Personalized Recommendation System

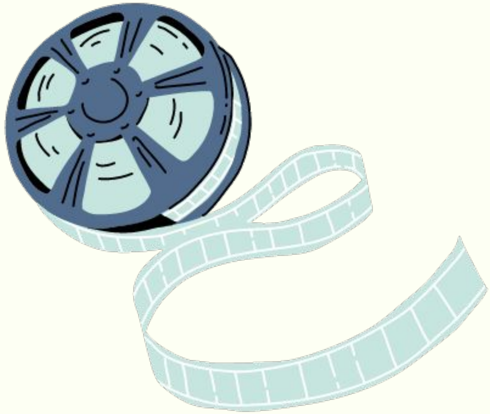
Under Supervision of AST ACADEMY

Team members:

- Ahmed Salah Emam
- Amir Tarek Ali
- Aya AbdElnaser Muhammed
- Dina Saed Farag
- Shadan Mohamed Aref
- Youssef Ezzeldin Mahmoud

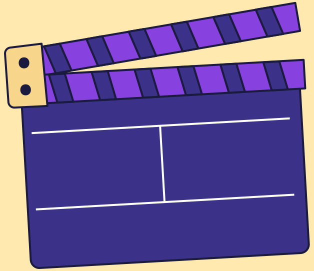
رواد مصر الرقمية

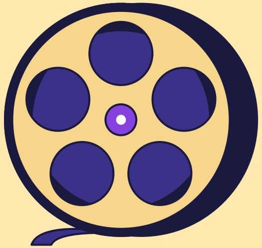
Table of contents



01

Introduction



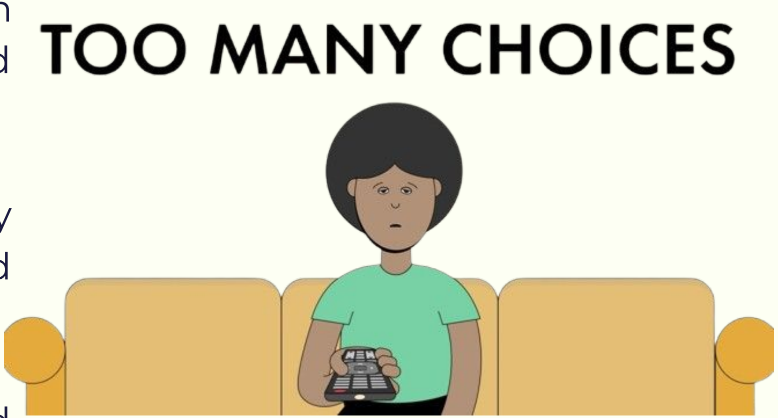


Problem Statement

- The number of movies and TV shows on platforms like Netflix and Disney has increased dramatically.

- Users face the challenge of having too many options, which leads to a lot of time being wasted searching for the right content.

- The traditional solution of browsing and searching is often inefficient and leads to user fatigue.



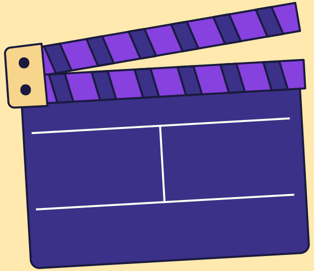
Recommendation Solution

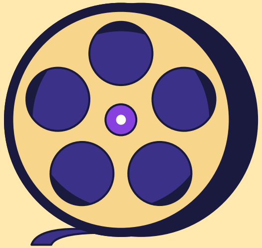
- We have developed a personalized recommendation system based on content characteristics such as genre, actors, and directors.

- This system provides accurate and personalized recommendations, saving users time and increasing their satisfaction with the experience.

02

Data Collection and Preprocessing





Data sources

Netflix Data

- **Description:** We utilized Netflix's publicly available dataset, which includes ratings,title, and metadata for a vast collection of movies and TV shows.

Disney+ Data

- **Description**: Similar to Netflix, we accessed Disney+'s public dataset, focusing on Disney-owned content, including Marvel, Pixar, Star Wars, and more.

TMDB (The Movie Database) API

- **Description**: To enrich our dataset with additional information, we leveraged the TMDB API to obtain detailed metadata such as genres, posters, overviews, and ratings and Trailers for movies and TV shows.

Challenges faced in data collection and solution

Handling Duplicates

- **Removed Duplicates :**

  Detected and removed duplicate content based on title, release year, rating and we find there **are 23 rows that Duplicated Then Drop Them**

- **Ensured Data Quality:**

  Improved data accuracy and consistency by eliminating redundant entries.

Challenges faced in data collection and solution

Handling Outlier

- **Corrected Outlier:**

  Detected and corrected an outlier movie with an unusually long duration **(e.g., "312 min") by updating it to a more plausible value (e.g., "90 min").**

- **Improved Data Accuracy:**

  Ensure reliable duration data for better recommendations

Challenges faced in data collection and solution

Handling Missing Values

- **Mode Imputation:**

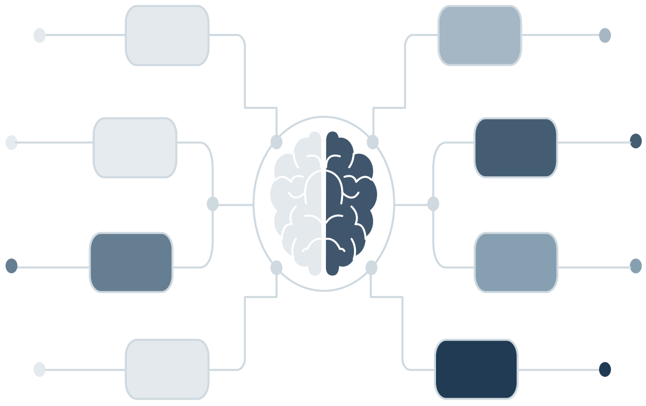
  Filled missing values in the "country" column with the most frequent value (mode) to preserve data integrity.

