It could involve techniques like data cleaning, data normalization, feature extraction, and dimensionality reduction.
3. Machine Learning Algorithms Module: This module would be responsible for developing and training the machine learning algorithms that will power the movie recommender system. Common algorithms used for recommendation systems include collaborative filtering, content-based filtering, and hybrid methods that combine both approaches.
4. Recommendation Generation Module: This module would be responsible for generating personalized movie recommendations for users based on their viewing history, preferences, and other relevant data. It would involve applying the trained machine learning algorithms to the user's data to generate a list of recommended movies.
5. User Interface Module: This module would be responsible for presenting the recommended movies to the user through a user interface. It could include features like search bars, filter options, and personalized recommendations.

DETAILED EXPLANATION OF ALGORITHM UTILIZED

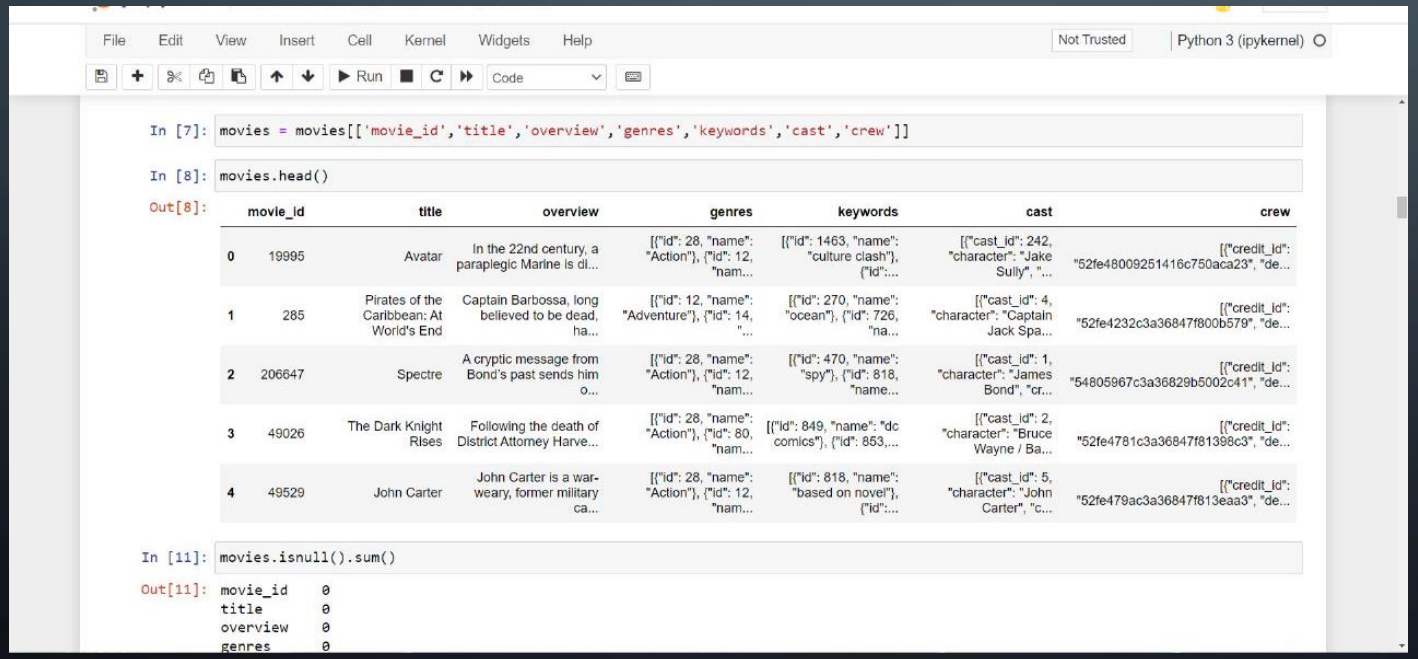
- Content-based filtering is a recommendation algorithm that suggests items to users based on their similarity to the items that the user has already shown an interest in. In the context of a movie recommender system, content-based filtering works by analyzing the features of movies that the user has previously watched and enjoyed and recommending other movies with similar features.
- To implement content-based filtering, the system needs to identify the relevant features of movies that the user might be interested in. These features could include movie genre, cast, director, plot, and other metadata. Once these features are identified, the system assigns weights to them based on how important they are to the user's viewing preferences. For example, if a user has watched and enjoyed several action movies, the system might assign a higher weight to the action genre when making recommendations.
- The system then uses these features and weights to calculate a similarity score between the user's past viewing history and the features of other movies. The higher the similarity score, the more likely it is that the user will enjoy the recommended movie. The system can use various machine learning algorithms to generate these similarity scores, such as cosine similarity, Euclidean distance, or Pearson correlation.

DATA SET AND FEATURE DETAILS

- We have used movielens review dataset
- Consist 100k ratings
- 6k movies
- 600 users
- Integrated the dataset with IMDB and TMDB data set publicly available

PRE PROCESSING WITH SCREENSHOTS

- The preprocessing phase of a movie recommender system involves preparing the raw data for analysis and generating the features that will be used by the recommendation algorithms.
- data cleaning



File Edit View Insert Cell Kernel Widgets Help — Not Trusted — Python 3 (ipykernel)

Run — Code

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

In [8]: movies.head()
```

Out[8]:

	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, "nam...	[{"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, "...	[{"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, "nam...	[{"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, "nam...	[{"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, "nam...	[{"id": 818, "name": "based on novel"}, {"id":...	[{"cast_id": 5, "character": "John Carter", "c...	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...

