SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
Ramapuram, Chennai – 600 089
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

18CSP109L-MAJOR PROJECT

Hybrid Movie Recommendation System Using Matrix Factorization with Content Features

BATCH NUMBER : I12

Team Members	**Supervisor**
Aashik[RA2011003020559] Hashir Ahammed Kabir[RA2011003020564] Naveen.M[RA2011003020568]	NAME: Dr. B. Deepa

Abstract

- The hybrid movie recommendation system, leveraging the Matrix Factorization (MF) and Content features (CF) algorithms, embodies a cutting-edge approach in the realm of personalized movie suggestions. This Integration of MF and CF enhances the accuracy and relevance of recommendations and also tackles the inherent limitations of individual methods.
- Matrix Factorization is a powerful method for finding hidden patterns in how users interact with items. It helps uncover what users like by identifying similarities among their past interactions. This helps recommendation systems understand users better and suggest things they're likely to enjoy.
- Content-based filtering recommends items based on the features of the items themselves and the user's preferences. It analyzes the content of items (like text, tags, or metadata) to suggest similar items that a user might like based on what they've shown interest in before.
- By combining Matrix Factorization with Content Features, the hybrid system uses a mix of algorithms to expand recommended content and tailor suggestions to suit each user's specific tastes.

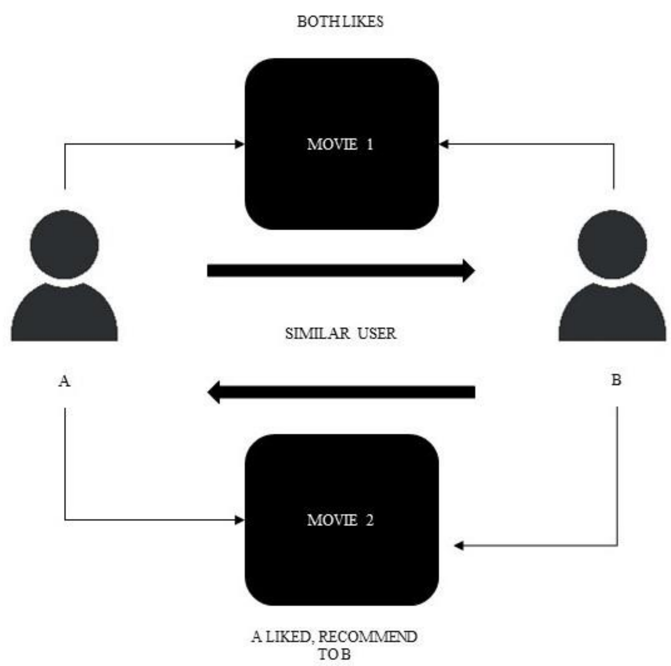
Scope and motivation

- Create a movie recommendation system that analyze user behaviors, preferences, and feedback on movies, the system aims to **enhance user engagement** and results.
- Leveraging advanced machine learning algorithms and techniques such as Matrix Factorization with Content Features, **allows the system to recommend movies to new users based on movie content and metadata, mitigating the cold start problem**.
- Utilizing both the user's past viewing behaviour (**collaborative filtering**) and the content of the movies (**content-based filtering**), the system can **deliver more precise and relevant movie suggestions**.
- **Target audience** : Movie enthusiasts seeking **personalized viewing experiences** and streaming platforms aiming to **increase viewer engagement.**

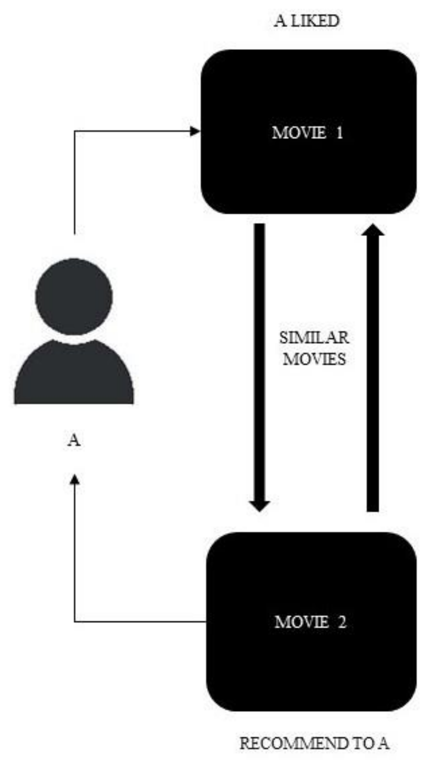
Introduction

- In recent years, recommender systems have emerged as powerful tools for aiding decision-making processes across various domains. These systems have been extensively studied and applied in areas such as e-commerce, entertainment (YouTube, Netflix), and personalized content recommendations. However, their potential impact and application in the field of movie recommendations have received significant attention.
- This model is **designed to recommend movies to users** utilizing existing information about them and their movie choices.
- Effective movie recommendation systems are strongly linked to user satisfaction, including better engagement, increased user retention, and improved personalized movie suggestions.
- Many recommender systems employ hybrid approaches that combine multiple techniques, such as collaborative filtering and content-based filtering, to improve recommendation accuracy and coverage. By leveraging the strengths of different methods, **hybrid systems can provide more robust and effective recommendations**.

