



ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.12 n.58, p.1030-1042. March/April, 2026.

DOI: <https://doi.org/10.5935/jetia.v12i58.3248>



RESEARCH ARTICLE

OPEN ACCESS

C2DR-VAE: ADAPTIVE RECOMMENDATION FRAMEWORK VIA CLUSTER-CONDITIONED AND DYNAMICALLY REFINED VARIATIONAL AUTOENCODERS

R. Navin Kumar^{1*}, J. Srimathi²

^{1,2}Department of Information Technology, KPR College of Arts Science and Research, Coimbatore, India

¹<http://orcid.org/0009-0007-2754-201X>^{id}, ²<http://orcid.org/0000-0003-0693-3496>^{id}

Email: *navinsoccer07@gmail.com, sriphd2020@gmail.com

ARTICLE INFO

Article History

Received: January 6, 2026

Reviewed: February 6, 2026

Accepted: March 10, 2026

Published: April 30, 2026

Keywords:

Recommendation Systems,
Variational Autoencoder (VAE),
Dynamic Clustering,
Cold Start Problem,
Adaptive Personalization.

ABSTRACT

Recommendation systems are crucial to customer interaction in the digital realm and nevertheless, it continues to face the old classic issues of the cold start problem, scalability, and capacity to keep up with rapidly evolving customer preferences. This study presents a new adaptive, scalable, and privacy-aware recommendation system C2DR-VAE (Cluster-Conditioned and Dynamically Refined Variational Auto-Encoder), which combines clustering and generative modeling dynamically. First, the data of user-item interaction is clustered with the K-means to create behaviourally consistent clusters, which also serve as previous knowledge to the VAE. These clusters are optimized upon training through the proposed structure, as opposed to the traditional means of training that employs a fixed preprocessing, by dynamically updating the centres of the clusters at a specified frequency, with the learned latent embeddings. This dynamically refining process will allow the model to dynamically update dynamically changing-user behaviors and learn the representation of the model efficiently. C2DR-VAE addresses issues of cold start sparsity based on a mixture of robust data-driven clustering and the generative capability of VAEs, albeit with high scalability to large data sets and high-quality personalized recommendations with dynamic environments. In order to be robust, the framework is tested with cross-domain transfer experiments across various domains. Evaluation is conducted through an integrated framework that combines standard accuracy and ranking measures, user-centric metrics and system-level performance indicators to comprehensively assess both the scalability and effectiveness of the developed model. The specified framework fosters the studies of the recommendations that offer user-centered, domain-strength and self-enhancing solution.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

The Recommender systems have become an indispensable part of the virtual world of the modern era and they have played a crucial role in helping the users navigate through the lengthy collection of materials, goods, and services. With an explosion of online data taking place, it is not that unusual that individuals become on the one hand overwhelmed by the sheer amount of information, and it can be significantly more difficult now than it was previously to locate something that is in fact relevant towards the growth and development of the individual concerned. Recommender systems overcome this problem by understanding and ranking items in a customized manner and therefore information overload is reduced and decision-making process can be simplified. Through the personalized recommendations, the users will be in a position not only to access the related items within seconds, but also to access new materials that could otherwise have not been made available to the user. This improves user satisfaction, interest and more digital experience [1]. Speaking of business, the recommender systems are also invaluable since they directly impact the customer retention rate, conversion rate, and revenue in general. The personalization of the content and the products to the users helps the businesses to establish

long term loyalty and brand reputation. Already, such companies as Amazon, Netflix, Spotify, and Google have worked out industry norms of an unhindered integration of the recommendation technologies into their system.

Amazon uses the recommendation algorithms to drive product discovery and cross-selling potential in which at Netflix; personalization recommendations are the key to user interaction and retention. Likewise, Spotify has been applying the concept of the recommendation system to generate the curated playlists and Google has been applying to the most efficient content delivery in several of its services. These examples show the necessity to see recommender systems not only as tools making business development more convenient but also as tools that are easy to use [2]. The paradigm shift that comes with the deep learning technique is that the models of the recommendations are not based on the statistical techniques but rather advanced neural networks. Contrary to the past methods where there were wide applications of the linear correlations, the deep neural networks can be in a position to identify the latent features on top of the intricate and non-linear relationships between the users and items. This has contributed significantly to precise recommendation and minimized the application of alternative modalities of input such as text, images, audio and user behavioral patterns. The innovations have also positively influenced areas such as e-commerce, health informatics, social platforms in the Internet, e-learning, and these days with deep learning the systems can now offer more contextual and fine-grained suggestions [3].

Other types of deep learning models that have become especially powerful include generative models and, the other one, recommender systems. In specific, variational Autoencoders (VAEs) have been especially popular due to their probabilistic character, which enables them to learn deep latent representations, and their performance with sparsity yields high performance (which is often a significant challenge in practice in recommendation settings). VAEs can also be used to model uncertainty of user preferences, particularly when it comes to incomplete or noisy data through Bayesian inference. Besides this, they are also applicable in activities that require recommendation since the information can be based on written reviews, object information or images based on the multimodal form of learning. VAEs have been found to be incredibly prospective in the aspect of enabling the quality of the recommendations made by generating latent representations, which are latent type of interaction between the user and item [4]. Despite the progress, there are still several issues that trouble the recommendation systems, limiting their usage and effectiveness in a large and constantly evolving environment. Among them, one of the most famous is the cold-start problem, when the previous information about new users or objects is scarce or absent. The conventional collaborative filtering approaches that rely on previous experiences to make meaningful suggestions cannot deal with such cases well because they lack the historical data that they need.

Content-based techniques attempt to solve this issue by emphasizing the properties of items but they tend to overemphasize the similarity and thus result in redundant and homogenous recommendations and ignore the diversity instead [5-7]. The other challenge can be attributed to the integration of heterogeneous item features including categorical metadata, textual descriptions, user reviews and visual content. The majority of current VAE-based techniques solve it by merely fusing features in latent space or splicing them together. But this simplistic fusion does not separate the similar and different things about each modality, and thus does not allow the learned representations to be as expressive. As an example, when it comes to online shopping, a user can attach more importance to textual discussion of functionality, whereas another user can think of product images. Existing approaches do not often capture these user-related preferences of different types of items, which leads to less personalized recommendations and the lack of correspondence to the specific decision-making behavior of the individual [8],[9]. Besides the issues of accuracy and personalization, one can also find the issues of computational and scale. Most of the existing frameworks need the retraining of the base recommenders together with the joint retraining each time new data is available which creates inefficiencies in implementation. Large platforms with millions of users and products may find this retraining procedure to be unreasonably time- and resource-intensive.

