ruth-kitasi / Movielens-Recommendation-System

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

Movielens-Recommendation-System

Public

1 Branch 0 Tags Go to file t Go to file + Add file Code ...

ruth-kitasi Update README.md 8e8afac · 23 minutes ago

.ipynb_checkpoints	added conclusion and re...	yesterday
ml-latest-small	creating index file	last week
work_folder	adjusting the notebook	1 hour ago
Movie.jpeg	updating image	yesterday
README.md	Update README.md	23 minutes ago
index.ipynb	adjusting data understan...	1 hour ago
presentation.pptx	last changes	yesterday

README

About

No description, website, or topics provided.

- Readme
- Activity
- 0 stars
- 1 watching
- 0 forks

Releases

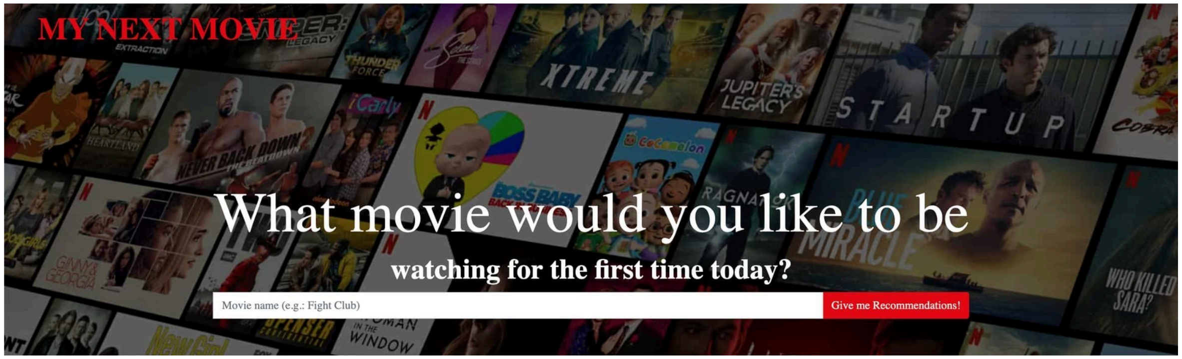
No releases published
Create a new release

Packages

No packages published
Publish your first package

Contributors 6

MOVIELENS RECOMMENDATION SYSTEM.



1.0 BUSINESS UNDERSTANDING

Overview

In today's data-driven world, recommendation systems are crucial for filtering information and enhancing user experiences. These systems help users discover relevant content by analyzing their past interactions, such as search queries or viewing histories.

Major platforms like Netflix and YouTube utilize recommendation algorithms to suggest movies and videos that cater to individual preferences, enhancing user engagement.

Aligned with our project objectives, we aim to leverage the power of data analysis to build an efficient movie recommendation system. Our goal is to deliver personalized movie suggestions by analyzing users' previous movie ratings and interactions.

Languages

- Jupyter Notebook 100.0%

By doing so, the system will generate top 5 tailored movie recommendations for each user, improving their viewing experience and aligning with their unique preferences.

Problem Statement

With the vast amount of content available on streaming platforms, users often feel overwhelmed by choices, making it difficult to discover movies that align with their preferences. Traditional search methods fall short in addressing this challenge, resulting in a less satisfying user experience and decreased engagement.

MovieLens has tasked our team of data scientists with optimizing their recommendation system through data-driven approaches. By analyzing user behaviors and preferences, we aim to enhance the system's ability to deliver personalized movie recommendations.

Objectives

- Personalized Recommendations: Collaborative filtering provided tailored suggestions. Outcome: Boosted user engagement.

- Content-Based Filtering: Users find similar movies by title. Outcome: Enhanced personalization.

- Cold Start Mitigation:

  - Popularity: High-rated movies for new users.
  - Genre Selection: Genre-specific recommendations for new users.

Outcome: Improved satisfaction for all user types. Model Evaluation:

- RMSE = 0.50: Predictions within 0.50 points of actual ratings. Outcome: Accurate recommendations.

Rating Frequency:

- Average rating: 4.0/5.0. Outcome: Positive trends for future recommendations.

Model Implementation

- KNNbasic Model.
- Singular Value Decomposition(SVD)

Evaluation

The project's success criteria is determined by achieving a Root Mean Squared Error (RMSE) of 0.50.

Findings

- On a scale of 0.5 to 5.0, the analysis shows that most movies received an average rating of 4.0, indicating that users generally rated the majority of films positively. This suggests a tendency for users to favorably evaluate the available content, with few movies receiving extremely low ratings of 0.5.
- There is a noticeable spike around the year 2000, where the number of ratings peaked at over 10,000.This might imply that there was a surge in user activity, perhaps due to the popularity of certain movies or the increased availability of the platform at the time.
- The genres "Drama" and "Comedy" are the most popular among the movies in the dataset, with significantly higher counts compared to other genres.
- Despite the relatively lower number of Western movies produced, this genre stands out due to its impressive average ratings, surpassing those of other genres. This observation suggests that while Western films may not be as prolific as others, they resonate more strongly with audiences, gaining higher appreciation and positive feedback.
- The analysis indicates that the majority of films were produced in 1994. One might expect this year to reflect the highest ratings. However, the ratings distribution graph reveals that 2000 actually had the highest number of ratings, despite a lower volume of films produced in that year compared to 1994.

Conclusions:

Personalized Recommendations: Collaborative filtering provided tailored suggestions. Outcome: Boosted user engagement.

Content-Based Filtering: Users find similar movies by title. Outcome: Enhanced personalization.

Cold Start Mitigation:

- Popularity: High-rated movies for new users.
- Genre Selection: Genre-specific recommendations for new users.

Outcome: Improved satisfaction for all user types. Model Evaluation:

- RMSE = 0.50: Predictions within 0.50 points of actual ratings. Outcome: Accurate recommendations.

Rating Frequency:

- Average rating: 4.0/5.0. Outcome: Positive trends for future recommendations.

Recommendations

- Recent Movies: Update dataset to include the latest releases for relevance.
- Filmmaker Info: Add screenwriter and producer data for deeper personalization.
- Broaden Genres: Expand to niche genres for wider audience appeal.
- Location Data: Utilize user location for region-specific recommendations.
- External Sources: Integrate reviews and social media sentiment to enhance quality.

Author and Acknowledgement:

Special thanks to our Moringa School Data Science Technical Mentors for their