IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.15536

CONTENT FILTER BASED MOVIE RECOMMENDATION SYSTEM USING AI AND ML

Asharf Khan[1], Alok Kumar Singh[2], Vishal Singh[3], Akrar Khan[4], Pankaj Kumar Gupta[5]

Student, Department of Computer science and Engineering,

Goel Institute of technology and Management, Lucknow, U.P, India[1-4]

Assistant Professor, Department of Computer Science And Engineering, Goel Institute of technology and Management, Lucknow, U.P, India[5]

Abstract: Recommendation system refers to the system which, dependent on the particular data set, provides users with recommendations for certain resources, such as books, movies, songs and so forth. Generally speaking, movie recommendation system usually suggests films which the User would love to watch due to particular attributes of the User, as well as previously accumulated data. Recommendation systems are very beneficial for companies which collect their data from a large number of customers and need to provide them with the most appropriate recommendations possible. Several aspects can be considered when building a movie recommendation system, such as the movie genre, the actors, its plot and even the director. Recommendation systems can offer movies based on a single aspect or several combined ones. The recommended system used in our paper is constructed based on the tags generated from the combinations of the film genre, actors and description which may be interesting for the User to watch. We decided to apply Content-based Filtering approach in our work.

Keywords: Movie Recommendation Systems, Content-Based Filtering, Movie recommendation, ai and machine learning project.

I.INTRODUCTION

With the ever-rising demand for machine automated solutions, machine learning has been seen to develop very rapidly today. In relation to the current 21st century era and increased usage of E-commerce via the internet, online buying and entertainment activities are at an all-time high.
Everything Online is the way of life for the future decades. Let us assume that you are shopping online in places like Amazon.com; there are millions of products to buy here, and the same is applicable to other e-commerce websites like Flipkart and more. There are about 10 million movies and series available at entertainment sites like Netflix and Hotstar. While you may be able to search everything you want to know from Amazon, what about others? If you are looking out for something better than what you got in your search result, it is like searching for a needle in a haystack full of trees. You get completely lost and confused in this forest. This is when Recommendation Systems come into the picture. Recommendation Systems can help you in such cases. In the same manner, recommendation systems can assist you while surfing Amazon, Netflix and more e-commerce websites. They would otherwise just be databases.

Recommendation Systems are commonly applied to facilitate the consumer in getting customized results depending upon his/her preferences. Recommendation Systems can also be considered a filtration technique that can be used to choose the optimal result out of a list of potential results through Machine Learning.

The classification of films can be done according to their genre, like thriller, animation, comedy, action, drama, etc. The other way could be classifying films according to metadata like the actor, director, year of production, etc.

Nowadays, most online streaming sites offer numerous similar tv programs and movies to users depending on their previous search history. While developing a movie recommendation system, its reliability and efficiency to suggest suitable options to users are the most important objectives. Normally, there are three types of recommendation systems: Collaborative/ User Filtering, Content-Based Filtering, and Hybrid Filtering.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.15536

A Recommendation System is essentially a sophisticated filtration framework powered by Artificial Intelligence (AI) and Machine Learning (ML). Unlike static search engines, these systems are proactive; they analyze historical data, behavioral patterns, and item metadata to predict and suggest the most 'optimal' content to a user. By leveraging ML algorithms, these systems transform raw data into actionable insights, allowing platforms to transition from being mere storage repositories to intelligent, personalized assistants. In the film industry, movies are categorized through diverse dimensions—ranging from intrinsic genres like thriller, animation, and drama to metadata such as director, cast, and production era. The ultimate objective of an AI-driven movie recommender is to maximize 'reliability' and 'relevance,' ensuring that the suggestions are both accurate and serendipitous.

The primary objective of this research is to design and evaluate a robust movie recommendation framework that addresses the persistent challenges of accuracy and user satisfaction in an increasingly crowded digital space. While traditional models have laid the foundation, this paper explores the integration of advanced computational techniques to refine the filtering process. By analyzing diverse datasets, the study aims to demonstrate how personalized algorithms can effectively reduce 'search fatigue' and enhance the discovery of niche content.

The remainder of this paper is structured as follows: **Section II** provides a comprehensive review of existing literature and related work in the field of recommendation systems. **Section III** details the proposed methodology, including data preprocessing and the architectural design of the recommendation model. **Section IV** presents the experimental results and a comparative analysis of the system's performance. Finally, **Section V** concludes the research with a discussion on potential future enhancements and the evolving role of AI in digital curation.

II. LITERATURE REVIEW: EVOLUTION OF MOVIE RECOMMENDATION SYSTEMS

"The landscape of recommendation systems has transitioned from basic heuristic-based approaches to sophisticated deep learning architectures. Early research primarily focused on Content-Based Filtering, which suggests items based on item descriptors, and Collaborative Filtering (CF), which leverages the collective wisdom of users to predict preferences. While Matrix Factorization techniques significantly improved prediction accuracy, they often struggled with the 'Cold-Start' problem and data sparsity. To overcome these limitations, recent literature highlights the integration of Hybrid Models and Deep Learning techniques, such as Neural Collaborative Filtering (NCF) and Graph Neural Networks (GNNs). These modern approaches allow for the extraction of complex, non-linear relationships between users and movies, providing more personalized and context-aware suggestions. Furthermore, current research trends are shifting towards improving the explainability and fairness of algorithms to ensure unbiased recommendations."

Detailed Literature Review: Technological Paradigms

1. The Era of Neighborhood-Based Models (Classical Approach)

Early research in the 1990s and early 2000s was dominated by User-based and Item-based Collaborative Filtering. These models relied on calculating similarity scores using metrics like Pearson Correlation and Cosine Similarity. While foundational, these systems faced the 'Search-Density' problem. Researchers like *Sarwar et al. (2001)* demonstrated that item-item similarity was often more stable and scalable than user-user similarity, leading to the architectures used by early E-commerce giants.

2. Matrix Factorization and Latent Factor Models

The 'Netflix Prize' competition marked a significant shift toward Latent Factor Models. Literature from this period (notably *Koren et al., 2009*) introduced Singular Value Decomposition (SVD). Instead of just looking at raw ratings, these models decomposed the user-item matrix into lower-dimensional vectors representing hidden features (e.g., a movie's "darkness" or "humor" levels). This significantly improved prediction accuracy but remained a 'Black Box'—users couldn't understand why a specific movie was suggested.

3. Knowledge-Based and Semantic Approaches

To solve the 'Cold Start' problem for new movies, researchers integrated Knowledge Graphs and Ontologies. By mapping movies into a structured network of actors, directors, and historical contexts, systems began to understand the 'meaning' behind a film. Studies showed that by using Semantic Analysis, a system could recommend a movie to a new user based on their interest in a specific historical event or a director’s unique style, even without prior rating data.

4. The Deep Learning Revolution (Neural RS)

The most significant leap in recent literature is the transition to Deep Neural Networks (DNN). Traditional models assumed linear relationships, but user behavior is often non-linear and chaotic. Neural Collaborative Filtering (NCF)

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.15536

replaced the inner product in matrix factorization with a neural architecture, allowing the system to learn much more complex patterns. Furthermore, the use of Autoencoders has been explored for 'Denoising' user data, effectively handling the 'Data Sparsity' problem where users have only rated a handful of movies.

5. Sequence Modeling and Temporal Dynamics

Modern research now treats movie watching as a 'sequence' rather than a static set of interests. Using Recurrent Neural Networks (RNNs) and LSTMs (Long Short-Term Memory), researchers have modeled the 'Temporal Dynamics' of user taste—recognizing that a user's preference on a Monday morning (perhaps news or documentaries) differs drastically from a Friday night (action or comedy). This 'Context-Aware' literature is currently the gold standard for real-time streaming platforms.

III. THE PROBLEM STATEMENT

"Despite the success of existing movie recommendation engines, several critical gaps remain. Most current systems suffer from 'Over-specialization,' where the user is repeatedly recommended similar items, leading to a lack of variety. Furthermore, the 'Shilling Attack' or profile injection—where fake ratings are used to manipulate a movie's popularity—remains a security concern. This research focuses on optimizing the trade-off between Exploitation (recommending what the user already likes) and Exploration (introducing new genres) to ensure a more dynamic user experience."

IV. COMPARISON OF RECOMMENDATION TECHNIQUES

Technique	Core Logic	Advantage	Disadvantage
Content-Based	Based on movie features (Genre, Actor)	No Cold-Start for movies	Limited novelty (Overspecialization)
Collaborative	Based on user-user similarity	Discovers new interests	Struggles with New Users (Cold-Start)
Matrix Factorization	Latent features (SVD)	Very accurate and scalable	Difficult to explain to the user
Deep Learning	Neural Networks (NCF/GNN)	Handles complex, non-linear data	Requires huge computational power

V. METHODOLOGY

1. **Requirement/Data Gathering**:

Data forms the most crucial and basic element of any machine learning project. Collecting data from different data sources is very important for any recommendation engine as well as its processing. More data means better results.

2. **Pre-processing**:

During the pre-processing step, Filtering & getting data ready for the project, we would make certain modifications like creating tags that could be used to define our data and assist in calculating its similarity with other data.

2.1 **System designing**:

In the process of designing this system, the system is designed such that it is user-friendly.
We design some UML diagrams and data flow diagrams to be able to comprehend the flow and modules of the system.

