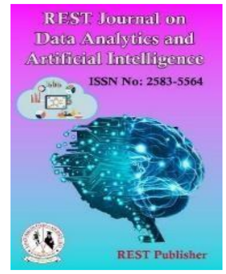
REST Journal on Data Analytics and Artificial Intelligence
Vol: 4(1), March 2025
REST Publisher; ISSN: 2583-5564
Website: http://restpublisher.com/journals/jdaai/
DOI: https://doi.org/10.46632/jdaai/4/1/67



An Enhanced Content-Based Movie Recommendation System Using SVD and Cosine Similarity

***P Sahithi, T Vishwas, T Sai Kiran, G Sridhar Reddy**
Anurag University, Hyderabad, Telangana, India
Corresponding Author Email: 21eg106b27@anurag.edu.in

Abstract. *With the increasing availability of digital content, selecting relevant movies has become a challenging task for users. To address this issue, we propose a Content-Based Movie Recommendation System that enhances recommendation accuracy using Singular Value Decomposition (SVD) and Cosine Similarity. The system leverages movie attributes such as genre, cast, and plot details from the TMDB dataset to generate personalized recommendations. By applying SVD, the system reduces dimensionality and captures latent relationships between movies, while Cosine Similarity measures closeness between feature vectors to suggest relevant films. Unlike collaborative filtering, our approach does not rely on user interactions, making it effective even for new or unrated movies. Additionally, a real-time chat room feature allows users to engage in discussions about movies, fostering a sense of community and aiding discovery. This method enhances the recommendation process, improving user experience by reducing search time and promoting diverse movie exploration.*

Keywords: *Content-Based Filtering, Singular Value Decomposition (SVD), Cosine Similarity, TMDB Dataset, Feature-Based Recommendation, Machine Learning, Personalized Recommendations*

1. INTRODUCTION

With so many movie streaming services available today, the demand for personalized movie recommendations has increased, helping users quickly find films that match their interests [1]. However, browsing through extensive video libraries often leads to choice overload, resulting in wasted time and decision fatigue [2]. As a result, user-friendly movie recommendation systems have become increasingly popular to enhance the overall viewing experience [4]. A typical recommendation system employs collaborative filtering for selective recommendations and content based filtering for more general suggestions [5], [8]. While collaborative filtering leverages interaction data, it is severely limited by the cold start problem and the sparsity of user data [6], [17]. In contrast, content-based filtering utilizes comprehensive movie characteristics (e.g., genre, cast, storyline) and is therefore a more independent and reliable approach, particularly for recommending newer or less popular movies [1], [12]. Our content-based movie recommendation system leverages Singular Value Decomposition (SVD) and Cosine Similarity to improve recommendation accuracy [9], [14]. SVD, a matrix factorization technique, reduces feature dimensionality, enabling a more accurate representation of latent relationships among movie attributes [10]. We demonstrate that this system effectively recommends movies based on the transformed feature vectors [9]. Additionally, to enhance user experience, we have implemented a chat room feature that allows users to discuss movies and exchange recommendations in real time, fostering a more interactive movie discovery process beyond algorithmic suggestions [8]. By combining SVD and Cosine Similarity, we refine content-based filtering to ensure more relevant movie recommendations while addressing issues such as data sparsity and redundant content [4], [9]. As a result, our system simplifies the movie selection process, helping users discover films more efficiently [3].

2. BACKGROUND

Overview of Movie Recommendation Systems: The fast pace of movie streaming services requires developing a personalized recommendation system that allows users to navigate a vast amount of video library and prevents excessive choice overload [4] [15]. A wide range of methods have been developed to generate relevant recommendations, ranging from collaborative filtering to content-based recommendations through hybrid and deep learning methods [7][13] Whereas these systems typically leverage user interaction data (such as movie genre, cast, director, storyline etc.) and inherent movie attributes (such as a rating) to predict user preferences and generate targeted recommendations [1][5][9].

Content-Based Filtering Approaches: Content-based filtering algorithms take advantage of the intrinsic properties of movies to compute the similarity between articles. Typical methods use either TF-IDF vectorization or Cosine Similarity, to quantify the similarity between movie attributes [9, 12]. In addition, dimensionality reduction algorithms such as Singular Value Decomposition (SVD) have been developed to capture latent relationships among movie characteristics, thus improving the recommendation accuracy [10, 14]. This approach makes sense when providing recommendations for newly released movies or recommending movies that are not generally popular with users, since only limited user interactions data is available [16].