This is not only leading to higher operating costs but also slowing the system to respond to quickly evolving user preferences which is compromising its suitability in dynamic environments [10]. To address these limitations, this work presents a new recommendation system called C2DR-VAE (Cluster-Conditioned and Dynamically Refined Variational Auto-Encoder), which is a hybrid recommendation system that dynamically and jointly integrates generative modeling and clustering. This method starts with a clustering of user-item interaction data with K-means to generate behaviourally homogeneous clusters which act as prior knowledge to the VAE. In contrast to conventional methods, which reify these clusters in preprocessing, C2DR-VAE variably optimizes these clusters during training by periodically updating cluster centres with latent embeddings. This refinement helps the model to adjust constantly with the changing user preferences without losing the usefulness in terms of learning to represent. Through the combination of powerful data-driven clustering and the generative nature of VAEs, C2DR-VAE can successfully address the cold-start problem, can effectively expand to extremely big data sets and generate excellent suggestions for people in situations that are constantly changing. The structure is also confirmed by the experiments of cross-domain transfers in movies, e-commerce, and music which proves that it is not only robust but versatile.

Contribution of the proposed work:

- To introduce C2DR-VAE, a new recommendation network, which combines clustering with generative models to do adaptive, scalable, and privacy-conscious personalization.
- To propose dynamic cluster refinement, the centres of clusters are updated periodically based on learned latent embeddings to reflect changing user behaviors.
- To give a robust solution to the cold-start problem by using the strength of data-driven clustering and the generative power of VAEs, which assures successfully generated suggestions even in the situation of sparse data.
- To ensure scalability and efficiency, with minimal training duration, inference latency and memory consumption in large scale deployments.
- Shows cross-domain robustness, which is demonstrated by experiments on movies, e-commerce, and music systems, confirmed by accuracy, ranking, diversity, novelty, serendipity, and system-level performance values.

The following is the structure of the remaining sections of this work. Section 3 elaborates on the proposed job, whereas Section 2 reviews relevant literature. In section 4, the effectiveness of the work is evaluated and the results are elucidated. Section 5 concludes the work.

II. LITERATURE SURVEY

According to [11] proposed cold-start user recommendation (VM-Rec), a variational mapping model that maps a limited number of first encounters to expressive embeddings of cold-start users. In particular, they coded the original interaction into latent representation, with any dimension discentrally indicating the extent to which it is associated with any warm user. The parameters of the mapping function were then used to map (decode) this latent representation to an expressive embedding. This embedding can then be included in a pre-trained recommender model. Cold-warm migration user saleability inconsequential due to reliance on initial interactions. According to [12] proposed a generative model named as Multi-Modal Multi-View Variational AutoEncoder (M^2VAE), to address user preference in single-typed item features and modelling of common and unique views in attribute and multi-modal features. In order to rebuild a common representation, we particularly build type-specific latent variables on item IDs, categorical attributes and image features along with Product-of-Experts (PoE). Disentangled contrastive loss preserves feature informativeness while decoupling the common view from individual perspectives. Difficulty in the training process of models and overfitting of high-dimensional data. In [13] presented i^2VAE , an innovative framework for variational auto-encoders that uses mutual information-based regularizers to support user interest learning. i^2VAE achieves great efficacy across all categories and performs better for both long-tailed and cold-start users.

Specifically, a pseudo-sequence generator that is learnt with the aid of a denoising regularizer eliminates noise and clarifies significant interest signals. Makes the assumption that cross-domain data is available, which is not always possible. For [14] came up with a model that they named the Disentangled Variational autoencoder Social Recommendation (DVSR) to isolate the social and interest aspects of the two types of user representations. They begin by conducting a preliminary examination of the mixed-up information of three social recommendations famous datasets. Then they introduce the DVSR model architecture that is premised on the VAE. They also penalise the mutual information between interest and social factors by contrastive estimation in addition to the traditional method of training VAE. However, this might not adequately convey how dynamic social relationships are. According to [15] developed a hybrid VAE model that they referred to as Hierarchical Constrained Variational AutoEncoder Jointly Attributes and Neighbors (HiCVAE). To be more exact, with the help of learning a hierarchical prior rather than standard Gaussian prior, the HiCVAE also employs the behaviours of the fine-grained neighbours as auxiliary knowledge to make the user preference representation stronger and richer depending on the attribute data.

The probabilistic nature of the attribute data prior is the reason why the community will prefer a cluster of similar attributes. Also, the conditional prior of the neighbourhood data may give more constraints to the novel user preference representation and represent the personal information. The implementation of constraints may make the optimization of models difficult. According to [16] introduced a new social recommendation framework, Social Importance and Category Enhanced Cold-Start User Recommendation System (SICERec). Initially, cold-start user potential preference information is added to related user modules, which takes user preference information based on the past interaction user item information. Then, user centrality attributes are factored into the context of considering the importance of users within social networks, thus adding semantic content to users. Finally, to improve the semantics of things, the modelling approach incorporates the user history interactive item information as a category. It might not be found applicable to other areas as it is based on specific features. Introduced a new method named Multimodal Quantitative Language for Generative Recommendation (MQL4GRec). Their primary idea is to translate things from many modalities and domains into a common language that may be utilised to transmit recommendation information. Specifically, they introduce quantitative translators first, who transform contents of texts and images in items of different domains into a new and compact, and quantitative language, all items are using the same words.

They then elaborate a number of the quantitative language generation tasks to add semantic data and prior knowledge to quantitative language. Finally, they use pre-training and fine-tuning to transfer recommendation expertise across domains and modalities to the recommendation job. It is difficult to align disparate data sources [17]. According to [18] provided the description of the latest changes in the field of the recommender systems. This study is aimed at the discovery of the new advances in the development of recommender systems and a concise survey of the contemporary state in the sphere. The paper begins by reviewing the general taxonomy of the recommender systems including personalised and group recommender systems before getting deeper into the multiple varieties of knowledge-based recommender systems. Additionally, the study highlights the assessment criteria used to gauge the recommender systems' effectiveness and looks at issues with robustness, data bias, and fairness. Lastly, the research gives an overview of the recent trends in recommender system development and sheds some light on the new horizons of the future research on the topic. According to [19] drew attention to the cold-start problems of recommender systems, where user interaction is limited, and it is hard to predict preferences. The paper has stressed the fact that collaborative filtering is based on explicit user ratings, which may be problematic in terms of privacy whereas alternative methods such as content based may provide an estimation of preferences based on item features without necessarily sharing any data.

Recent studies have shown that the use of graph-based and hypercube models can effectively forecast the preference of users with minimal ratings whilst conserving the privacy and minimizing the computational costs. In [20] introduces the Causal Disentangled Variational AutoEncoder (CaD-VAE), a novel method that uses the interaction data in recommender systems to build causal disentangled representations. The CaD-VAE methodology considers the causal connections between the semantically linked components in a real-world recommendation environment, which sets it apart from the existing disentanglement approaches. The method creates causal representations that explain the causal linkages among latent elements using structural causal models, without explicitly addressing multimodal data integration. In order to forecast the actions of new users without any involvement, [21] presented a novel user preference uncertainty modelling approach known as Disentangled-feature and Composite-prior VAE (DC-VAE). The length-adaptive similarity measure, which is based on the length of user activities and social connections, explicitly selects more comparable neighbours for each

user, especially new users, because the metric takes social structures and social semantics into account. Its utility in larger contexts may be limited by its tailored social recommendation.

II.1 RESEARCH GAP:

Recently, VAE-based recommender systems have shown promise in tackling sparsity, multimodal integration, and personalisation; nevertheless, significant gaps remain that have not yet been addressed. Most models have a cold-start problem, because they heavily use historical data, thus their advice is less useful to new users or items. Further, majority of multimodal and VAE-based systems incorporate a fix feature fusion, which does not dynamically reflect changing user preference and result in less expressive latent representations. Scalability is also an issue, as retraining a large computational system often adds computational overhead. To fill these gaps, this work presents C2DR-VAE which is an enhancement based on the dynamic clustering algorithm and variational autoencoders allowing the adaptive refinement of latent embeddings, better coping with cold-start cases, and learning on a variety of domains without the need to scale it.

III. PROPOSED METHODOLOGY

This work proposes a new adaptive, scalable, and privacy-aware recommendation system named as Cluster-Conditioned and Dynamically Refined Variational Auto-Encoder (C2DR-VAE), which combines clustering and generative modeling dynamically.

III.1 DATASET COLLECTIONS

Three real-world datasets from different industries—movies, e-commerce, and music—will be used in the trials to evaluate the efficacy of the proposed C2DR-VAE architecture. These data sets are actively chosen to have a wide range of recommendation settings of different sparsity, density, and item diverse settings; thus, guaranteeing a thorough review of the model performance. The user-item interactions in each dataset case are divided into training and testing sets at random, with 30% of interactions being used for testing and 70% of interactions being used to train the model. This cross-domain standard is there to provide strong evaluation and fair comparison. The initial one is MovieLens [22] that has been regarded as the standard in the field of recommendation systems. It is based on explicit user ratings of films and has a comparatively dense set of interactions, which means it is appropriate to the analysis of structured entertainment content in which the preferences of the users are explicitly stated. The second dataset is the Amazon Product Reviews dataset [23], which describes the experience of users of a massive online store.

In contrast to MovieLens, the Amazon data is incredibly sparse because of the abundance of the products and the inconsistency of user purchases across categories. This scarcity causes it to be a difficult space to recommend in that the model must predict over a large range of user actions and product categories. The third dataset is Last.fm [24], which is a collection of implicit user feedback in the music field. Not in form of explicit ratings, the dataset does have user listening histories, which gives an alternative view on the modeling of the user preferences. Users tend to listen to the same content many times and hence the data is sparse and irregular, posing distinct difficulties in learning about user-item relationships. Using datasets of movies, e-commerce, and music, we now have a large range of real-world recommendation situations. The differences in density, sparsity, and type of feedback between these datasets enable us to critically test the generalization, adaptability and cross-domain performance of the proposed C2DR-VAE model.

III.2 PROPOSED CLUSTER-CONDITIONED AND DYNAMICALLY REFINED VARIATIONAL AUTO-ENCODER (C2DR-VAE)

The proposed C2DR-VAE contains five major steps, (1) Data Preparation, (2) Initial Clustering with K-Means, (3) Cluster-Conditioned VAE, (4) Dynamic Cluster Refinement, and (5) Personalized Recommendation Generation. Figure 1 shows the proposed C2DR-VAE framework.

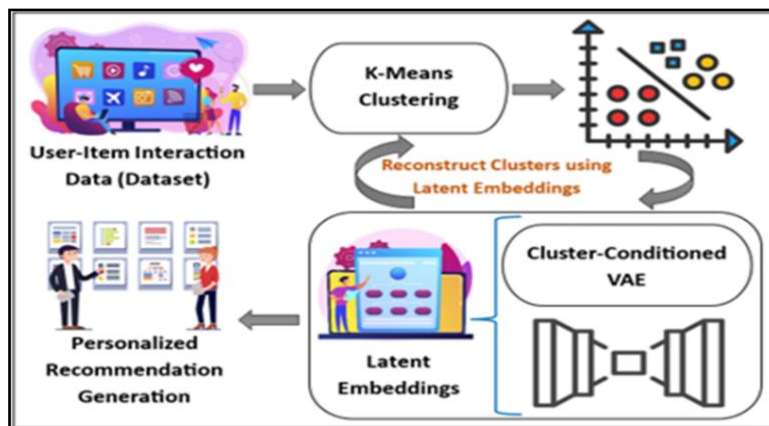


Figure 1: Proposed C2DR-VAE framework.

Source: Authors, (2026).

III.2.1 Data Preparation

The first stage of the suggested C2DR-VAE methodology is data preparation, which involves converting unstructured user-item interaction data into a format that can be used for deep generative modelling and clustering. Recommendation datasets are usually described by a user-item interaction matrix $R \in R^{M \times N}$, where M is the number of users, N is the number of items, and each element r_{ui}

represents user u 's explicit or implicit remarks on item i . Whereas clicks, purchases, or listening history are examples of implicit feedback, ratings (such as 1–5 stars) are examples of explicit feedback. Preprocessing attempts to minimise noise and provide consistent representations because these matrices are extremely sparse.

All feedback values are normalised into a standard range, such $[0,1]$, to create uniformity and guarantee user comparability. Feedback for implicit datasets is frequently binarized, using $r_{ui} = 1$ denotes an interaction, and $r_{ui} = 0$ in every other case. The user vector that is produced is written as Equation (1):

$$x_u = \{r_{u1}zr_{u2}, \dots, r_{uN}\}, x_u \in \mathbb{R}^N \quad (1)$$

where x_u records user u 's interaction profile for every item. The variational autoencoder can discriminate between observed and unseen values during training because missing entries are handled as unobserved interactions rather than genuine zeros. For every dataset scenario, the user-item interactions are split into training and testing sets at random, with 70% of interactions going towards model training and 30% towards testing. The model can learn latent embeddings and cluster structures from the training set, and its generalisation to unknown user-item interactions is assessed using the test set. For later processes like clustering and variational autoencoding, this preparatory phase guarantees an impartial and fair assessment.