- **Consistent Handling:**

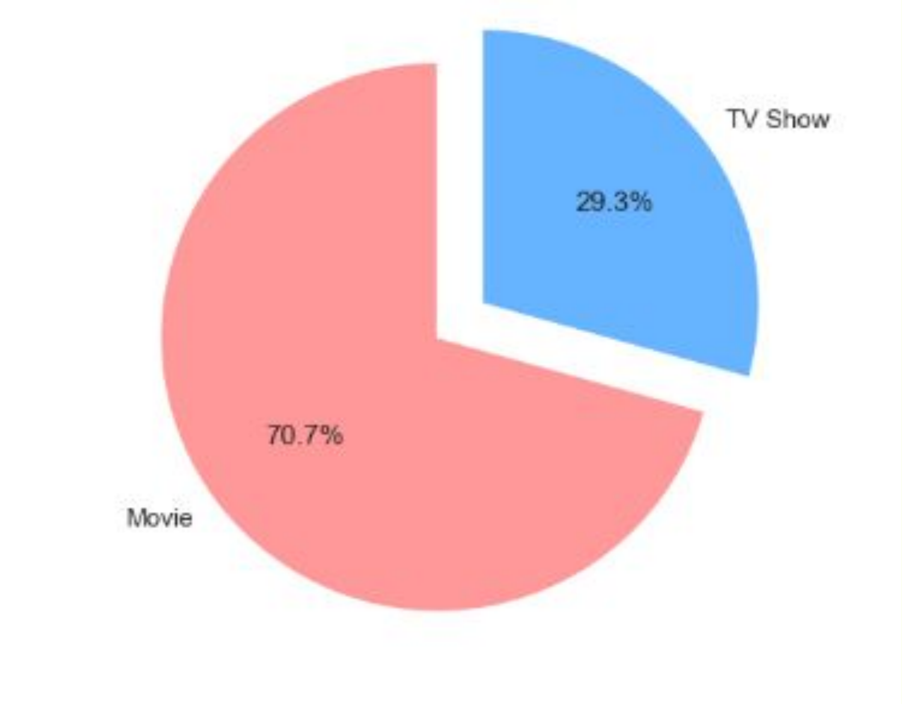
  Replaced NaN values in the "cast" and "director" columns with "No Data" for consistent representation and that we use it appears in part of modeling and handle it

03

Exploratory Data Analysis (EDA)

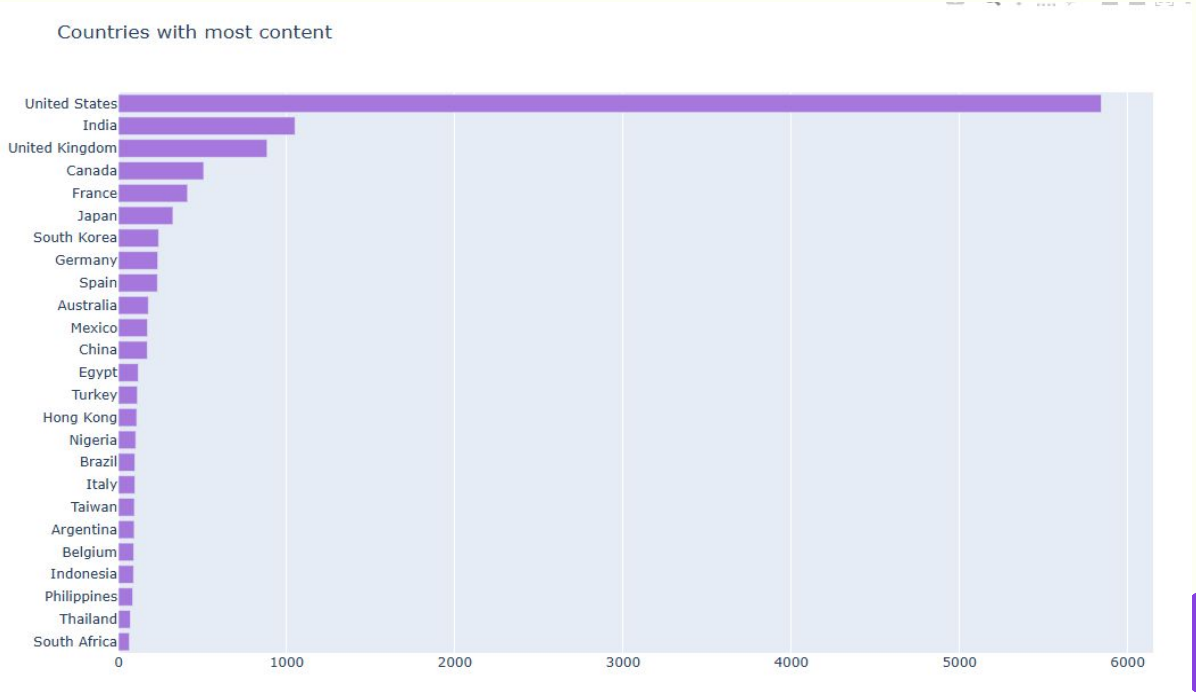


- Ratio of movies and TV shows in the concatenated Data and from that we find most data is Movies