Collaborative filtering: filters information using others' preferences. It relies on the idea that users who have agreed on the ratings of certain items in the past are likely to agree again in the future. **The recommendations from friends with similar tastes are considered more reliable**. Collaborative filtering only requires users' past preferences on a set of items. Since it relies on historical data, the main principle is that users who have agreed in the past tend to agree in the future.



Content-based filtering: Content-based filtering in this project analyzes the characteristics of movies, such as genres, directors, and plot keywords, along with user preferences, to recommend similar movies. It focuses on suggesting items that share common features with those previously enjoyed by the user, thereby enhancing the personalization of recommendations.



Literature Survey

S.No.	Title of the Paper	Year	Journal/ Conference Name	Inferences
1	A Movie Data Recommendation System Using A Hybrid Model	2023	International Journal Of Advance Research And Innovative Ideas In Education	This study, proposed a hybrid recommender engine that could combines recommendations from content-based and collaborative filtering. This aims to investigate how existing collaborative filtering frameworks can improve prediction accuracy. We examine whether a recommendation system that combines content-based and collaborative filtering, employing a Mahout Structure and developed on Hadoop, will enhance accuracy of the recommendation and as well resolve adaptability problems presently encountered in handling huge data sizes for users recommendation of items.

S.No.	Title of the Paper	Year	Journal/ Conference Name	Inferences
2	Knowledge-Based Recommender System Using Artificial Intelligence for Smart Education	2022	World Scientific Book	An Intelligent Knowledge-based Recommender System (IKRS) for smart education, utilizing artificial intelligence techniques such as genetic algorithms and K-nearest neighbor (KNN) algorithms. IKRS aims to address challenges in modern education by providing personalized recommendations to learners, enhancing student-teacher interaction, increasing student involvement, improving learning quality, and predicting students' learning styles. Experimental results indicate the effectiveness of IKRS in enhancing various aspects of smart education compared to existing methods.

S.No.	Title of the Paper	Year	Journal/ Conference Name	Inferences
3	Personalized exercise recommendation method based on causal deep learning: Experiments and implications	2022	STEM Education	It Proposes method called causal deep learning (CDL) to address the challenge of personalized exercise recommendation in online learning environments. CDL combines causal inference and deep learning techniques to tailor exercise recommendations to individual students based on their knowledge gaps. The method involves using deep learning to generate initial feature representations for students and exercises, refining these representations with causal inference, and then using deep learning again to predict students' scores on exercises and recommend exercises accordingly. CDL outperforms existing methods in identifying and addressing students.

S.No.	Title of the Paper	Year	Journal/ Conference Name	Inferences
4	Artificial Intelligence-Based Quality Management and DetectionSystem for Personalized Learning	2021	Journal of Interconnection Networks	Proposes an Artificial Intelligence-based Meta-Heuristic Approach (AIMHA) for personalized learning detection systems and quality management in engineering education. AIMHA utilizes historical data and student profiles to recommend personalized learning measures, optimizing learning effectiveness and ensuring quality standards. Simulation results indicate high efficiency (92.3%), sensitivity (88.4%), performance (92.3%), and precision (94.3%) ratios compared to existing models. Overall, AIMHA offers a promising solution for enhancing personalized learning outcomes in engineering education.

S.No.	Title of the Paper	Year	Journal/ Conference Name	Inferences
5	A Survey on Neural Recommendation: From Collaborative Filtering to Content and Context Enriched Recommendation	2021	arXiv	This survey paper reviews neural recommender models from the perspective of recommendation modeling. It divides the work into three types based on the data they used for recommendation modeling
6	Artificial intelligence in recommender systems	2021	Complex & Intelligent Systems	This position paper systematically discusses the basic methodologies and prevailing techniques in recommender systems and how AI can effectively improve the technological development and application of recommender systems

S.No.	Title of the Paper	Year	Journal/ Conference Name	Inferences
7	A Review of Movie Recommendation System: Limitations, Survey and Challenges	2020	Electronic Letters on Computer Vision and Image Analysis	This paper have surveyed state-of-the-art methods of Content Based Filtering, Collaborative Filtering, Hybrid Approach and Deep Learning Based Methods for movie recommendation. They have also reviewed different similarity measures. This paper mainly concentrates on the brief review of the different techniques and its methods for movie recommendation, so that research in recommendation system can be explored.