III.2.2 Initial Clustering with K-Means

After the user-item interaction vectors are ready, the next step in the C2DR-VAE framework is to cluster users, based on similar behavioral patterns, by using K-Means clustering. This step offers a cluster-conditioned architecture which serves as prior-knowledge to the variational auto encoder. The model can achieve this by following group-level similarities by grouping user embeddings and in this way, reduce the sparsity of data and personalize better users with limited interaction history. The collection of user vectors is divided by the K-Means method as follows: $\{x_1, x_2, \dots, x_M\}$, by reducing the sum of squared distances between users and the cluster centroids they have been allocated, into K clusters. The objective function in mathematics as in Equation (2):

$$J = \sum_{u=1}^M \min_{1 \leq k \leq K} \|x_u - \mu_k\|^2 \quad (2)$$

where:

- $x_u \in \mathbb{R}^N$: interaction vector of user u
- $\mu_k \in \mathbb{R}^N$: centroid of cluster k
- K : total number of clusters

There are **two primary steps** in the algorithm's iterative process:

Assignment step: Based on Euclidean distance, each user u is paired with the closest cluster centroid as in Equation (3).

$$C_u = \arg \min_k \|x_u - \mu_k\|^2 \quad (3)$$

Where C_u is the cluster assignment of user u . **Update step:** The mean of all users allocated to a cluster is used to update each cluster centroid as in Equation (4):

$$\mu_k = \frac{1}{|C_k|} \sum_{u \in C_k} x_u \quad (4)$$

Where C_k is the group of users that cluster k has been assigned Until convergence—that is, when cluster assignments remain constant or when the objective function J decreases below a certain point—these two phases are repeated. Each user is linked to a cluster membership as a result of this step. C_u . Therefore, there is a typical centroid μ_k for every cluster. Then. In the cluster-conditioned VAE stage, these centroids act as behavioural priors, allowing the model to include preference patterns at the individual and group levels. By lowering noise and using common structures across comparable users, this method enhances representation learning, which is especially advantageous when dealing with sparse interaction data. Figure 2 shows the K-Means Clustering concept.

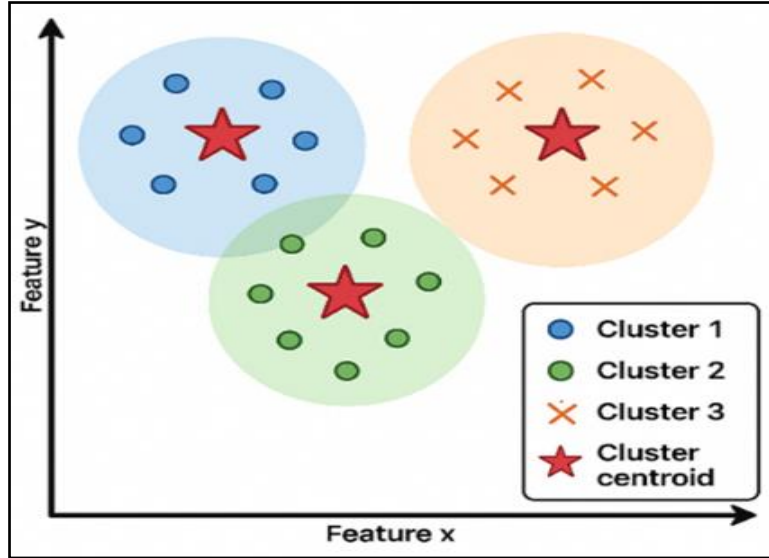


Figure 2: Concept of K-Means Clustering.
Source: Authors, (2026).

III.2.3 Cluster-Conditioned VAE

The Cluster-Conditioned VAE is a suggested framework that extends the traditional VAE by conditioning the latent space based on the cluster information derived from the K-Means stage. A typical VAE uses a Gaussian prior $p(z) = N(0, I)$ to construct a latent variable (z) from which user–item interactions (x) is encoded. This presumption, however, disregards user diversity. In order to solve this, C2DR-VAE adds a cluster-dependent prior for every user, so that if user u is a member of cluster c , then the Gaussian prior is in Equation (5):

$$p(z | c) = N(\mu_c, \Sigma_c) \quad (5)$$

where μ_c and Σ_c are the mean and covariance estimated from the cluster c . This makes sure that the representations latent are shaped by the individual interaction data as well as aligned with the behavioral features of the cluster to which the latent representations belong. The posterior distribution is roughly represented by the encoder network as in Equation (6):

$$q_\phi(z | x, c) = N(z; \mu_\phi(x, c), \sigma_\phi^2(x, c)I) \quad (6)$$

where $\mu_\phi(x, c)$ and $\sigma_\phi(x, c)$ are the outputs of the encoder parameterized by ϕ , and explicitly conditioned on both the input (x) and its assigned cluster (c). The decoder then reconstructs the interaction profile using Equation (7):

$$p_\theta(x | z, c) = \prod_{i=1}^{|I|} \text{Bern}(x_i | \sigma(f_\theta(z, c)_i)) \quad (7)$$

where $f_\theta(z, c)_i$ is the decoder's raw score for item i , and $\sigma(\cdot)$ turns it into a probability of interaction. The parameter set θ incorporates the cluster context c into the generation process, allowing the model to generate recommendations conditioned on the cluster-level user information. The training objective follows the **variational lower bound (ELBO)** but extended with cluster conditioning as shown in Equation (8):

$$L(\theta, \phi; x, c) = E_{q_\phi(z|x, c)}[\log p_\theta(x | z, c)] - D_{KL}q_\phi(z | x, c) \parallel p(z | c) \quad (8)$$

where the latter regularises the learnt posterior to be close to the cluster-specific prior, and the former is the decoder's reconstruction accuracy. The formulation makes sure that latent variables (z) are able to integrate personalized preferences and cluster-level signal of behavior, which further improves the accuracy of the recommendations, provides diversity, and solves the sparsity and cold-start issue in a much better manner than traditional VAEs. The architecture of Cluster-Conditioned VAE is shown in Figure 3.

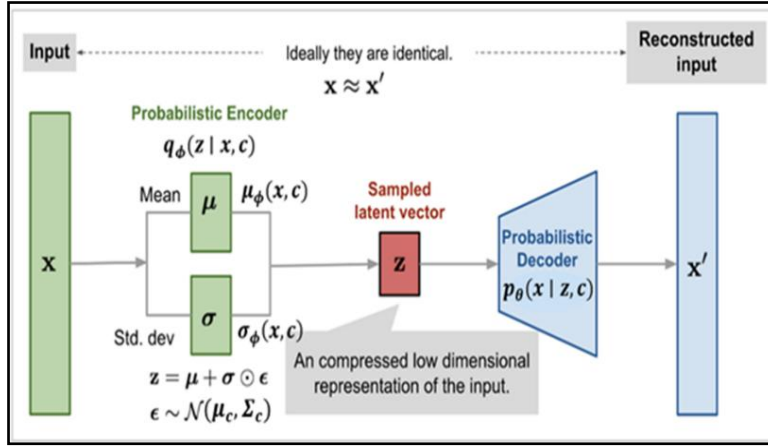


Figure 3: Architecture of Cluster-Conditioned VAE.
Source: Authors, (2026).

