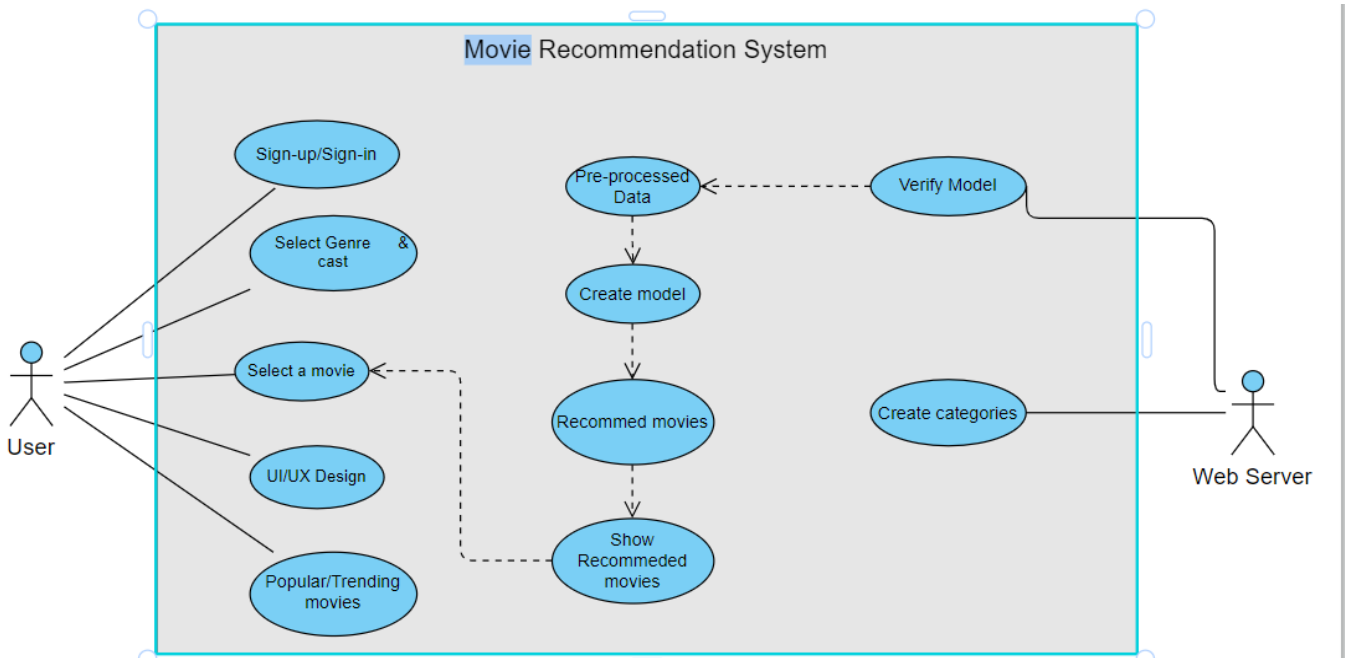
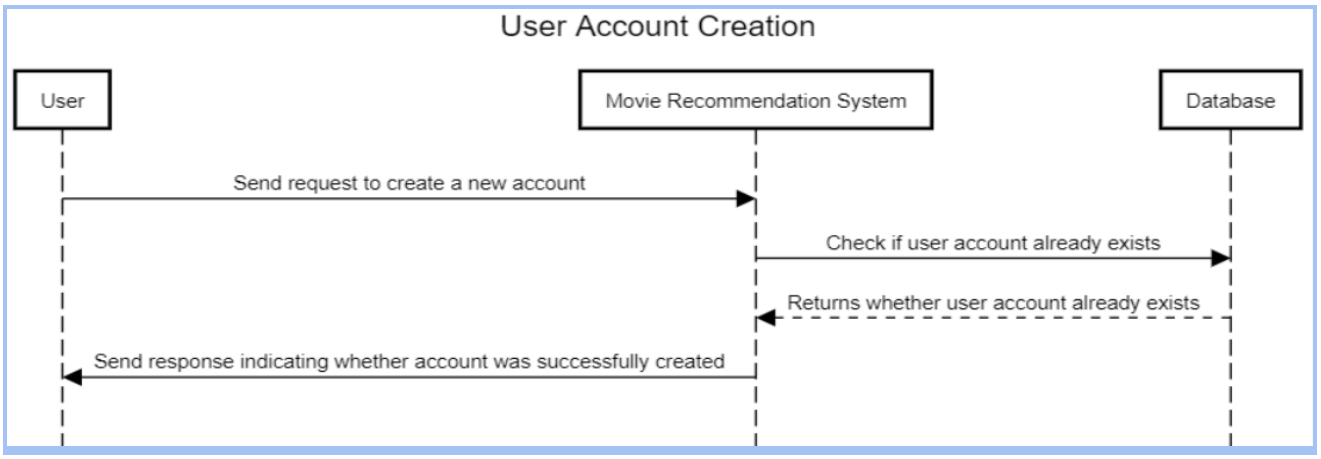
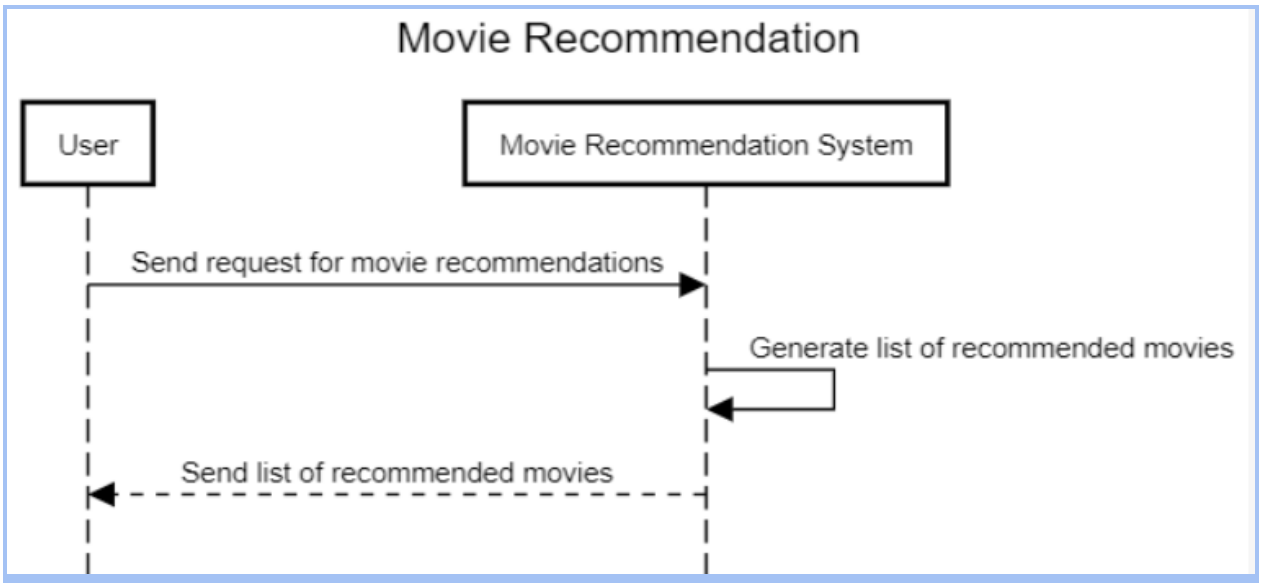
