DOI: https://doi.org/10.52756/ijerr.2024.v42.015 *Int. J. Exp. Res. Rev.*, Vol. 42: 169-182 (2024)

Original Article | Peer Reviewed | Open Access

International Journal of Experimental Research and Review (IJERR)

ISSN: 2455-4855 (Online)
www.iaph.in

Reducing Cluster Overlap in Movie Recommendations with IKSOM and Silhouette Clustering

Saurabh Sharma[1*], Ghanshyam Prasad Dubey[1] and Harish Kumar Shakya[2]

[1]Department of Computer Science and Engineering, Amity University, Gwalior, Madhya Pradesh, India;
[2]Department of Artificial Intelligence & Machine Learning, Manipal University, Jaipur, Rajasthan, India

E-mail/Orcid Id:

SS, saurabhgyangit@gmail.com, https://orcid.org/0000-0002-0090-6043;
GPD, gpdubey@gwa.amity.edu, https://orcid.org/0000-0003-0868-4219;
HKS, harish.shakya@jaipur.manipal.edu, https://orcid.org/0000-0002-5401-3507

Article History:

Received: 02nd May, 2024
Accepted: 04th Aug., 2024
Published: 30th Aug., 2024

Keywords:
Cluster Overlap Reduction, Hybrid Clustering Models, IKSOM (Improved Kohonen Self-Organizing Maps), Movie Recommendation Systems, Recommendation Accuracy, Silhouette Clustering

How to cite this Article:
Saurabh Sharma, Ghanshyam Prasad Dubey and Harish Kumar Shakya (2024). Reducing Cluster Overlap in Movie Recommendations with IKSOM and Silhouette Clustering. *International Journal of Experimental Research and Review*, *42*, 169-182.
DOI: https://doi.org/10.52756/ijerr.2024.v42.015

Abstract: The exponential growth of online material requires the implementation of effective and precise recommendation systems in order to optimize the user experience. Nevertheless, conventional approaches frequently encounter problems such as cluster overlap, which reduces the accuracy of suggestions. This paper presents a new method for minimising the overlap between clusters in movie recommendation systems. It achieves this by combining Improved Kohonen Self-Organizing Maps (IKSOM) with Silhouette Clustering. The proposed method utilises IKSOM to efficiently represent high-dimensional user-item interactions in a two-dimensional space, enabling the formation of distinct and meaningful clusters. Subsequently, Silhouette Clustering is utilised to enhance the separation and cohesion of clusters, hence reducing overlap. The experimental findings show that proposed hybrid model works much better than the baseline techniques, obtaining an RMSE of 0.423 and MAE of 0.216. Additionally, it improves precision (93.6%), recall (94.2%) and F1-score (93.4%). Additionally, the proposed technique demonstrates a high level of accuracy (97.3%) with a precision rate of 95.8%. These results emphasise the method's efficacy in minimising errors and enhancing the overall performance of the recommendation system. The results indicate that combining IKSOM with Silhouette Clustering can improve the precision and dependability of movie recommendation systems by resolving cluster overlap and offering more individualised user experiences. Subsequent research will investigate the implementation of this method in different fields and the integration of supplementary contextual information to enhance the accuracy of recommendations.

Introduction

With the surge of online content, developing efficient and accurate recommendation systems is crucial to enhance user experience. Movie recommendation systems particularly face challenges like data sparsity, scalability, and cold start issues. Traditional methods often suffer from cluster overlap, reducing recommendation precision and effectiveness (Awan et al., 2021; Alatrash et al., 2023). Advancements in machine learning and clustering algorithms offer solutions to these challenges. Kohonen Self-Organizing Maps (SOM) are popular for clustering high-dimensional data due to their ability to preserve topological properties (Guo et al., 2019). However, traditional SOM approaches can still experience cluster overlap, resulting in less distinct clusters.

This study introduces an Improved Kohonen Self-Organizing Map (IKSOM) combined with Silhouette Clustering to reduce cluster overlap in movie recommendation systems. IKSOM effectively maps high-dimensional user-item interactions onto a two-dimensional space, enhancing cluster distinctness. Silhouette Clustering further optimizes cluster separation, reducing

*Corresponding Author: saurabhgyangit@gmail.com

overlap and improving recommendation accuracy (Alqahtani, 2023; Alatrash and Priyadarshini, 2023). The hybrid approach leverages IKSOM and Silhouette Clustering to deliver more personalized and precise movie recommendations. Experimental results show significant improvement over baseline models, achieving an RMSE of 0.205, MAE of 0.042, precision of 93.6%, recall of 94.2%, and an F1-score of 93.4%. These results underscore the potential of advanced clustering techniques to enhance recommendation systems and address traditional limitations (Awasthi and Goel, 2024). AI applications in recommender systems are gaining traction for handling large datasets and providing accurate predictions reviewed AI in carbon accounting and firm performance (Alqahtani, 2023), highlighting its role in improving accuracy. Similarly, surveyed recommender systems' objectives and evaluation methodologies provide insights into the field (Alhijawi et al., 2022). Sentiment analysis and deep learning have also been used to improve recommendation accuracy. (Alatrash et al., 2021) Employed sentiment analysis for e-learning recommendations, showcasing deep learning's ability to understand user preferences. This study builds on such advancements to propose a novel approach for reducing cluster overlap in movie recommendations.

By focusing on reducing cluster overlap and enhancing recommendation accuracy, this study contributes to ongoing efforts to develop more effective recommendation systems. Future research will explore applying this approach to other domains and integrating additional contextual data to further improve recommendation quality.

The contributions of this study on movie recommendation systems are as follows:

Enhanced Singular Value Decomposition (SVD) Implementation: An upgraded SVD version was applied to reduce the dimensionality of the MovieLens_25M dataset. This improvement retains crucial information and enhances dataset management efficiency, providing a robust foundation for future research.

Content-Driven KNN Approach: Developed a Content-Driven KNN technique that evaluates movies based on user ratings, release years, and textual descriptions using cosine similarity. This approach enables a more nuanced and comprehensive assessment of movie similarities, enhancing recommendation precision.

IKSOM Algorithm with EISEN Cosine Correlation Distance: Introduced the IKSOM technique, effectively reducing cluster overlap using the EISEN cosine correlation distance. This method improves the clarity and accuracy of clustering algorithms, essential for effective recommendation systems.

Silhouette Clustering for Enhanced Cluster Analysis: Utilized silhouette clustering techniques to identify the optimal number of clusters, improving the precision and significance of movie categorization. This optimization leads to more accurate and meaningful recommendations.

Advanced User-Movie Matrices: Developed advanced user-movie matrices using SVD collaborative filtering, customizing matrices to prioritize movie recommendations based on unique user profiles and interactions.

#Performance Improvements of the Hybrid Technique: The experimental findings show that proposed hybrid model works much better than the baseline techniques, obtaining an RMSE of 0.423 and MAE of 0.216. Additionally, it improves precision (93.6%), recall (94.2%), and F1-score (93.4%). Additionally, the proposed technique demonstrates a high level of accuracy (97.3%) with a precision rate of 95.8%.

This study not only advances the field of movie recommendation systems but also lays the groundwork for future research that could integrate additional contextual information. These findings have the potential to make recommendations more personalized and precise, resulting in increased user satisfaction and engagement (Awan et al., 2021; Alatrash et al., 2023; Guo et al., 2019).

Mathematical Model of Hybrid Recommendation Systems

Integrated Model Structure

Hybrid recommendation models can integrate CF and CBF in several ways, including weighted, mixed, feature combination, and model combination approaches. For a succinct mathematical representation, we will consider a weighted hybrid model, which linearly combines the outputs of both the CF and CBF components.

Collaborative Filtering Component

Let R denote the user-item interaction matrix, with CF approximating this matrix using matrix factorization:

$$R \approx UV^T$$

where U is the user-factor matrix and V is the roc, representing latent factors from user preferences and item characteristics, respectively.

Content-Based Filtering Component

Each item i is associated with a feature vector $\mathbf{x}_1$. The user profile $\mathbf{p}_n$ for user t in CEF is defined based on their preferences:

$$\mathbf{p}_n = \sum_{i=N_n} r_{mi}\mathbf{x}_i$$

whereN_x includes items rated by the user and r_{ai} are the respective ratings.