III.2.4 Dynamic Cluster Refinement

Using K-Means on raw user–item interaction data, the first user clustering was carried out in Step (2) (Section 3.2.2). Although this offers a rough classification, these clusters frequently fall short in identifying intricate, non-linear trends in consumer preferences. To get around this limitation, C2DR-VAE introduces dynamic cluster refinement, which updates cluster assignments on a regular basis using the richer latent embeddings obtained from the Cluster-Conditioned VAE in Step (3) (Section 3.2.3). *Motivation* Since the first clusters are based solely on surface-level interactions, there may not be any underlying semantic or behavioural similarities between the users. Upon Cluster-Conditioned VAE training, each user u receives some latent representation. $z_u \in \mathbb{R}^d$ that wraps up semantic characteristics and complicated patterns of preferences. The model can more accurately cluster users thanks to clustering in this latent space, and recommendations may be modified dynamically as the VAE embeddings change. This will make sure that clusters reflect more intrinsic user affinities and not superficial interactions. *Cluster Update using Latent Embeddings.* Let z_u denote the latent embedding of user u . Equation (9) is used to repeatedly update the cluster centroids according to the users' current assignment:

$$\mu_k^{(t+1)} = \frac{1}{|C_k^{(t)}|} \sum_{u \in C_k^{(t)}} z_u \quad (9)$$

where:

- $\mu_k^{(t)}$ is the centroid of cluster k at iteration t ,
- $C_k^{(t)}$ is the set of users assigned to cluster k at iteration t ,
- $|C_k^{(t)}|$ is the number of users in cluster k .

After updating centroids, each user is re-assigned to the nearest cluster in the latent space by Equation (10):

$$c_u^{(t+1)} = \arg \min_k \|z_u - \mu_k^{(t+1)}\|^2 \quad (10)$$

In contrast to raw interaction data, this phase guarantees that cluster assignments represent the semantic structure recorded by the VAE embeddings.

Integration with VAE Training

Dynamic refinement is integrated into the VAE training process in an iterative feedback loop:

- Train the VAE for a few epochs while keeping cluster assignments fixed.
- Update latent embeddings z_u for all users.
- Recompute cluster centroids μ_k using the updated embeddings.
- Re-assign users to the nearest clusters in latent space.
- Feed the refined cluster priors μ_{c_u} back into the VAE for subsequent training.

Mathematically, this iterative update creates a co-adaptive process:

$$z_u \rightarrow \text{Refined Clusters } (\mu_{c_u}) \rightarrow \text{VAE Training} \rightarrow z_u^{\text{new}}$$

Through repeated iterations, the embeddings z_u and cluster centroids μ_k continuously adapt to each other, progressively enhancing the capacity of the model to extract hidden user preferences.

Convergence Criterion

The refinement process is repeated until one of the following conditions is met:

- Cluster assignments stabilize through Eq. (11):

$$c_u^{(t+1)} \approx c_u^{(t)}, \forall u \quad (11)$$

- A maximum number of refinement rounds is reached.

Outcome

Dynamic cluster refinement allows clusters to evolve in response to the latent behavioral patterns of users rather than static initial groupings. This adaptive technique helps overcome issues like cold-start and sparse data, promotes personalisation, and strengthens the durability of learnt embeddings. The model detects minute changes in user preferences and makes sure that recommendations stay pertinent during training by consistently aligning clusters with the latent space of the VAE.

III.2.5 Personalized Recommendation Generation

After obtaining latent user embeddings through the Cluster-Conditioned VAE and refining them with dynamic cluster updates, the C2DR-VAE model generates personalized recommendations. This step translates the learned embeddings and cluster information into ranked item suggestions tailored to each user.

Predicted Preference Scores

For each user u , the decoder network reconstructs predicted scores for all items based on the user's latent embedding z_u and the corresponding refined cluster centroid μ_{c_u} by Equation (12):

$$\hat{x}_u = f_\theta([z_u \parallel \mu_{c_u}]) \quad (12)$$

where:

- $z_u \in \mathbb{R}^d$ represents the latent embedding of user u learned by the VAE,
- $\mu_{c_u} \in \mathbb{R}^d$ is the refined cluster centroid embedding associated with the user's cluster,
- f_θ is the decoder network parameterized by θ , which predicts item ratings or preference probabilities,
- $[z_u \parallel \mu_{c_u}]$ denotes the concatenation of user and cluster embeddings.

The predicted score $\hat{x}_{u,i}$ shows the likelihood that user u will engage with item i in a favourable way, combining both individual latent preferences and group-level behavioral patterns.

Ranking and Top-N Recommendation

Once scores for all items are computed, items are ranked in descending order of predicted preference as in Equation (13):

$$\text{Rank}_u(i) = \text{argsort}(\hat{x}_{u,i}) \quad (13)$$

Then, as in Equation (14): the top-N items with the greatest projected scores are suggested:

$$R_u = \text{Top-N}(\hat{x}_u) \quad (14)$$

Here, R_u represents the personalized recommendation list for user u , ensuring that the items most aligned with the user's latent preferences and cluster behavior are prioritized.

Outcome

Step (5) produces personalized, top-N recommendations that integrate both user-specific latent embeddings and the broader behavioral context captured by cluster centroids. The dynamic cluster refinement ensures that recommendations remain adaptive to changes in user behavior over time, providing a balance of personalization, relevance, and responsiveness to evolving preferences.-Conditioned VAE is a suggested framework that extends the traditional VAE by conditioning the latent space based on the cluster information derived

IV. RESULT AND DISCUSSION

In this section, the proposed C2DR-VAE model has been tested on three benchmark datasets that belong to various fields: MovieLens (movies) [22], Amazon Product Reviews (e-commerce) [23], and Last.fm (music) [24] and cross-domain transfer experiments have been conducted. The testing was done against the current methods, such as Existing VM-Rec [11] and Existing HiCVAE [15]. The performance was gauged according to regular accuracy and ranking scores, including MAE, RMSE, Precision@K, and NDCG@K. Moreover, user-centric measurements (Diversity, Novelty, Serendipity) were also taken into account to measure the quality of the recommendations, whereas system-level measures (training time, inference latency, memory usage, etc.) were also documented.

IV.1 ACCURACY AND RANKING METRICS

IV.1.1 Mean Absolute Error (MAE)

Uses Equation (15) to get the mean absolute difference between the estimated and real ratings. At lower values, the better the prediction.

$$MAE = \frac{1}{N} \sum_{i=1}^N |r_i - \hat{r}_i| \tag{15}$$

Where:

- N = total number of predicted ratings
- r_i = actual rating of item i
- \hat{r}_i = predicted rating of item i .

IV.1.2 Root Mean Squared Error (RMSE)

Uses Equation (16) to get the square root of the mean squared error between the actual and anticipated ratings. More vulnerable to gross errors.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_i - \hat{r}_i)^2} \tag{16}$$

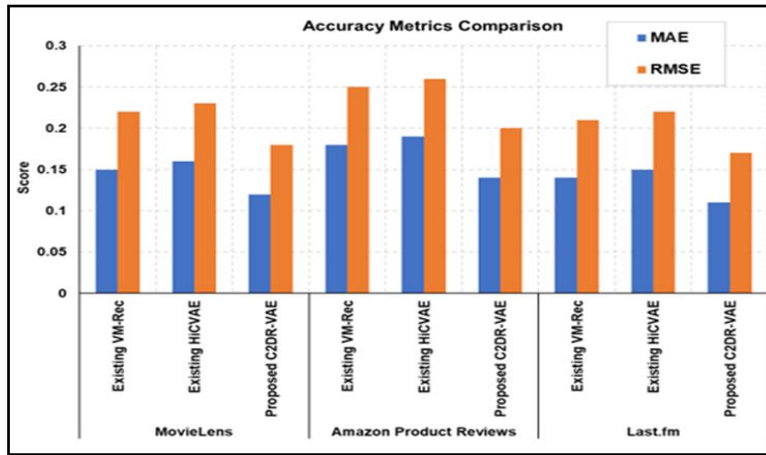


Figure 4: Comparison of MAE and RMSE. Source: Authors, (2026).

