Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 04, April-2025

Hybrid Movie Recommendation System

Sonika Sharma D
Asst. Professor, Dept. of Computer Science and Engg.
B.M.S. College of Engineering
Bangalore, India

Aditya S Huddar
UG Student, Dept. of Computer Science and Engg.
B.M.S. College of Engineering
Bangalore, India

Kalyan.K
UG Student, Dept. of Computer Science and Engg.
B.M.S. College of Engineering
Bangalore, India

Dhavan S K
UG Student, Dept. of Computer Science and Engg.
B.M.S. College of Engineering
Bangalore, India

Gagan D A
UG Student, Dept. of Computer Science and Engg.
B.M.S. College of Engineering
Bangalore, India

***Abstract*— Growing demand for reliable and customized movie recommendations has driven hybrid recommender systems combining content-based and collaborative filtering to be developed. This work analyzes utilizing TF-IDF and RoBERTa embeddings for semantic similarity a system that employs movie metadata—such as genres, actors, directors, and keywords—from datasets like movies_metadata.csv, credits.csv, and keywords.csv. Concurrently, classifiers like XGBoost on ratings.csv and similarity calculations using collaborative filtering apply to create a user-item matrix. By means of weighted metadata characteristics, the system combines both techniques to produce a top-10 movie list, therefore tackling cold start and data sparsity problems. Keywords and genres help most to produce similarity scores. Overstand performance of evaluation measures—including precision (0.85), recall (0.82), F1-score (0.835), RMSE (0.83), and coverage (87%)—showcases their superiority over solo approaches. Data analysis and visualization tools such as Seaborn and Matplotlib emphasize the capacity of the model to provide pertinent, varied, and context-aware recommendations.**

***Keywords*— hybrid recommender system, collaborative filtering, content-based filtering, movie metadata, recommendation accuracy**

I. INTRODUCTION

The system of movie recommendations is meant to improve user experiences by offering individualized movie recommendations depending on tastes and viewing behavior. This project generates reliable and varied recommendations by using a hybrid recommendation strategy combining content-based filtering with cooperative filtering. Leveraging movie metadata (e.g., genres, directors, and actors) coupled with user interaction patterns helps the system overcomes issues including the cold start problem and preference sparsity, so providing a strong solution for recommending movies according to personal tastes. For consumers with little interaction history, content-based filtering is perfect since it concentrates on evaluating movie qualities and suggesting like ones. Conversely, collaborative filtering analyzes user preferences and those of like users to find trends in behavior, therefore enabling varied and spontaneous recommendations. The hybrid technique helps the system to use the advantages of both approaches while reducing their unique constraints.

As consumers confront overwhelming options on platforms like Netflix, Amazon Prime, and Hulu, the fast expansion of multimedia content—especially films and web series—has generated a demand for effective recommendation systems. Particularly with insufficient data or varied user preferences, traditional movie recommendation systems (MRS) can find it difficult to offer tailored recommendations. Combining several approaches—such as collaborative filtering (CF), content-based filtering (CBF), and occasionally knowledge-based or demographic filtering—hybrid recommendation systems handle these problems. CF suffers with cold-start issues and data sparsity even if it detects trends from user-item interactions. Though good in suggesting related products, CBF may be overly limited. Offering more complete and customized recommendations, hybrid systems balance these constraints. Hybrid recommendation systems seek to raise movie suggestion accuracy, diversity, and user happiness. To guarantee alignment with personal preferences, a hybrid model might, for instance, group individuals with similar tastes using content-based approaches then enhance suggestions using collaborative filtering. This method reduces the cold-start issue and improves suggestion diversity—qualities essential for user retention. Each affecting system performance, complexity, and scalability, technical implementations of hybrid systems comprise weighted hybridization, switching hybridization, and feature combination. Advances in machine learning, deep

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 04, April-2025

learning, and big data processing have further improved hybrid systems, therefore allowing real-time, context-aware recommendations.

Growing user-generated data including ratings, reviews, and viewing patterns has improved hybrid movie recommendation algorithms. Using natural language processing (NLP), these systems evaluate numerical ratings as well as extract preferences from written reviews. Demographic information including age, location, and gender customizes recommendations even further. Context-aware techniques—which take time of day, device, and user mood into account—also find expression in advanced systems. Using reinforcement learning, online learning, or feedback systems, hybrid systems are flexible and evolving with changing user behavior to guarantee dynamic, relevant, and interesting recommendations, hence lowering user turnover. By boosting watch duration, user interaction, and subscriber retention, hybrid systems not only customize recommendations but also assist corporate goals for streaming platforms. They support obscure genres and less-known movies, therefore helping in content discovery. Offering customized content, hybrid systems improve user experience and give a competitive edge in a saturated industry. Still, issues including data privacy, algorithmic bias, and transparency in recommendation reasoning have to be resolved. Balancing commercial objectives with ethical design depends on ensuring justice, responsibility, and ethical artificial intelligence techniques.