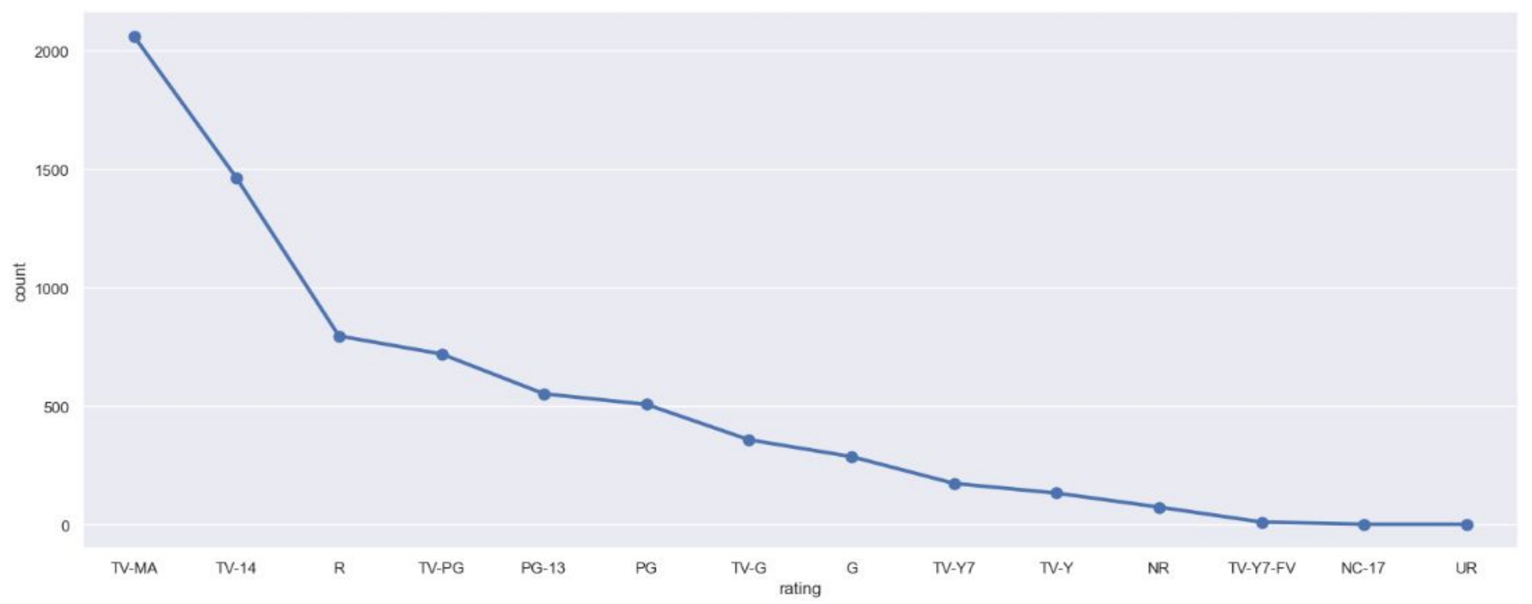


- **Horizontal bar chart to visualize the number of content pieces by country looks great and get most five counters are :**

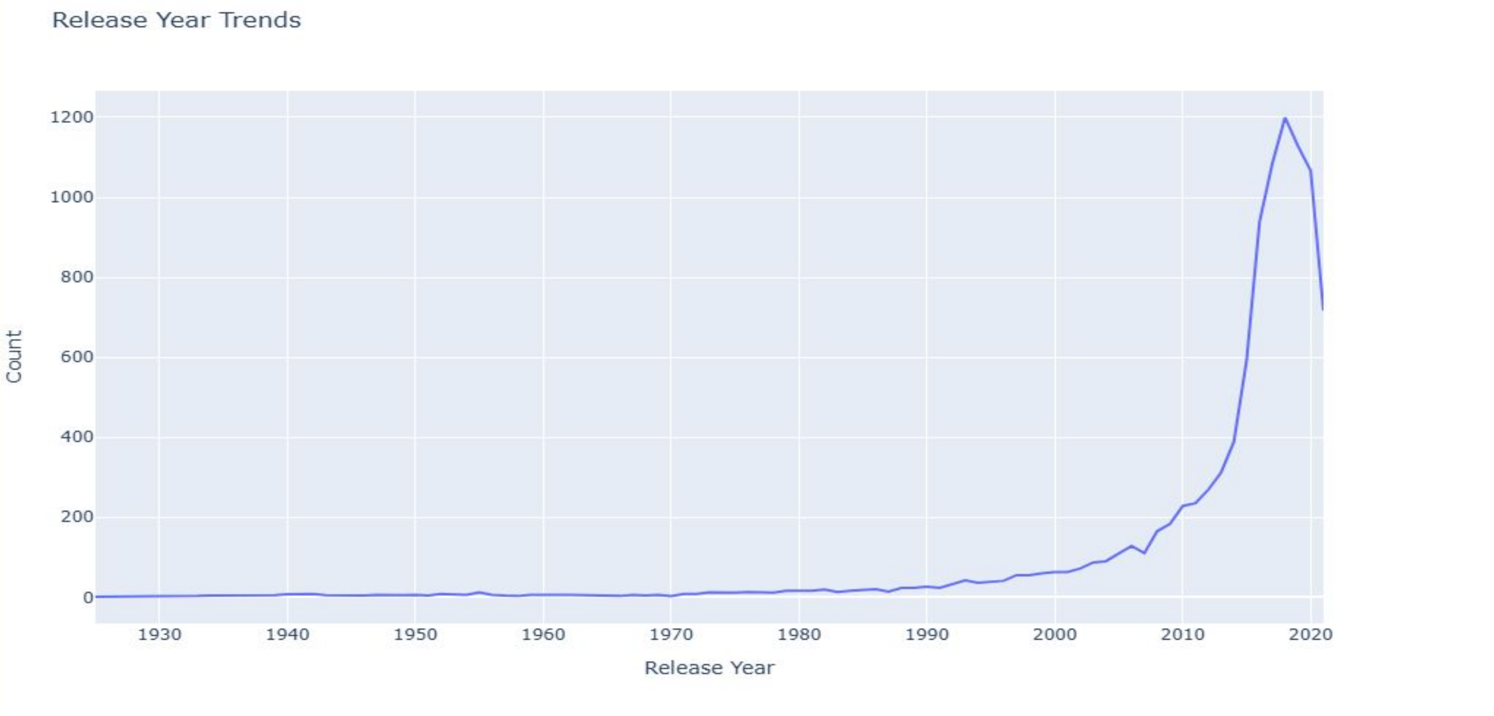
1- Usa
2- India
3- United kingdom
4- Canada
5- Franca



