

Original Article

Novel Movie Recommendation System using K-Means and Mean Shift Clustering

Nisha Bhalse¹, Ramesh Thakur², Archana Thakur³

¹Department of Computer Engineering, Institute of Engineering & Technology, DAVV, Indore, Madhya Pradesh, India.

¹Department of Computer Science & Engineering, IPS Academy, Institute of Engineering & Science, Indore, Madhya Pradesh, India.

²International Institute of Professional Studies, DAVV, Indore, Madhya Pradesh, India.

³School of Computer Science & IT, DAVV, Indore, India.

¹Corresponding Author : nishabhalse@gmail.com

Received: 19 June 2025

Revised: 22 January 2026

Accepted: 06 February 2026

Published: 28 March 2026

Abstract - Online shopping has risen during the COVID-19 pandemic. Nowadays, recommendation systems are important for providing personalized suggestions. Recommendation Systems (RS) face the challenges of efficiently and relevantly providing suggestions from the large volume of information. Many fields use recommendation systems, such as movies, e-commerce, and news. A Collaborative Filtering (CF) algorithm is an effective RS technique that recommends items that are similar to the active user's items. CF is caused by data sparsity, cold start, and scalability problems. The proposed Novel Hybrid K-means and Mean-Shift Clustering (NHMM) algorithm for recommending movies based on user preferences. Based on users' past preferences, the input is collected from the MovieLens 1M dataset. The NHMM model is preprocessed and trained on the MovieLens dataset, and it recommends the top-k movies to the user based on the user's interest preferences. The proposed NHMM model performance was evaluated with different recommendation techniques: k-means clustering, collaborative filtering, and matrix factorization. The experiment shows that the comprehensive results of the proposed NHMM model achieve the highest accuracy of 92.4%, precision 93.3%, recall 90.8%, and F1-score 91.9%. The proposed NHMM model recommends accurate, relevant top-k movies to users. The proposed NHMM model achieves the lowest RMSE (0.798) and MAE (0.633). The results show that the proposed NHMM model recommends accurate and robust, as well as diverse and serendipitous, top-k movies to users compared to other traditional models.

Keywords - Recommendation System, Collaborative Filtering, K-means Clustering, Mean Shift Clustering.

1. Introduction

The rapid development of digital platforms has revolutionized the way users interact with online available movies. Due to the large volume of online movies available, users are overwhelmed. It is very difficult for users to find new, relevant movies based on their interests. Nowadays, it is a critical area in the field of machine learning to develop a sophisticated movie recommendation system [1]. The recommendation system filters relevant movies from a large volume of data and recommends the most efficient and accurate product based on their previous experience and choices. Traditional recommendation systems primarily employ Collaborative Filtering (CF), Content-Based Filtering (CBF), and Hybrid Approaches. CF recommends items to active users based on the past behavior of other users with similar tastes [2,3]. CF faces persistent problems of cold start, scalability, and noisy rating patterns [4]. Due to this issue, the system is not recommending an accurate recommendation to the new users because of a lack of sufficient data. To address

such limitations, researchers have progressively assessed clustering-based methods that group similar users and products prior to generating recommendations[5]. K-means clustering is used for computational efficiency and the ability to group users or movies into similar clusters. The number of clusters is predefined for initial centroid selection [6]. Although there are various clustering techniques, just like DBSCAN, agglomerative, K-means clustering, and mean shift clustering have received significant attention. The K-means clustering method assumes spherical cluster forms, and it does not always reflect the inherent structure of data. A non-parametric mean shift clustering technique does not need to redefine the number of clusters [7]. In which cluster can be any shape, and automatically figure out the number of clusters, and they are identified by regions of high data point density. It is computationally more demanding and less scalable than K-means clustering. It is effective in capturing the local density structure, efficient in handling large datasets, and able to fine-tune cluster boundaries based on local density. A



hybrid approach is developed to overcome the shortcomings of cold start and scalability, noisy rating patterns, and to enhance clustering quality. Hybrid approaches combine several clustering-based techniques to capitalize on their individual strengths [8]. In this paper, a Novel Hybrid Movie recommendation system (NHMM) model incorporates the k-means clustering method with the mean shift clustering method. The proposed NHMM model overcomes the issue of collaborative filtering's sparsity and cold start problem and handles non-linear behavior patterns of users [9].

The incorporation of clustering algorithms with collaborative filtering can improve the performance and system capacity for recognizing and exploiting hidden patterns in the data. The clustering method effectively handles the large volume of data and performs better in finding similar items and users, and provides efficient and diverse recommendations [10]. The mean shift clustering method minimizes parameter tuning and easily adapts to changing user preferences, which is ideal for real-world applications. In this paper, the proposed NHMM model improves the granularity and relevance of recommendations by leveraging. K-means clustering handles large datasets, and Mean Shift's clustering is fine-tuned cluster boundaries based on local density.

Our key contributions include the following:

- The NHMM model integrates the K-means clustering and the mean shift clustering technique. This hybrid design leverages the computational efficiency of K-Means clustering to initially make a group of similar interests and mean shift clustering to refine clusters without predefining the number of clusters and handling heterogeneous user preferences.
- The NHMM model captures the linear and nonlinear user-movie rating pattern of behaviour. It generates accurate and coherent user groups, enhancing similar computations and relevant movie recommendations to users.
- The NHMM model, in which K-Means clustering reduces scalability and computational load. It balances the computational efficiency and accuracy for large-scale recommendations.