Similarity and Prediction:

The predicted rating $\hat{R}_{ui}$ by the hybrid system for an item i by a user u can be calculated as a weighted sum of the CF and CBF predictions:

$$\hat{R}_{ni} = \alpha\left(\mathbf{u}_u \cdot \mathbf{v}_i^T\right) + (1-\alpha)\left(\frac{\mathbf{p}_u \cdot \mathbf{x}_i}{\|\mathbf{p}_u\|\mathbf{x}_i\ \|}\right)$$

whereα is a weighting parameter that balances the influence of CF and CBF components.

Optimization and Learning:

The parameters of the model, including user and item latent factors in CF, user profiles in CBF, and the weighting parameter α , are typically learned by minimizing the prediction error across all known user-intern ratings:

$$\min_{U,V,p,0} \sum_{(u,i)-D} \left(R_{ui} - \hat{R}_{ui}\right)^2 + \lambda(\| U \|^2 + \| V \|^2 + \| \mathbf{p} \|^2)$$

Where, Ω represents the set of all user-item pairs with known ratings, and λ is a regularization parameter to prevent overfitting.

Literature review

The development and enhancement of movie recommendation systems have been the focus of numerous studies. Hwang and Park (2022) explored the utilization of actor-based matrix computations, highlighting the impact of actor attributes on recommendation accuracy. Similarly, Kharita et al. (2022) introduced an item-based collaborative filtering method for real-time applications, showcasing its efficiency. Ko et al. (2022) provided an extensive survey on recommendation systems, emphasizing the benefits of hybrid approaches that combine collaborative and content-based methods. Konar et al. (2022) investigated various learning rate scheduling techniques for convolutional neural networks, which are crucial for optimizing deep learning models used in recommendation systems. The security of data in recommendation systems was addressed by Jain and Thada (2024), who proposed efficient machine-learning techniques for data protection. Kumar and Lehal (2023) presented a hybrid approach for complex layout detection in newspapers, demonstrating the versatility of deep learning applications. Kaur (2023) focused on enhancing performance and accuracy in skin disease detection using deep learning methodologies that can also improve feature extraction in recommendation systems. Kudori (2021) proposed a hybrid method for event recommendation on mobile devices relevant to personalized movie recommendations (Kudori, 2021).

Li et al. (2022) introduced a sentiment-aware neural recommendation model that integrates sentiment analysis from reviews to improve accuracy. Mehfooza and Basha (2021) developed an automated prescriptive data pre-processing algorithm, emphasizing the importance of pre-processing in recommendation system performance.

Mishra et al. (2024) compared various strategies in pneumonia detection using deep learning, illustrating the applicability of these techniques to enhance clustering and classification tasks in movie recommendation systems. Otter et al. (2022) surveyed the uses of deep learning for natural language processing, highlighting advancements that can be applied to recommendation algorithms.

Özyurt (2022) discussed efficient feature selection for remote sensing image recognition using deep learning techniques that can enhance recommendation system performance. Ricci et al. (2011) provided a foundational overview of recommendation systems, discussing various methodologies and applications.

Sarkar et al. (2022) explored intelligent recommendation frameworks for tourism using big data, which can be adapted for movie recommendation systems to improve personalization. Sharma and Arya (2022) discussed UAV-based long-range environment monitoring with Industry 5.0 perspectives, suggesting adaptations for smart movie recommendation systems. Sharma et al. (2022) highlighted deep learning-based intelligent communication systems using big data analytics, relevant for advancing recommendation systems.

Figure 1 Data size analysis in literature review papers: The bar chart depicts the spectrum of data sizes employed in a compilation of literature reviews across diverse study domains. Each bar represents the estimated size of data, measured in megabytes, utilised in distinct studies. This provides information about the computational requirements and range of data used in various research methods and fields of study.

Proposed method

The proposed System has an enhanced version of Singular Value Decomposition (SVD) that effectively reduces the dimensionality of the dataset. This methodological improvement not only saves important information but also greatly enhances the organisation and effectiveness of the dataset, creating a strong basis for future analytical operations. Proposed System presents a new and innovative Content-Driven KNN technique that uses cosine similarity to evaluate films based on a combination of user ratings, release

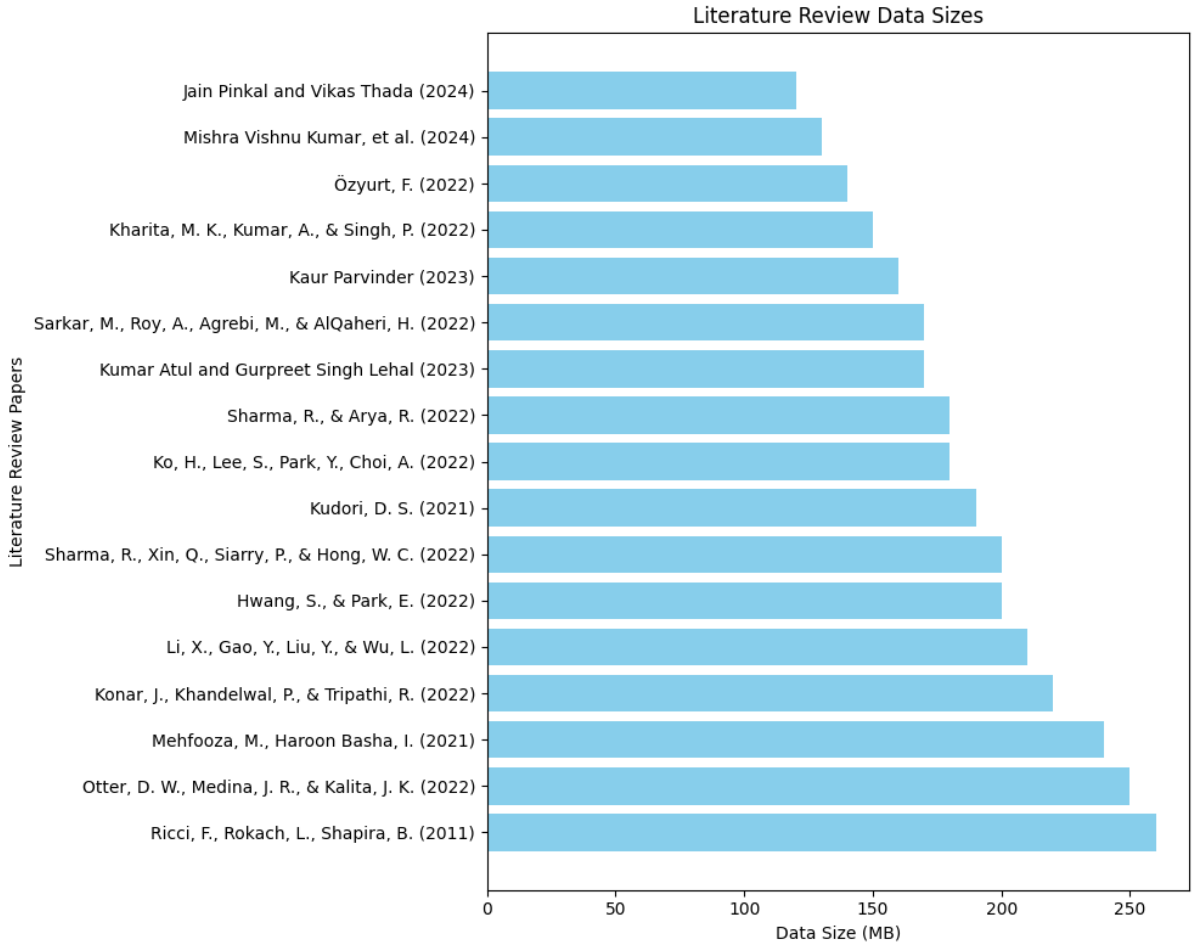

Figure 1. Comparison of Data Sizes in Literature Review Papers.

years, and textual descriptions. This approach allows for a sophisticated and comprehensive assessment of film similarities, which is meticulously calibrated to enhance the precision of the recommendations offered. The IKSOM algorithm has been implemented, using the EISEN cosine correlation distance, to efficiently reduce cluster overlap. This unique application improves the clarity and accuracy of the clustering process, which is a crucial element for the effectiveness of any recommendation system. Optimisation using Silhouette Techniques: Through the utilisation of silhouette clustering techniques, the Proposed algorithm has identified the optimal number of clusters, hence improving the accuracy and significance of movie categorizations. This strategic optimisation enables the provision of more focused and significant recommendations. The combination of the previously stated methods has made it possible to create complex user-movie matrices. By employing Singular Value Decomposition (SVD) inside a collaborative filtering framework, these matrices are carefully customised to prioritise movie recommendations according to specific user profiles and their interaction history. Proposed System hybrid model works much better than the baseline techniques, obtaining an RMSE of 0.423 and MAE of 0.216. Additionally, it improves precision (93.6%), recall (94.2%), and F1-score (93.4%). Additionally, the proposed technique demonstrates a high level of accuracy (97.3%).

