CountVectorizer in scikit-learn

The code then gets the index of the movie that the user likes (`movie_index`) and finds the list of similar movies (`similar_movies`) by enumerating over the row of the cosine similarity matrix corresponding to the movie index. Finally, the list of similar movies is sorted in descending order of similarity score (cosine similarity).



The `get_index_from_title` function takes in a movie title as an input argument and returns the index of the movie in the `df` dataframe. It does this by filtering the dataframe using the movie title, selecting the "index" column, and then returning the first value of this column using the `values` attribute.

For example, if you have a dataframe `df` with columns "title" and "index", and you want to get the index of the movie with title "Avatar", you can use the following code:

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

movie_index = get_index_from_title("Avatar")
```

This will return the index of the movie with the title "Avatar" in the dataframe. Note that this function will only work if the movie title is unique in the dataframe. If there are multiple movies with the same title, this function will only return the index of the first movie with that title.

e why is this piece of code important

This piece of code is important because it allows you to get the index of a movie in the dataframe based on its title. This index is used to access the row of the cosine similarity matrix corresponding to the movie, and to get the list of similar movies.

For example, in the code you provided, the index of the movie that the user likes is obtained using the `get_index_from_title` function. This index is then used to get the row of the cosine similarity matrix corresponding to the movie:

```
movie_index = get_index_from_title(movie_user_likes)
similar_movies = list(enumerate(cosine_sim[movie_index]))
```

The `similar_movies` list is then sorted in descending order of similarity score (cosine similarity) and the top-n movies with the highest similarity scores are recommended to the user.

Without the `get_index_from_title` function, it would not be possible to get the index of the movie that the user likes and to find the list of similar movies.

e can't one use just the title to do that

It is not possible to use just the title to get the list of similar movies, because the cosine similarity matrix is based on the count matrix

e >

ChatGPT Dec 15 Version. Free Research Preview. Our goal is to make Al systems more natural and safe to interact with. Your feedback will help us improve.

6 9

凸 豆

Copy code

Copy code