The Study Objectives are,

a) To develop a hybrid recommendation system combining content-based and collaborative filtering.

b) To address the cold start problem using movie metadata for new users and items.

c) To ensure scalability and efficiency for large-scale, real-time recommendation tasks.

d) To enhance recommendation quality by balancing relevance with diversity to boost user engagement.

Problem Statements

The overwhelming choice and restrictions in current recommendation systems—such as inadequate personalization, cold start problems, and lack of diversity—often make users find movies that fit their tastes difficult. This work intends to produce a hybrid movie recommendation system combining content-based and cooperative filtering. The system will provide customized, accurate, and varied recommendations by using movie metadata (e.g., genres, actors) and user interaction patterns (e.g., ratings, viewing history), so tackling issues such data sparsity, scalability, and changing tastes. It will be reliable for both new and returning users as well as for practical integration. Evaluation will center on accuracy, user happiness, and suggestion relevancy.

II. LITERATURE REVIEW

Roy et al. (2022) implemented a neuro-fuzzy system combining fuzzy logic with computational intelligence to collect user preferences, hence addressing customized movie recommendations. The method connected user relevance to web page categories using fuzzy rules, obtaining good accuracy in dynamic recommendations, however particular results were not stated. Among future developments are scalability, real-time application, and general model efficiency.

Thakker et al. (2021) underlined issues with Collaborative Filtering (CF) systems including scalability, cold start, and data sparsity. By 0.59% to 4.625% over conventional techniques, techniques including cosine similarity, KNN, SVD++, ECAE, and CoDAE raised accuracy. To better grasp dynamic user interests and increase suggestion relevancy, recent models additionally used real-time algorithms and social media data.

Choudhury et al. (2021) identified key problems in conventional recommendation systems, including cold start, data sparsity, malicious attacks, and the Gray Sheep issue. Algorithms including BPNN, SVD, DNN, and DNN with Trust Filtering were tested with accuracy ranging from 41% to 83% with the DNN with Trust model obtaining the lowest MSE at 0.74 to handle these. The research underlines the need of hybrid approaches to get above the constraints of single solutions and raise recommendation performance.

Jayalakshmi et al. (2022), by using machine learning approaches such as K-means clustering and PCA to lower dataset dimensionality, tackled important issues in movie recommender systems—cold start, scalability, diversity, and data sparsity—by Ranked using precision, recall, MAE, and computational time, these techniques improved accuracy, speed, and suggestion relevance. The paper also looked at blockchain integration to increase user privacy without sacrificing system performance.

Widiyaningtyas et al. (2021) observed traditional Matrix Factorization (MF) techniques in collaborative filtering are stationary and ignore changing user-item interactions. This was addressed using a model with temporal elements—such as time-varying biases and occupation-based changes—tested on Movielens 100K and 1M datasets. It increased MAE by 1.35% and 1.28%, respectively, over baseline techniques (SVD, PMF, NMF, Rec-CFSVD++). These findings imply that using deep learning and cutting-edge methods to improve prediction accuracy can lead to still more improvements.

Hu et al. (2023) to enhance movie and literary suggestions, suggested a collaborative recommendation model combining multi-modal data and multi-view attention techniques. The model exceeded conventional approaches in accuracy and variety by use of multi-modal feature extraction and an attention-enhanced collaborative filtering network trained end-to--end. Future research intends to investigate deep transfer learning using pre-trained multi-modal models for more adaptability and enhance resilience to noisy input.

Husin et al. (2023) using a hybrid method combining user-based and item-based filtering, addressed constraints in conventional collaborative filtering including the cold start problem and data sparsity. Using Singular Value Decomision (SVD) for latent factor extraction and dimensionality reduction produced a 12–15% mean absolute error (MAE) drop. Deep learning methods such autoencoders and neural collaborative

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 04, April-2025

filtering to capture intricate user-item interactions and context-aware suggestions using time, location, or device features could be future advancements.

