Creating a Netflix Recommendation Engine Using NLP
Juliana McCausland

The goal of this project was to create a netflix recommendation engine using content-based filtering. I attempted to discover similarities between descriptions of nearly 8,000 Netflix movies and shows to produce a recommendation system that suggests the top ten most relevant shows/movies to a selected title. The engine originally produced weak results, but after concatenating various features and calculating cosine similarities on the resulting feature, I was able to improve the engine to an acceptable level.

Design
There are tons of video streaming platforms out there, but a new startup aims to provide the same service with a very different approach. This startup is founded on the principle that quality services can be provided without collecting data about individual people. This company hopes to pave the way into a new era of privacy beyond the scope of video streaming. The purpose of this project is to help said startup develop methods for creating recommendation engines using privacy-friendly techniques (e.g. they won't allow you to "like" content on their platform). Because they are just starting out, they want to use data from an existing platform before launching theirs.

Data
I used a dataset I found on Kaggle that contains information for nearly 8,000 Netflix movies and TV shows. The dataset contains columns for the titles, actors, directors, duration, descriptions, and more.

Algorithms
I created a count vectorizer and an TF-IDF matrix on the movie/show descriptions after preprocessing the text. I then used the TF-IDF matrix in conjunction with NMF to perform some topic modeling. For the actual recommendation system, I calculated cosine similarities using sklearn's linear kernel function on the TF-IDF matrix of the descriptions. I improved the recommendation engine by creating a new column in the data that was a combination of the descriptions, ratings, actors, and categories, and performed the same cosine similarity function on a TF-IDF matrix of the combined text. This resulted in a more accurate engine. I then used Stremlit to create a web application for the engine.

Tools
I primarily used NLTK, pandas, scikit-learn, numpy, WordCloud, regex, streamlit

Communication
Below is a screenshot of the recommendation engine web application:

Netflix Recommendation Engine

Enter the name of a netflix show or movie you like:

The Universe

Recommend

You might like:

1. The Universe: Ancient Mysteries Solved

2. Babies

3. Connected

4. Moving Art

5. Absurd Planet

6. The Surgeon's Cut

7. Tiny Creatures

8. The Beginning of Life: The Series

9. Blue Planet II

10. The Mind, Explained