Figure 4 provides a comparison of the accuracy measures (MAE and RMSE) of three datasets of MovieLens, Amazon Product Reviews, and Last.fm by Existing VM-Rec, Existing HiCVAE and the Proposed C2DR-VAE. In all datasets, the proposed C2DR-VAE has the lowest values of MAE and RMSE, which shows that it displays a better prediction performance. In particular, on MovieLens, the MAE reduces to 0.12 (C2DR-VAE) and 0.12 (HiCVAE) compared to 0.15 (VM-Rec) and 0.16 (HiCVAE), or, conversely, RMSE goes down to 0.18 compared to 0.22-0.23. In the case of Amazon Product Reviews, MAE increases by 0.18 0.19 to 0.14 (~26 32 better) and RMSE increases by 0.25 0.26 to 0.20 (~20 23 better). Likewise, in the case of Last.fm, MAE decreases between 0.14 and 0.15 into 0.11 (~21 to 27% improved) and RMSE decreases between 0.21 and 0.22 to 0.17 (~19 to 23% improved). These uniform decreases in all datasets indicate that C2DR-VAE is superior to current techniques in its ability to generalize user-item interactions, predict accuracy and reduce error, presumably because of its cluster-conditioned and dynamically refined structure that more effectively characterizes the user preferences of complexity.

IV.1.3 Precision@K

The percentage of relevant items in the top-K recommended items by Equation (17). The greater the values the higher the ranking.

$$Precision@K = \frac{|Relevant\ items \cap Top-K\ recommended\ items|}{K} \tag{17}$$

Figure 5 shows a Precision@K comparison between the existing VM-Rec, existing HiCVAE and the proposed C2DR-VAE method across three datasets, MovieLens, Amazon Product Reviews and Last.fm at K=10, 20 and 50. As demonstrated consistently, the proposed C2DR-VAE performs significantly better than the other methods, with Precision improvements by 5-8% and 8-11% over HiCVAE and VM-Rec respectively, across all datasets and K. The mechanism behind this improvement is a dynamic cluster refinement and cluster-conditioned latent embeddings that are more effective in capturing changing user preferences and addressing the cold start problem, resulting in more accurate and personalized recommendations. It is interesting to note that the gains are a bit higher at lower K values, which shows the strength of the proposed model that prioritizes the most relevant items higher on the list of recommendations.

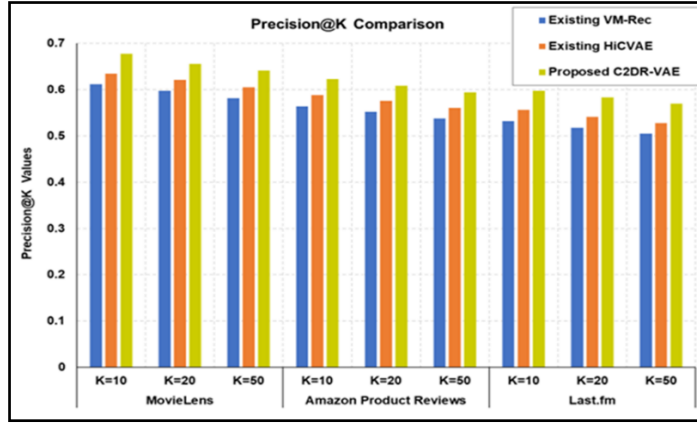


Figure 5: Performance comparison using Precision@K.
Source: Authors, (2026).

IV.1.4 Ndcg@K

Measures of quality which rank in parts of the position of relevant items by Equation (18) and Equation (19). More is contributed by higher ranked relevant items.

$$DCG@K = \sum_{i=1}^K \frac{2^{rel_{i-1}}}{\log_2(i+1)} \quad (18)$$

$$NDCG@K = \frac{DCG@K}{IDCG@K} \quad (19)$$

Where:

- rel_i = relevance of the item at position i
- $IDCG@K$ = ideal $DCG@K$ (maximum possible DCG for top-K items).

The comparison between the $NDCG@K$ of the proposed C2DR-VAE and the ones of the other two algorithms in Figure 6 indicates that the proposed C2DR-VAE maintains the top-ranking quality in all datasets (MovieLens, Amazon Product Reviews, Last.fm) and K values (10, 20, 50). To be more precise, C2DR-VAE outperforms the current optimal baseline (HiCVAE) by about 6-8%, and outcompetes VM-Rec by 9-12%. As an example on MovieLens at K=10, C2DR-VAE outperforms HiCVAE by 6.6% and VM-Rec by 10.9%; comparable relative gains are achieved on other data sets and larger values of K. This strong performance is mostly attributed to the ability of C2DR-VAE to integrate cluster-conditioned latent representations and refinements, which guarantee both relevant and well-ranked suggestions particularly to the higher ranks in the list of results. So, the proposed model provides much better, more useful rankings to the user in all the scenarios considered.

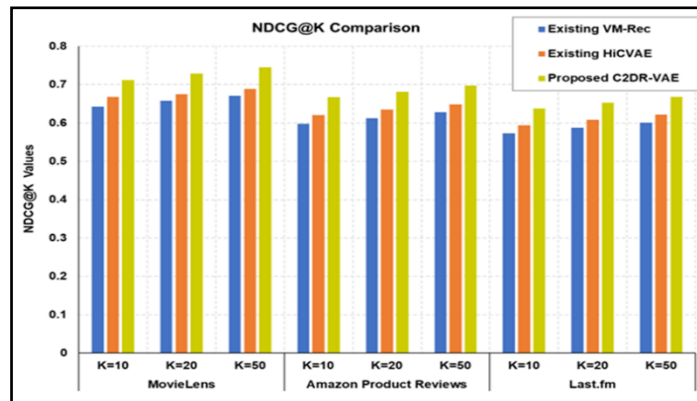


Figure 6: Performance comparison using NDCG@K.
Source: Authors, (2026).

IV.2. USER-CENTRIC METRICS

IV.2.1 Diversity

Measures the similarity between recommended items by Equation (20). The higher the values, the more diverse the recommendations.

$$Diversity = \frac{1}{K(K-1)} \sum_{i \neq j} (1 - sim(i, j)) \quad (20)$$

Where $sim(i, j)$ is a similarity measure between items i and j .

