

Rgmoogachiri / DSC-phase-4-project

<> Code

Issues

Pull requests 1

Actions

Projects

Wiki

Security

Insights

Settings

0 stars

1 fork

1 watching

2 Branches

0 Tags

Activity

Public repository

Your main branch isn't protected

Protect this branch from force pushing or deletion, or require status checks before merging. [View documentation.](#)

Protect this branch

Dismiss

main

2 Branches

0 Tags










Go to file

Go to file

+

Add file

<> Code

 Rgmoogachiri	Final	4 minutes ago
 .ipynb_checkpoints	Final	4 minutes ago
 Data	Data	2 weeks ago
 Notebooks	Harmonizing	3 days ago
 Presentation	Updating the Presentation Slides	19 minutes ago
 images	Add files via upload	yesterday
 README.md	Updating README.md	48 minutes ago
 annoy_index.ann	95% notebook done	yesterday
 movie.ipynb	Final	4 minutes ago

README

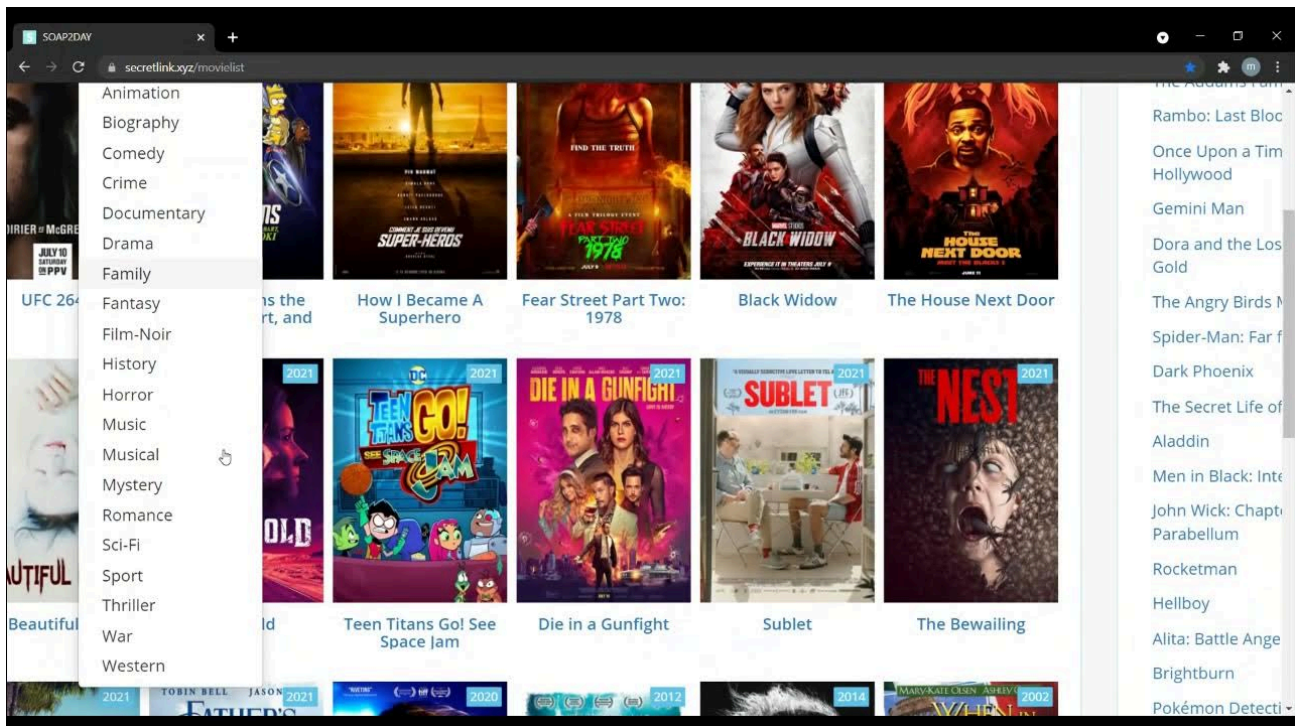
Stream254:

Enhance Your Movie Watching Experience with Uncharted Cinematic Treasures

Introduction

The rapid development of data collection has heralded a new era of information. Data is being used to create more efficient systems, which is where Recommendation Systems come into play. Recommendation Systems are a sort of information filtering system that improves search results by displaying items that are more relevant to the search query or to the user's search history.

They estimate a user's rating or preference for an item. Almost every major internet company has utilized them in some form or another: Amazon uses them to recommend products to customers, YouTube uses them to choose which video to play next on autoplay.



Documentation

Dataset from (<https://grouplens.org/datasets/movielens/latest/>)

We used the "small" dataset containing 100,000 user ratings

Business Problem

Stream254 recognizes that traditional one-size-fits-all approaches to content delivery are no longer sufficient to meet the evolving demands of Kenyan viewers. Without personalized recommendations, users may face decision fatigue and frustration when navigating the platform's extensive content library. In a highly competitive market, failure to engage users effectively could lead to decreased retention rates and loss of market share.

To address this challenge, Stream254 seeks to implement a recommendation system that leverages machine learning algorithms to analyze user data and deliver highly relevant movie recommendations. By doing so, Stream254 aims to enhance user satisfaction, drive long-term engagement, and establish itself as a leader in the Kenyan streaming space.

The primary stakeholders of this project include:

1. **Online Streaming Platforms:** Companies like Netflix, Amazon Prime Video, and Hulu can integrate this recommendation system to enhance user experience, increase viewer engagement, and reduce churn rates by providing tailored movie suggestions.
2. **Users:** Individual users of streaming services stand to benefit from a more personalized viewing experience, leading to improved satisfaction and discovery of new content aligned with their tastes.
3. **Content Creators and Distributors:** By promoting movies that are likely to be well-received, the system can help content creators and distributors reach their target audience more effectively, thereby increasing viewership and revenue from underexposed content.

Business Objectives

Develop and implement state-of-the-art recommendation algorithm that will analyze user behaviour, preferences, and viewing patterns to generate personalized movie recommendations with the intention to:

1. **Enhance User Satisfaction:** By providing users with tailored movie recommendations, we aim to enhance user satisfaction and deliver an immersive viewing experience that keeps users coming back for more.

2. Increase User Retention: Through personalized recommendations, we seek to increase user retention rates and encourage users to spend more time on the platform, ultimately driving revenue growth and profitability.
3. Establish Competitive Differentiation: By offering a recommendation system that surpasses those of competitors, we aim to differentiate Stream254 as the go-to destination for personalized entertainment experiences in the Kenyan market.
4. Implement Performance Metrics: We will implement robust metrics and analytics to measure the performance of the recommendation system. This will enable us to continuously optimize algorithms and enhance user engagement over time.

Feature Description

The features used are:

1. MovieId
2. UserId
3. Rating
4. Genres
5. Title

Data Understanding

The ml-latest-small dataset is a collection of data from MovieLens, a movie recommendation service. It comprises user ratings and free-text tagging activities, offering insights into user preferences. This dataset contains a total of 100,836 movie ratings and 3,683 tag applications across 9,742 movies.

The data are contained in the files links.csv, movies.csv, ratings.csv and tags.csv and are contained in the 'Data' folder of this repository.

Dataset Details:

1. Ratings: There are 100,836 user ratings, primarily based on a 5-star scale.
2. Tags: The dataset includes 3,683 instances of user-generated tags for movies. Users: All 610 users captured were chosen at random and each user has rated at least 20 movies.
3. Timeframe: The data spans a period from March 29, 1996, to September 24, 2018.
4. File Structure: The dataset is organized into four main files: links.csv, movies.csv, ratings.csv, and tags.csv.

Out of the four datasets, only movie.csv and ratings.csv will be used for the analysis

Limitations Of The Data

1. Cold Start Problem: New movies or users with few ratings pose a challenge, as there's limited data to base recommendations on.
2. Bias in User Ratings: Ratings are subjective and can be biased, affecting the system's ability to accurately predict preferences.
3. Sparsity of the Ratings Matrix: With a large number of movies and users but relatively fewer ratings, the ratings matrix is sparse, which can complicate the modeling process.

Cold Start Problem

Collaborative filtering relies solely on user-item interactions inside the utility matrix. The difficulty with this strategy is that it removes new individuals and things from the recommendation system that have not had any interactions. This is referred to as the "cold start" problem.

Model Performance

The models performance are:

No	Model	Train RMSE	Test RMSE	Train MAE	Test MAE
1	Baseline Model(KNN)	0.71190	0.9428	0.5318	0.7245
2	KNN Basic	0.66003	0.9719	0.4934	0.7518
3	Tuned KNN Basic	0.83660	0.9679	0.6322	0.7468
4	KNN Baseline	0.75250	0.8715	0.5627	0.6701
5	Tuned KNN Baseline	0.75250	0.8715	0.5627	0.6701
6	SVD Model	0.50000	0.7903	0.3887	0.6042
7	Bayesian SVD	0.15370	0.1833	0.1044	0.1275
8	Tuned Bayesian SVD	0.08310	0.1452	0.0581	0.0774
9	NMF Model	0.71620	0.7162	0.5322	0.5322

More information can be found in the description below.

[Jupyter Notebook](#)

Recommendations

1. Incorporate Real-time User Feedback: Implement a feedback loop that captures users' ratings, likes, dislikes, and viewing habits in real time. Use this data to adjust recommendations accordingly, making the system more responsive and personalized.
2. Enhance Content Diversity and Fairness: Implement algorithms that consciously diversify the content shown to users, preventing filter bubbles. Monitor recommendation outcomes for fairness and adjust algorithms to ensure broad representation of genres, cultures, and perspectives.
3. User Interface and Experience Improvements: Gather user feedback on the interface and experience. Implement design improvements that make browsing recommendations more intuitive and enjoyable. Consider features like personalized watchlists, user reviews, and interactive content discovery tools.
4. Continuous Monitoring and Evaluation: Establish a framework for ongoing monitoring of system performance, user engagement metrics, and satisfaction levels. Use these insights to guide regular updates and refinements to the recommendation algorithms and user interface.

Conclusions

1. The project delivers a robust foundation for a movie recommendation system that can significantly impact user engagement and platform growth.
2. By addressing both immediate recommendations for new users and items and providing a pathway for continuous improvement and scalability, the system is well-poised to meet and exceed stakeholder expectations.
3. Future developments should focus on refining these models and incorporating real-world feedback to ensure the recommendations remain relevant, personalized, and engaging for all users.

Contributors

1. Caroline Ngabu
2. Muhsin Ahmed







Releases

No releases published
[Create a new release](#)

Packages

No packages published
[Publish your first package](#)

Contributors 4

-  **Rgmoogachiri** Richard Gachiri Muriithi
-  **Beatrice-Kariuki**
-  **Damuki007**
-  **Cngabu**

Languages