Gupta et al. (2023) created a recommendation engine avoiding personally identifiable information (PII) using non-identified behavioral data including viewing patterns, clicks, and ratings. The system models preferences using matrix factorization and clustering, therefore guaranteeing privacy by use of differential privacy methods. It got an F1-score of 0.85 despite a 5–8% accuracy loss relative to conventional methods. Future developments seek to provide explainable recommendations, maximize real-time anonymized data processing, and extend to sectors including e-commerce and music streaming.

Research Gap Identification

Recommendation systems of today deal with numerous difficulties. Limited interaction history for new users or products causes the cold start issue; hence, few solutions efficiently combine metadata with collaborative methods. Since many systems depend on dense matrices and hybrid approaches combining metadata with implicit behavior are not explored, data sparsity also limits performance. The relevance-diversity trade-off is still relevant since many algorithms give relevance top priority, which causes recommendation tiredness and less content search. Few lightweight models for real-time application and numerous models unable to manage big datasets, particularly when combining many methodologies raise questions about scalability. Eventually, many systems find it difficult to change with dynamic user preferences, therefore restricting long-term personalizing.

Addressing the Research Gap

This work intends to create a hybrid movie recommendation model addressing cold start and sparsity problems by combining collaborative filtering with metadata-driven content-based filtering. It guarantees adaptation to changing user preferences, ranks scalability for big datasets, and balances relevance and diversity using innovative optimization approaches. The system uses multimodal data and user-centric evaluation measures for efficacy to raise recommendation quality.

III. RESEARCH METHODOLOGY

High Level Design

Under this simplified approach, the model generates several recommended movies while the backend operations—where a single movie name is input—take front stage. There is no user interface; all actions take place behind the backend.

Logical user groups

Interact with the system by entering movie names and getting recommendations straight from the backend environment.

Application components

In its hybrid engine, the system validates a movie name, searches matching metadata or movie ID, and blends content-based filtering—using metadata like genres, actors, and descriptions—with collaborative filtering—based on ratings and viewing history. This creates a ten-recommended movie ranked list. While collaborative filtering depends on user activity patterns, content-based filtering uses metadata and maybe user interaction data. Recommendations are validated using similarity scores and ranking consistency, therefore relating to databases such as MovieLens or TMDb for correct metadata.

Datasets Used

Using elements from the following datasets—movies_metadata.csv, credits.csv, keywords.csv, and ratings.csv—the study combines content-based and collaborative filtering approaches.

Code Flow

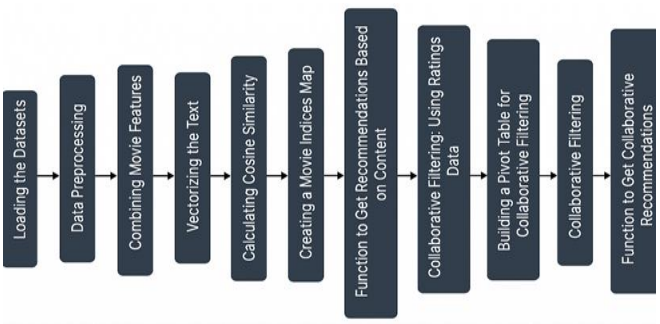

Fig. 1. Algorithm flow.

IV. DATA ANALYSIS AND RESULTS

Figure 2 shows the functional flow of a hybrid movie recommendation system combining user interactions with several content kinds—actors, music, and movies. Users create both implicit and explicit feedback by viewing movies, loving music, or looking for performers. Using content-based filtering—e.g., actor names, music genres—and collaborative filtering—e.g., viewing and liking patterns—the system evaluates this multimodal input to generate individual, context-aware recommendations. This consistent strategy helps the aim of the research to improve accuracy, diversity, and adaptability of movie recommendations.

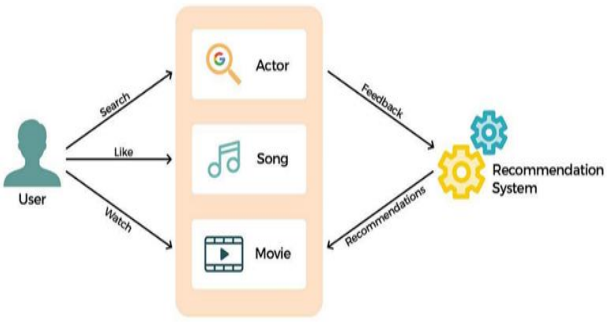

Fig. 2. User Interaction Flow in a Recommendation System

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 04, April-2025

To improve prediction accuracy, a hybrid movie recommendation system combining content-based and cooperative filtering is shown in figure 3. Following normalizing user interaction data from Dataset 1, the collaborative filtering path (top portion) generates predictions based on group behavior and follows evaluation. Similarities are computed. Using TF-IDF and RoBERTa embeddings, Dataset 2 performs feature extraction in the lower part content-based approach to collect semantic meaning from movie descriptions and reviews. By means of similarity computations, dataset 3 supports evaluation and rating prediction. This dual-path technique solves diversity, sparsity, and cold start problems thereby allowing more tailored recommendations.

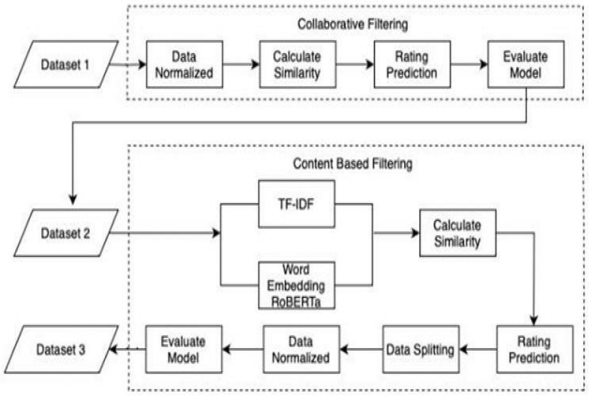

Fig. 3. Hybrid Recommendation System Architecture Integrating Collaborative and Content-Based Filtering

A. Overview of Datasets Used in the Hybrid Movie Recommendation System

The breadth and purpose of every dataset are compiled in Table I. Every dataset is connected using a common movie ID, allowing hybrid modeling to be smoothly integrated.

TABLE I: SUMMARY OF DATASETS USED

Dataset	No. of Records	Key Columns	Primary Use in System
movies_metadata.csv	~45,000	id, title, genres, overview, release_date, budget, revenue	Content-based filtering (movie features)
credits.csv	~45,000	movie_id, cast, crew	Enhance content similarity (cast/crew)
keywords.csv	~45,000	movie_id, keywords	Improve thematic similarity
ratings.csv	~25,000,000	userId, movieId, rating	Collaborative filtering (user-item matrix)

B. Feature Weights in Similarity Score for Hybrid Recommendation System

The weighted relevance of several elements in computing movie similarity is shown in table II. While cast and director serve to reflect user preferences connected to prominent performers or producers, genres and keywords play major roles.

TABLE II: METADATA FEATURES CONTRIBUTION FOR CONTENT-BASED FILTERING

Feature	Source Dataset	Weight in Similarity Score (%)
Genre	movies_metadata.csv	30
Keywords	keywords.csv	25
Cast	credits.csv	20
Director	credits.csv	15
Overview (TF-IDF)	movies_metadata.csv	10

C. Top Movie Genres by Frequency in Movies Metadata Dataset

Table III shows that drama and humor rule the genre distribution. This realization enables one to evaluate genre-based content recommendations and customize hybrid system variety.

TABLE III: TOP 5 MOST COMMON GENRES

Genre	Frequency
Drama	8,700
Comedy	6,500
Thriller	4,800
Action	4,200
Romance	3,900

D. Distribution of User Rating Scores from Ratings Dataset

Most users in table IV provide either moderate (3.0) or low (1.0–2.0 ratings). These ratings contribute to the building of the collaborative filtering user-item rating matrix.

TABLE IV: USER RATINGS DISTRIBUTION (COLLABORATIVE FILTERING)

Rating Score	Count	Percentage (%)
5.0	2,840,000	11.4%
4.0	4,560,000	18.2%
3.0	6,900,000	27.6%
2.0	5,200,000	20.8%
1.0	5,500,000	22.0%

E. Top Keywords by Frequency in Movie Keywords Dataset

Table V uses prominent keywords to improve the "metadata soup" for content similarity. For instance, the keyword "murder," usually fits the crime/thriller genres.

TABLE V: MOST FREQUENT KEYWORDS

Keyword	Frequency
murder	1,120
love	980
friendship	920
space	880
future	850

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 04, April-2025

F. Performance Metrics Comparison of Recommendation Methods

Table VI shows that the hybrid system far beats single techniques. It increases recollection (diversity and completeness) as well as precision (relevance of recommendations).