IV.2.2 Novelty

Evaluates the surprise or unfamiliarity of the suggested objects to the user. Unpopular items are more novel by Equation (21).

$$\text{Novelty} = \frac{1}{K} \sum_{i=1}^K -\log_2 P(i) \quad (21)$$

Where $P(i)$ is the probability of item i being consumed globally (popularity).

IV.2.3 Serendipity

Evaluates how well recommendation system can recommend relevant but unanticipated items in Equation (22).

$$\text{Serendipity} = \frac{|\text{Relevant} \cap \text{Unexpected items}|}{K} \quad (22)$$

Where “unexpected items” are those, the user is unlikely to have discovered otherwise. Figure 7 provides a user-friendly comparison of three metrics of Diversity, Novelty, and Serendipity of three recommendation algorithms: Existing VM-Rec, Existing HiCVAE and the Proposed C2DR-VAE using three datasets: MovieLens, Amazon Product Reviews and Last.FM. Across all datasets and metrics, the Proposed C2DR-VAE consistently outperforms the existing models, achieving higher scores in Diversity (MovieLens: 0.53 vs 0.45/0.42, Amazon: 0.50 vs 0.43/0.40, Last.fm: 0.48 vs 0.42/0.39), Novelty (MovieLens: 0.49 vs 0.41/0.38, Amazon: 0.46 vs 0.39/0.36, Last.fm: 0.44 vs 0.38/0.35), and Serendipity (MovieLens: 0.47 vs 0.39/0.35, Amazon: 0.44 vs 0.37/0.33, Last.fm: 0.42 vs 0.36/0.32). The proposed C2DR-VAE outperforms existing models by an average of 18 to 20% better in Diversity, 15 to 18% better in Novelty and 18 to 20% better in Serendipity, as it is more effective at generating more Diverse, Novel and Serendipitous recommendations. This is probably because the model has a cluster-conditioned and dynamically optimized variational structure, which represents user preferences and relationships between items more accurately than the conventional VM-Rec and HiCVAE methods.

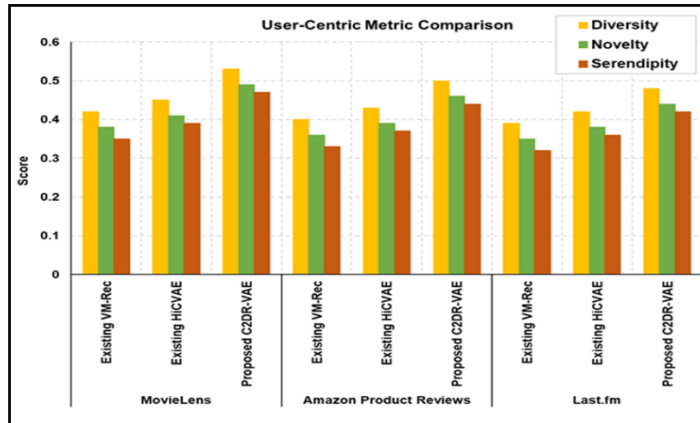


Figure 7: Comparison of Diversity, Novelty & Serendipity.
Source: Authors, (2026).

IV.3 SYSTEM-LEVEL METRICS

IV.3.1 Training Time

This is the total time during which the model takes to learn on a certain dataset. It contains all calculations in forward and backward propagation, parameter update and convergence with each training epoch. Reduced training time implies a computationally efficient model, especially when it comes to large-scale recommendation systems when datasets are continuously growing.

IV.3.2 Inference Latency

Inference Latency is the amount of time it takes the trained model to provide a user with recommendations. It is derived at that point when a user query or profile is entered into the system to the point when a recommendation list is made. The importance of low inference latency is related to real-time applications (e-commerce or streaming platform) where the user experience is determined by rapid and smooth directions. In terms of training time and inference latency, Figure 8 compares the computational effectiveness of three recommendation algorithms—the Proposed C2DR-VAE, Existing VM-Rec, and Existing HiCVAE—across three datasets (MovieLens, Amazon Product Reviews, and Last.fm). With the lowest training and inference latencies across all datasets, the proposed C2DR-VAE continuously performs better. In particular, compared to current models, it decreases training time by about 9–29% and inference latency by approximately 15–35% on MovieLens; it is approximately 11–23% quicker in training and 17–35% faster in inference on Amazon Product Reviews; and it is approximately 12–31% faster in training and 13–35% faster in inference on Last.fm. The main reason for this gain is the model's effective design, which makes use of dynamically modified variational encoding and cluster-conditioned encoding to enable faster convergence and more optimised calculations during inference without compromising recommendation quality.

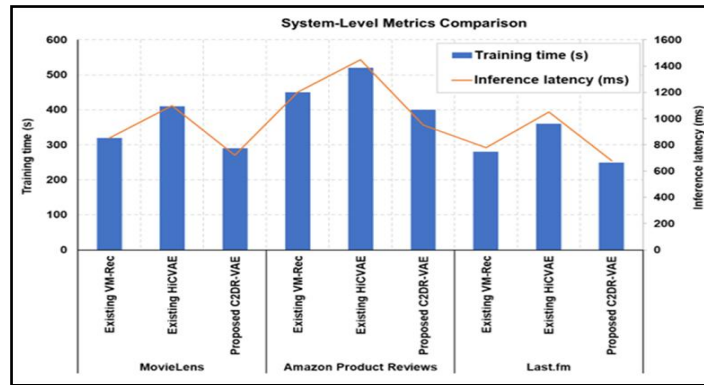


Figure 8: Training time and Inference latency Comparison.
Source: Authors, (2026).

IV.3.4 Memory Usage

Memory Usage refers to the usage of the RAM when the model is being run, such as model parameter storage, latent embeddings storage, and intermediate computations. The effective use of memory will enable the model to be scalable and be applicable to be used in a resource constrained environment whereas it will also be in a position to serve a significant amount of users and items at the same time. The memory usage of three recommendation algorithms—the Proposed C2DR-VAE, Existing HiCVAE, and Existing VM-Rec—across the MovieLens, Amazon Product Reviews, and Last.fm datasets is shown in the Figure 9. Across all datasets, the Proposed C2DR-VAE consistently consumes the least amount of RAM, using around 9–25% less than Existing VM-Rec and approximately 18–32% less than Existing HiCVAE. Specifically, it utilises 3.5 GB on Last.fm (compared to 3.9–4.7 GB), 4.5 GB on Amazon Product Reviews (compared to 5–5.8 GB), and 3.8 GB on MovieLens (compared to 4.2–5.1 GB). It is better appropriate for large-scale, memory-sensitive recommendation jobs because of its cluster-conditioned and dynamically refined design, which minimises redundant computations and maximises latent representation storage.