The layout of a research paper is as follows: Section 2 discusses a review of literature on recommendation systems. Section 3 presents the existing algorithms and works of recommendation systems: the proposed Novel Hybrid Movie Recommendation System (NHMM) and pseudo-code. Section 4 presents the experimental results conducted on the MovieLens dataset. Section 5 discusses the conclusion and highlights future work.

2. Literature Review

Many researchers have significantly contributed to the field of movie recommendation systems. Recommendation

Systems (RS) assist in identifying items that are likely to interest clients. Nowadays, they are mainly used in web applications, where product and customer data are generated daily. RS uses this data to identify similarities between products or consumers, thereby inferring client preferences [11]. The RS contains the user's set, which is $U = \{u_1, u_2, u_3, u_4, \dots, u_m\}$ and item sets are $I = \{i_1, i_2, i_3, i_4, \dots, i_n\}$ and a collection of evaluations. Users provided the ratings for possible evaluations among them. User i rated item j , which is called the rating; it is represented as $R(i,j)$. These ratings form a matrix of $m \times n$ size. These matrices extrapolate two types of data [12].

- Prediction: The main aim is to identify the unknown rating $R(i,j)$ of
- The rating matrix R , which indicates the user's rating of item j .
- Recommendation: The aim is to generate a list of the most relevant products for a user that have not yet been purchased.

Recommender systems can achieve their goals through various methods, which are generally categorized into three main [4]:

Content-Based Filtering (CB): The CB recommends items to users based on the features of items and their past preferences.

Collaborative Filtering (CF): This approach recommends items based on the user's interests by analyzing other users' behavior. CF is further divided into:

Memory-Based CF: Use the historical ratings or user interactions to generate new recommendations.

Model-Based CF: A model is developed from the existing data and generates the new recommendations.

The hybrid approach integrates CB and CF methods to improve recommendation accuracy. Since collaborative filtering is a widely used and effective approach. The Singular Value Decomposition (SVD) is based on rating predictions. In SVD, two matrices are used, where one is V , an $m \times f$ matrix, and D is an $f \times n$ matrix, where f is a user-chosen dimensionality factor. This method verifies the equation of the utility matrix R . The user i is predicting the rating for item j , which is calculated by multiplying the i th row of V with the j th column. The matrix V and matrix D are calculated by using normally distributed random numbers. The error for each known value is calculated by using $E(i, j) = R(i, j) - D_j^T V_i$, where V is the row of user i and D is the column of item j . These errors use the stochastic gradient descent technique, and according to these techniques, the V matrix and D matrix values are adjusted. The V and D matrices are used for calculating the clusters. These clusters are used for the

evaluation of the prediction of collaboration. According to genres, an initial prediction is generated. The overall prediction is generated after the combination of all the local predictions.

Sarwar et al. presented research on dimensionality reduction techniques in recommendation systems. The SVD enhanced the efficiency and scalability by compressing sparse rating matrices while retaining significant latent features. Their results showed neighborhood formation and CF prediction accuracy. The main challenge is extensive computing and pre-processing [13].

Wang et al. developed a multi-clustering collaborative recommendation system that incorporates K-means, hierarchical clustering techniques for capturing diverse user-item patterns. Compared to single clustering techniques, these multi-clustering strategies increased computing complexity but generated more stable and refined clusters, improving accuracy, scalability, and noise reduction [5]. Anwar et al. compared the CF and kNN-based recommendation systems. The kNN variants-based RS employing Pearson correlation and cosine using similarity measures, produce more accurate predictions and perform better than traditional collaborative filtering, particularly when dealing with new users or items [14].

Teng Li et al. proposed an ensemble agglomerative hierarchical clustering algorithm that improves the clustering quality. The cluster clustering technique, in which initial clusters are further reclustered relying on a novel similarity measurement to capture deep structural associations between data points. They achieve higher accuracy and better cluster compactness [15]. Selamat and Moghaddam demonstrated an enhanced collaborative filtering that is applied to smoothing density-based user clustering. Inaccurate similarity computation and data sparsity are handled. Within each cluster, it produces rating predictions that are more dependable and efficient [11].

Hu et al. introduced the LK-means algorithm that addresses the challenges of big data clustering. It is enhancing the search results, distributing cluster centroids more evenly, and enhancing global search capability and big data processing efficiency, particularly in handling unevenly distributed centroids, outperforming the traditional k-means algorithm [16].

Cariou et al. introduced an improved mean-shift clustering technique that combines adaptive stopping criteria with multi-scale bandwidth selection with the mean shift algorithm. This approach maintains clustering accuracy and achieves faster convergence. However, the primary issue is performance degradation in highly high-dimensional data [17].

Frémal et al. developed a recommendation system based on item clustering that enhances collaborative filtering

effectiveness by utilizing genre information and weighting approaches. In which item belongs to multiple genre-based clusters. Recommendation accuracy increases by combining cluster-specific predictions using weighting techniques [18].