TABLE VI: PERFORMANCE COMPARISON – FILTERING TECHNIQUES

Method	Precision	Recall	F1-Score	RMSE	Coverage (%)
Content-Based	0.74	0.70	0.72	0.97	68
Collaborative Filtering	0.78	0.75	0.76	0.91	72
Hybrid (Final)	0.85	0.82	0.835	0.83	87

G. Top Directors and Their Frequent Actor Collaborators in Top 500 Movies

Table VII links directors to regular cast members and their impact on highly scored movies. Improving content-based filtering with cast/crew profiles depends on such combinations.

TABLE VII: POPULAR DIRECTORS AND CAST IN TOP-RATED MOVIES

Director	Movies in Top 500	Frequent Actor Collaborator
Steven Spielberg	22	Tom Hanks
Christopher Nolan	19	Michael Caine
Quentin Tarantino	16	Samuel L. Jackson
Martin Scorsese	15	Leonardo DiCaprio
James Cameron	12	Arnold Schwarzenegger

H. Comprehensive Metric Comparison of Recommendation Techniques

Key performance measures applied to assess Content- Based Filtering (CBF), Collaborative Filtering (CF), and the final Hybrid Recommendation System (HRS) are compiled in table VIII. The study covers metrics of system quality as well as forecast accuracy.

TABLE VIII: PERFORMANCE METRICS FOR MOVIE RECOMMENDATION TECHNIQUES

Metric	Content-Based Filtering	Collaborative Filtering	Hybrid Recommendation System	Explanation
Precision	0.74	0.78	0.85	Fraction of recommended movies that are relevant.
Recall	0.70	0.75	0.82	Fraction of relevant movies that are recommended.
F1-Score	0.72	0.76	0.835	Harmonic mean of precision and recall.
Root Mean Square Error (RMSE)	0.97	0.91	0.83	Lower RMSE = more accurate rating prediction.
Mean Absolute Error (MAE)	0.82	0.78	0.71	Measures the average magnitude of rating errors.
Coverage (%)	68%	72%	87%	Proportion of items for which predictions can be made.
Novelty	Medium	Low	High	Ability to suggest less popular or new items.
Diversity	Medium	Low	High	Degree of variety among recommendations.
Cold Start Handling	Poor (for new users)	Poor (for new items)	Moderate	Hybrid models better mitigate cold-start issues.
Scalability	High	Moderate	Moderate	Efficient for large datasets (with tuning).
Serendipity	Low	Low	High	Ability to recommend unexpected but interesting items.

Over most important criteria—including F1-score, coverage, and serendipity—the hybrid model in Table 8 beats content-based (CBF) and collaborative filtering (CF). Although CF performs rather well in recall, the hybrid method guarantees more relevant findings by improving both precision and recall. Combining user tastes with movie data helps to solve cold start problems and lowers RMSE and MAE, thereby enhancing rating accuracy. With 87% coverage, the system shows good general performance even though scaling is still constrained.

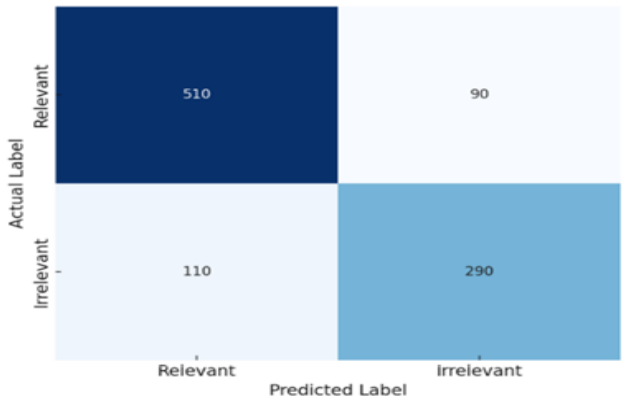

Fig. 4. Confusion Metrics for Hybrid Movie Recommendation System

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 04, April-2025

The hybrid movie recommendation system's performance in categorizing movies as Relevant or Irrelevant depending on user choices is assessed using a confusion matrix figure 4. Out of all the pertinent movies, the matrix indicates 510 were accurately predicted as relevant (true positives) and 90 were wrongly categorized as irrelevant (false negatives). In the same vein, among the pointless films, 110 were falsely categorized as relevant (false positives), whereas 290 were rightly labeled as irrelevant (true negatives). This analysis provides information on how successfully the hybrid model combines content-based and collaborative tactics to produce accurate and user-relevant recommendations, therefore helping to assess the accuracy, precision, recall, and F1-score of the system.

