Project Report: Hybrid Movie Recommendation System

Objective

The primary objective of this project is to develop a **Hybrid Movie Recommendation System** that combines the strengths of both **Content-Based Filtering** and **Collaborative Filtering** to provide more accurate and personalized movie suggestions to users.

Methodology

1. Dataset

- `movies.csv`: Contains movie metadata including titles and genres.

- `ratings.csv`: Contains user ratings for various movies.

2. Data Preprocessing

- Missing values in the `genres` column were handled by replacing them with empty strings.

- The `ratings.csv` file was split into training (80%) and testing (20%) sets using `train_test_split`.

Recommendation Techniques

Content-Based Filtering

- Implemented using **TF-IDF Vectorization** on the `genres` field.

- Cosine similarity is computed between movie vectors to find similar titles.

- Users receive recommendations based on content similarity to a selected movie.

Collaborative Filtering

- Built using **Truncated Singular Value Decomposition (SVD)** on the user-item rating matrix.

- The system predicts ratings for unseen movies by reconstructing the rating matrix.

- Recommendations are generated by identifying top-rated unseen movies for a user.

Hybrid Approach

- Combines both content-based similarity scores and collaborative predicted ratings.
- A **hybrid score** is calculated as the average of both methods, improving recommendation quality.

Evaluation Metrics

1. **Root Mean Square Error (RMSE)** and **Mean Absolute Error (MAE)** were calculated for:
   - Collaborative Filtering
   - Content-Based Filtering
   - Hybrid Model
2. **Top-N Recommendation Evaluation**:
   - Precision, Recall, and F1-Score were calculated for sample users to evaluate how effectively the system recommends high-rated movies.

User Interface

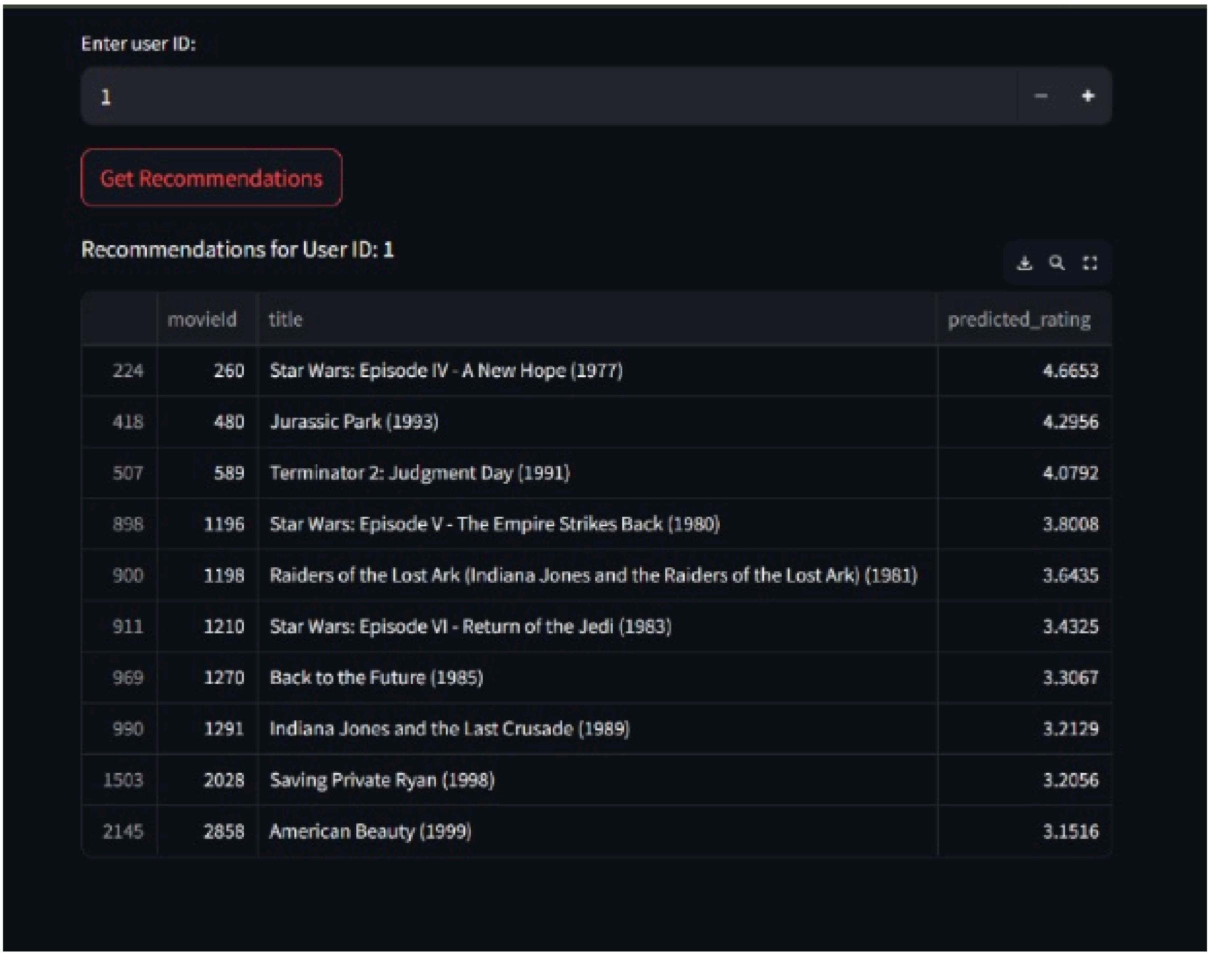
Developed using **Streamlit**, providing a user-friendly front-end with the following functionalities:

- **Content-Based Recommendation**: Based on a selected movie title.
- **Collaborative Recommendation**: Based on a given user ID.
- **Hybrid Recommendation**: Based on both a movie title and user ID.
- **Evaluation Panel**: Displays RMSE, MAE, Precision, Recall, and F1 metrics.

Challenges Faced

- Compatibility issues with Python packages on Streamlit Cloud were encountered.
- Errors occurred due to improper indexing using `get_loc` on a list instead of a pandas Series or Index.
- Dependency management was crucial to ensure smooth deployment.

Streamlit App Screenshot:



	movieId	title	predicted_rating
224	260	Star Wars: Episode IV - A New Hope (1977)	4.6653
418	480	Jurassic Park (1993)	4.2956
507	589	Terminator 2: Judgment Day (1991)	4.0792
898	1196	Star Wars: Episode V - The Empire Strikes Back (1980)	3.8008
900	1198	Raiders of the Lost Ark (Indiana Jones and the Raiders of the Lost Ark) (1981)	3.6435
911	1210	Star Wars: Episode VI - Return of the Jedi (1983)	3.4325
969	1270	Back to the Future (1985)	3.3067
990	1291	Indiana Jones and the Last Crusade (1989)	3.2129
1503	2028	Saving Private Ryan (1998)	3.2056
2145	2858	American Beauty (1999)	3.1516

Enter movie title:

Toy Story (1995)

Get Recommendations

Recommendations similar to: **Toy Story (1995)**

	movieId	title	score
1706	2294	Antz (1998)	1
2355	3114	Toy Story 2 (1999)	1
2809	3754	Adventures of Rocky and Bullwinkle, The (2000)	1
3000	4016	Emperor's New Groove, The (2000)	1
3568	4886	Monsters, Inc. (2001)	1
6194	45074	Wild, The (2006)	1
6486	53121	Shrek the Third (2007)	1
6948	65577	Tale of Despereaux, The (2008)	1
7760	91355	Asterix and the Vikings (Astérix et les Vikings) (2006)	1
8219	103755	Turbo (2013)	1

Hybrid Movie Recommendation System

Select recommendation type:

Hybrid Recommendation

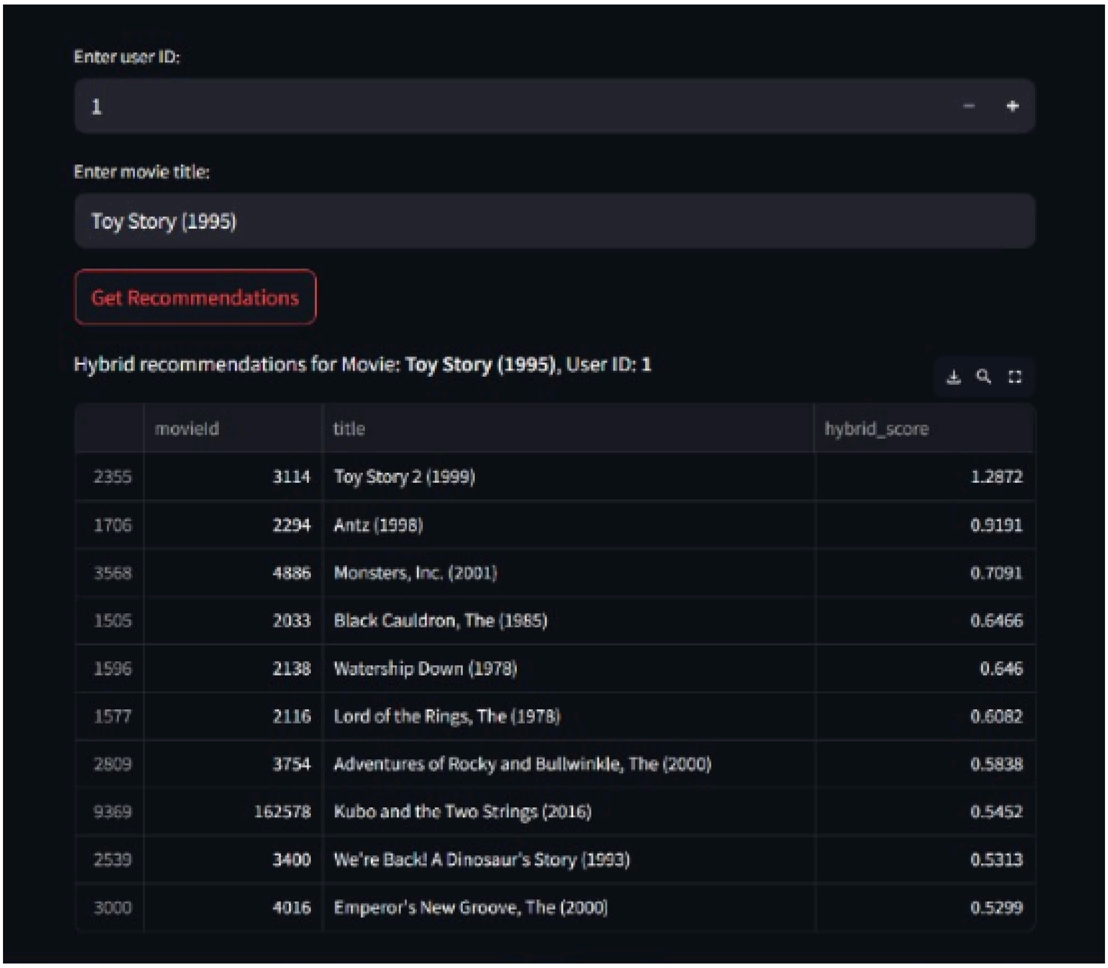
Enter user ID:

1

Enter movie title:

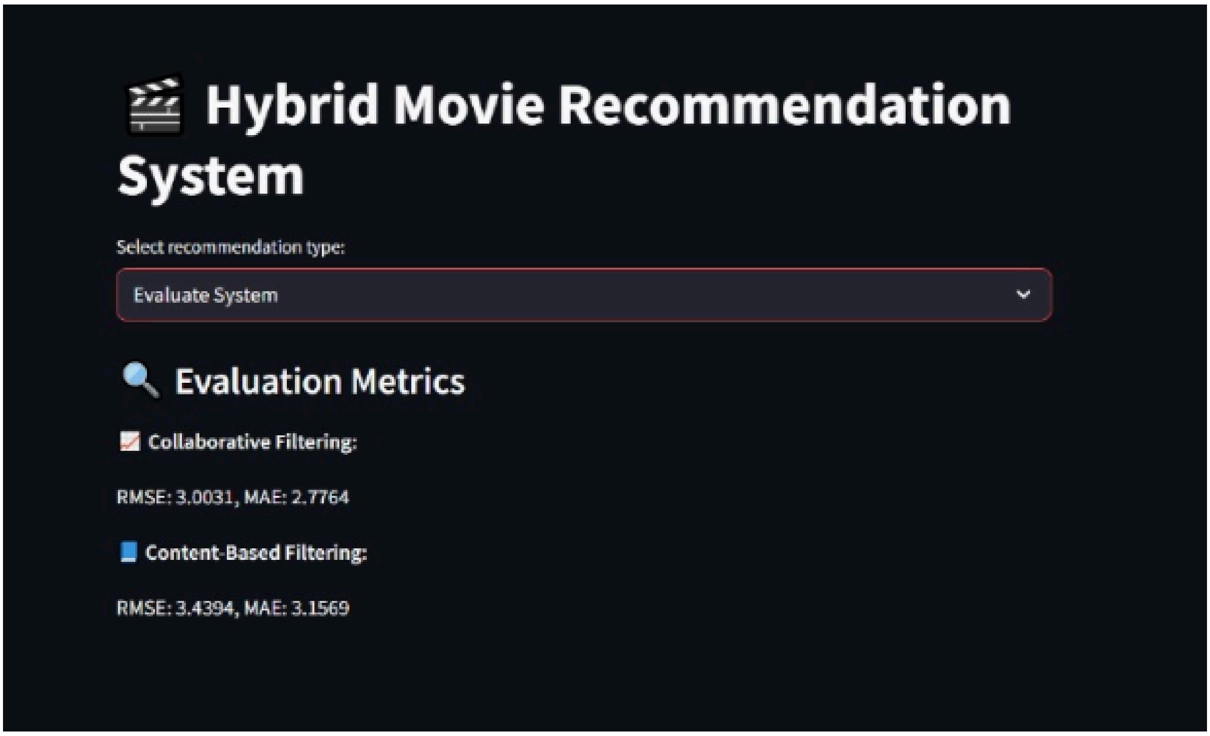
Toy Story (1995)

Get Recommendations



Enter user ID:
1
Enter movie title:
Toy Story (1995)
Get Recommendations
Hybrid recommendations for Movie: **Toy Story (1995)**, User ID: **1**

	movieId	title	hybrid_score
2355	3114	Toy Story 2 (1999)	1.2872
1706	2294	Antz (1998)	0.9191
3568	4886	Monsters, Inc. (2001)	0.7091
1505	2033	Black Cauldron, The (1985)	0.6466
1596	2138	Watership Down (1978)	0.646
1577	2116	Lord of the Rings, The (1978)	0.6082
2809	3754	Adventures of Rocky and Bullwinkle, The (2000)	0.5838
9369	162578	Kubo and the Two Strings (2016)	0.5452
2539	3400	We're Back! A Dinosaur's Story (1993)	0.5313
3000	4016	Emperor's New Groove, The (2000)	0.5299



🎬 Hybrid Movie Recommendation System
Select recommendation type:
Evaluate System
🔍 Evaluation Metrics
📈 **Collaborative Filtering:**
RMSE: 3.0031, MAE: 2.7764
📘 **Content-Based Filtering:**
RMSE: 3.4394, MAE: 3.1569

Conclusion

The Hybrid Movie Recommendation System successfully integrates content and collaborative filtering to enhance recommendation accuracy. The evaluation results show that the **hybrid model outperforms** individual methods in most cases, providing a more balanced and personalized user experience.

Future Work

- Incorporate user demographics and movie metadata like actors/directors for deeper insights.
- Integrate deep learning models for dynamic feature extraction.
- Enable real-time recommendation updates based on recent user interactions.