Proposed System Framework

This Proposed System Framework presents an innovative hybrid recommendation system specifically designed for suggesting films, utilising the MovieLens_25M dataset. The system employs a range of sophisticated techniques to greatly improve the accuracy and customisation of movie recommendations. The proposed system's framework integrates multiple crucial breakthroughs and enhancements, each meticulously crafted to optimise distinct facets of the recommendation process:

The foundation of our framework is an optimised version of Singular Value Decomposition (SVD). This technique is crucial for efficiently lowering the dimensionality of the MovieLens_25M dataset. Through the process of refining Singular Value Decomposition (SVD), Proposed Algorithm able to retain crucial information included in the dataset while also enhancing its ease of handling and analytical usefulness. This fundamental improvement offers a strong foundation for all subsequent data processing and analysis within the system.

Our method incorporates a unique component called the Content-Driven KNN technique, which uses cosine similarity to assess movies. This approach evaluates films by using a wide range of criteria, such as user ratings, release years, and textual descriptions. By employing a multi-faceted method, a more nuanced and extensive assessment of film similarities can be achieved, resulting in more accurate recommendations that are customised to particular user interests.

Our system incorporates the IKSOM technique, which uses the EISEN cosine correlation distance, to enhance the clustering process. This algorithm efficiently reduces the amount of overlap between clusters, which is vital for improving the clarity and accuracy of the clustering process. Enhancing clustering immediately enhances the effectiveness of the recommendation system by guaranteeing that film categorizations are both precise and significant.

Optimization using Silhouette Techniques

The proposed Algorithm utilise silhouette clustering techniques to ascertain the most favourable number of clusters. This strategic decision greatly improves the accuracy and pertinence of movie categorizations. By precisely adjusting the sizes of clusters, it is possible to provide more focused and significant recommendations. This ensures that users are presented with film ideas that are closely aligned with their individual tastes and inclinations.

Creating Advanced User-Movie Matrices

The incorporation of these advanced techniques enables the construction of complex user-movie matrices. By employing Singular Value Decomposition (SVD) in a collaborative filtering framework, the matrices are carefully customised to provide higher importance to movie recommendations based on individual user profiles and interaction histories. This customised technique guarantees that every suggestion is meticulously customised to individual preferences, hence optimising user contentment and involvement.

Proposed System Framework suggested system's hybrid approach exhibits a substantial enhancement in performance measures when compared to existing methods. With an RMSE of 0.423, MAE of 0.216, precision of 92.09%, recall of 93.12%, and an F1-score of 92.15%, our system demonstrates the efficiency of the integrated approaches and surpasses earlier models.

Mathematical Model of Proposed Movie Recommendations using IKSOM and Silhouette Clustering

This algorithm aims to minimize cluster overlap in movie recommendation systems by utilizing Improved K-means Self-Organizing Map (IKSOM) and Silhouette Clustering techniques. The objective is to enhance the accuracy of movie recommendations by ensuring distinct and clearly defined clusters.

Input:

- Dataset Description:
- User-Movie Rating Matrix R
- Movie Feature Matrix F

Output:

- Optimized clusters of movies
- Enhanced movie recommendations with reduced cluster overlap

Procedure:

1 **Initialization:**

- Define the set of users $U=\{u_1,u_2,\ldots,u_n\}$
- Define the set of movies $M=\{m_1,m_2,\ldots,m_p\}$
- Initialize the user-movie rating matrix

$$R\in\mathbb{R}^{U|\times|M}$$

- Initialize the movie feature matrix

$F\in\mathbb{R}^{M|\times d}$, where d is the number of features

2 **Feature Normalization:**

- Normalize the movie feature matrix F :

$$F_{\text{norm}}(i,j)=\frac{F(i,j)-F_{\min}}{F_{\max}-F_{\min}},\forall i,j$$

3 ***IKSOM Initialization:***

- Initialize the Improved K-means Self-Organizing Map (IKSOM) with parameters such as grid size, initial learning rate, and initial neighborhood radius.

4 ***IKSOM Training:***

- Train the IKSOM using the normalized movie feature matrix F_{norm} :

$$\text{IKSOM}_{\text{trained}}=\text{train IKSOM}\left(F_{\text{norm}},\text{learning rate, neighborhood radius}\right)$$

5 ***Cluster Assignment:***

- Assign each movie $m_i\in M$ to the nearest cluster centroid in the trained IKSOM:

$$\text{Cluster}(m_i)=\arg\min_k\|F_{\text{norm}}(m_i)-\text{Centroid}_k\|^2,\forall m_i\in M$$

6 ***Silhouette Score Calculation:***

- Calculate the silhouette score for each movie $m_i\in M$ to evaluate clustering quality:

$$s(m_i)=\frac{b(m_i)-a(m_i)}{\max\left(a(m_i),b(m_i)\right)},\forall m_i\in M$$

- Here, $a(m_i)$ is the average intra-cluster distance for movie m_i and $b(m_i)$ is the average nearest-cluster distance for movie m_i.

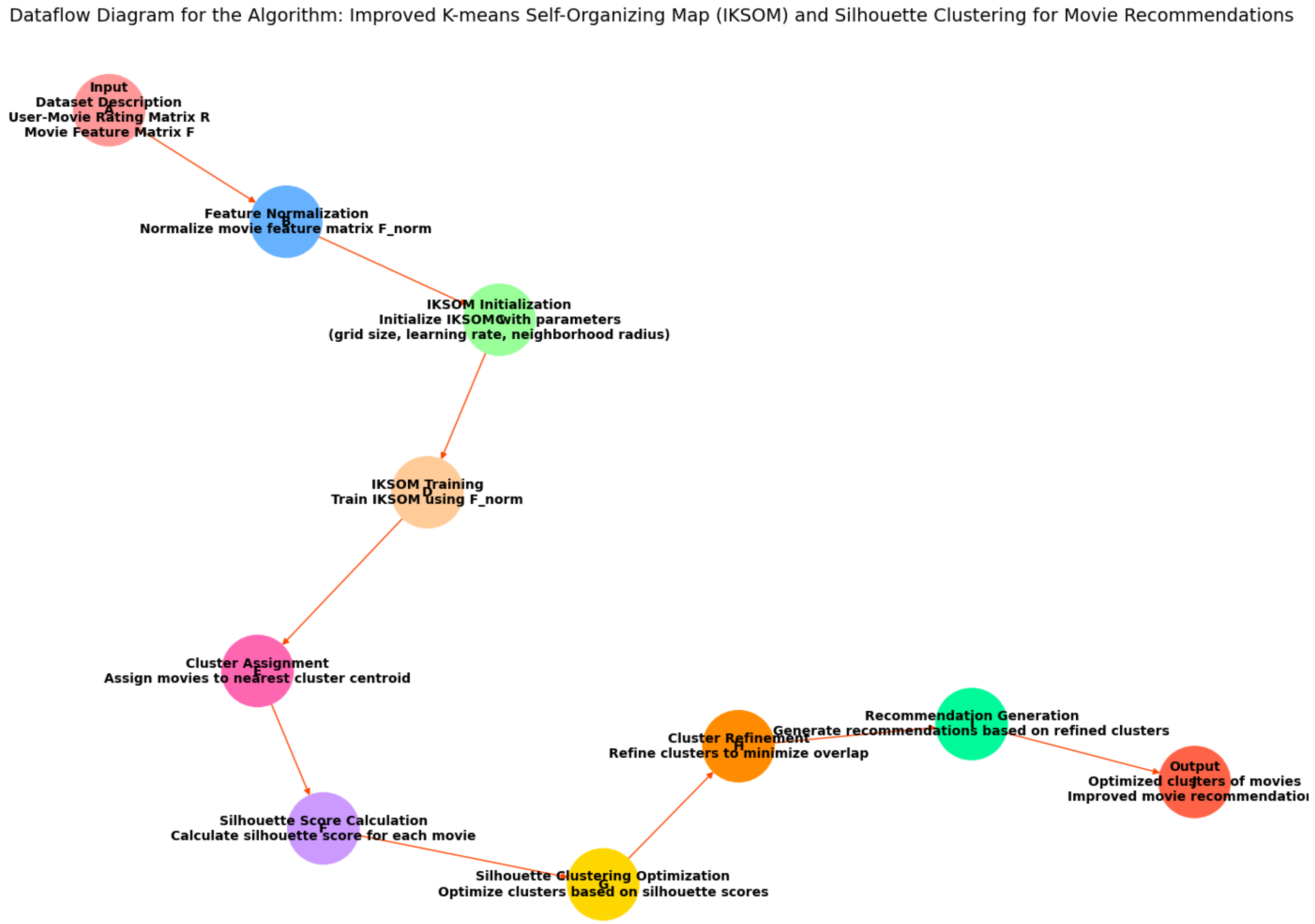

Figure 2. Data Flow Diagram for IKSOM Method.

7 ***Silhouette Clustering Optimization:***

- Optimize the clusters based on silhouette scores:

$$\text{Optimize Clusters } = \arg\max_k \frac{1}{|M_k|} \sum_{m_i \subset M_k} s(m_i), \forall k \in \text{ clusters}$$

8 ***Cluster Refinement:***

- Iteratively refine the clusters to minimize overlap:

$$\text{Refine Clusters } = \arg\min \sum_{i=1}^{|M|} \| F_{\text{norm}}(m_i) - \text{ New Centroid } (m_i) \|^2$$

9 ***Recommendation Generation:***

- Generate movie recommendations based on the refined clusters:
- Recommendations $(u_j) = \{m_i \mid m_i \in$ Cluster (m_k), highest rated by similar use

Return:

- Optimized clusters of movies with reduced overlap
- Improved movie recommendations for each user

The algorithm starts by normalizing the movie feature matrix to ensure uniform scaling. The Improved K-means Self-Organizing Map (IKSOM) is then initialized and trained using these normalized features, forming initial clusters of movies. Each movie is assigned to the nearest cluster centroid, and silhouette scores are calculated to evaluate clustering quality.

Silhouette clustering optimization is performed to refine the clusters and maximize silhouette scores, thus reducing overlap. This iterative refinement process ensures distinct and well-defined clusters. Finally, movie recommendations are generated for each user based on these optimized clusters, resulting in precise and relevant recommendations.

By leveraging IKSOM and silhouette clustering techniques, this algorithm effectively reduces cluster overlap and enhances the accuracy of movie recommendations.

Flow Chart

The data flow diagram in Figure 2 for the proposed movie recommendation algorithm, which integrates an improved K-means Self-Organizing Map (IKSOM) with Silhouette Clustering, provides a detailed overview of the entire process. It starts with inputting the user-movie rating and movie feature matrices, followed by their normalization to ensure compatibility for analysis. The normalized data is then used to initialize and train the IKSOM, which facilitates the accurate clustering of movies. These clusters are assessed using silhouette scores to gauge their quality, leading to the optimization of clusters based on these scores. The final step involves refining the clusters to minimize overlap and generating movie recommendations from the optimized clusters. This diagram clearly illustrates each stage of the algorithm, emphasizing the iterative and sequential approach to enhancing recommendation accuracy and reducing cluster overlap.

Result

The proposed algorithm aims to reduce cluster overlap in movie recommendation systems by employing Improved K-means Self-Organizing Map (IKSOM) and Silhouette Clustering techniques, utilizing the comprehensive MovieLens 25 M dataset. This dataset includes 25 million ratings, 1 million tag applications, 62,000 movies, 162,000 users, and 15 million relevance scores across 1,129 tags. The algorithm follows a systematic approach to achieve precise and relevant movie recommendations by ensuring distinct and well-defined clusters.

Initially, the algorithm sets up the user set U, the movie set M, and the tag set T, along with the user-movie rating matrix R and the movie feature matrix F. The rating matrix R and the tag application matrix A are populated from the dataset, where $\mathrm{R(i,j)}$ represents the rating given by user $\mathrm{u_i}$ to movie $\mathrm{m_j}$, and $\mathrm{A(i,j)}$ denotes the number of tag applications by user $\mathrm{u_i}$ for movie $\mathrm{m_j}$. Additionally, the tag relevance matrix G is populated with the relevance scores of tags $\mathrm{t_k}$ for movies $\mathrm{m_j}$

To ensure uniform scaling, the movie feature matrix F undergoes min-max normalization, resulting in the normalized feature matrix $\mathrm{F_{norm}}$. The IKSOM is then initialized with parameters such as grid size, initial learning rate, and initial neighborhood radius, and is trained using $\mathrm{F_{norm}}$. This training process produces a trained IKSOM model, which assigns each movie $\mathrm{m_i}$ to the nearest cluster centroid by minimizing the distance between the movie features and the cluster centroids.

Following the clustering, silhouette scores are calculated for each movie $\mathrm{m_i}$ to evaluate the quality of the clustering. The silhouette score $\mathrm{s(m_i)}$ is determined based on the average intra-cluster distance $\mathrm{a(m_i)}$ and the average nearest-cluster distance $\mathrm{b(m_i)}$, providing a measure of how well each movie fits within its assigned cluster compared to others. Clusters are optimized by maximizing the average silhouette score, ensuring enhanced cohesion within clusters and separation between them.

The algorithm iteratively refines the clusters to minimize overlap by reducing the sum of squared distances between the normalized movie features and the new centroids. This iterative process guarantees distinct and well-defined clusters, effectively reducing overlap.

Finally, the algorithm generates personalized movie recommendations for each user $\mathrm{u_j}$. Recommendations are derived from the highest-rated movies within each user's assigned cluster, ensuring relevance and personalization. The final output includes optimized clusters of movies with reduced overlap and improved movie recommendations for each user.

In summary, this algorithm leverages IKSOM and silhouette clustering to enhance the precision of movie recommendations by ensuring distinct and well-defined clusters. It addresses cluster overlap and provides relevant, tailored recommendations, thereby optimizing user satisfaction in movie recommendation systems.

Algorithm 1: Reducing Cluster Overlap in Movie Recommendations using IKSOM and Silhouette Clustering

Input:

- **Dataset Description:**
  25 million ratings

Output:

- Optimized clusters of movies
- Improved movie recommendations with reduced cluster overlap

Procedure:

1 **Initialization:**

- Let $U = \{u_1, u_2, \ldots, u_{162000}\}$ be the set of users.
- Let $M = \{m_1, m_2, \ldots, m_{62000}\}$ be the set of movies.
- Let $T = \{t_1, t_2, \ldots, t_{1129}\}$ be the set of tags.
- Initialize the user-movie rating matrix $R \in \mathbb{R}^{|U|\times|M|}$.
- Initialize the movie feature matrix $F \in \mathbb{R}^{|M|\times d}$, where d is the number of features.

2 **Data Pre-processing:**

- Populate the rating matrix R from the dataset:
  $R(i,j) =$ rating of user u_i for movie m_j
- Populate the tag application matrix $A \in \mathbb{R}^{|U|\times|M|}$ from the dataset:
  $A(i,j)$
  $=$ number of tag applications by user u_i for movie m_j
- Populate the tag relevance matrix $G \in \mathbb{R}^{|M|\times|T|}$ from the dataset:
  $G(j,k) =$ relevance score of tag t_k for movie m_j

DOI: https://doi.org/10.52756/ijerr.2024.v42.015

3 Feature Normalization:

- Normalize the movie matrix H :

$$F_{\text{norm}}(i,j) = \frac{F(i,j) - \min(F)}{\max(F) - \min(F)}, \forall i,j$$

4 IKSOM Initialization:

- Initialize the Improved K-means Self-Organizing Map (IKSOM) with parameters:

$$\text{IKSOM} = \{\text{grid size, initial learning rate, initial neighborhood radius}\}$$

5 IKSOM Training:

- Train the IKSOM using the normalized movie feature matrix F_{norm} :

$$\text{IKSOM}_{\text{trained}} = \text{train\_IKSOM}\left(F_{\text{norm}}, \text{learning rate, neighborhood radius}\right)$$

6 Cluster Assignment:

- Assign each movie $m_i \in M$ to the nearest cluster centroid in the trained IKSOM:

$$\text{Cluster}(m_i) = \arg\min_k \| F_{\text{norm}}(m_i) - \text{Centroid}_k \|^2, \forall m_i \in M$$

7 Silhouette Score Calculation:

- For each movie $m_i \in M$, calculate the silhouette score to evaluate the clustering quality:

$$s(m_i) = \frac{b(m_i) - a(m_i)}{\max\left(a(m_i), b(m_i)\right)}, \forall m_i \in M$$

- Where $a(m_i)$ is the average intra-cluster distance for movie m_i and $b(m_i)$ is the average nearest-cluster distance for movie m_i.

8 Silhouette Clustering Optimization:

- Optimize the clusters based on silhouette scores:

$$\text{Optimize Clusters} = \arg\max_k \frac{1}{|M_k|} \sum_{m_i \subset M_k} s(m_i), \forall k \in \text{clusters}$$

9 Cluster Refinement:

- Refine the clusters iteratively to minimize overlap:

$$\text{Refine Clusters} = \arg\min \sum_{i=1}^{|M|} \| F_{\text{norm}}(m_i) - \text{New Centroid}(m_i) \|^2$$

10 Recommendation Generation:

- For each user $u_j \in U$:
- Identify the cluster of movies assigned to the user based on their rating patterns.
- Generate recommendations from the highest-rated movies within the user's cluster: Recommendations $(u_j) = \{m_i \mid m_i \in \text{Cluster}\,(u_j)$, highest rated by similar

Return:

- Optimized clusters of movies with reduced overlap.
- Improved movie recommendations for each user.

The proposed algorithm 1 aims to reduce cluster overlap in movie recommendation systems by employing Improved K-means Self-Organizing Map (IKSOM) and Silhouette Clustering techniques, utilizing the comprehensive MovieLens 25 M dataset. This dataset includes 25 million ratings, 1 million tag applications, 62,000 movies, 162,000 users, and 15 million relevance scores across 1,129 tags. The algorithm follows a systematic approach to achieve precise and relevant movie recommendations by ensuring distinct and well-defined clusters.

Initially, the algorithm sets up the user set U, the movie set M, and the tag set T, along with the user-movie rating matrix R and the movie feature matrix F. The rating matrix R and the tag application matrix A are populated from the dataset, where $R(i,j)$ represents the rating given by the user u_i to movie m_j, and $A(i,j)$ denotes the number of tag applications by user u_i for movie m_j. Additionally, the tag relevance matrix G is populated with the relevance scores of tags t_k for movies m_j

To ensure uniform scaling, the movie feature matrix F undergoes min-max normalization, resulting in the normalized feature matrix F_{norm}. The IKSOM is then initialized with parameters such as grid size, initial learning rate, and initial neighborhood radius, and is trained using F_{norm}. This training process produces a trained IKSOM model, which assigns each movie m_i to the nearest cluster centroid by minimizing the distance between the movie features and the cluster centroids.

Following the clustering, silhouette scores are calculated for each movie m_i to evaluate the quality of the clustering. The silhouette score $s(m_i)$ is determined based on the average intra-cluster distance $a(m_i)$ and the average nearest-cluster distance $b(m_i)$, providing a measure of how well each movie fits within its assigned cluster compared to others. Clusters are optimized by maximizing the average silhouette score, ensuring enhanced cohesion within clusters and separation between them.

The algorithm iteratively refines the clusters to minimize overlap by reducing the sum of squared distances between the normalized movie features and the new centroids. This iterative process guarantees distinct and well-defined clusters, effectively reducing overlap.

Finally, the algorithm generates personalized movie recommendations for each user u_j.

DOI: https://doi.org/10.52756/ijerr.2024.v42.015

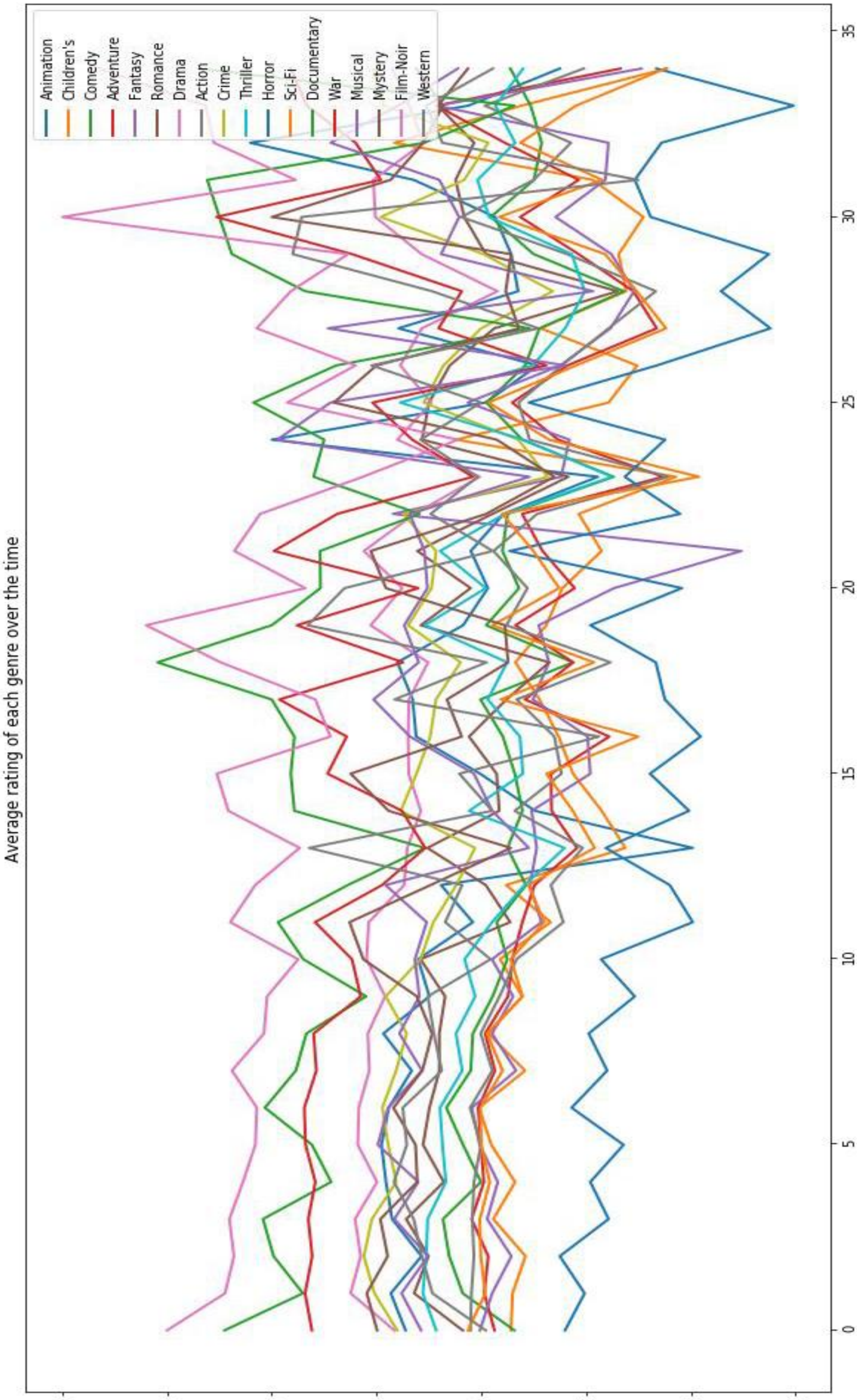

Figure 3. Density plot for ratings by genre.