3. **Model implementation and testing**:

Our aim is to design a dataset framework that will work in harmony with the coding involved in this phase. Coding and design of models are the essential aspects of this project. Models ensure that the project works efficiently on a local scale.
These test cases are used to verify the validity of the output of the project module.

4. **Website designing**:

After we have created a working model, we will create the same into a website. This stage will involve designing an immersive UI.

5. **Deployment of System**:

After performing functional and non-functional testing on the product, deployment is carried out either virtually or by releasing to the local hosting such as Heroku.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.15536

VI. MODELLING AND ANALYSIS

Types of Recommendation System

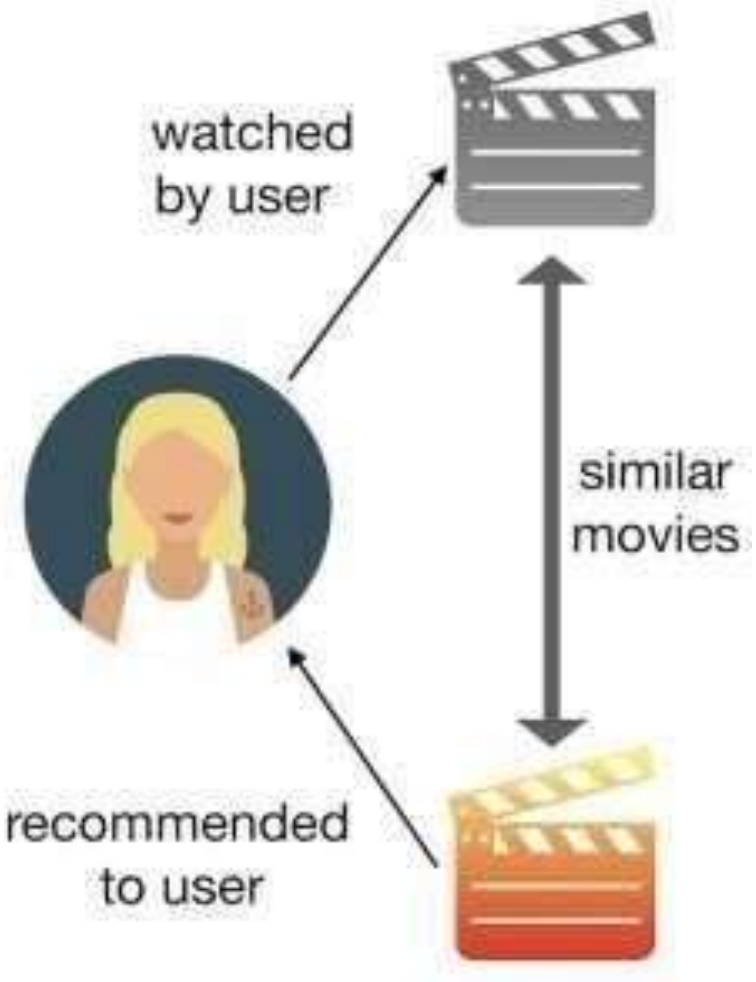
The recommendation system is usually classified on rating estimation

[1] Collaborative Filtering system
[2] Content-based system
[3] Hybrid system

Content-Based Recommending System:

As can be judged from the name, a content-based recommending system makes recommendations on the basis of similarities between content/products. In a content-based recommending system, the similarities between content/products are compared to determine similar contents.

Example: If you wish to watch an action movie, then the system will recommend to you similar action movies.



We have used the cosine similarity method to calculate a numeric value, that calculates the similarity among other movies. The cosine similarity function calculates by independent of magnitude and it is easy to calculate. Mathematically, it is defined as follows:

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.15536

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}},$$

To implement the recommendation process these are the following steps,: - ● Get the movie's index based on its title. ● Get a list of cosine similarity ratings for that movie compared to all other movies. Convert it to a tuple list, with the first member being the position and the second being the value. the score of similarity.
Get top 5 similar results and display.

Since our system has successfully been able to authenticate movies and recommend movies similar to the searched movies by finding similarities in the tag categories, our system is not robust enough to be considered a Perfect recommendation system. For instance, when someone searches "The Avengers," the system will give all movies associated with the Avengers,

whereas there are chances that people who watch the Avengers movies would prefer watching other Marvel movies as well.

VII. RESULTS

VII. FUTURE SCOPE

While the current research provides a solid framework for movie recommendations, the field is rapidly evolving. The following areas present significant opportunities for future investigation and system enhancement:

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.15536

Cross-Domain Recommendation Systems: Future work can explore 'Cross-Domain' mapping, where the system learns a user's preferences from other platforms—such as music (Spotify), books (Goodreads), or social media—to provide highly accurate movie suggestions even for new users (mitigating the Cold-Start problem).