- **We get sum insight from that most movies Rating is TV-MA and that mean :** Suitable for mature audiences
- **Audience**:
  Adults, typically 17 years and older.

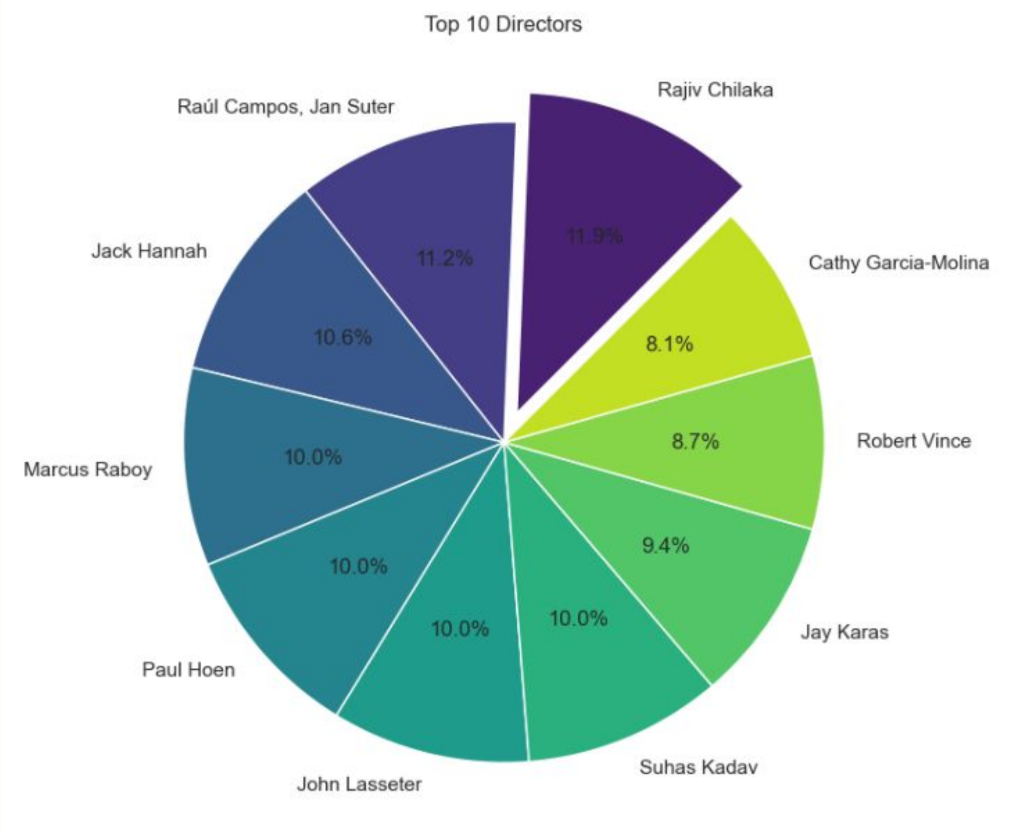


As we Say before Most Trend Movies from 2013 to 2020

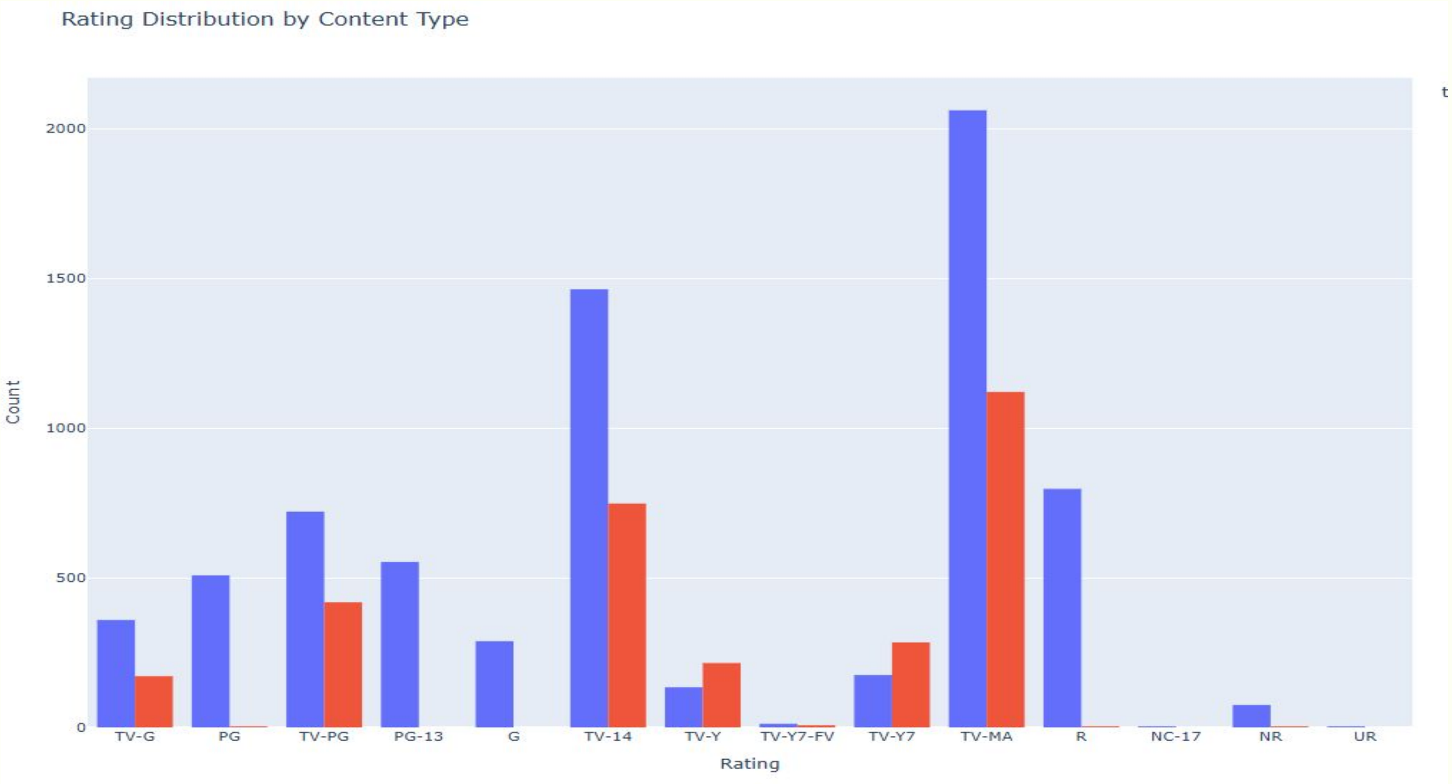


First top 10 directors are :

1 >> Rajiv Chilaka
2 >> Raúl Campos, Jan Suter
3 >> Jack Hannah
4 >> Marcus Raboy
5 >> Paul Hoen
6 >> John Lasseter
7 >> Suhas Kadav
8 >> Jay Karas
9 >> Robert Vince