Beel, S. Langer et al. demonstrated a movie recommendation that integrates collaborative filtering with weighting strategies. Every cluster generates its local rating prediction and then combines weighting strategies into a unified evaluation [19].

The investigated literature emphasizes that clustering approaches are becoming more essential to enhance recommendation systems. It deals with issues of data sparsity, scalability, and handling heterogeneous user preferences. The proposed NHMM model is a hybrid incorporating K-means and mean shift clustering to exploit both partition-based and density-based clustering. The proposed NHMM model aims to generate movie recommendations that are more reliable, scalable, and contextually relevant than stand-alone methods. The proposed NHMM model promises improved performance in resolving issues identified in the literature review, such as handling heterogeneous user preferences and enhancing clustering granularity in high-dimensional ratings.

3. Methods

3.1. Background of Movie Recommendation Systems

Recommender Systems (RS) are indeed effective knowledge management systems that play a crucial role in helping users filter through large amounts of data and avoid information overload [20]. The main purpose of RS is to help users discover useful and relevant information from a potentially overwhelming sea of data by leveraging various techniques such as CF, CBF, and hybrid approaches. CF is an effective and widely used technique for movie recommender systems.

The nearest-neighbor method is employed in CF-based movie recommender systems. This technique finds the nearest neighbors similar according to the active user's choice. The two most popular techniques adopted by CF are memory-based CF and model-based CF. These techniques are used in recommender systems to recommend personalized suggestions based on users' preferences and similarities. The memory-based methods handle the problem of data sparsity for recommendation systems.

3.2. K-Means Clustering Technique

The K-means clustering algorithm is used for partitioning the dataset into k distinct and non-overlapping clusters [21].

- Initialization: Choose the first k cluster centroids at random or using a heuristic.
- Assignment: Form k clusters nearest to the centroid. Assign x_i to each data point in the dataset, calculate the distance between the centroid and x_i , and find the cluster nearest to the centroid.

- Update: Recalculate the cluster centroids by averaging all points assigned to each cluster.
- Iteration: Repeat steps 2 until the centroids are no longer changed. The number of iterations is at its maximum.

3.3. Mean Shift Clustering Technique

The mean shift clustering method is a non-parametric approach in which the number of clusters is not predefined [24]. It works by relocating data points to the mode (peak) of data density. Here is an overview of the mean shift clustering process:

Initialization: Choose a kernel function (often Gaussian) and a bandwidth parameter.

- Choose a kernel function (x) (commonly a Gaussian kernel).
- Select a bandwidth parameter h, which determines the radius of the window used for shifting.

Shift Points: Compute for each point and weight the mean for each neighbor within the bandwidth, and shift the point to the mean. Mean shift vector m(x) is computed and shifted to modes.

3.4 Methodology of the Proposed NHMM Movie Recommendation System

The proposed personalized movie recommendation system integrates the K-means clustering and mean shift clustering algorithms and is called the Novel Hybrid K-Means Clustering and Mean Shift Clustering Movie Recommendation System (NHMM).

K-means clustering makes a group of users or movies into clusters based on their similar features. Mean shift helps to refine these groups by spotting patterns or sub-clusters that K-means missed. The NHMM model uses the MovieLens dataset. The methodology of the NHMM model is depicted in Figure 1.

Input-user-item rating matrix R

Output-Top-N movie recommendations to the user

Step 1-Load the user-item rating matrix

$$R=[r_{u,i}]_{m \times n}$$

Step 2- normalized the rating matrix

Step 3: SVD applies for dimensional reduction

$$R'=U_k \Sigma_k V_k^T$$

Step 4- construct feature vectors

$$x_u=U_k(u, :)$$

Step 5- Initialize K centroids

$$C=\{c_1, \dots, c_k\}$$

Step 6: Assign the user to the nearest centroid

$$\text{cluster}(u) = \arg \min_j \|x_u - c_j\|$$

Step 7: centroids update

$$c_j = \frac{1}{|S_j|} \sum_{u \in S_j} x_u$$

Step 8: Steps 6-7 repeat until

$$\|c_j^{(t)} - c_j^{(t-1)}\| < \epsilon$$

Step 9: Apply mean shift on K-means centroids with a Gaussian kernel

$$m(x) = \frac{\sum_i x_i K(\frac{x-c_i}{h})}{\sum_i K(\frac{x-c_i}{h})}$$

Step 10: Shift the centroid until convergence

$$x^{(t+1)}=m(x^{(t)})$$

Step 11- converged points merge to form refined clusters and map the user's final cluster of centroids

Step 12: Calculate user similarity within the refined cluster.

$$\text{sim}(u, v) = \frac{\sum_i r'_{u,i} r'_{v,i}}{\sqrt{\sum_i (r'_{u,i})^2} \sqrt{\sum_i (r'_{v,i})^2}}$$

Step 13- select top-k neighbors

Step 14: rating prediction for users

$$\hat{r}_{u,i} = \bar{r}_u + \frac{\sum_{v \in N_k(u)} \text{sim}(u,v) (r_{v,i} - \bar{r}_v)}{\sum_{v \in N_k(u)} |\text{sim}(u,v)|}$$

Step 15- items ranked by prediction score

$$\text{Rank}_u = \text{argsort}(\hat{r}_{u,i}) \text{ Rank}_u$$

Step 16: Top-N recommendation list generated

Step 17: evaluate performance for metrics RMSE, Accuracy, etc.

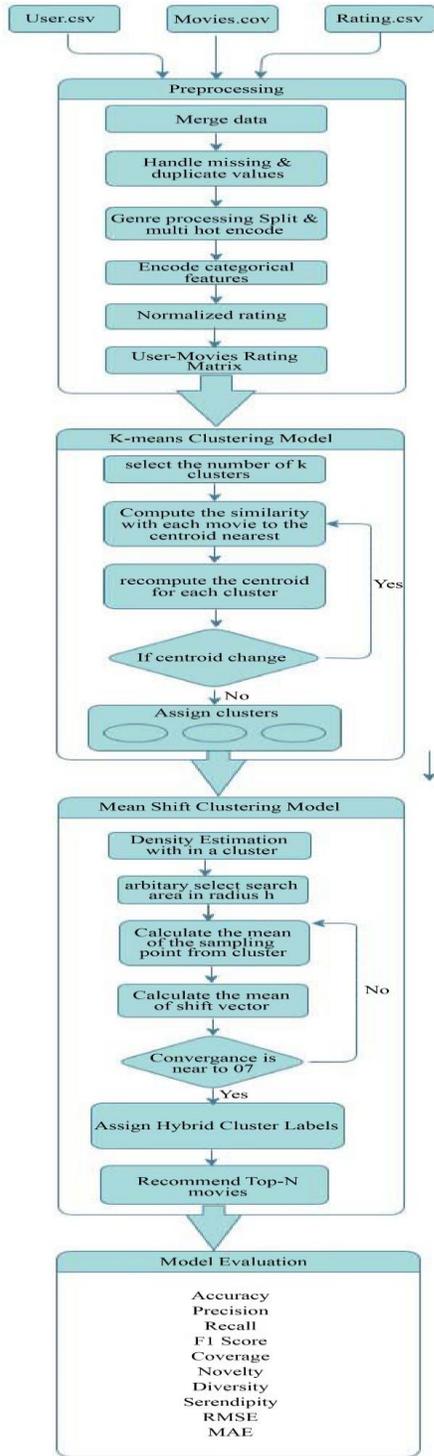


Fig. 1 Methodology of the NHMM movie recommendation system

4. Results and Discussion

4.1. Dataset

The proposed NHMM model utilizes the 1M MovieLens dataset, which includes 1 million movies, users, and ratings metadata. It consists of three main files: ratings.dat, which

includes the ratings provided by users for movies; users.dat, which includes demographic information about the users; and movies.dat, which contains information about the movies. Table 1 depicts the description of the MovieLens dataset.

Table 1. MovieLens dataset

Dataset	Parameters	Value
MovieLens 1 M	Users	6040
	Movies	3952
	Ratings	1000209

4.2. Implementation

The proposed NHMM model is implemented in MATLAB R2014a. The system is equipped with 8 GB of RAM and an Intel® Core™ i3-4005U CPU @ 1.70 GHz on WINDOWS (10 Pro) OS. The evaluation of the proposed NHMM model and traditional methods of recommendation systems is implemented in a specific computing environment with properties. The experiments are performed on the real-world datasets of 1M MovieLens [23]. The proposed NHMM movie recommendation system uses hyperparameters, described in Table 2.

Table 2. Parameters

Parameters	Value
Number of Clusters	10-100
Max Iterations	500
Distance Metric	Euclidean distance
Convergence Threshold	1e-4 to 1e-6
Bandwidth	0.2-2.0
Kernel Type	Gaussian
Cluster Merge Threshold	0.1-1.0

The proposed NHMM incorporates K-means clustering with the mean shift clustering method. K-means clustering is used for user segmentation, and mean shift clustering for local movie clustering. The model preprocesses data, creates feature vectors, clusters users and movies, maps clusters for recommendations, and periodically updates based on feedback. The NHMM model is demonstrated in the performance analysis in terms of showing the ability to cluster users and recommend movies effectively. The proposed model provides personalized and relevant movie recommendations to active users.

4.3 Results Analysis

The evaluation of performance and comparison of the proposed NHMM model with the K-means clustering method, matrix factorization, and collaborative filtering [14]. The proposed NHMM model is more effective than traditional methods. Preprocessed 2000 samples of the 1 M MovieLens dataset are used for the experiment. Results are evaluated through the metrics of accuracy, precision, recall, F1 score, coverage, diversity, and serendipity. The comparison analysis of K-means clustering, collaborative filtering, and matrix factorization with the proposed model is depicted in Table 3.

