Project Proposal: Movie Recommendation System using Python and Streamlit(web)

Sharath Kumar Karnati (011852253)
CptS 475/575: Data Science

October 7, 2024

Project Overview

For this project, I will develop a movie recommendation system based solely on content-based filtering. The system will analyze movie features such as genre, cast, and director to recommend movies that are similar to ones the user has previously liked and want to explore more on that content. The recommendation engine will be deployed as a web application using Streamlit, providing users with personalized movie suggestions.

Methods

The system will employ a content-based filtering approach, which focuses on the properties of movies to make recommendations. The primary steps include:

- **Feature Extraction:** Metadata like genre, cast, directors, and keywords from movie plots will be used to construct feature vectors for each movie.
- **Similarity Calculation:** Cosine similarity will be calculated between movies based on their feature vectors. This will allow the system to measure how similar one movie is to another.
- **Recommendation Generation:** Using the similarity scores, a ranked list of top-5 movie recommendations will be presented to the user based on their searched movie.

Data

The dataset for this project was sourced from **Kaggle**, and it includes a comprehensive collection of movie metadata such as titles, genres, cast, and ratings. The dataset has already been preprocessed to handle missing values and standardize metadata for each movie. Additional external metadata from the Kaggle TMDB Dataset may be used to further enrich the features used in the model.

Background Work

Content-based recommendation systems have been used extensively in various domains. For instance, music recommendation services like Pandora use content-based filtering to suggest songs based on attributes such as genre and mood. In this project, the same principles will be applied to recommend movies based on their metadata. This method is particularly useful for providing recommendations for new movies, as it doesn't rely on user ratings but instead on movie attributes.

Tentative Plan

a) **Week 1-2:** Preprocess and explore the Kaggle movie dataset to understand the key features.

b) **Week 3-4:** Implement the content-based filtering model using cosine similarity.

c) **Week 5-6:** Build and integrate the recommendation system into a Streamlit web application.

d) **Week 7:** Conduct user testing, refine the recommendation model based on working of it.

e) **Week 8:** Finalize and deploy the web application.

Team Members

- Sharath Kumar Karnati(011852252)(No other team members, as I am gonna do it alone!)