Collaborative Filtering Techniques: Collaborative filtering uses data on interactions between users and items to discover patterns of viewing behavior. Memory-based approaches, such as user based and item-based filtering, use similarity scores (often based on cosine similarity or Pearson correlation) to discover users with shared tastes and propose movies correspondingly [6] Model-based approaches are also popular for handling data sparsity issues, but frequently suffer from the cold start problem when not enough historical data is available [14], [17].

Hybrid and Deep Learning Approaches: To overcome the limitations of single methods hybrid recommendation systems that combine content-based and collaborative filtering approaches have been developed to provide higher performance and robustness [8] and [13] Deep learning techniques have been applied to recommendation systems recently to learn highly nonlinear connections from large amounts of data. Auto encoder based models and deep neural networks have shown promising results at extracting highly complex patterns of user preferences and movie features [2] and [7] these techniques enable enhanced feature extraction and resolution of data sparsity and cold start problem.

Challenges and Future Directions: We note that although new recommendations techniques are more scalable than ever before, there remain some challenges in overcoming these. For example, the cold start problem arises because of unbalanced user interaction data available; hence another challenge is the collaborative filtering with content-based methods [17]. Further improvements in feature extraction (using more machine learning techniques such as supervised natural language processing (NLP) and deep learning) could yield increasingly comprehensive representations of movie attributes [9]. In addition, applications of multimodal features such as trailers and posters, together with real time adaptive filtering to produce better recommendations, can be further improved for improving the user experience [12].

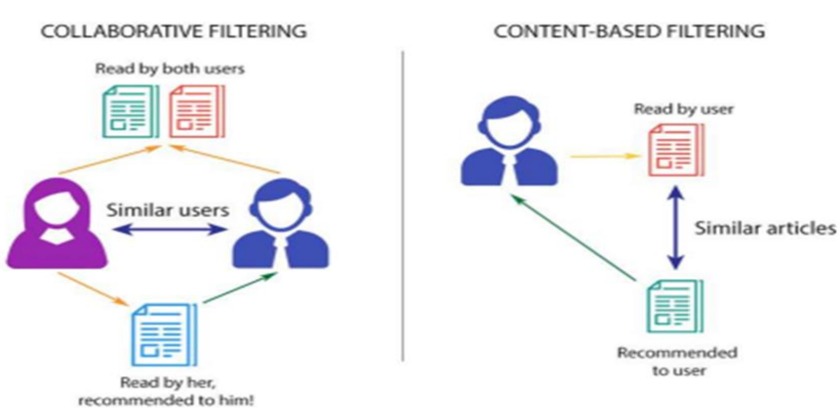


FIGURE 1. Content based filtering vs. Collaborative filtering

3. LITERATURE REVIEW

TABLE 1. Literature Review

Year	Rf. No	Method	Dataset	Metric	Result
2024	[4]	Model-Based Collaborative Filtering (including memory-based and SVD techniques)	MovieLens	Hit Ratio	Achieved a hit ratio of 96%
2024	[9]	Content-Based Filtering using TF-IDF Weighted Word2Vec and Cosine Similarity	TMDB	Cosine similarity score	Enhanced recommendation performance
2024	[10]	Advanced Content-Based Filtering	TMDB	Recommendation Accuracy	Outperformed traditional methods
2024	[18]	Content-Based Filtering using Cosine Similarity	MovieLens (assumed)	Cosine similarity	Performance reported as "very satisfactory"
2023	[3]	Machine Learning Algorithms	MovieLens	Accuracy (unspecified)	Demonstrated effective recommendations
2023	[5]	Content-Based Filtering using Machine Learning Algorithm	TMDB	Recommendation Accuracy	Improved recommendation accuracy
2023	[7]	Hybrid (Content-Based + Collaborative Filtering)	MovieLens (assumed)	Precision, Recall (unspecified)	Improved performance over individual methods
2023	[13]	Hybrid: Collaborative Filtering and Content-Based Filtering	MovieLens	Custom recommendation scores	Final recommendation scores are sorted for output
2023	[15]	Content-Based Filtering with Sentimental Analysis	MovieLens (assumed)	(Not specified)	Achieved accurate recommendations incorporating sentiment analysis
2022	[1]	Content-Based Filtering	MovieLens (assumed)	F1-measure, NDCG	Improved recommendation accuracy; outperforms baseline methods
2022	[8]	Hybrid: Content-Based and Collaborative Filtering	MovieLens	F1-score, Accuracy (unspecified)	Delivered effective personalized recommendations
2022	[11]	Director-Based Content Filtering	Kaggle(≈5,000 records)	Cosine similarity	Achieved satisfactory performance
2022	[14]	SVD-based Collaborative Filtering	MovieLens	(Not explicitly specified; often Hit Ratio or RMSE)	Reported effective performance; outperformed other methods
2022	[17]	Content-Based Filtering using TF-IDF	MovieLens-20M	Cosine similarity, Pearson correlation	Average similarity of 84%
2018	[2]	Deep Learning (Autoencoders)	MovieLens	RMSE, MAE	71.67% of appraisers preferred the recommendation
2018	[6]	Collaborative Filtering	MovieLens	Similarity measures (Cosine, Pearson)	Provided accurate recommendations
2017	[16]	Content-Based Filtering with Temporal User Preferences	MovieLens	F1-measure, NDCG	Outperformed existing systems