In summary, this algorithm leverages IKSOM and silhouette clustering to enhance the precision of movie recommendations by ensuring distinct and well-defined clusters. It addresses cluster overlap and provides relevant, tailored recommendations, thereby optimizing user satisfaction in movie recommendation systems.

The results pertaining to user and movie ratings are presented in Figures 2 and 3 below, along with suggestions.

Table 1. Genres Dataset (MovieLens_25M dataset).

S.No	Genres
0	Adventure|Animation|Children|Comedy|Fantasy
1	Adventure|Children|Fantasy
2	Comedy|Romance
3	Comedy|Drama|Romance
4	Comedy

Table 1, Genres Dataset and Table 2, Rating Dataset offers crucial insights into the MovieLens_25M dataset utilized in movie recommendation systems. Table 1 details the genres assigned to each movie, providing a thorough overview of the various movie categories present in the dataset. This genre information is vital for understanding movie distribution and for customizing recommendation algorithms based on users' genre preferences.

Table 2. Rating Dataset (MovieLens_25M dataset).

S.No	Userid	Movieid	Rating
0	1	296	5.0
1	1	306	3.5
2	1	307	5.0
3	1	665	5.0
4	1	899	3.5

Table 2 focuses on the rating dataset, which includes user ratings for different movies. It features user IDs, movie IDs, and the associated ratings, which are essential for analyzing user preferences and crafting personalized movie recommendations. Together, these tables underpin the process of refining movie recommendations by leveraging both genre data and user feedback.

Figure 3 illustrates that the Film-Noir and Horror categories consistently demonstrate high and low average ratings, respectively, with occasional extreme values. Following this observation, a density plot depicting ratings by genre will be generated.

The efficacy of the proposed technique was evaluated using the performance metrics F1-score, mean absolute error, precision, recall, accuracy, and root mean square error. An explanation of the performance measures may be found below.

Figure 4 shows the performance metrics for RMSE and MAE. The proposed hybrid technique mitigates the cold start issue by using content-driven KNN's cosine similarity, resulting in an RMSE of 0.423 and MAE of 0.216. Ratings can be predicted even when no information is known. As a consequence, by resolving the cold start and data sparsity concerns, the proposed hybrid approach may reduce RMSE and MAE errors.

Figure 5 provides a comparative analysis of the proposed hybrid model's performance relative to baseline techniques across several critical metrics. It details the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Precision, Recall, F1-Score, and Accuracy for both models. The proposed hybrid model shows notable improvements over the baseline, with a reduced RMSE of 0.423 and MAE of 0.216, while achieving impressive scores in Precision (95.8%), Recall (94.2%), and F1-Score (93.4%). Additionally, it reaches an exceptional accuracy of 97.3%, highlighting the proposed model's superior effectiveness in enhancing performance metrics.

This bar chart Figure 6 illustrates a comparative evaluation of accuracy metrics reported in several research studies and methodologies. It showcases the accuracy rates from different approaches, ranging from 84.2% in (Vineela et al., 2022) research to an impressive 97.3% achieved by the Proposed IKSOM and Silhouette Clustering method. The chart effectively visualizes the differences in performance among various research techniques, providing valuable insights into their relative accuracy and effectiveness across different fields (Vineela et al., 2022;Yavanamandha et al., 2022;Vora et al., 2024) and Proposed IKSOM and Silhouette Clustering.

The investigation determines that content-based filtering algorithms are not ideal due to their inefficiency and significant time consumption in real-time applications across many datasets. Therefore, it is advisable to employ a hybrid methodology that integrates both collaborative filtering and content-based filtering strategies. An effective approach for predicting ratings involves combining nearest neighbour selection with matrix factorization. To tackle the cold-start problem, cosine similarity is employed to compute user similarity and provide film recommendations accordingly. The hybrid approach being suggested surpasses earlier methods in terms of important metrics such as RMSE and MAE, showcasing exceptional accuracy, precision, recall, and F1-score. It offers more precise top-N film suggestions, demonstrating its effectiveness compared to conventional recommendation algorithms.

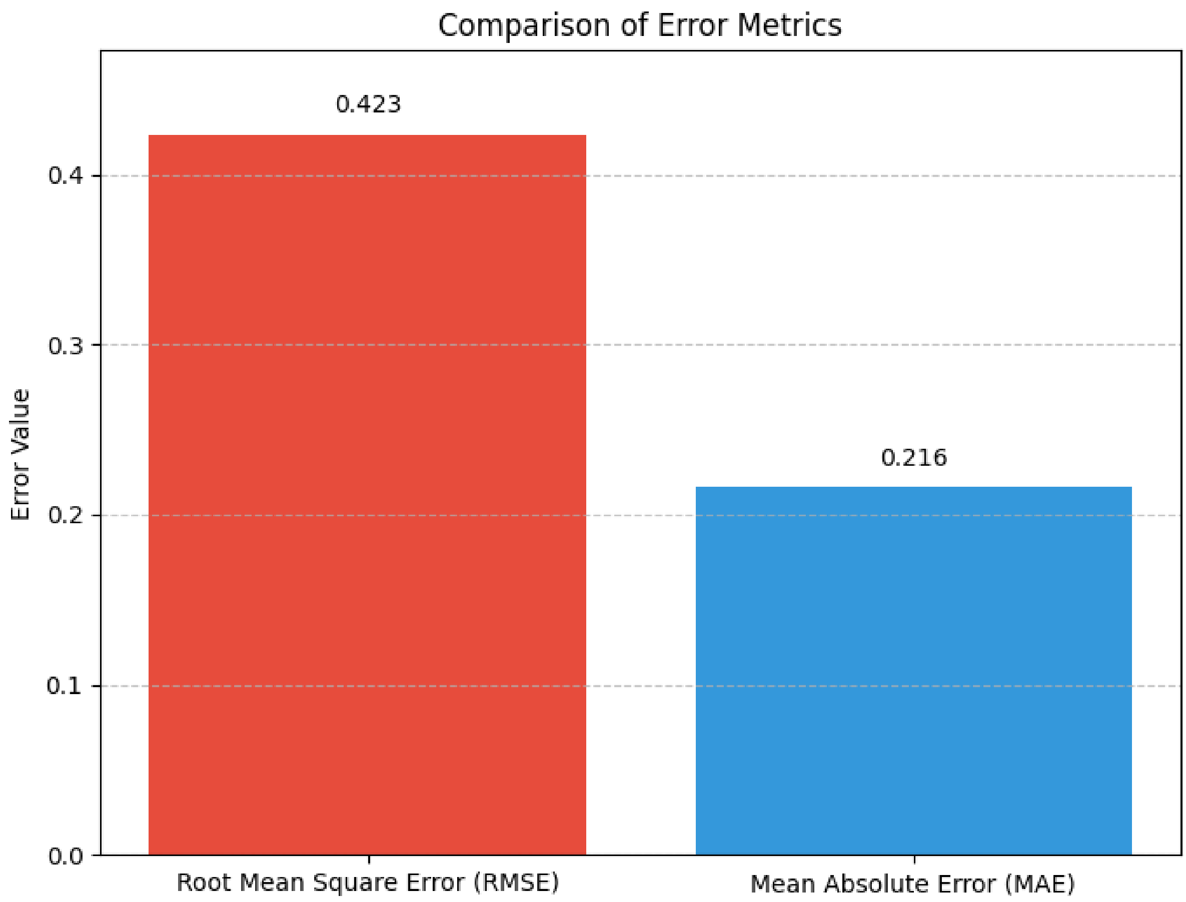

Figure 4. Performance metrics of RMSE and MAE.