10 >> Cathy Garcia-Molina



❑ **histogram that visualizes the distribution of ratings by content type (Movies and TV Shows)**



04

Machine Learning and Recommendation Modeling

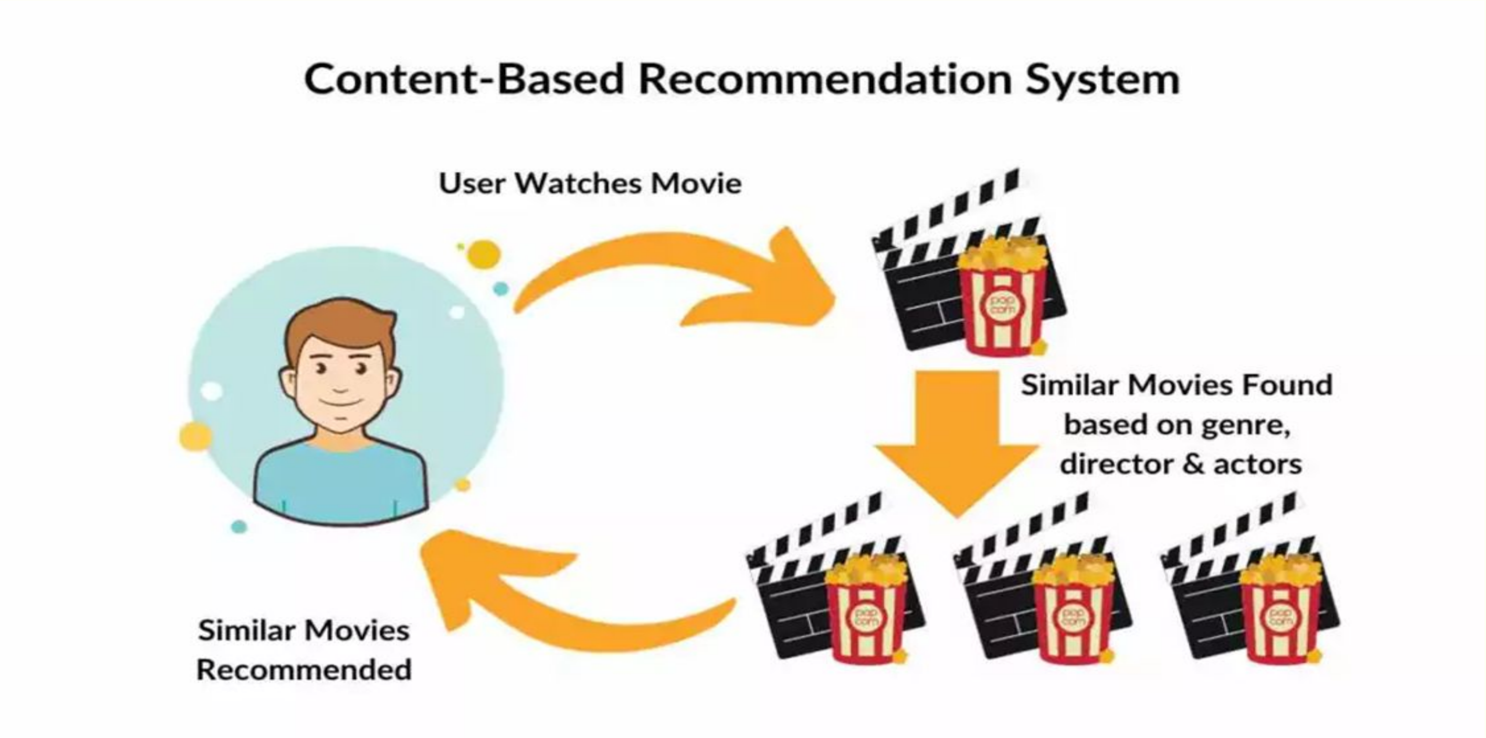
Suitable Models For Recommendation systems

- **Content-Based Filtering:** Recommends items based on the features of the items themselves and user preferences. Personalization based on specific user tastes. Use Cases such as, Suggesting movies based on previously liked genres and directors.