4. FINDINGS AND LIMITATIONS

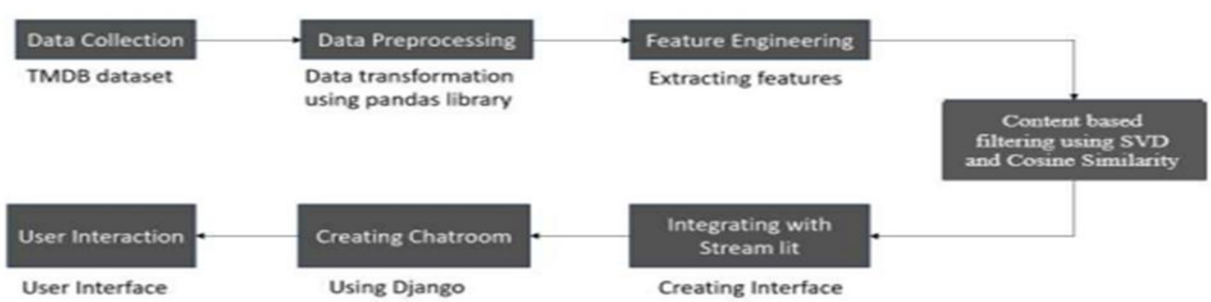


FIGURE 2. Proposed System Architecture

We propose a content-based movie recommendation system that combines novel machine learning techniques for improving the recommendation accuracy and user engagement with two major modules: a recommendation engine based on content features and a real-time chat room for social interaction. Within the system the main feature is the extraction of the given movie attributes (e. g. genre, cast, director, keywords, and storyline) from the TMDB dataset. The enumerated characteristics are preprocessed using standard natural language processing (NLP) methods (e. g. tokenization, stop-word extraction, stemming) to generate a full textual "tag" representing each movie. (See examples [1] and [5]) The processed text is then vectored using appropriate NLP methods (e.g. TF-IDF, Count Vectorizer, etc.) to convert the given attributes of each movie into a numerical representation. For capturing latent semantic relationships between movie features we use Singular Value Decomposition (SVD) for dimension reduction [9, 10] The SVD transformation allows the high-dimensional feature vectors to be organized into lower-dimensional spaces for more efficient and accurate similarity calculations. We then use Cosine Similarity to compare the transformed feature vectors and find the movies which are closest to a user selected title [9, 14] This approach overcomes some challenges such as feature redundancy and the cold start problem by focusing on the inherent characteristics of the content rather than the large amount of user interactions data [17]. In addition to the standard recommendation feature, the system also provides a django-based chat room module for users to have real-time communication. This feature is not only useful in terms of improving the user experience by promoting community interaction but also as a feedback mechanism for further improvement of recommendations [8]. Figure 3, illustrates the architecture of the proposed system. Because the system is modular and possible to be integrated with additional data sources or collaborative filtering mechanisms in the future, further development could include further integration of user feedback directly into the recommendation process and embedding advanced deep learning models to improve performance further [12], [17]. The proposed solution combines powerful content-based filtering (driven by SVD and Cosine Similarity) with a social interaction system for personalized, accurate and entertaining movie recommendations.