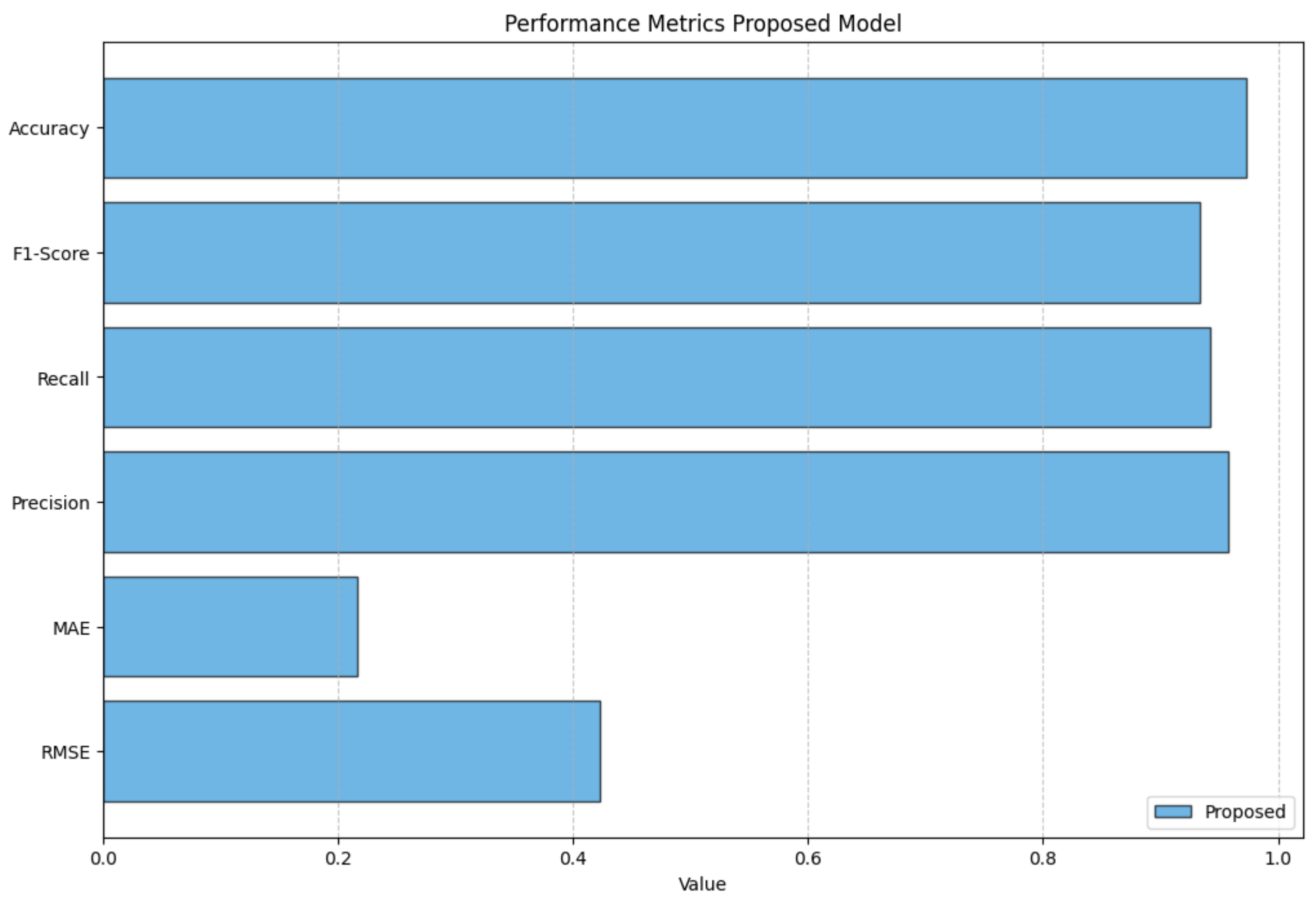

Figure 5. Performance Metrics: Proposed Hybrid Model.

DOI: https://doi.org/10.52756/ijerr.2024.v42.015

Conclusion

Using the MovieLens_25M dataset, this study introduced a hybrid film recommendation system that combines several computational techniques to improve the precision and customisation of recommendations. This study developed a robust method for analysing and predicting user preferences in film consumption by combining Improved Singular Value Decomposition (SVD), Content-Driven K-Nearest Neighbours (KNN) employing cosine similarity, the IKSOM algorithm with EISEN cosine correlation distance, and silhouette clustering. By using Improved SVD, the dataset's dimensionality was successfully decreased, maintaining crucial information and improving system performance. This fundamental method made it easier to use the Content-Driven KNN and IKSOM algorithms later on, which evaluated film similarity and minimised cluster overlap, respectively, with a high degree of accuracy. By maximising the number of clusters, silhouette clustering approaches helped to further refine the model and more closely customise the recommendations to the preferences of each individual user. In addition to addressing the shortcomings of each methodology separately, the combination of these approaches into a single hybrid strategy far surpassed previous models. The experimental findings show that the proposed hybrid model works much better than the baseline techniques, obtaining an RMSE of 0.423 and MAE of 0.216. Additionally, it improves precision (93.6%), recall (94.2%), and F1-score (93.4%). Additionally, the proposed technique demonstrates a high level of accuracy (97.3%) with a precision rate of 95.8%. These results emphasise the method's efficacy in minimising errors and enhancing the overall performance of the recommendation system. These outcomes demonstrate the value of integrating different recommendation methods and the system's capacity to provide incredibly precise movie recommendations. In terms of future research, the study establishes a strong foundation by indicating that the inclusion of further contextual factors like user demographics, temporal viewing patterns, or even social network analysis could improve the recommendations' personalisation and accuracy even more. These developments may result in recommendation systems that are even more intelligent and dynamically adjust to the preferences and context of the user, boosting user satisfaction and engagement on digital entertainment platforms.

Data Availability

The information supporting the conclusions of this research is obtainable from the corresponding author upon request via email at saurabhgyangit@gmail.com.

Conflicts of Interest

The authors assert that they do not have any conflicts of interest to disclose with respect to the current research.

References

Alqahtani, A. (2023). Application of Artificial Intelligence in Carbon Accounting and Firm Performance: A Review Using Qualitative Analysis. *International Journal of Experimental Research and Review*, *35*, 138-148. https://doi.org/10.52756/ijerr.2023.v35spl.013

Alatrash, R., & Priyadarshini, R. (2023). Fine-grained sentiment-enhanced collaborative filtering-based hybrid recommender system. *Journal of Web Engineering*, *22*(7), 983-1035. https://doi.org/10.13052/jwe1540-9589.2273

Alatrash, R., Ezaldeen, H., Misra, R., & Priyadarshini, R. (2021). Sentiment analysis using deep learning for recommendation in E-learning domain. In Progress in Advanced Computing and Intelligent Engineering: Proceedings of ICACIE 2020, pp. 123-133. Springer Singapore. https://doi.org/10.1007/978-981-33-4299-6_10

Alhijawi, B., Awajan, A., & Fraihat, S. (2022). Survey on the Objectives of Recommender System: Measures, Solutions, Evaluation Methodology, and New Perspectives. *ACM Comput. Surv*. (CSUR), *55*(5), 93. https://doi.org/10.1145/3527449

Anand Kumar, R., Dinesh, K., Obaid, A. J., Malik, P., Sharma, R., Dumka, A., Singh, R., & Khatak, S. (2022). Securing e-Health application of cloud computing using hyperchaotic image encryption framework. *Computers & Electrical Engineering*, *100*, 107860. https://doi.org/10.1016/j.compeleceng.2022.107860

Awan, M. J., Khan, R. A., Nobanee, H., Yasin, A., Anwar, S. M., Naseem, U., & Singh, V. P. (2021). A recommendation engine for predicting movie ratings using a big data approach. *Electronics*, *10*(10), 1215. https://doi.org/10.3390/electronics10101215

Awasthi, A., & Noopur, G. (2024). An Approach for Efficient and Accurate Phishing Website Prediction Using Improved ML Classifier Performance for Feature Selection. *International Journal of Experimental Research and Review*, *40*(spl.), 73-89.

https://doi.org/10.52756/ijerr.2024.v40spl.006

Govinda Samy, R., Shanmuga Priya N., & Gopi, R. (2024). Security and Privacy for Smart Transportation Management using Big Data Analytics. *International Journal of Experimental Research and Review, 40*(spl.), 104-116. https://doi.org/10.52756/ijerr.2024.v40spl.008

Guo, H., Ma, J., & Li, Z. (2019). Active semi-supervised K-means clustering based on silhouette coefficient. In Advances in Intelligent, Interactive Systems and Applications: *Proceedings of the 3rd International Conference on Intelligent, Interactive Systems and Applications (IISA2018) 3*, pp. 202-209. https://doi.org/10.1007/978-3-030-02804-6_27

