Report: Recommendition Engine

Reem Abanmi - Norah Tamsan

ABSTRACT

On the Internet, where the number of choices is overwhelming, there is need to filter, prioritize and efficiently deliver relevant information in order to alleviate the problem of information overload, which has created a potential problem to many Internet users. Recommender systems solve this problem by searching through large volume of dynamically generated information to provide users with personalized content and services. In this project we well build a recommendation system for streaming services using NLP.

DESIGN

Using the dataset provided from Kaggle we create a content-based recommendation engine using NLP unsupervised learning. After writing the name of a movie or show, the engine will return ten relevant movies and shows using cosine similarities of the text descriptions of all movies/tv shows provided in the database.

DATA

The dataset has over 8000 movies or tv shows. This tabular dataset consists of listings of movies and tv shows, along with details such as cast, directors, ratings, release year, duration, and description. Our focus is on the description as we are going to perform the NLP and unsupervised learning on it.

ALGORITHM

1. We started by loading the data from the csv file.
2. Using NLP to understand and clean the description text.
3. we vectorize the description text using TfidfVectorizer.
4. after vectorization we started topic modeling using LSA and NMF.
5. lastly, we perform Cosine similarity to build recommendation engine.

TOOIS

- Seaborn library
- Pandas library
- Matplot library
- Scikit-learn library
- Gensim library
- spaCy library
- NLTK library
- Regular Expression library
- Wordcloud library
- Streamlit library