Mini Project 10: Movie Recommendation System

Step 1: Import library

```
import pandas as pd
```

```
import numpy as np
```

Step 2: Import Data

```
df=pd.read_csv('https://github.com/YBI-Foundation/Dataset/raw/main/Movies%20Recommendation.csv')
```

```
df.head()
```

	Movie_ID	Movie_Title	Movie_Genre	Movie_Language	Movie_Budget	Movie_Popularity	Movie_Release_Date	Movie_Revenue	Movie
0	1	Four Rooms	Crime Comedy	en	4000000	22.876230	09-12-1995	4300000	
1	2	Star Wars	Adventure Action Science Fiction	en	11000000	126.393695	25-05-1977	775398007	
2	3	Finding Nemo	Animation Family	en	94000000	85.688789	30-05-2003	940335536	
		Forrest	Comedy						

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4760 entries, 0 to 4759
Data columns (total 21 columns):
 #   Column                  Non-Null Count  Dtype  
---  ------                  --------------  -----  
 0   Movie_ID                4760 non-null   int64  
 1   Movie_Title             4760 non-null   object 
 2   Movie_Genre             4760 non-null   object 
 3   Movie_Language          4760 non-null   object 
 4   Movie_Budget            4760 non-null   int64  
 5   Movie_Popularity        4760 non-null   float64
 6   Movie_Release_Date      4760 non-null   object 
 7   Movie_Revenue           4760 non-null   int64  
```

```
 8   Movie_Runtime             4758 non-null   float64
 9   Movie_Vote                4760 non-null   float64
 10  Movie_Vote_Count          4760 non-null   int64
 11  Movie_Homepage            1699 non-null   object
 12  Movie_Keywords            4373 non-null   object
 13  Movie_Overview            4757 non-null   object
 14  Movie_Production_House    4760 non-null   object
 15  Movie_Production_Country  4760 non-null   object
 16  Movie_Spoken_Language     4760 non-null   object
 17  Movie_Tagline             3942 non-null   object
 18  Movie_Cast                4733 non-null   object
 19  Movie_Crew                4760 non-null   object
 20  Movie_Director            4738 non-null   object
dtypes: float64(3), int64(4), object(14)
memory usage: 781.1+ KB
```

```
df.shape
```

```
(4760, 21)
```

```
df.columns
```

```
Index(['Movie_ID', 'Movie_Title', 'Movie_Genre', 'Movie_Language',
       'Movie_Budget', 'Movie_Popularity', 'Movie_Release_Date',
       'Movie_Revenue', 'Movie_Runtime', 'Movie_Vote', 'Movie_Vote_Count',
       'Movie_Homepage', 'Movie_Keywords', 'Movie_Overview',
       'Movie_Production_House', 'Movie_Production_Country',
       'Movie_Spoken_Language', 'Movie_Tagline', 'Movie_Cast', 'Movie_Crew',
       'Movie_Director'],
      dtype='object')
```

Step 3: Feature Selection

```
df_features=df[['Movie_Genre','Movie_Overview','Movie_Tagline','Movie_Cast', 'Movie_Director']].fillna('')
```

```
df_features.shape
```

```
(4760, 5)
```

```
df_features
```

	Movie_Genre	Movie_Overview	Movie_Tagline	Movie_Cast	Movie_Director
0	Crime Comedy	It's Ted the Bellhop's first night on the job....	Twelve outrageous guests. Four scandalous requ...	Tim Roth Antonio Banderas Jennifer Beals Madon...	Allison Anders
1	Adventure Action Science Fiction	Princess Leia is captured and held hostage by ...	A long time ago in a galaxy far, far away...	Mark Hamill Harrison Ford Carrie Fisher Peter ...	George Lucas
2	Animation Family	Nemo, an adventurous young clownfish, is unexp...	There are 3.7 trillion fish in the ocean, they...	Albert Brooks Ellen DeGeneres Alexander Gould ...	Andrew Stanton
3	Comedy Drama Romance	A man with a low IQ has accomplished great thi...	The world will never be the same, once you've ...	Tom Hanks Robin Wright Gary Sinise Mykelti Wil...	Robert Zemeckis
4	Drama	Lester Burnham, a depressed suburban father in...	Look closer.	Kevin Spacey Annette Bening Thora Birch Wes Be...	Sam Mendes
...	...	...	...	...	...
		A Broadway producer	The hot spot where	Lisa Hart Carroll	

```
X=df_features['Movie_Genre'] + ' ' + df_features['Movie_Overview']+' '+df_features['Movie_Tagline']+' '+df_features['Movie_Cast']+' '
```

```
X
```

```
0       Crime Comedy It's Ted the Bellhop's first nigh...
```

```
1       Adventure Action Science Fiction Princess Leia...
2       Animation Family Nemo, an adventurous young cl...
3       Comedy Drama Romance A man with a low IQ has a...
4       Drama Lester Burnham, a depressed suburban fat...
                            ...
4755    Horror A Broadway producer puts on a play with...
4756    Comedy Family Drama In 1979, an Indian family ...
4757    Thriller Drama After sneaking to a party with ...
4758    Family After being estranged since her mother'...
4759      Documentary   Tony Oppedisano Simon Napier-Bell
Length: 4760, dtype: object
```

```
X.shape
```

```
(4760,)
```

Step 4: Text Conversion to Tokens

[ ] ↳ *5 cells hidden*

Step 5: Similarity Score

[ ] ↳ *4 cells hidden*

Step 6: Accept Movie Name as Input and Validate for Closest Match

[ ] ↳ *8 cells hidden*

Step 7: Sorting all Movies based on Recommendation Score(r_score)

[ ] ↳ *2 cells hidden*

Step 8: Top 10 Movie Recommendation System

[ ] ↳ *1 cell hidden*

Link of the same:

https://colab.research.google.com/drive/1_JXMpac7LdKDZJ4LLpqr3a27FaWfDUyN?usp=sharing