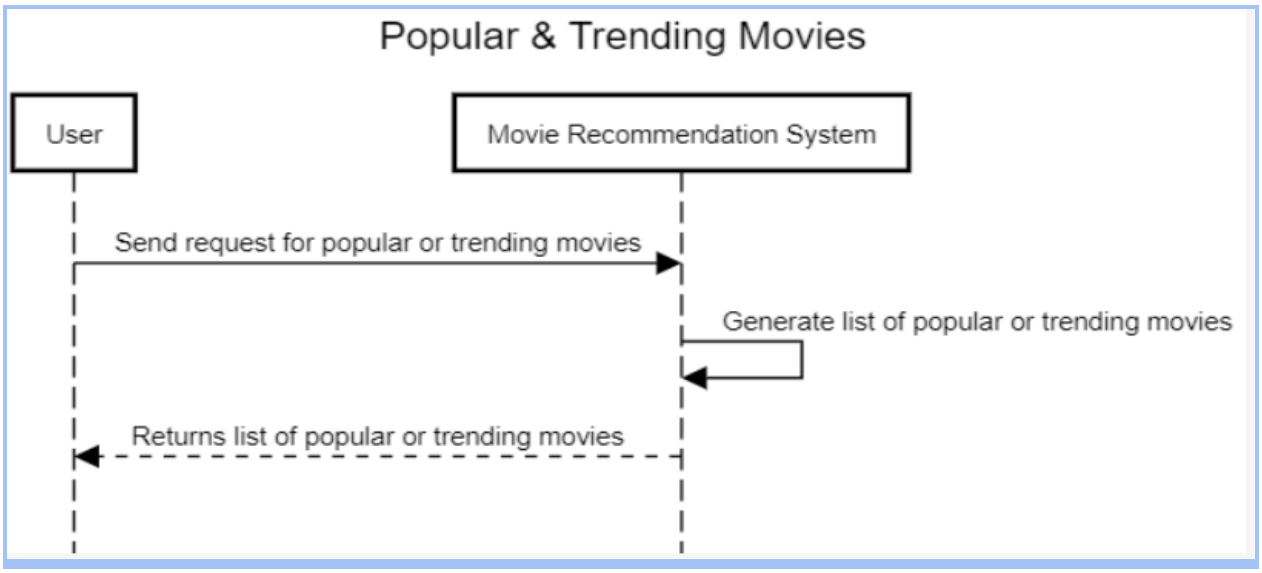
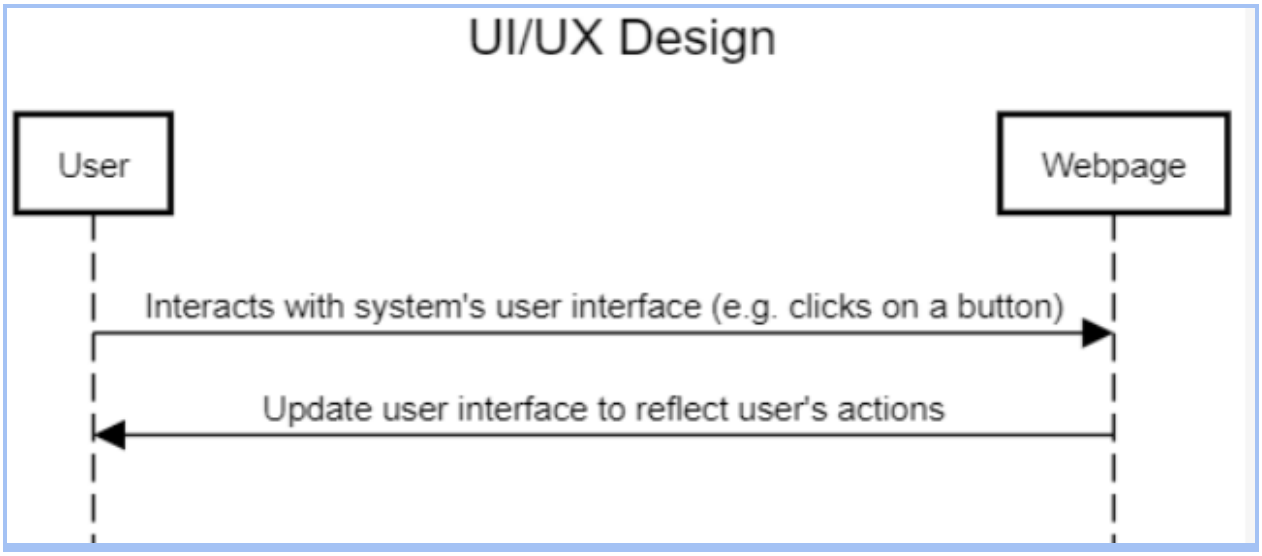