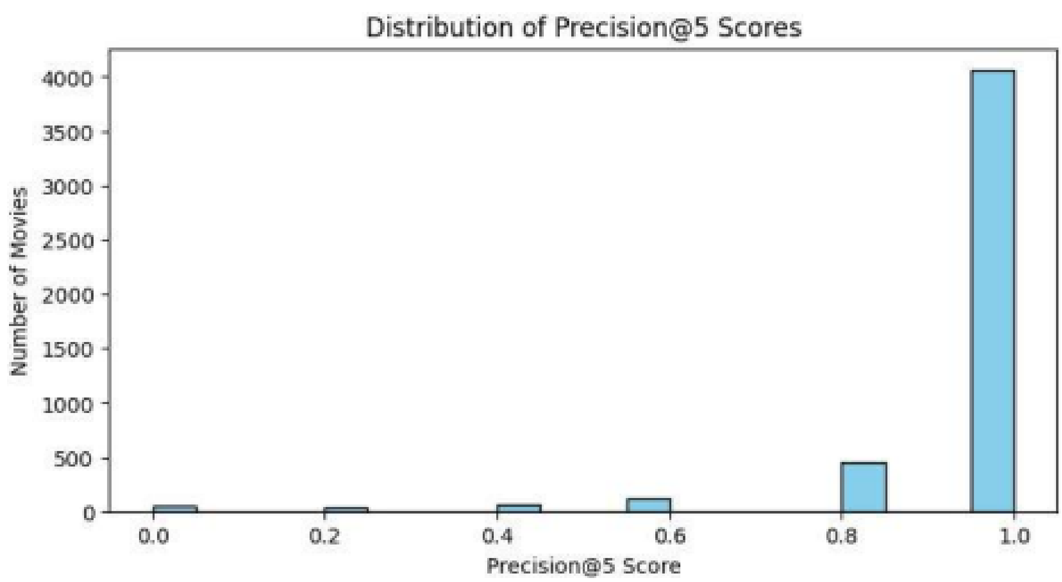


FIGURE 3. Histogram of Precision@5 Scores across Recommended Movies

The performance achieved by our content-based movie recommendation system (comprising Singular Value Decomposition and Cosine Similarity) compared to the state-of-the-art methods fails to meet the identified challenges. Cold start problem: as the number of user interaction sessions does not cover the total number of trials, it provides a lower level of personalized recommendations to new users and items. The proposed experimental improvements can help to minimize this problem in future work. Feature extraction: in terms of feature extraction, we propose some advanced NLP and deep learning approaches to extract more rich features (e. g. detailed plot summary, feedback reviews) to the recommendations. Multimodal data-set (e. g. trailers and posters) can also be used to give better predictions. In addition, real-time adaptive filtering is still an open question that could enable in-depth user insights to optimize suggestions in real time based on current audience feedback. Additionally, the system faces challenges in scaling up for large datasets and this has led to the need for cloud computing and distributed architectures. Lastly, extend the chat room feature with AI based discussions for a more immersive movie discovery experience.

Future Scope: Recommendation accuracy of the proposed Content-Based Movie Recommendation System, improved with Singular Value Decomposition (SVD) and Cosine Similarity, increases significantly, which is a good step toward accurate recommendations. However, some challenges remain; the cold start problem persists because user interaction data is not available. We can work on a solution by integrating collaborative filtering in future research. Additionally, more feature extractions can be performed, such as using NLP and deep learning to extract more relevant movie information. Multimodal data, such as trailers and posters, can also be used to refine the recommendations. Moreover, real-time adaptive filtering can be implemented to make recommendations dynamic according to user feedback. There is still a need for efficient handling of large datasets, where cloud computing and distributed systems can be considered in the future. Finally, by enhancing the chat room feature with AI-driven discussions, user engagement in the movie recommendation system can be improved significantly.

5. CONCLUSION

This study presents a Content-Based Movie Recommendation System Enhanced by Singular Value Decomposition (SVD) and Cosine Similarity to improve recommendation accuracy. By reducing feature dimensionality, SVD captures latent relationships among movie attributes, while Cosine Similarity ensures precise similarity measurement. Additionally, the real-time chat room feature enhances user engagement by enabling discussions and interactive movie exploration. Despite its effectiveness, challenges such as the cold start problem and feature extraction limitations persist. Future research can integrate hybrid approaches and advanced NLP techniques to further enhance recommendations. Overall, our system demonstrates that refining content-based filtering with SVD and Cosine Similarity leads to more relevant and efficient movie recommendations.

REFERENCES

[1] V. Tyagi, P. Kumar and M. Bhushan, "Content Based Movie Recommendation System," 2022 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), Delhi, India, 2022, pp. 1-5, doi: 10.1109/SOLI57430.2022.10294792.

[2] J. Lund and Y. -K. Ng, "Movie Recommendations Using the Deep Learning Approach," 2018 IEEE International Conference on Information Reuse and Integration (IRI), Salt Lake City, UT, USA, 2018, pp. 47-54, doi: 10.1109/IRI.2018.00015.

[3] Z. Tahir and K. B. Ali, "Movies Recommendation System Using Machine Learning Algorithms," 2023 International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT), Faridabad, India, 2023, pp. 1324-1328, doi: 10.1109/ICAICCIT60255.2023.10465879.