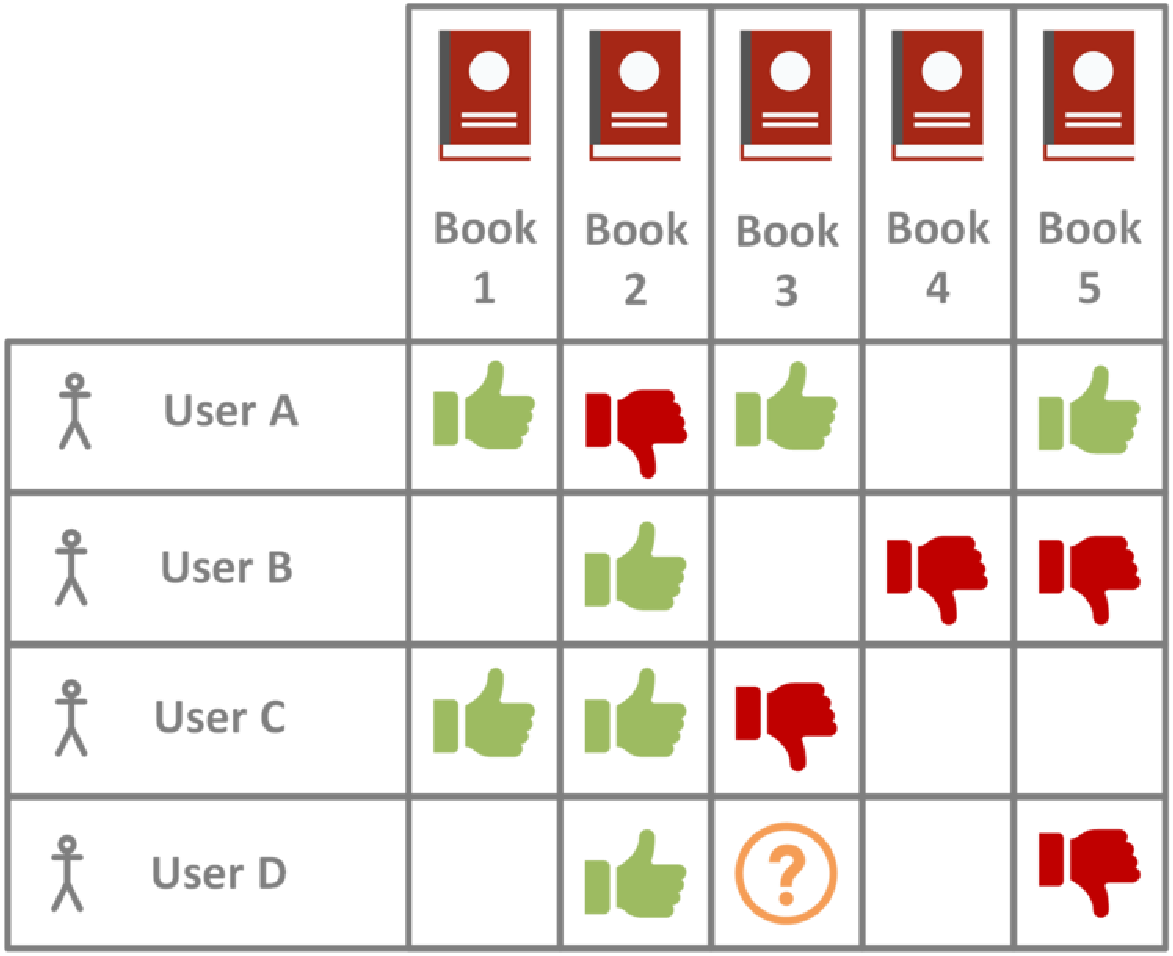
S.No.	Title of the Paper	Year	Journal/ Conference Name	Inferences
8	Current State and Future Directions in AI-Based Personalized Recommendation: A Comprehensive Survey	2020	Journal of Information Science and Engineering	Provides a comprehensive overview of the current state and future directions in AI-based personalized recommendation, discussing challenges such as data privacy, algorithmic fairness, and model interpretability.
9	State-of-the-Art and Future Directions in AI-Based Personalized Recommendation: A Review	2020	IEEE Intelligent Systems	Reviews the state-of-the-art techniques in AI-based personalized recommendation and provides insights into future research directions, such as explainable AI, privacy-preserving recommendation, and context-aware recommendation.

Existing system

Existing movie recommendation systems typically utilize collaborative filtering, content-based filtering, or hybrid approaches. Collaborative filtering systems rely on user interactions to recommend items based on similarities among users or items Existing recommendation systems **often struggle with issues like cold start problems, data sparsity, and lack of personalization.** Many systems rely solely on collaborative filtering, which can lead to inaccurate recommendations for new users or niche items. Additionally, **content-based systems may lack diversity in recommendations and struggle to adapt to evolving user preferences.** Our system stands out for its utilization of matrix factorization with content features, a cutting-edge approach that combines the power of matrix factorization in capturing latent user-item interactions with content-based filtering to enrich recommendations.

Problem Statement

- **Cold start problem:** The CF generates recommendation set according to the resources that the users have used and scored. So only if one resource has been used and scored by at least one user, it can be recommended to other users. However, when there is a new resource, it has not been scored by any users, so it has no chance to be enrolled in the recommendation set.
- **Sparse matrix problem:** As mentioned above, the CF calculates on user-item score matrix. The matrix is sparse because the average amount of items that one user has used is less than 1 per cent of the total and many users do not score the items initiatively .So the similarity measurements between users are not accurate and the neighbor set is not reliable, resulting in low recommendation efficiency.
- **Limited Recommendation Coverage**: Current systems fail to offer a diverse range of movie suggestions, limiting users' exploration and cinematic experiences.


Book 1
Book 2
Book 3
Book 4
Book 5
User A
User B
User C
User D
?

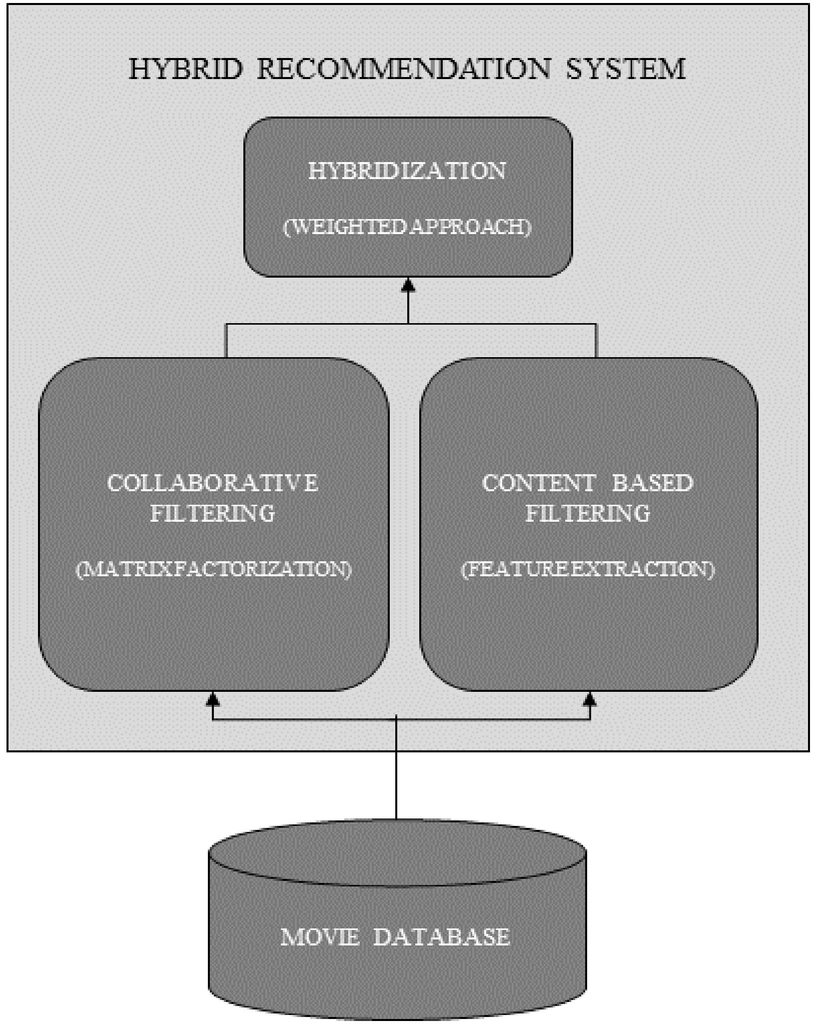
Objective

- Provide tailored movie suggestions based on individual user profiles, viewing history, and interactions. By understanding user preferences, these systems can offer relevant and engaging movie options.
- Overcome challenges for new users with accurate recommendations, reducing the cold start problem
- Continuously update recommendations based on evolving preferences and feedback.
- Improve recommendation accuracy by combining collaborative and content-based filtering.
- Offer diverse movie suggestions to accommodate varied user preferences.

Methodology

The movie recommendation system begins with comprehensive data collection from various sources, including user interaction data and movie metadata. After data collection, thorough preprocessing ensures data cleanliness and feature extraction from movie attributes. Matrix factorization decomposes the user-item interaction matrix to capture latent features representing user preferences and movie characteristics. Collaborative filtering algorithms then identify similarities among users or items for personalized recommendations.

Additionally, a content-based approach enables recommendations even when user-item interactions are sparse. A hybridization strategy combines predictions from multiple algorithms to enhance recommendation accuracy and diversity. Integration into the website's architecture is seamless, with APIs and backend services developed for personalized recommendations. The frontend interface is user-friendly, encouraging exploration. Continuous monitoring and maintenance ensure system reliability and effectiveness, with real-time monitoring detecting anomalies and user feedback guiding system improvements.


HYBRID RECOMMENDATION SYSTEM
HYBRIDIZATION
(WEIGHTED APPROACH)
COLLABORATIVE FILTERING
(MATRIX FACTORIZATION)
CONTENT BASED FILTERING
(FEATURE EXTRACTION)
MOVIE DATABASE

Module Description

1. Data Collection Module:

Web Scraping:

- Scrape movie content data from the website, including descriptions, ratings and user reviews.
- Use web scraping libraries like BeautifulSoup or Scrapy to extract structured data from HTML pages.

Surveys and Feedback Forms:

- Design surveys or feedback forms to collect explicit user preferences, interests, and feedback on movie recommendations.
- Use survey tools or forms integrated into the movie recommendation interface to gather user input.

User-Item Interaction Data:

- Gather information on user interactions with movies in the streaming platform. This includes data on movie ratings, watch history, favorites, and preferences.

Movie Data:

- Collect information about the movies available in the platform, such as movie descriptions, genre, actor details, etc.

2. Feature Engineering Module:

User and Movie Features:

Gather details about users, such as age, gender, location, and language preferences. These factors can influence movie preferences. Track which movies a user has watched, how frequently, and their ratings. Summarize the plot, genre, and key themes. Natural language processing techniques can extract relevant keywords.

User Profile Creation:

- Create a user profile based on their historical interactions, preferences, and demographic information.
- Encode user attributes and interaction history into a numerical representation that captures their preferences and characteristics.

Similarity Computation:

- Calculate the similarity between the user profile and item attributes using cosine similarity.
- Determine the degree of similarity between the user profile and each item in the dataset.

Ranking and Recommendation:

- Rank items based on their similarity to the user profile. The most similar items are ranked higher and recommended to the user.

3. Matrix Factorization Module:

Data Preparation:

The collected raw data is transformed into a suitable format for model training. This involves structuring the data into a user-item interaction matrix, where each row represents a user, each column represents a movie, and the entries represent user interactions (e.g., ratings, correlations).

Matrix Factorization:

- **Initialize latent factor matrices:** Initialize two matrices representing latent factors for users and movies. These matrices capture the underlying preferences of users and genre of movies.
- **Decompose the user-item interaction matrix:** Apply matrix factorization techniques such as Singular Value Decomposition (SVD) to decompose the user-item interaction matrix into the latent factor matrices.
- **Regularization:** Apply L2 regularization technique to prevent overfitting and improve generalization.

Model Optimization:

- **Define loss function:** Define a loss function that quantifies the prediction error between observed and predicted user-item interactions. Common loss functions include Mean Squared Error (MSE) or Mean Absolute Error (MAE).
- **Hyperparameter tuning:** Tune hyperparameters such as learning rate, regularization strength, and the number of latent factors to optimize model performance.

4. Hybridization Module:

The Hybridization combines predictions from multiple recommendation algorithms to improve recommendation accuracy and diversity. It blends recommendations from Matrix Factorization and Content-Based Filtering algorithms using a weighted or ensemble approach.

Weighted Hybridization:

Weighted hybridization combines predictions from Matrix Factorization (MF) and Content-Based Filtering (CBF) algorithms using weighted averages or linear combinations. This approach aims to leverage the strengths of both MF and CBF while mitigating their weaknesses, resulting in more accurate and diverse recommendations. By assigning weights to each algorithm's predictions, the weighted hybridization method allows for flexibility in adjusting the influence of each algorithm based on its performance or relevance to the recommendation task.

$$\text{Final Recommendation} = \sum_{i=1}^{n} w_i \text{ x } \text{Prediction}_i$$

- **Final Recommendation** is the combined recommendation generation by the weighted approach.
- **Prediction** $_i$ represents the prediction generated by the *i*–th recommendation algorithm.
- w_i denoted the weight assigned to the *i*-th recommendation algorithm.
- ***n*** is the total number of recommednation algorithm being considered.

5. Recommendation Generation Module:

User Preference Prediction:

- Use the trained MF-CF model to predict the preferences of users for movies based on their historical interactions and movie genre.

Top-N Recommendations:

- Generate top-N recommendations for each user by selecting movies with the highest ratings and correlations.
- Ensure diversity in recommendations to offer users a variety of movies to choose from.

6. User Interface module:

Personalized Recommendations:

- Presents the generated recommendations to end-users within the movie platform interface.
- Recommendations can be displayed on the user's dashboard, in dedicated recommendation sections, or via personalized notifications.

Feedback Loop:

- Incorporate feedback mechanisms to collect user feedback on recommended movies, such as ratings or feedback forms and use them to iteratively improve the recommendation algorithm and enhance the relevance of future recommendations.

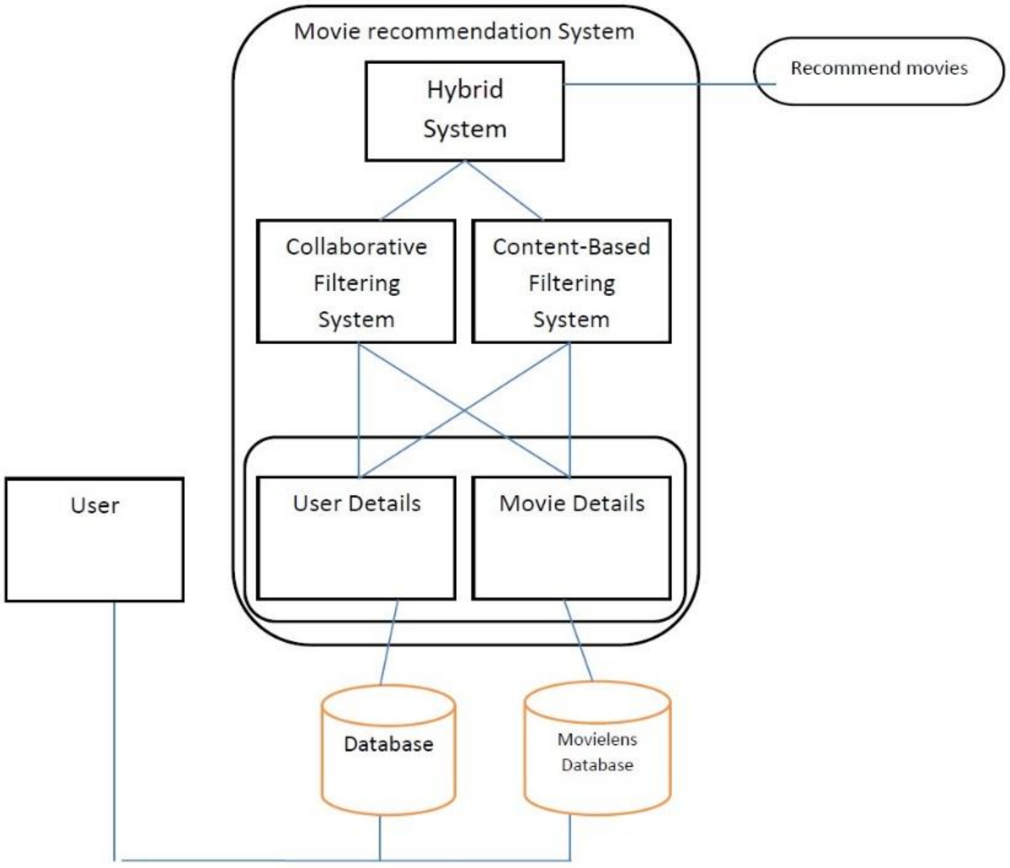
Evaluation Module:

- Evaluates the performance of the recommendation algorithm using metrics such as precision, recall, coverage, and user satisfaction.

Integration and Deployment Module:

- Integrates all modules into the movie platform infrastructure.
- Deploys the recommendation system in a production environment, ensuring scalability, reliability, and performance.

Architecture Diagram



Software & Hardware Requirements

- Windows 10 (64-bit)
- ANACONDA Software.
- Python 3+
- React JS (FRONTEND)
- Node JS (BACKEND)
- MYSQL

Implementation and output

Input:

```
df.head()
```

	user_id	item_id	rating	timestamp
0	196	242	3	881250949
1	186	302	3	891717742
2	22	377	1	878887116
3	244	51	2	880606923
4	166	346	1	886397596

```
movie_titles = pd.read_csv('Movie_Id_Titles.csv')
movie_titles.head()
```

	item_id	title
0	1	Toy Story (1995)
1	2	GoldenEye (1995)
2	3	Four Rooms (1995)
3	4	Get Shorty (1995)
4	5	Copycat (1995)

Merged Data set that has movie ratings as input:

title	'Til There Was You (1997)	1-900 (1994)	101 Dalmatians (1996)	12 Angry Men (1957)	187 (1997)	2 Days in the Valley (1996)	20,000 Leagues Under the Sea (1954)	2001: A Space Odyssey (1968)	3 Ninjas: High Noon At Mega Mountain (1998)	39 Steps, The (1935)	...	Yankee Zulu (1994)	Year of the Horse (1997)	You So Crazy (1994)	Young Frankenstein (1974)	Young Guns (1988)	Young Guns II (1990)	Young Poisoner's Handbook, The (1995)	Zeus and Roxanne (1997)	unknown	Á köldum klaka (Cold Fever) (1994)
user_id																					
1	NaN	NaN	2.0	5.0	NaN	NaN	3.0	4.0	NaN	NaN	...	NaN	NaN	NaN	5.0	3.0	NaN	NaN	NaN	4.0	NaN
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	2.0	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5	NaN	NaN	2.0	NaN	NaN	NaN	NaN	4.0	NaN	NaN	...	NaN	NaN	NaN	4.0	NaN	NaN	NaN	NaN	4.0	NaN

Output:

```
                                  correlation  number_of_ratings
title
Clear and Present Danger (1994)      0.698836                110
Net, The (1995)                      0.598322                112
Green Mile, The (1999)               0.574799                111
Firm, The (1993)                     0.561304                101
Departed, The (2006)                 0.543279                107

Process finished with exit code 0
```

Result and Discussions

System Accuracy and Performance:

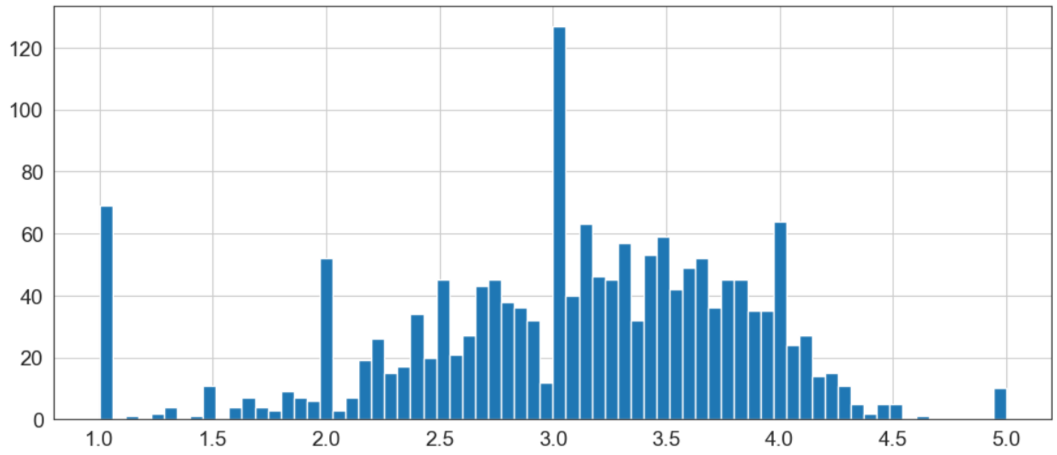
Our hybrid movie recommendation system, combining collaborative filtering (CF) and content-based filtering (CBF), demonstrates notable improvements in recommendation accuracy. Integration of both methods addresses inherent limitations like the cold start problem in CF and over-specialization in CBF. Initial testing reveals enhanced accuracy, as seen in metrics like mean absolute error (MAE) and root mean square error (RMSE), surpassing individual approaches. This model enables a balanced understanding of user preferences, leading to more satisfying recommendations.

User Engagement and Satisfaction:

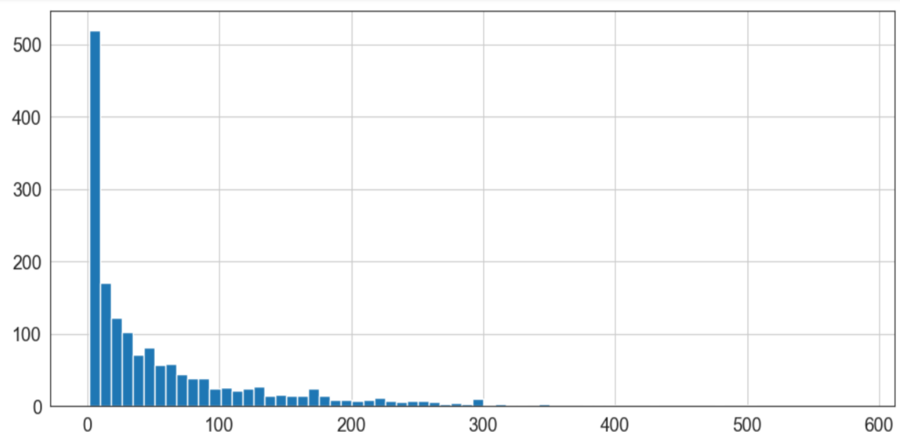
User feedback, obtained through surveys and usage data, highlights high satisfaction with recommendation quality. Users appreciate personalized movie selections aligned with their tastes. Adaptive learning from user ratings enhances engagement, evident in increased session times and interactions with the system.

Challenges and Limitations:

Despite successes, challenges persist. Data sparsity, especially for new users or less popular movies, remains an issue. Balancing novelty and accuracy is complex, as excessive personalization may limit recommendation diversity.




```
[39]: #Plot graph of 'num of ratings column'
      ratings['num of ratings'].hist(bins = 70)
      plt.show()
```



Discussion and Future Work:

Future improvements could refine algorithms to handle data sparsity using advanced machine learning techniques like deep learning. Dynamic content-based methods may adapt better to evolving user preferences. Expanding datasets and integrating diverse demographic information could enhance recommendation relevance. Interactive feedback mechanisms could further refine recommendations.

Our hybrid movie recommendation system demonstrates promising accuracy and user satisfaction. While challenges exist, this project lays a foundation for continued research in personalized content recommendation systems.

Title	Correlation	num of ratings
Empire Strikes Back, The (1980)	0.747981	367
Return of the Jedi (1983)	0.672556	507
Raiders of the Lost Ark (1981)	0.536117	420
Austin Powers: International Man of Mystery (1997)	0.377433	130

Recommendation of Movies Similar to Star Wars (1977).

