Netflix Recommendation System

This code creates a recommendation system using the Netflix dataset. The dataset includes information such as titles, country, cast, description, director etc.

It includes a basic recommendation model which just uses title of the movie and more appropriate function which includes a list of features found in the dataset.

Data Preprocessing

The code loads the data from a CSV file and performs some preprocessing steps. Missing values are filled with the string 'missing'. The 'listed_in' column is renamed to 'Genre' and its values are split by comma and only the first value is kept.

TF-IDF Vectorization

A TF-IDF Vectorizer is defined to remove English stop words such as 'the' and 'a'. This is important as stop words are repeated and can affect statistical analysis. The 'description' column is filled with an empty string if it contains NaN values. The TF-IDF matrix is then constructed by fitting and transforming the data.

Cosine Similarity

The cosine similarity between the descriptions is computed using the linear_kernel function from scikit-learn. A reverse map of indices and movie titles is also constructed.

Recommendation Function

The get_recommendations function takes in a title and returns 10 recommended movies based on their similarity to the input title. It does this by getting the pairwise similarity scores of all movies with the input movie and sorting them based on their similarity scores.