```
In [11]: movies.isnull().sum()

Out[11]: movie_id    0
         title       0
         overview    0
         genres      0
```

```
In [11]: movies.isnull().sum()

Out[11]: movie_id    0
         title       0
         overview    0
         genres      0
         keywords    0
         cast        0
         crew        0
         dtype: int64

In [10]: movies.dropna(inplace=True)

In [12]: movies.duplicated().sum()

Out[12]: 0

In [13]: movies.iloc[0].genres

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

In [14]: # 'action', 'adventure', 'scifi', 'fantasy'

In [15]: import ast
         ast.literal_eval('[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "nam
```

Data normalization

```
In [34]: movies['overview'] = movies['overview'].apply(lambda x:x.split())
```

In [35]:
```
movies.head()
```

Out[35]:

	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 [36]: movies['genres'].apply(lambda x:[i.replace(" ","") for i in x])

Out[36]: 0       [Action, Adventure, Fantasy, ScienceFiction]
         1                      [Adventure, Fantasy, Action]
         2                       [Action, Adventure, Crime]
         3                    [Action, Crime, Drama, Thriller]
         4                 [Action, Adventure, ScienceFiction]
                                        ...
         4804                     [Action, Crime, Thriller]
```

```
In [37]: 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])
```

In [38]:
```
movies.head()
```

Out[38]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[Action, Adventure, Fantasy, ScienceFiction]	[cultureclash, future, spacewar, spacecolony, ...	[SamWorthington, ZoeSaldana, SigourneyWeaver]	[JamesCameron]
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]
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...	[Action, Adventure, Crime]	[spy, basedonnovel, secretagent, sequel, mi6, ...	[DanielCraig, ChristophWaltz, LéaSeydoux]	[SamMendes]
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...	[Action, Crime, Drama, Thriller]	[dccomics, crimefighter, terrorist, secretiden...	[ChristianBale, MichaelCaine, GaryOldman]	[ChristopherNolan]
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...	[Action, Adventure, ScienceFiction]	[basedonnovel, mars, medallion, spacetravel, p...	[TaylorKitsch, LynnCollins, SamanthaMorton]	[AndrewStanton]

```
In [39]: movies['tags'] = movies['overview'] + movies['genres'] + movies['keywords'] + movies['cast'] + movies['crew']
```

In [40]:
```
movies.head()
```

Out[40]:

Recommendation algorithm

1

```
4808    Ever since the second grade when he first saw ...
Name: tags, Length: 4806, dtype: object
```

In [43]:
```
new_df['tags'] = new_df['tags'].apply(lambda x:" ".join(x))
```

```
C:\Users\vansh\AppData\Local\Temp\ipykernel_6724\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 [44]:
```
new_df.head()
```

Out[44]:

	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 [45]:
```
from nltk.stem.porter import PorterStemmer
ps = PorterStemmer()
```

2

4	49529	John Carter	John Carter is a war-weary, former military ca...

In [45]:
```
from nltk.stem.porter import PorterStemmer
ps = PorterStemmer()
```

In [46]:
```
def stem(text):
    y = []

    for i in text.split():
        y.append(ps.stem(i))

    return " ".join(y)
```

In [47]:
```
new_df['tags'] = new_df['tags'].apply(stem)
```

```
C:\Users\vansh\AppData\Local\Temp\ipykernel_6724\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 [48]:
```
new_df['tags'][0]
```

Out[48]:
```
'in the 22nd century, a parapleg marin is dispatch to the moon pandora on a uniqu mission, but becom torn between follow order
and protect an alien civilization. action adventur fantasi sciencefict cultureclash futur spacewar spacecoloni societi spacetra
vel futurist romanc space alien tribe alienplanet cgi marin soldier battl loveaffair antiwar powerrel mindandsoul 3d samworthin
```

3

In [58]:
```
from sklearn.metrics.pairwise import cosine_similarity
```

In [61]:
```
cosine_similarity(vectors)
```

Out[61]:
```
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.        ]])
```

In [65]:
```
similarity
```

Out[65]:
```
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,
```

4

```
rsus-a-copy
  new_df['tags'] = new_df['tags'].apply(lambda x:x.lower())
```

In [50]:
```
new_df.head()
```

Out[50]:

	movie_id	title	tags
0	19995	Avatar	in the 22nd century, a parapleg marin is dispa...
1	285	Pirates of the Caribbean: At World's End	captain barbossa, long believ to be dead, ha c...
2	206647	Spectre	a cryptic messag from bond' past send him on a...
3	49026	The Dark Knight Rises	follow the death of district attorney harvey d...
4	49529	John Carter	john carter is a war-weary, former militari ca...

In [51]:
```
from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer(max_features=5000,stop_words='english')
```

In [52]:
```
cv.fit_transform(new_df['tags']).toarray()
```

Out[52]:
```
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)
```