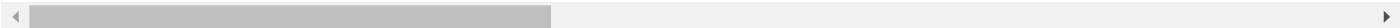


```
import pandas as pd
import numpy as np
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity
```

```
df=pd.read_csv('/content/movies1.csv')
df
```

	index	budget	genres	homepage	id	keywords	original_language	original_title
0	0	237000000	Action Adventure Fantasy Science Fiction	http://www.avatarmovie.com/	19995	culture clash future space war space colony so...	en	Av
1	1	300000000	Adventure Fantasy Action	http://disney.go.com/disneypictures/pirates/	285	ocean drug abuse exotic island east india trad...	en	Pirates o Caribbea World's
2	2	245000000	Action Adventure Crime	http://www.sonypictures.com/movies/spectre/	206647	spy based on novel secret agent sequel mi6	en	Spe
3	3	250000000	Action Crime Drama Thriller	http://www.thedarkknightises.com/	49026	dc comics crime fighter terrorist secret ident...	en	The Dark Kr R
4	4	260000000	Action Adventure Science Fiction	http://movies.disney.com/john-carter	49529	based on novel mars medallion space travel pri...	en	John Ci
...
4798	4798	220000	Action Crime Thriller		NaN	united stateslu2013mexico barrier legs arms pa...	es	El Mari
4799	4799	9000	Comedy Romance		NaN	NaN	en	Newlyw
4800	4800	0	Comedy Drama Romance TV Movie	http://www.hallmarkchannel.com/signedsealeddel...	231617	date love at first sight narration investigati...	en	Signed, See Deliv
4801	4801	0	NaN	http://shanghaicalling.com/	126186	NaN	en	Shanghai Ca
4802	4802	0	Documentary		NaN	obsession camcorder crush dream girl	en	My Date C

4803 rows × 24 columns



df.shape

(4803, 24)

```
features=['genres', 'keywords', 'original_language', 'title', 'cast', 'director']
for feature in features:
```

```
df[feature]=df[feature].fillna('')

def combined_features(row):
    return row['title']+","+row['genres']+","+row['keywords']+","+row['original_language']+","+row['cast']+","+row['director']
df['combined_features']=df.apply(combined_features,axis=1)
df['combined_features']
```


 **combined_features**

0	Avatar,Action Adventure Fantasy Science Fictio...
1	Pirates of the Caribbean: At World's End,Adven...
2	Spectre,Action Adventure Crime,spy based on no...
3	The Dark Knight Rises,Action Crime Drama Thril...
4	John Carter,Action Adventure Science Fiction,b...
...	...
4798	El Mariachi,Action Crime Thriller,united state...
4799	Newlyweds,Comedy Romance,,en,Edward Burns Kerr...
4800	Signed, Sealed, Delivered,Comedy Drama Romance...
4801	Shanghai Calling,,,en,Daniel Henney Eliza Coup...
4802	My Date with Drew,Documentary,obsession camcor...


4803 rows × 1 columns

dtype: object


```
tfidf=TfidfVectorizer()
tfidv=tfidf.fit_transform(df['combined_features'])
tfidv.toarray()
```

 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.]])

tfidv.shape

 (4803, 17502)

```
cosine_sim=cosine_similarity(tfidv)
cosine_sim
```

 array([[1.00000000e+00, 2.48272960e-02, 4.22228416e-02, ...,
1.12200797e-03, 1.34747833e-03, 1.11917020e-03],
[2.48272960e-02, 1.00000000e+00, 1.42630716e-02, ...,
4.07129084e-02, 1.11032424e-03, 9.22197982e-04],
[4.22228416e-02, 1.42630716e-02, 1.00000000e+00, ...,
1.15436886e-03, 5.90455012e-02, 1.15144924e-03],
...,
[1.12200797e-03, 4.07129084e-02, 1.15436886e-03, ...,
1.00000000e+00, 1.17574481e-03, 5.83911241e-02],
[1.34747833e-03, 1.11032424e-03, 5.90455012e-02, ...,
1.17574481e-03, 1.00000000e+00, 1.17277112e-03],
[1.11917020e-03, 9.22197982e-04, 1.15144924e-03, ...,
5.83911241e-02, 1.17277112e-03, 1.00000000e+00]])

cosine_sim.shape