- **Collaborative Filtering:** Recommends items based on the preferences and behaviors of similar users. Can recommend items that a user might not have considered.Use Cases such as, Recommending movies based on what similar users have watched and enjoyed.

- **Demographic Filtering:** Recommends items based on demographic information (age, gender, location). Effective for targeting specific user segments. Use Cases such as, Recommending family-friendly movies to users identified as parents.

Content-Based Filtering for Movie Recommendation Systems



Models used

Content-Based Filtering for Movie Recommendation Systems

- **Content-based filtering** is a method used to recommend items based on their features. Unlike collaborative filtering, which relies on user interactions, content-based filtering focuses on the attributes of the items themselves.

- It plays a crucial role in recommendation systems by personalizing user experiences based on item attributes and user preferences.

Overview of how model works.

This process goes through main steps some of them are:

- ❑ **Feature Extraction:** Identifying key attributes of movies, like genre, director, and actors.

- ❑ **Recommendation Generation:** Matching movies to users by finding those that share similar features with their preferences."

Feature Extraction

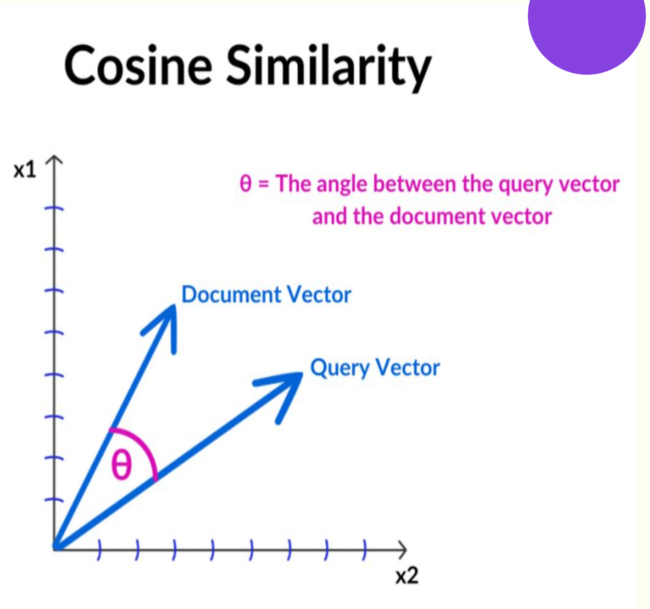
Feature extraction involves gathering various types of data:

- **Metadata:** Information such as title, genre, and release year.
- **Descriptive Features:** Plot summaries, cast details, and description. We can also utilize natural language processing to analyze textual data, enhancing our understanding of a movie's content."

Similarity Measurement

To recommend movies, we need to measure similarity between items. Common techniques include:

- ❑ **Cosine Similarity:** Measures the cosine of the angle between two non-zero vectors.
- ❑ In our project we used Cosine Similarity.



Advantages of Content-Based Filtering

Content-based filtering offers several benefits:

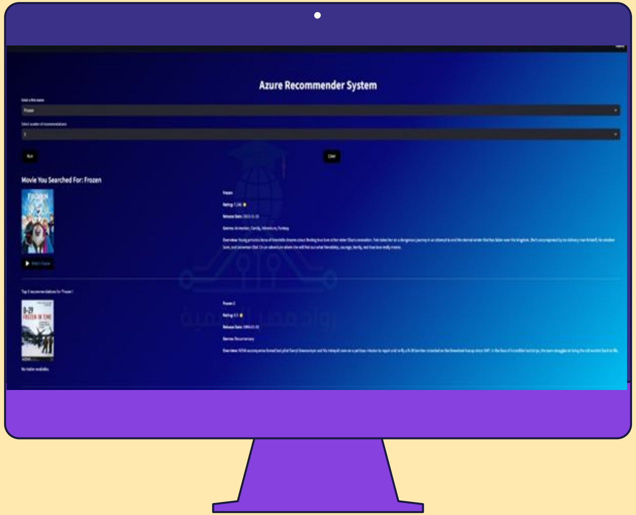
- ❑ Personalized recommendations based on individual preferences.
- ❑ Independence from data about other users, making it effective in niche markets.
- ❑ The ability to adapt to user preferences over time, improving the relevance of suggestions.

Limitations of Content-Based Filtering

- **Cold Start Problem:** Strategies for New Users: in Content-based recommendation systems address this by leveraging demographic information, preferences from similar user groups, or hybrid approaches
- **Over-Specialization:** Diversification Techniques: Content-based systems risk over-specialization, where users consistently recommend similar items, limiting their exposure to diverse content.
- **Scalability:** Optimization and Parallelization: Scalability becomes a concern as user and item databases grow. Content-based systems must efficiently handle the increasing volume of data to provide real-time recommendations.

