Film Recommendation System

Group name: By a time zone united
Name: Olena Panchenko
Email: hel.panchenko@gmail.com
Country: United Kingdom
College/Company: Taras Shevchenko National University of Kyiv
Specialisation: Data Science
Link:

Problem Description

The video-on-demand streaming service is looking to develop a machine learning algorithm that can predict which movies a user will enjoy based on various factors such as genre, online ratings, and previous decisions. The primary objective is to create a system for movie recommendations.

Business understanding

Objective: The goal of our team is to create a personalized movie recommendation system that enhances user engagement, satisfaction, and retention. Below is a list of key stakeholders who would benefit from this model:

Key Stakeholders:

1. **Users**: Consumers of streaming platforms who seek personalized content recommendations to discover new titles and highly rated movies aligned with their interests and preferences.

2. **Streaming Platforms**: Any streaming platform aiming to increase user satisfaction, retention, and overall consumption by offering accurate and effective content recommendations.

Scope:

• Our recommendation system will leverage the data from "https://grouplens.org/datasets/movielens/latest/". This dataset consists of 6 files with different features which include but are not limited to movie titles, tags, ratings, genres, users, etc.

• Machine learning algorithms and collaborative filtering techniques will be used on movie ratings and user preferences, to generate personalized recommendations.

• The system will operate across web and mobile to ensure a seamless user experience.

Key Functional Requirements:

• **User Profiling**: Develop profiles for individual users based on their viewing history and ratings

• **Movie Representation**: Here we are going to analyse content attributes such as genre and tags to accurately represent each title in the recommendation system.

• **Recommendation Generation**: We would employ algorithms such as collaborative filtering, content-based filtering, and hybrid models to generate personalised recommendations for users.

• **Real-Time Updates**: We would implement mechanisms to continuously update user profiles and recommendations based on new ratings and movie additions.

• **Feedback Mechanism**: We would enable users to provide explicit feedback (ratings, likes, dislikes) to refine future recommendations and improve algorithm performance.

Key Non-Functional Requirements:

• **Scalability**: We will make sure the system can handle a growing user base and a vast library of content without compromising performance.

• **Accuracy**: We will aim for high accuracy in recommendation predictions to enhance user satisfaction.

• **Privacy**: We will implement robust privacy measures to protect user data and comply with relevant regulations (e.g., GDPR, CCPA).

• **Responsiveness**: We will deliver recommendations promptly, minimizing latency to enhance the user experience across different devices and network conditions.

Success Metrics:

• **Retention Rate**: We will monitor the percentage of users who continue their subscription on streaming platforms that will utilize our recommendation system over time, indicating the effectiveness of personalized recommendations in retaining customers.

• **Click-Through Rate (CTR**): We will evaluate the percentage of users who click on recommended titles, indicating the relevance and appeal of the recommendations.

• **User Satisfaction Surveys**: We will conduct periodic surveys to gather feedback on the recommendation system's performance and user satisfaction levels.

• **Content Consumption Diversity**: We will assess the variety of content consumed by users, aiming to ensure recommendations promote exploration and discovery across different genres.

Project life cycle

1. Business Understanding	13-19 March
2. Data Acquisition and Understanding	20-26 March
3. Data Cleaning and Processing	27 March - 2 April
4. Exploratory Data Analysis	3-9 April
5. EDA Presentation	10-16 April
6. Model Building and Evaluation	17-23 April
7. Project Presentation	24-30 April

Data Intake Report

Name: Film Recommendation System
Report date: 2024-03-19
Internship Batch: LIMSUM30
Version: 1.0
Data intake by: Daniel Emakporuena, Olena Panchenko, Tehesuma Imoro
Data intake reviewer: Data Glacier
Data storage location:
https://drive.google.com/file/d/1A5E6BSvJIdENAn2f2DUn-mKKBEw32ruK/view?usp=drive_link

Tabular data details: **genome-scores**

Total number of observations	18472128
Total number of files	1
Total number of features	3
Base format of the file	.csv
Size of the data	497 MB

Tabular data details: **genome-tags**

Total number of observations	1128
Total number of files	1
Total number of features	2
Base format of the file	.csv
Size of the data	17.6 KB

Tabular data details: **links**

Total number of observations	86537
Total number of files	1
Total number of features	3
Base format of the file	.csv
Size of the data	1.83 MB

Tabular data details: **movies**

Total number of observations	86537
Total number of files	1
Total number of features	3
Base format of the file	.csv
Size of the data	3.99 MB

Tabular data details: **ratings**

Total number of observations	33832162
Total number of files	1
Total number of features	4
Base format of the file	.csv
Size of the data	1.44 GB

Tabular data details: **tags**

Total number of observations	2328315
Total number of files	1
Total number of features	4
Base format of the file	.csv
Size of the data	81.4 MB