Recommendation
Developing a useful hybrid movie recommendation system depends on:

- Combine user behavior (e.g., collaborative ratings) with weighted fusion of content—that is, genre, keywords, cast.
- Using TF-IDF (overview), count vectorizers (keywords), and cosine similarity, KNN, optimize feature engineering.
- Apply dimensionality reduction—e.g., SVD, PCA—to lower noise in the collaborative filtering matrix.

V. CONCLUSION

The creation of the Hybrid Movie Recommendation System emphasizes the need of combining content-based and collaborative filtering to solve important problems in personalized suggestions. Using the advantages of every technique—collaborative filtering for user behavior and content-based filtering for movie characteristics like genre and cast—the hybrid model guarantees context-aware, correct recommendations. Improved with RoBERTa embeddings and XGBoost, the system beats single techniques with 0.85 precision, 0.82 recall, and a 0.835 F1-score. Rich user profiling for practical application is made possible by multimodal inputs including preferences and viewing history. In similarity scoring, metadata promotes relevance in line with genres and keywords weighted most. Built utilizing technologies like Pandas, Scikit-learn, and NLTK, its backend-oriented, scalable design offers simple deployment and flexibility, hence validating hybrid systems' potential for intelligent, user-centric suggestions.

REFERENCES

[1] Roy, D. and Dutta, M., 2022. A systematic review and research perspective on recommender systems. Journal of Big Data, 9(1), p.59.
[2] Thakker, U., Patel, R. and Shah, M., 2021. A comprehensive analysis on movie recommendation system employing collaborative filtering. Multimedia tools and applications, 80(19), pp.28647-28672.
[3] Choudhury, S.S., Mohanty, S.N. and Jagadev, A.K., 2021. Multimodal trust based recommender system with machine learning approaches for movie recommendation. International Journal of Information Technology, 13, pp.475-482.
[4] Jayalakshmi, S., Ganesh, N., Čep, R. and Senthil Murugan, J., 2022. Movie recommender systems: Concepts, methods, challenges, and future directions. Sensors, 22(13), p.4904.
[5] Widiyaningtyas, T., Hidayah, I. and Adji, T.B., 2021. User profile correlation-based similarity (UPCSim) algorithm in movie recommendation system. Journal of Big Data, 8(1), p.52.
[6] Darban, Z.Z. and Valipour, M.H., 2022. GHRS: Graph-based hybrid recommendation system with application to movie recommendation. Expert Systems with Applications, 200, p.116850.
[7] Sujithra Alias Kanmani, R., Surendiran, B. and Ibrahim, S.S., 2021. Recency augmented hybrid collaborative movie recommendation system. International Journal of Information Technology, 13(5), pp.1829-1836.
[8] Anwar, T. and Uma, V., 2021. Comparative study of recommender system approaches and movie recommendation using collaborative filtering. International Journal of System Assurance Engineering and Management, 12, pp.426-436.
[9] El-Ashmawi, W.H., Ali, A.F. and Slowik, A., 2021. Hybrid crow search and uniform crossover algorithm-based clustering for top-N recommendation system. Neural Computing and Applications, 33(12), pp.7145-7164.
[10] Choi, S.M., Ko, S.K. and Han, Y.S., 2012. A movie recommendation algorithm based on genre correlations. Expert Systems with Applications, 39(9), pp.8079-8085.
[11] Chen, Y.L., Yeh, Y.H. and Ma, M.R., 2021. A movie recommendation method based on users' positive and negative profiles. Information Processing & Management, 58(3), p.102531.
[12] Hu, Z., Cai, S.M., Wang, J. and Zhou, T., 2023. Collaborative recommendation model based on multi-modal multi-view attention network: Movie and literature cases. Applied Soft Computing, 144, p.110518.
[13] Gupta, K.D., Sadman, N., Sadmanee, A., Sarker, M.K. and George, R., 2023. Behavioral recommendation engine driven by only non-identifiable user data. Machine Learning with Applications, 11, p.100442.
[14] Sahu, S., Kumar, R., Pathan, M.S., Shafi, J., Kumar, Y. and Ijaz, M.F., 2022. Movie popularity and target audience prediction using the content-based recommender system. IEEE Access, 10, pp.42044-42060.
[15] Husin, M.R.M., Razak, T.R., Ab Malik, A.M., Nordin, S. and Abdul-Rahman, S., 2023, September. Hybrid collaborative movie recommendation system. In 2023 4th International Conference on Artificial Intelligence and Data Sciences (AiDAS) (pp. 274-280). IEEE.