05

Azure Integration & Final Deployment

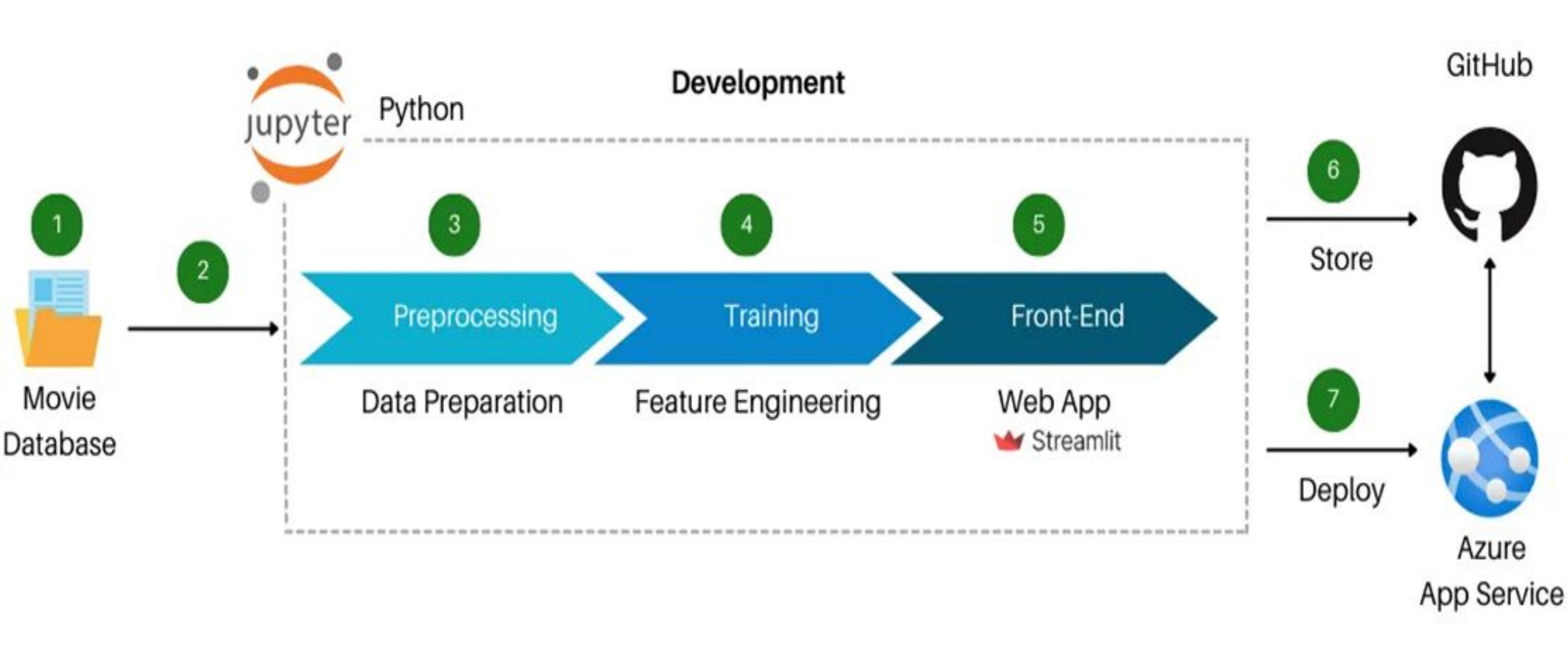


Introduction to Azure OpenAI

- **Definition:** A cloud-based service providing access to OpenAI’s models.

- **Purpose:** Enhancing applications with AI-driven capabilities.
- Azure OpenAI is a service that allows developers to integrate OpenAI's advanced models into their applications, enabling them to leverage natural language processing, code generation, and more

How It Works



Development
jupyter
Python
GitHub
1
Movie
Database
2
3
Preprocessing
Data Preparation
4
Training
Feature Engineering
5
Front-End
Web App
Streamlit
6
Store
7
Deploy
Azure
App Service

How It Works

- ❑ Azure OpenAI operates through a robust architecture where users access OpenAI's models via API. This allows for easy integration into existing applications, leveraging Azure's cloud infrastructure.

- ❑ By using Azure OpenAI, organizations can enhance productivity through automation, improve user engagement with personalized interactions, and significantly reduce development time for AI-driven features.

Integration between Model and Azure OpenAI

- Using Azure OpenAI, we can extract features from movie metadata effectively. Natural language processing allows us to analyze sentiment in reviews and classify genres, enriching our recommendation engine.

- User profiles are created by gathering preferences from viewing history and ratings. Azure OpenAI can dynamically update these profiles, adapting recommendations as user tastes evolve.

- Recommendations can be generated using the OpenAI API by querying user profiles and applying algorithms that consider both past behavior and current trends, continuously refining suggestions based on user feedback.

How Azure OpenAI Can Enhance Recommendations

- ❏ Azure OpenAI enhances recommendation systems by utilizing natural language understanding. It can analyze user preferences through text input and generate personalized movie descriptions, leading to more accurate recommendations.

- ❏ Implementing Azure OpenAI in movie recommendations offers numerous benefits, including improved personalization, enhanced user engagement through relevant suggestions, and the scalability and security of Azure's infrastructure.

06

Challenges and Solutions

Recommendation systems help users discover content aligned with their preferences. Content-based models focus on the attributes of items, while collaborative filtering relies on user interactions.

In our Project, We faced some challenges such as:

- ❑ limited availability of item features. In a content-based model, the quality and richness of metadata—like genre, plot, and cast—are crucial for generating accurate recommendations. **To handle this challenge,** we can utilize Azure OpenAI to analyze movie reviews and synopses to enhance feature extraction. This allows us to extract deeper insights and semantic features, enriching our recommendation model with nuanced data.

- Another challenge is that user profiles can become static, relying on initial interactions. This can lead to outdated recommendations that don’t reflect current user preferences. **To handle this challenge,** we integrate Azure OpenAI, we can dynamically update user profiles based on ongoing interactions and feedback. This ensures that recommendations remain relevant and personalized as user tastes evolve.