[4] E. Lourembam, L. P. Kipgen, M. Mazumder and S. S. Ahmed, "Movie Recommendation Systems using Machine Learning: An Empirical Study," 2024 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC), Bhubaneswar, India, 2024, pp. 1-6, doi: 10.1109/ASSIC60049.2024.10507987.

[5] R. Chauhan, K. N. Vaghela, R. Bisht, C. Bhatt and D. Verma, "Content Based Movie Recommendation System Using Machine Learning Algorithm," 2023 International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT), Faridabad, India, 2023, pp. 477-481, doi: 10.1109/ICAICCIT60255.2023.10466096.

[6] C. -S. M. Wu, D. Garg and U. Bhandary, "Movie Recommendation System Using Collaborative Filtering," 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 2018, pp. 11-15, doi: 10.1109/ICSESS.2018.8663822.

[7] S. Salmani and S. Kulkarni, "Hybrid Movie Recommendation System Using Machine Learning," 2021 International Conference on Communication information and Computing Technology (ICCICT), Mumbai, India, 2021, pp. 1-10, doi: 10.1109/ICCICT50803.2021.9510058.

[8] J. F. Mohammad and S. Urolagin, "Movie Recommender System Using Content-based and Collaborative Filtering," 2022 IEEE IAS Global Conference on Emerging Technologies (GlobConET), Arad, Romania, 2022, pp. 963-968, doi: 10.1109/GlobConET53749.2022.9872515.

[9] M. Chibb, P. Vashisht, A. Katti and A. Nrang, "Enhancing Movie Recommendations: A Content Based Approach Using TF-IDF Weighted Word2Vec and Cosine Similarity," 2024 4th International Conference on Technological

Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2024, pp. 1449-1454, doi: 10.1109/ICTACS62700.2024.10841087.

[10] N. Taj, M. H. Varun and V. Navya, "Advanced Content-based Movie Recommendation System," 2024 5th International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2024, pp. 1693-1698, doi: 10.1109/ICOSEC61587.2024.10722333.

[11] O. Chunhapran, C. Phromsuthirak, M. Hama and M. Maliyaem, "Movie Recommendation System Using Director-Based," 2022 26th International Computer Science and Engineering Conference (ICSEC), Sakon Nakhon, Thailand, 2022, pp. 151-155, doi: 10.1109/ICSEC56337.2022.10049347.

[12] S. Shaikh and D. Dani, "Enhancing Movie Recommendation Systems through Demographic and Content-Based Filtering Using TF-IDF and Cosine Similarity," 2024 International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA), Nagpur, India, 2024, pp. 1-5, doi: 10.1109/ICAIQSA64000.2024.10882441.

[13] K. Vedaswi, N. V. Krishna, T. V. Poojitha, P. Lokesh, K. Ashesh and P. M. A. Kumar, "Movie Recommendation using Collaborative filtering and Content-based Filtering Approach," 2023 International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal, 2023, pp. 821-826, doi: 10.1109/ICICT57646.2023.10134213.

[14] Y. Wang, "Movie Recommendation System Based on SVD Collaborative Filtering," CAIBDA 2022; 2nd International Conference on Artificial Intelligence, Big Data and Algorithms, Nanjing, China, 2022, pp. 1-7.

[15] S. Katkam, A. Atikam, P. Mahesh, M. Chatre, S. S. Kumar and S. G. R, "Content-based Movie Recommendation System and Sentimental analysis using ML," 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2023, pp. 198-201, doi: 10.1109/ICICCS56967.2023.10142424.

[16] B. R. Cami, H. Hassanpour and H. Mashayekhi, "A content-based movie recommender system based on temporal user preferences," 2017 3rd Iranian Conference on Intelligent Systems and Signal Processing (ICSPIS), Shahrood, Iran, 2017, pp. 121-125, doi: 10.1109/ICSPIS.2017.8311601.

[17] E. J. Chia and M. K. Najafabadi, "Solving Cold Start Problem for Recommendation System Using Content-Based Filtering," 2022 International Conference on Computer Technologies (ICCTech), Melaka, Malaysia, 2022, pp. 38-42, doi: 10.1109/ICCTech55650.2022.00015.

[18] N. P. Sable, A. Yenkikar and P. Pandit, "Movie Recommendation System Using Cosine Similarity," 2024 IEEE 9th International Conference for Convergence in Technology (I2CT), Pune, India, 2024, pp. 1-5, doi: 10.1109/I2CT61223.2024.10543873.