Context-Aware and Spatio-Temporal Modeling: Beyond just "what" a user likes, future systems should focus on "when" and "where." Integrating spatio-temporal data (time of day, day of the week, or current location) can help the system distinguish between a user looking for a "Friday night party movie" versus a "Sunday morning documentary."
Hybridization with Generative AI and LLMs: The integration of Large Language Models (LLMs) like GPT-4 or Gemini can transform recommendation engines into Conversational Agents. Future systems will not just display a grid of posters but will engage in natural language dialogue, allowing users to provide complex queries like, *"Show me a psychological thriller with a plot twist similar to Inception but with a more hopeful ending."*

Emotional Intelligence and Bio-Feedback: With the rise of wearable technology, future research could incorporate biometric data (heart rate, skin conductance) or facial expression analysis via the user's camera to detect real-time emotional responses. This would allow the system to adapt its recommendations based on whether the user is stressed, happy, or bored.

Privacy-Preserving AI (Federated Learning): As data privacy regulations (like GDPR) become stricter, the future scope includes implementing Federated Learning. This allows the model to be trained across multiple decentralized devices holding local data samples, without ever exchanging them, thus ensuring high personalization without compromising user privacy.

Addressing Algorithmic Bias and Filter Bubbles: A critical area for future study is the development of "DiversityPromoting" algorithms. Current systems often trap users in an echo chamber of similar content. Future models will aim for Serendipity, intentionally introducing users to high-quality international or indie cinema to broaden their cultural horizons while maintaining relevance.

Explainable AI (XAI) for User Trust: Future systems will move away from 'Black Box' models toward Explainable AI. By providing transparent reasons for a recommendation (e.g., *"We recommended this because you enjoy non-linear storytelling and 90s noir aesthetics"*), systems can build deeper trust and higher retention rates with the user.

VIII. CONCLUSION

[1] The recommendation systems could be improved further to meet current as well as future needs in improving quality and providing better recommendations.
[2] Recommendation system will prove to be an ideal guide for you in the field of e-commerce if equipped with AI capabilities.
[3] There would be significant loss for such big brands like Amazon and Netflix if the customers do not purchase or view their products
[4] With the rising demand for machine automated solutions 'ML' has come to prominence in recent years, alongside AI and data science.
[5] The future use of recommender systems will include prediction of demands for certain products, linking sellers with buyers as well as becoming the basic requirement of the supply chain.

REFERENCES

[1]. Gayen DK. Ai-based movie recommendation system. InIndustry 4.0 Key Technological Advances and Design Principles in Engineering, Education, Business, and Social Applications 2024 Aug 21 (pp. 221-232). CRC Press.
[2]. Singh G, Saini AK, Singh VP, Abedin M, Alam H. Movie recommendation system using AI. InIntelligent Computing and Communication Techniques 2025 Jun 10 (pp. 450-456). CRC Press.
[3]. Furtado F, Singh A. Movie recommendation system using machine learning. International journal of research in industrial engineering. 2020 Mar 1;9(1):84-98.
[4]. Keshava MC, Reddy PN, Srinivasulu S, Naik BD. Machine learning model for movie recommendation system. International Journal of Engineering Research & Technology (IJERT). 2020 Apr;9(04):800-5.
[5]. Smitha N, Anusha D, Chaithanya C, Sindhu J, Tanuja R, Kumar HH. A review on movie recommendation system using machine learning. In2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV) 2021 Feb 4 (pp. 769-773). IEEE.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.15536

[6]. Surendran A, Yadav AK, Kumar A. Movie recommendation system using machine learning algorithms. International Research Journal of Engineering and Technology. 2020 Apr;7(4).

[7]. Singh RH, Maurya S, Tripathi T, Narula T, Srivastav G. Movie recommendation system using cosine similarity and KNN. International Journal of Engineering and Advanced Technology. 2020 Jun;9(5):556-9.

[8]. Rajarajeswari S, Naik S, Srikant S, Sai Prakash MK, Uday P. Movie recommendation system. InEmerging Research in Computing, Information, Communication and Applications: ERCICA 2018, Volume 1 2019 May 3 (pp. 329-340). Singapore: Springer Singapore.

[9]. Benjamin M, Kanjirappally K, Joseph A. Movie recommendation system using content-based filtering and cosine similarity. InProceedings of the Proceedings of the National Conference on Emerging Computer Applications (NCECA) 2022 (Vol. 4, p. 405).

[10]. Abhinav N, Sujatha K. Content-based movie recommendation system using cosine similarity measure. InAIP Conference Proceedings 2023 Dec 15 (Vol. 2901, No. 1, p. 060035). AIP Publishing LLC.

