

Similaridade dos Cossenos

Gustavo Pinho - Matheus ILeck

May 2023

1 Bibliotecas

- `import pandas as pd`: Importa a biblioteca Pandas e a renomeia como "pd".
- `from sklearn.feature_extraction.text import CountVectorizer`: Importa a classe CountVectorizer da biblioteca Scikit-learn (sklearn) para criar uma matriz de frequência de termos.
- `from sklearn.metrics.pairwise import cosine_similarity`: Importa a função cosinesimilarity da biblioteca Scikit-learn para calcular a similaridade do cosseno entre duas matrizes.

2 Leitura do arquivo

- `netflix_movies = pd.read_csv('/content/n_movies.csv')`: Lê o arquivo csv 'n_movies.csv' e o armazena em um DataFrame chamado "netflix_movies".
- `netflix_movies = netflix_movies[['title', 'certificate', 'description', 'genre']].dropna()`: Seleciona apenas as colunas relevantes (título, certificado, descrição e gênero) do DataFrame "netflix_movies" e remove as linhas que possuem valores nulos.

3 Criação da matriz de frequência de termos e matriz de similaridade

- `vectorizer = CountVectorizer().fit_transform(netflix_movies['certificate'] + netflix_movies['description'])`: Cria um objeto CountVectorizer e aplica-o aos dados de certificado, descrição e gênero do DataFrame "netflix_movies", criando uma matriz de frequência de termos.
- `cosine_similarities = cosine_similarity(vectorizer)`: Calcula a similaridade do cosseno entre a matriz de frequência de termos e ela mesma, produzindo uma matriz de similaridade.

4 Seleção do filme de referência e filmes similares

- `reference_movie = input('Enter the name of the reference movie: '):`
Solicita ao usuário que digite o nome do filme de referência.
- `reference_index = netflix_movies[netflix_movies['title'] == reference_movie].index[0]:`
Localiza o índice do filme de referência no DataFrame "netflix_movies".
- `similarities = cosine_similarities[reference_index]:` Selecciona a linha correspondente ao filme de referência na matriz de similaridade, que contém as similaridades do cosseno entre o filme de referência e todos os outros filmes.
- `similar_movies_indices = similarities.argsort()[-11:-1][::-1]:`
Ordena as similaridades em ordem decrescente e selecciona os índices dos 10 filmes mais similares (excluindo o próprio filme de referência).
- `similar_movies = netflix_movies.iloc[similar_movies_indices]:` Selecciona as informações dos 10 filmes mais similares no DataFrame "netflix_movies".

5 Impressão dos resultados

- `print('Recommended movies for', reference_movie, ':'):` Imprime o título do filme de referência.
- `print(similar_movies):` Imprime as informações dos 10 filmes mais similares.