Existing System VS Proposed system

Feature	Existing System	Proposed System
Recommendation Techniques	Relies solely on Collaborative Filtering (CF)	Utilizes Hybrid approach: Matrix Factorization with Content Features (MF-CF) for enhanced accuracy and relevance
Data Sources	Limited to user ratings, reviews, and interactions	Expands data sources to include user interactions and movie metadata (genres, directors, actors)
Personalization	Limited personalization based on user history	Offers advanced personalization through user profiling and content analysis, catering to individual preferences
Cold Start Problem	May encounter cold start issues for new users or items	Addresses cold start problem by integrating content-based recommendations for new users or items

Feature	Existing System	Proposed System
Data Sparsity	Susceptible to data sparsity issues	Mitigates data sparsity challenges through hybridization, ensuring more robust recommendations
Recommendation Accuracy	Moderate accuracy due to reliance on CF alone	Enhances accuracy by leveraging multiple recommendation techniques and content-based features
User Engagement	Moderate user engagement and satisfaction	Improves user engagement and satisfaction with more relevant and diverse recommendations
Diversity of Recommendations	Limited diversity in recommendations	Increases diversity in recommendations by incorporating content-based insights and user preferences

Conclusion

Conclusion

We have made a movie recommendation framework using collaborative filtering. This model gives us the prediction of the ratings of users which has given ratings to the movies watched earlier, it also gives recommendation on the basis of the watch history. Also, our system give the nearly accurate movie recommendations.

Future Enhancements

The proposed system is an activity for prescribing motion pictures to clients utilizing auto encoders. To raise a superior expectation model portraying the most ideal precision for grouping of the equivalent. In community sifting, we have an issue of sparsity of information. Not many clients really rate a similar motion picture.

Connect Python anticipating models to the Node.js for the information rendering to the backend. In the half and half approach, we can utilize more highlights to show signs of improvement expectations.

This project offers numerous avenues for further development. Initially, the content-based approach could be enhanced by integrating additional criteria to better categorize the movies. A straightforward improvement would involve incorporating attributes to recommend films featuring common actors, directors, or producers. Additionally, the likelihood of recommending movies released in the same era could also be increased, enhancing the recommendation system's relevance and accuracy.

References

1. Humin Yang et al., "Knowledge-Based Recommender System Using Artificial Intelligence for Smart Education", in World Scientific Book, 2022
2. Yuan Wang et al., "Innovative Research on English Teaching Model Based on Artificial Intelligence and Wireless Communication" in International Journal of Reliability, Quality and Safety Engineering, 2022
3. Li Zhang et al., "The Capture and Evaluation System of Student Actions in Physical Education Classroom Based on Deep Learning" in Journal of Interconnection Networks, 2022
4. Haixia Yu et al., "Artificial Intelligence-Based Quality Management and Detection System for Personalized Learning" in Journal of Interconnection Networks, 2021
5. Wang et al., "Smart Educational Learning Strategy with the Internet of Things in Higher Education System" in Jing International Journal on Artificial Intelligence Tools
6. Suhua Wang et al., "Personalized exercise recommendation method based on causal deep learning: Experiments and implications" in Mathematical Modelling and Control
7. Soheila Garshasbi et al., "Microlearning and computer-supported collaborative learning: An agenda towards a comprehensive online learning system" in Mathematical Modelling and Control
8. Tung-Kuang Wu et al., "Rough sets as a Knowledge Discovery and Classification Tool for the Diagnosis of Students with Learning Disabilities" in International Journal of Computational Intelligence Systems, 2011

9. Vaibhav Bagaria et al., "Augmented Intelligence in Joint Replacement Surgery: How can artificial intelligence (AI) bridge the gap between the man and the machine?" in Arthroplasty, 2022
10. Syed Sabir Hussain Rizvi et al., "A review on peak shaving techniques for smart grids" in AIMS Energy, 2023
11. Borges HL and Lorena AC., "A survey on recommender systems for news data" in Smart Information and Knowledge
12. Herlocker JL, Konstan JA, and Riedl J., "Explaining collaborative filtering recommendations" In Proc 2000 ACMConf Comput Supported Cooperative Work, 2000, 241-250p.
13. Subramaniyaswamy V, Logesh R, Chandrashekhar M, et al., "A personalized movie recommendation system based on collaborative filtering" in Int J High Perform Comput Netw, 2017, 10(1-2):54-63.
14. Bleier A, Harmeling C, and Palmatier RW., "Creating effective online customer experiences" in J Market, 2019, 83(2):98-119.
15. Cenci MP, Scarazzato T, Munchen DD, et al., "Eco-friendly electronics—A comprehensive review" in Advan MaterTechnol, 2022, 7(2):2001263.
16. Ashley-Dejo E, Ngwira S, and Zuva T., "A survey of context-aware recommender system and services" in 2015 IntConf Comput, Commun Securi (ICCCS). IEEE, 2015, 1-6p.
17. Kulkarni S, and Rodd SF., "Context Aware Recommendation Systems: A review of the state of the art techniques" in Comput Sci Rev, 2020, 37:100255