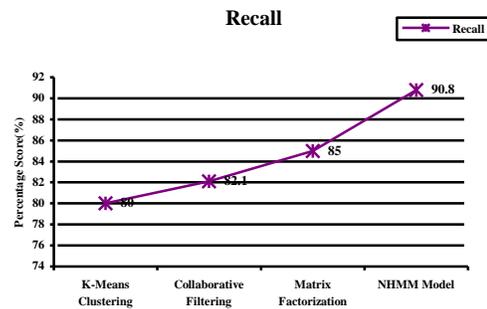
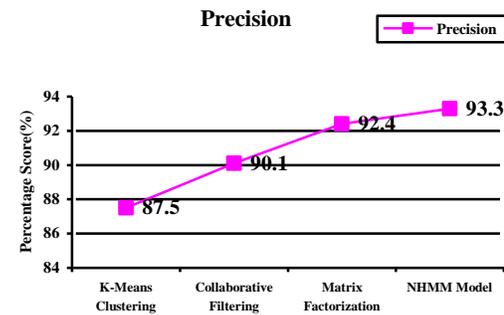
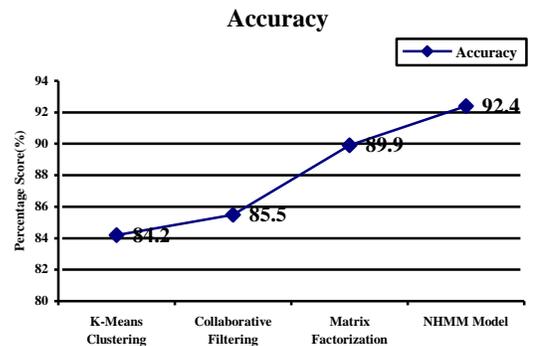
Table 3. Performance analysis of different techniques

Metric Techniques	Accuracy	Precision	Recall	F1 Score	Coverage	Novelty	Diversity	Serendipity
K-Means Clustering (%)	84.2	87.5	80	85.7	76.8	79.3	80.1	78.4
Collaborative Filtering (CF) (%)	85.5	90.1	82.1	88	78.9	80.5	82.3	80.7
Matrix Factorization (MF) (%)	89.9	92.4	85	90.1	80.7	85.0	84.6	88.0
NHMM (Proposed Model) (%)	92.4	93.3	90.8	91.9	84.2	86.9	88.9	89.8

The proposed NHMM model achieves the highest accuracy at 92.4%, surpassing CF 85.5 %, MF 89.90%, and K-means clustering 84.2%. The NHMM model is highly precise as compared to other techniques, and it is shown to be highly relevant according to the user preferences. The proposed model demonstrates outstanding performance in both precision, 93.3% and recall, 90.8% outperforming all other models. The NHMM model is highly precise as compared to other techniques, and it is shown to be highly relevant to the user’s interests. The NHMM model also achieves the highest F1 score of 91.9%; it is balanced in precision and recall. It achieved an F1 score of 91.9%, which means balanced precision and recall. The NHMM offers highly relevant recommendations to the users without sacrificing the number of correct recommendations.

Coverage refers to the proportion of items in the dataset that the system can recommend. The NHMM model achieves 84.2%, which means that it explores a wider range of movies and provides the users with wider recommendations. Traditional methods like K-means clustering, CF, and MF, with coverage scores of 76.8%, 78.9 % and 80.7% respectively, tend to focus on popular items, leaving lesser-known movies unexplored. Novelty measures how often the system suggests new and less popular movies. NHMM achieves a strong novelty score of 86.9% outperforming K-means clustering 79.3%, CF 80.5% and MF 85%. This is crucial in a recommendation system because users often seek fresh and undiscovered content. The proposed model in diversity achieved 88.9%, K-means Clustering 80.1%, CF 82.3%, and MF 84.6%. Diversity is exposed according to genres rather than similar movies. The high diversity score of NHMM reflects its ability to recommend movies across different categories, keeping the recommendation list varied and exciting for users. Traditional models tend to focus more on user history and popular content, which limits the variety of recommendations. Serendipity is recommending unexpected but relevant movies to active users. The NHMM model scored 89.8%, showing that it not only suggests popular or frequently watched movies but also surprises users with hidden gems that match their preferences. Serendipity is a key factor in user satisfaction, as it introduces new movies to users that might not be considered. It has solved the cold start problem of movie recommendation systems. Additionally, Figure 2 depicts the performance of the proposed

NHMM model and traditional techniques. The proposed NHMM model achieves better performance in terms of accuracy, precision, recall, F1 score, coverage, diversity, and serendipity.