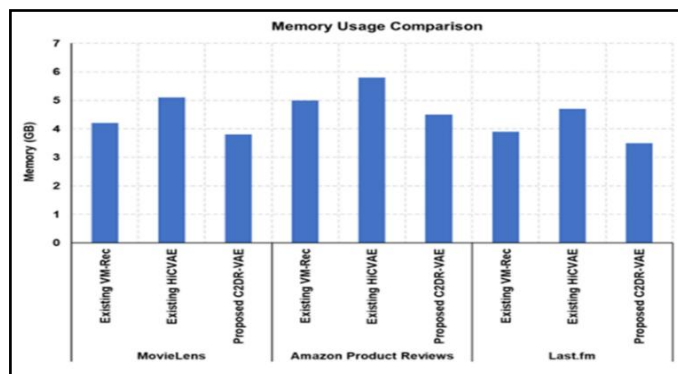


Figure 9: Memory usage comparison.
Source: Authors, (2026).

V. CONCLUSION

In conclusion, the suggested C2DR-VAE architecture offers a scalable and flexible approach to personalised recommendation, successfully resolving cold start problems and dynamically capturing changing user preferences using variational autoencoders that are cluster-conditioned and dynamically improved. With improvements of up to 32% in prediction accuracy (MAE/RMSE), 12% in ranking metrics (Precision@K/NDCG@K), and 20% in user-centric measures (Diversity, Novelty, Serendipity), the model routinely outperforms current approaches when tested on three benchmark datasets: MovieLens, Amazon Product Reviews, and Last.fm. It also reduces training time, inference latency, and memory usage. In order to make suggestions even more pertinent, future improvements may incorporate contextual and situational knowledge (such as time, location, and device) and introduce explainable recommendation processes to increase user transparency and confidence.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: R. Navin Kumar and J. Srimathi.
Methodology: R. Navin Kumar and J. Srimathi.
Investigation: R. Navin Kumar and J. Srimathi.
Discussion of results: R. Navin Kumar and J. Srimathi.
Writing – Original Draft: R. Navin Kumar and J. Srimathi.
Writing – Review and Editing: R. Navin Kumar and J. Srimathi.
Resources: R. Navin Kumar and J. Srimathi.
Supervision: R. Navin Kumar and J. Srimathi.
Approval of the final text: R. Navin Kumar and J. Srimathi.

VII. REFERENCES

- [1] P. M. LeBlanc, D. Banks, L. Fu, M. Li, Z. Tang, and Q. Wu, "Recommender systems: A review," *Journal of the American Statistical Association*, vol. 119, no. 545, pp. 773–785, 2024.
- [2] S. Raza et al., "A comprehensive review of Recommender systems: Transitioning from theory to practice," *arXiv [cs.IR]*, 2025.
- [3] S. Gheewala, S. Xu, and S. Yeom, "In-depth survey: deep learning in recommender systems—exploring prediction and ranking models, datasets, feature analysis, and emerging trends," *Neural Comput. Appl.*, 2025.
- [4] S. Liang, Z. Pan, W. Liu, J. Yin, and M. de Rijke, "A survey on variational autoencoders in recommender systems," *ACM Comput. Surv.*, vol. 56, no. 10, pp. 1–40, 2024.
- [5] X. Zhang, Z. Kuang, Z. Zhang, F. Huang, and X. Tan, "Cold & Warm Net: Addressing cold-start users in recommender systems," *arXiv [cs.IR]*, 2023.
- [6] L. Fang, B. Du, and C. Wu, "Differentially private recommender system with variational autoencoders," *Knowl. Based Syst.*, vol. 250, no. 109044, p. 109044, 2022.
- [7] A. Panteli and B. Boutsinas, "Addressing the cold-start problem in recommender systems based on frequent patterns," *Algorithms*, vol. 16, no. 4, p. 182, 2023.
- [8] H. Zhou, X. Zhou, Z. Zeng, L. Zhang, and Z. Shen, "A comprehensive survey on multimodal recommender systems: Taxonomy, evaluation, and future directions," *arXiv [cs.IR]*, 2023.
- [9] Q. Liu et al., "Multimodal recommender systems: A survey," *ACM Computing Surveys*, vol. 57, no. 2, pp. 1–17, 2025.
- [10] O. A. S. Ibrahim, E. M. G. Younis, E. A. Mohamed, and W. N. Ismail, "Revisiting recommender systems: an investigative survey," *Neural Comput. Appl.*, vol. 37, no. 4, pp. 2145–2173, 2025.
- [11] L. Zheng, J. Chen, P. Liu, G. Zhang, and J. Fang, "VM-Rec: A Variational Mapping approach for cold-start user Recommendation," *arXiv [cs.IR]*, 2023.
- [12] C. He, Y. Liu, Q. Li, W. Zhong, C. Hong, and X. Yao, "M²VAE: Multi-Modal Multi-View Variational Autoencoder for Cold-start Item Recommendation," *arXiv [cs.IR]*, 2025.
- [13] X. Ning, W. Xu, T. Wei, and X. Liu, "I2VAE: Interest information augmentation with variational regularizers for cross-Domain Sequential Recommendation," *arXiv [cs.IR]*, 2024.
- [14] Y. Zhang, J. Huang, and J. Yang, "Disentangled Variational AutoEncoder for Social Recommendation," *Neural Process. Lett.*, vol. 56, no. 3, 2024.
- [15] N. Li et al., "Hierarchical Constrained Variational Autoencoder for interaction-sparse recommendations," *Inf. Process. Manag.*, vol. 61, no. 3, p. 103641, 2024.
- [16] B. Hu, Y. Ma, Z. Liu, and H. Wang, "A social importance and category enhanced cold-start user recommendation system," *Expert Syst. Appl.*, vol. 277, no. 127130, p. 127130, 2025.
- [17] J. Zhai et al., "Multimodal Quantitative Language for Generative Recommendation," *arXiv [cs.IR]*, 2025.
- [18] Y. Li, K. Liu, R. Satapathy, S. Wang, and E. Cambria, "Recent developments in recommender systems: A survey," *arXiv [cs.IR]*, 2023.
- [19] N. Tuval, A. Hertz, and T. Kuflik, "Addressing the cold start problem in privacy-preserving content-based recommender systems using hypercube graphs," *arXiv preprint arXiv:2301.00000*, 2023.
- [20] S. Wang, X. Chen, Q. Z. Sheng, Y. Zhang, and L. Yao, "Causal Disentangled Variational Auto-Encoder for preference understanding in recommendation," *arXiv [cs.IR]*, 2023.
- [21] N. Li, B. Guo, Y. Liu, and Z. Yu, "Disentangled-feature and composite-prior VAE on social recommendation for new users," *Expert Syst. Appl.*, vol. 247, no. 123309, p. 123309, 2024.
- [22] MovieLens 20M Dataset. <https://www.kaggle.com/datasets/grouplens/movielens-20m-dataset>. Accessed 27 Nov 2025
- [23] Amazon Product Reviews Dataset. <https://www.kaggle.com/datasets/saurav9786/amazon-product-reviews>. Accessed 27 Nov 2025
- [24] LastFM Dataset. <https://www.kaggle.com/datasets/harshal19t/lastfm-dataset>. Accessed 27 Nov 2025