Movie Recommendation System
Sign-up/Sign-in
Select Genre & cast
Select a movie
UI/UX Design
Popular/Trending movies
Pre-processed Data
Create model
Recommed movies
Show Recommeded movies
Verify Model
Create categories
User
Web Server


User Account Creation
User
Movie Recommendation System
Database
Send request to create a new account
Check if user account already exists
Returns whether user account already exists
Send response indicating whether account was successfully created


Movie Recommendation
User
Movie Recommendation System
Send request for movie recommendations
Generate list of recommended movies
Send list of recommended movies





Feature-based description report of dataset:

TMDB 5000 movie dataset & TMDB 5000 credits dataset

The dataset is a collection of 5000 movies with credits data from the popular movie database TMDb. It contains information about the movies' title, genre, budget, release date, and popularity, among other features. In this feature-based description report, we will focus on 10 key features of the dataset:

1. Movie_id: This feature is a unique identifier for each movie in the dataset.
2. Title: This feature contains the title of each movie in the dataset.
3. Tagline: This feature contains a short tagline or slogan for each movie in the dataset.
4. Overview: Contains a brief summary of the plot for each movie in the dataset.

5. Genres: Contains a list of genres associated with each movie in the dataset. Each movie can be associated with multiple genres.

6. Keywords: This feature contains a list of keywords associated with each movie in the dataset. These keywords describe the themes, concepts, and ideas presented in each movie.

7. Cast: Contains list of actors & actresses who appear in each movie in the dataset.

8. Crew: This feature contains a list of crew members who worked on each movie in the dataset, including directors, writers, producers, and others.

9. Popularity: This feature measures the popularity of each movie in the dataset, based on user ratings, views, and other factors.

10. Release_date: This feature contains the date on which each movie was released.

Data Cleaning: First, we clean the dataset by removing any missing values or duplicate entries and check for any inconsistencies in the data & fix them accordingly.

Exploratory Data Analysis: Next, we perform EDA on the dataset to gain insights. We start by analyzing the distribution of various features in the dataset, such as genres, keywords, and release dates.

Feature Engineering: Based on the insights gained from the exploratory data analysis, we create new features and transform existing features to improve the performance of our machine learning model. For example, we can create a new feature that measures the similarity between different movies based on their genres and keywords.

Machine Learning Models: Finally, we use the CountVectorizer machine learning model to make predictions about the popularity and success of movies in the dataset. We train the models using a combination of features from the dataset.

Conclusion: The TMDB 5000 Movie Dataset is a rich and comprehensive dataset that can be used to analyze trends in the movie industry and make predictions about the success of movies.

More details of the dataset:

https://www.kaggle.com/datasets/tmdb/tmdb-movie-metadata