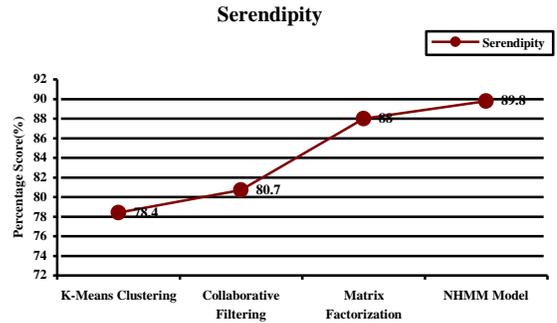
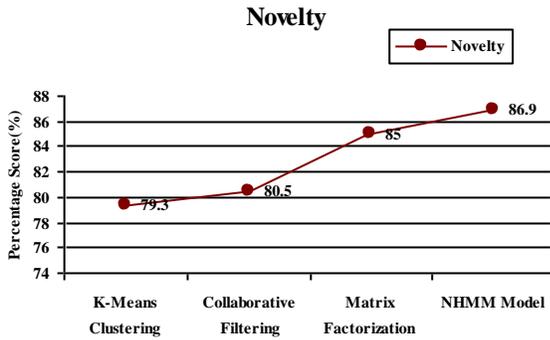
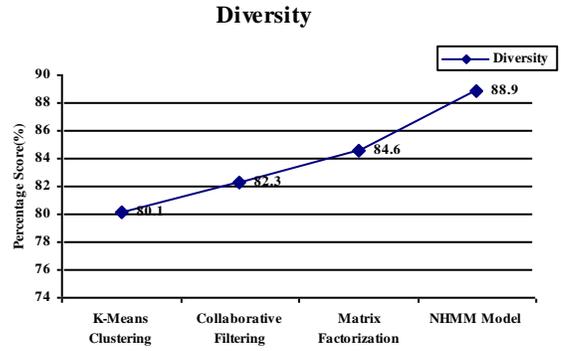
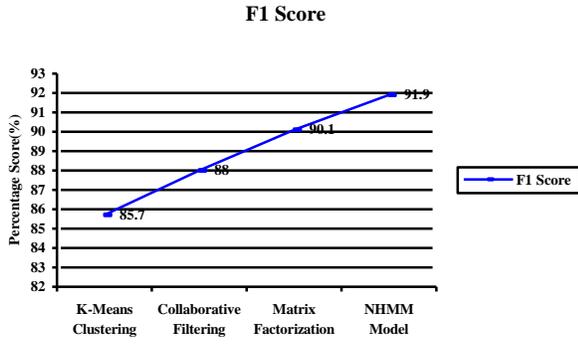


Fig. 2 Comparison result of different techniques on the movie recommendation system

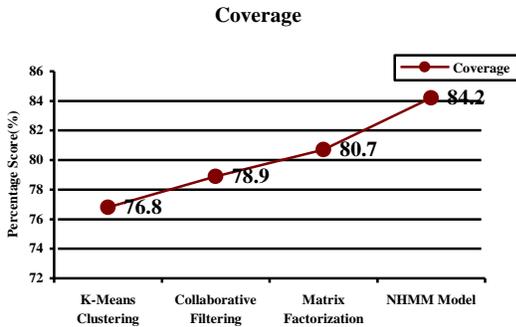


Table 4 Represents the comparison of the proposed NHMM model and different models of the movie recommendation system.

Table 4. Comparison of different techniques

Metric	RMSE	MAE
K-Means Clustering	0.889	0.701
Collaborative Filtering (CF)	0.87	0.696
Matrix Factorization (MF)	0.85	0.68
NHMM (Proposed Model)	0.798	0.633

Figure 3 depicts that the proposed NHMM model has less error of RMSE and MAE compared to K-means clustering, CF, and MF.

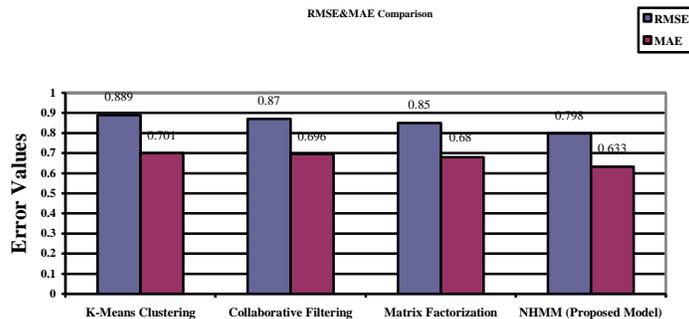


Fig. 3 RMSE & MAE comparison of different techniques on movie recommendation

Figure 4 Depicts the Comparison Analysis of Different Models on Different Parameters.

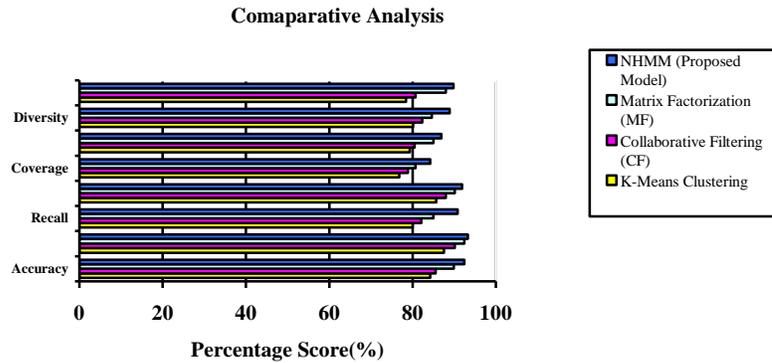


Fig. 4 Comparative analysis of different techniques of movie recommendation

The proposed NHMM model performance evaluation is superior in effectiveness for NDCG and Hit Ratio for top N = 5 and 10 movies. Across all iterations, 10, 50, 100, the NHMM model consistently outperforms as compared to K-means clustering, Collaborative Filtering (CF), and Matrix Factorization (MF). The K-means clustering model improves with more iterations, whereas CF, MF, and NHMM improve

moderately but plateau early. The NHMM model is most robust and accurate in generating top-N movies to users according to the user’s interest. Figures 5 and 6 show the comparison of performance evaluation of NDCG and hit ratio of 1M MovieLens for 5,10,50,100 iterations for top N=5, 10 movie recommendations.