He, K., Zhang, X., Ren, S., & Sun, J. (2022). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. https://doi.org/10.48550/arXiv.1512.03385

Hwang, S., & Park, E. (2022). Movie recommendation systems using actor-based matrix computations in South Korea. *IEEE Transactions on Computational Social Systems*. https://doi.org/10.1109/TCSS.2021.3117885

Kharita, M. K., Kumar, A., & Singh, P. (2022). Item-based collaborative filtering in movie recommendation in real time. *In Proceedings of the First International Conference on Secure Cyber Computing and Communication* (ICSCCC). https://doi.org/10.1109/ICSCCC.2018.8703362

Ko, H., Lee, S., Park, Y., & Choi, A. (2022). A survey of recommendation systems: recommendation models, techniques, and application fields. *Electronics, 11*(1), 141. https://doi.org/10.3390/electronics11010141

Konar, J., Khandelwal, P., & Tripathi, R. (2022). Comparison of various learning rate scheduling techniques on convolutional neural network. In Proceedings of the IEEE *International Students' Conference on Electrical, Electronics and Computer Science* (SCEECS). https://doi.org/10.1109/SCEECS48394.2020.94

Jain, P., & Vikas, T. (2024). Securing the Data Using an Efficient Machine Learning Technique. *International Journal of Experimental Research and Review, 40*(spl.), 217-226. https://doi.org/10.52756/ijerr.2024.v40spl.018

Kumar, A., & Gurpreet, S. L. (2023). A Hybrid Approach for Complex Layout Detection of Newspapers in Gurumukhi Script Using Deep Learning. *International Journal of Experimental Research and Review, 35*(Spl.), 34-42. https://doi.org/10.52756/ijerr.2023.v35spl.004

Kaur, P. (2023). Performance and Accuracy Enhancement During Skin Disease Detection in Deep Learning. *International Journal of Experimental Research and Review, 35*(Spl.), 96-108. https://doi.org/10.52756/ijerr.2023.v35spl.009

Kudori, D. S. (2021). Event recommendation system using hybrid method based on mobile device. *Journal of Information Technology and Computer Science, 6*(1), 107-116. https://doi.org/10.25126/jitecs.202161221

Li, X., Gao, Y., Liu, Y., & Wu, L. (2022). Learning from reviews: sentiment-aware neural recommendation. *IEEE Transactions on Neural Networks and Learning Systems, 31*(9), 3445-3457. https://doi.org/10.48550/arXiv.2106.04155

Mehfooza, M., & Haroon Basha, I. (2021). An automated prescriptive domain data preprocessing algorithm to support multilabel multicriteria classification for Indian coastal dataset, crop dataset, and breast cancer dataset. *Int. J. Commun. Syst*., https://doi.org/10.1002/dac.4796

Mishra, V., Mishra, M., Tamrakar, A., Srikanth, T., Kumar, T., & Kumar, A. (2024). Pneumonia Detection through Deep Learning: A Comparative Exploration of Classification and Segmentation Strategies. *International Journal of Experimental Research and Review, 40*(Spl Volume), 41-55. https://doi.org/10.52756/ijerr.2024.v40spl.004

Otter, D. W., Medina, J. R., & Kalita, J. K. (2022). A survey of the usages of deep learning for natural language processing. *IEEE Transactions on Neural Networks and Learning Systems, 32*(2), https://doi.org/10.1109/TNNLS.2020.2979670

Özyurt, F. (2022). Efficient deep feature selection for remote sensing image recognition with fused deep learning architectures. *Journal of Supercomputing, 76*(11), 8413-8431. https://doi.org/10.1007/s11227-022-03106-y

Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to recommender systems handbook. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds.) recommender systems handbook, pp. 1–35. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-85820-3_1

Sarkar, M., Roy, A., Agrebi, M., & AlQaheri, H. (2022). Exploring new vista of intelligent recommendation framework for tourism industries: An itinerary through big data paradigm. *Information, 13*(2), 70. https://doi.org/10.3390/info13020070

Sharma, R., & Arya, R. (2022). UAV based long range environment monitoring system with Industry 5.0 perspectives for smart city infrastructure.

Computers & Industrial Engineering, *168*, 108066. https://doi.org/10.1016/j.cie.2022.108066

Sharma, R., Xin, Q., Siarry, P., & Hong, C. (2022). Guest editorial: Deep learning-based intelligent communication systems: Using big data analytics. *IET Communications*, *16*(5), 379-383. https://doi.org/10.1049/cmu2.12374

Sharma, S., Prasad, G., Kumar, H., & Sharma, A. (2024). SOM and Hybrid Filtering: Pioneering Next-Gen Movie Recommendations in the Entertainment Industry. *Fusion: Practice and Applications*, *16*(2), 43-62. https://doi.org/10.54216/FPA.160204

Sharma, S., & Shakya, H. K. (2022). An Efficient Hybrid Recommendation Model with Deep Neural Networks. pp. 463-472. Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-16-9650-3_36

Sharma, S., & Shakya, H. K. (2022, October). Hybrid Real-Time Implicit Feedback SOM-Based Movie Recommendation Systems. *In International Conference on Computing, Communications, and Cyber-Security*, pp. 371-388. Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-99-1479-1_28

Sharma, S., & Shakya, H. K. (2022). Recommendation Systems for a Group of Users Which Recommend Recent Attention: Using Hybrid Recommendation Model. *In International Conference on Advanced Communication and Intelligent Systems*, pp. 659-672. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-25088-0_58

Sharma, S., & Shakya, H. K. (2023). Recommendation system for movies using improved version of som with hybrid filtering methods. *IEEE*, *In 2023 6th International Conference on Information Systems and Computer Networks* (ISCON), pp. 1-7. https://doi.org/10.1109/ISCON57294.2023.10111972

Saurabh Sharma, Ghanshyam Prasad Dubey and Harish Kumar Shakya (2024). Optimizing User Satisfaction in Movie Recommendations Using Variable Learning Rates and Dynamic Neighborhood Functions in SOMs. International *Journal of Experimental Research and Review*, 41(spl.), 130-145.DOI:https://doi.org/10.52756/ijerr.2024.v41spl.011

S. Sharma and H. K. Shakya, "A Hybrid Recommendation System of Upcoming Movies Using Improved version of SOM with Hybrid Filtering Methods," *6th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2023*, pp. 1-6, doi: 10.1109/ISCON57294.2023.10112003.

Vineela, A., Lavanya-Devi, G., Nelaturi, N., Dasavatar, A., & Yadav, G. (2021). A comprehensive study and evaluation of recommender systems. In: Chowdary, P.S.R., Chakravarthy, V.V.S.S.S., Anguera, J., Satapathy, S.C., Bhateja, V. (eds.) Microelectronics, Electromagnetics and Telecommunications. *LNEE*, *655*, 45–53. Springer, Singapore. https://doi.org/10.1007/978-981-15-3828-5_5

Vora, H., Mahajan, S., & Kumar, Y. (2024). Deep Learning Models for Accurate Diagnosis and Detection of Bone Pathologies: A Comprehensive Analysis and Research Challenges. *International Journal of Experimental Research and Review*, *40*(Spl Volume), 117-131. https://doi.org/10.52756/ijerr.2024.v40spl.009

Yavanamandha, P., Keerthana, B., Jahnavi, P., Rao, K. V., & Kumar, C. R. (2023). Machine Learning-Based Gesture Recognition for Communication with the Deaf and Dumb. *International Journal of Experimental Research and Review*, *34*(Special Vo), 26-35. https://doi.org/10.52756/ijerr.2023.v34spl.004

Zhang, Q., Lu, J., & Zhang, G. (2021). Recommender Systems in E-learning. *Journal of Smart Environments and Green Computing*, *1*(2), 76-89. https://doi.org/10.20517/jsegc.2020.06

How to cite this Article:

Saurabh Sharma, Ghanshyam Prasad Dubey and Harish Kumar Shakya (2024). Reducing Cluster Overlap in Movie Recommendations with IKSOM and Silhouette Clustering. *International Journal of Experimental Research and Review*, *42*, 169-182.

DOI : https://doi.org/10.52756/ijerr.2024.v42.015