- Over-specialization is a significant risk in content-based models. Users may receive recommendations that are too similar, limiting their exposure to diverse content. **To handle this challenge,**we can leverage Azure OpenAI to generate diverse recommendations. By balancing relevance with variety, we can introduce users to a broader range of movies they may enjoy.

07

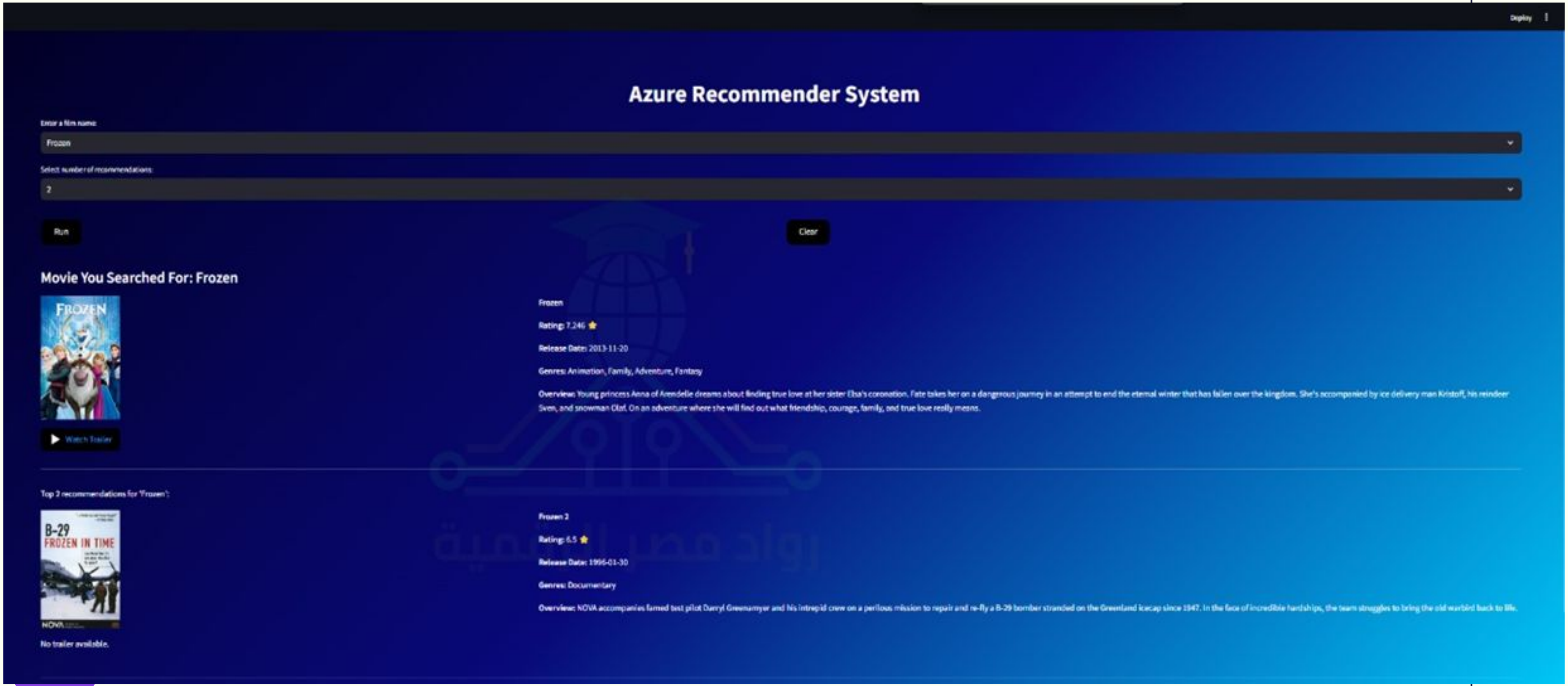
Conclusion and Future Work

- **To conclude our movie recommender system project:**

Based on content-based recommendation systems, it play a crucial role in providing personalized suggestions to users based on the intrinsic features of items and their expressed preferences. These systems utilize the cosine similarity metric to effectively measure the similarity between items, facilitating identifying and recommending things that align closely with user preferences.

- **Our recommendation for future recommendation systems:**

Content-based recommendation systems can be refined and improved over time by incorporating additional features, experimenting with different algorithms, and adapting to changing user preferences.


Deploy
Azure Recommender System
Enter a film name
Frozen
Select number of recommendations:
2
Run
Clear
Movie You Searched For: Frozen
Frozen
Rating: 7.246
Release Date: 2013-11-20
Genres: Animation, Family, Adventure, Fantasy
Overview: Young princess Anna of Arendelle dreams about finding true love at her sister Elsa's coronation. Fate takes her on a dangerous journey in an attempt to end the eternal winter that has fallen over the kingdom. She's accompanied by ice delivery man Kristoff, his reindeer Sven, and snowman Olaf. On an adventure where she will find out what friendship, courage, family, and true love really means.
Watch Trailer
Top 2 recommendations for 'Frozen':
Frozen 2
Rating: 6.5
Release Date: 1996-01-30
Genres: Documentary
Overview: NOVA accompanies famed test pilot Darryl Greenamyer and his intrepid crew on a perilous mission to repair and re-fly a B-29 bomber stranded on the Greenland icecap since 1947. In the face of incredible hardships, the team struggles to bring the old warbird back to life.
No trailer available.

Future Work

- Incorporation of User Interaction Data
- Hybrid Recommendation System
- Context-Aware Recommendations
- Natural Language Processing (NLP) for Reviews

Q&A