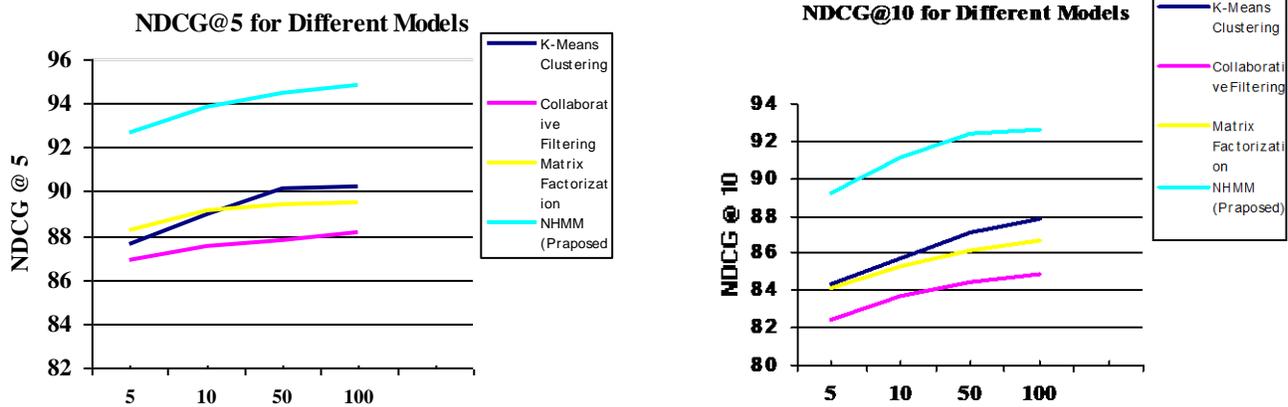


Fig. 5 Performance evaluation of NDCG of 1M for 5 to 100 iterations for Top N=5,10 movie recommendations

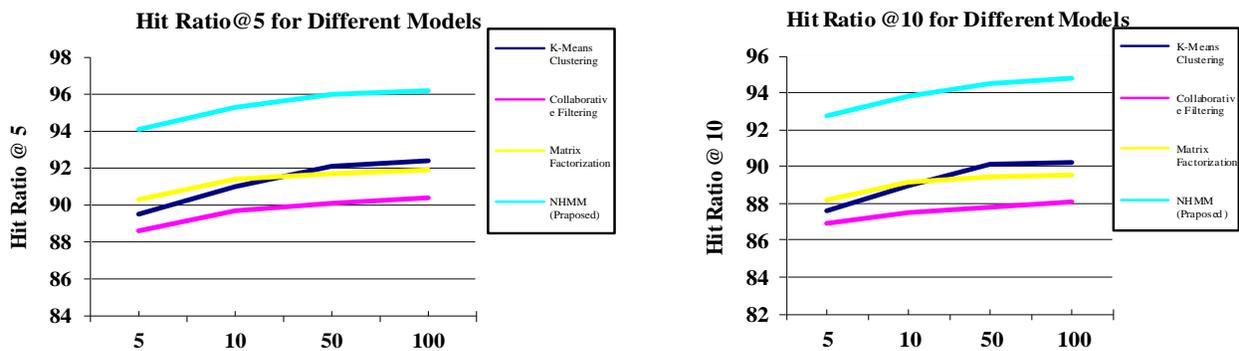


Fig. 6 Performance evaluation of the hit ratio of 1M for 5 to 100 iterations for Top N=5,10 movie recommendations

5. Conclusion

The proposed NHMM movie recommendation system incorporates the strengths of the K-means clustering technique and the means shift clustering technique into a cohesive framework that balances efficiency and adaptability. The K-means clustering provides clusters of users or movies in a predefined manner and minimizes computational complexity. Mean shift clustering refines clusters by discovering dense, non-linear preference patterns without the need for a predefined number of clusters.

By limiting collaborative filtering operations to linguistically consistent and locally dense clusters, the system minimizes noise, better controls multi-genre and overlapping user preferences, and improves personalization. The proposed NHMM movie recommendation model performance is better and more effective than K-means clustering, collaborative filtering, and matrix factorization methods, across several key metrics. The NHMM model achieved 92.4% accuracy, 93.3% precision, 90.8% recall, and 91.9% F1 score. It is showing the balance between relevancy and completeness of movie recommendations. Additionally, it demonstrated superior

coverage 84.2%, novelty 86.9%, diversity 88.9%, and serendipity 89.8%, compared to the competing algorithms, which reflects the ability to deliver a broader and more engaging range of movie recommendations. The proposed NHMM model has higher prediction accuracy, with the lowest error matrix, which is the lowest RMSE (0.798) and MAE (0.633). According to these findings, the proposed NHMM model improves the efficiency of movie recommendation systems by delivering more accurate, diverse, novel, and serendipitous recommendations. This model addresses the issue of the cold start problem and recommends better and more efficient recommendations to the active user. The proposed hybrid NHMM clustering approach enhances the recommendation quality and similarity computation. It is recommended to personalize the top-k movies to the active user. It is an efficient and relevant movie recommendations system. Future research could focus on enhancing the NHMM model in other areas like tourism, education, e-commerce, music streaming, etc. Any other optimization algorithm is applied with a hybrid model of a movie recommendation system for reducing the computational complexity and improving the execution time of the model.

References

- [1] Amin Beheshti et al., "Towards Cognitive Recommender Systems," *Algorithms*, vol. 13, no. 8, pp. 1-27, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [2] G. Suganeshwari, and S.P. Syed Ibrahim, "A Survey on Collaborative Filtering based Recommendation System," *Proceedings of the 3rd International Symposium on Big Data and Cloud Computing Challenges (ISBCC- 16')*, pp. 503-518, 2016. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [3] Patrick M. LeBlanc et al., "Recommender Systems: A Review," *Journal of the American Statistical Association*, vol. 119, no. 545, pp. 773-785, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [4] Gediminas Adomavicius, and Alexander Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734-749, 2005. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [5] Lili Wang et al., "Implementation of a Collaborative Recommendation System based on Multi-Clustering," *Mathematics*, vol. 11, no. 6, pp. 1-21, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [6] Abiodun M. Ikotun et al., "K-Means Clustering Algorithms: A Comprehensive Review, Variants Analysis, and Advances in the Era of Big Data," *Information Sciences*, vol. 622, pp. 178-210, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [7] Mahnoor Chaudhry et al., "A Systematic Literature Review on Identifying Patterns using Unsupervised Clustering Algorithms: A Data Mining Perspective," *Symmetry*, vol. 15, no. 9, pp. 1-44, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [8] Robin Burke, "Hybrid Recommender Systems: Survey and Experiments," *User Modeling and User-Adapted Interaction*, vol. 12, no. 4, pp. 331-370, 2002. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [9] Fethi Fkih, Delel Rhouma, and Mohamed Nazih Omri, "DemogCF Model of Personalized Recommendations based on Demographic Characteristics for Overcoming Data Sparsity and Cold Start Problems," *International Journal of Information Technology*, vol. 17, no. 1, pp. 169-177, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [10] Rui Xu, and Donald C. Wunsch, "Survey of Clustering Algorithms," *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645-678, 2005. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [11] Ali Selamat, and Siavash Ghodsi Moghaddam, "Improved Collaborative Filtering on Recommender based Systems using Smoothing Density-based User Clustering," *International Journal of Advancements in Computing Technology*, vol. 4, no. 13, pp. 352-359, 2012. [[Google Scholar](#)] [[Publisher Link](#)]
- [12] Sambandam Jayalakshmi et al., "Movie Recommender Systems: Concepts, Methods, Challenges, and Future Directions," *Sensors*, vol. 22, no. 13, pp. 1-22, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]

- [13] Badrul Sarwar et al., “Application of Dimensionality Reduction in Recommender System-A Case Study,” Computer Science & Engineering (CS&E) Technical Reports, University Digital Conservancy, 2000. [[Google Scholar](#)] [[Publisher Link](#)]
- [14] Taushif Anwar et al., “Collaborative Filtering and kNN based Recommendation to Overcome Cold Start and Sparsity Issues: A Comparative Analysis,” *Multimedia Tools and Applications*, vol. 81, no. 25, pp. 35693-35711, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [15] Teng Li et al., “An Ensemble Agglomerative Hierarchical Clustering Algorithm based on Clusters Clustering Technique and the Novel Similarity Measurement,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, pp. 3828-3842, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [16] Haize Hu et al., “An Effective and Adaptable K-means Algorithm for Big Data Cluster Analysis,” *Pattern Recognition*, vol. 139, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [17] Claude Cariou et al., “A Novel Mean-Shift Algorithm for Data Clustering,” *IEEE Access*, vol. 10, pp. 14575-14585, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [18] Sébastien Frémal, and Fabian Lecron, “Weighting Strategies for a Recommender System using Item Clustering based on Genres,” *Expert Systems with Applications*, vol. 77, pp. 105-113, 2017. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [19] Joeran Beel et al., “Research Paper Recommender System Evaluation: A Quantitative Literature Survey,” *Proceedings of the International Workshop on Reproducibility and Replication in Recommender Systems Evaluation*, pp. 15-22, 2013. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [20] Yashar Deldjoo et al., “Fairness in Recommender Systems: Research Landscape and Future Directions,” *User Modeling and User-Adapted Interaction*, vol. 34, no. 1, pp. 59-108, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [21] Mohiuddin Ahmed, Raihan Seraj, and Syed Mohammed Shamsul Islam, “The k-means Algorithm: A Comprehensive Survey and Performance Evaluation,” *Electronics*, vol. 9, no. 8, pp. 1-12, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [22] Loai AbdAllah, and Ilan Shimshoni, “Mean Shift Clustering Algorithm for Data with Missing Values,” *Data Warehousing and Knowledge Discovery*, pp. 426-438, 2014. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [23] MovieLens, [Online]. Available: <https://grouplens.org/datasets/movielens/>