

# movierecommendation

April 4, 2025

Develop a movie recommendation model using the scikit-learn library in python.

```
[1]: from sklearn.metrics.pairwise import cosine_similarity
import pandas as pd
import numpy as np
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.metrics.pairwise import cosine_similarity
```

```
[2]: df = pd.read_csv("https://raw.githubusercontent.com/rashida048/
↳Some-NLP-Projects/master/movie_dataset.csv")
```

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

```
[4]:
```

	index	budget	genres \
0	0	237000000	Action Adventure Fantasy Science Fiction
1	1	300000000	Adventure Fantasy Action
2	2	245000000	Action Adventure Crime
3	3	250000000	Action Crime Drama Thriller
4	4	260000000	Action Adventure Science Fiction

  

	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	culture clash future space war space colony so...	en
1	ocean drug abuse exotic island east india trad...	en
2	spy based on novel secret agent sequel mi6	en
3	dc comics crime fighter terrorist secret ident...	en
4	based on novel mars medallion space travel pri...	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	...	runtime	\
0	In the 22nd century, a paraplegic Marine is di...	150.437577	...	162.0	
1	Captain Barbossa, long believed to be dead, ha...	139.082615	...	169.0	
2	A cryptic message from Bond's past sends him o...	107.376788	...	148.0	
3	Following the death of District Attorney Harve...	112.312950	...	165.0	
4	John Carter is a war-weary, former military ca...	43.926995	...	132.0	

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

	cast	\
0	Sam Worthington Zoe Saldana Sigourney Weaver S...	
1	Johnny Depp Orlando Bloom Keira Knightley Stel...	
2	Daniel Craig Christoph Waltz L\u00e9a Seydoux ...	
3	Christian Bale Michael Caine Gary Oldman Anne ...	
4	Taylor Kitsch Lynn Collins Samantha Morton Wil...	

	crew	director
0	[{'name': 'Stephen E. Rivkin', 'gender': 0, 'd...	James Cameron
1	[{'name': 'Dariusz Wolski', 'gender': 2, 'depa...	Gore Verbinski
2	[{'name': 'Thomas Newman', 'gender': 2, 'depar...	Sam Mendes
3	[{'name': 'Hans Zimmer', 'gender': 2, 'departm...	Christopher Nolan
4	[{'name': 'Andrew Stanton', 'gender': 2, 'depa...	Andrew Stanton

[5 rows x 24 columns]

```

[15]: features = ['keywords','cast','genres','director']

[7]: def combine_features(row):
      return row['keywords']+" "+row['cast']+" "+row['genres']+" "+row['director']

[8]: for feature in features:
      df[feature] = df[feature].fillna('')

      df["combined_features"] = df.apply(combine_features,axis=1)

[9]: cv = CountVectorizer()
      count_matrix = cv.fit_transform(df["combined_features"])

[10]: cosine_sim = cosine_similarity(count_matrix)

[11]: def get_title_from_index(index):
      return df[df.index == index]["title"].values[0]
      def get_index_from_title(title):
      return df[df.title == title]["index"].values[0]

[12]: movie_user_likes = "Avatar"
      movie_index = get_index_from_title(movie_user_likes)
      similar_movies = list(enumerate(cosine_sim[movie_index]))

[13]: sorted_similar_movies = sorted(similar_movies,key=lambda x:x[1],reverse=True)[1:
      ↪]

[14]: i=0
      print("Top 5 similar movies to "+movie_user_likes+" are:\n")
      for element in sorted_similar_movies:
          print(get_title_from_index(element[0]))
          i=i+1
          if i>5:
              break

```

Top 5 similar movies to Avatar are:

Guardians of the Galaxy  
 Aliens  
 Star Wars: Clone Wars: Volume 1  
 Star Trek Into Darkness  
 Star Trek Beyond  
 Alien

[ ]:

[ ]: