IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 27, Issue 6, Ser. 4 (Nov. – Dec. 2025), PP 64-73
www.iosrjournals.org

Unified Benchmarking Of Machine Learning And Deep Learning Based Collaborative Filtering Techniques Across Multi-Domain Data

Akshay P. Deshpande, Bharat R. Naiknaware, Vishnu N. Dabhade
Dr. G.Y. Pathrikar College Of Computer Science & Information Technology, MGM University, Chhatrapati Sambhajinagar, Maharashtra, India

Abstract:
This research presents a cross-domain evaluation framework for collaborative filtering in e-commerce, movies, and books, integrating data preprocessing, synthetic interaction modelling, feature engineering, and multi-metric evaluation. The framework integrates regression and classification metrics and allows for threshold recommendation modelling to enhance performance. The results indicate that Neural Collaborative Filtering always exhibits better accuracy, generalization, and interpretability in underlying spaces, especially in sparse environments. Matrix factorization techniques, especially SVD and ALS, perform well in dense environments, while KNN performs consistently, if not somewhat conservatively, in the evaluation; and in terms the efficiency of data interaction, weak sparse environments yield poor performance. In the benchmark evaluation, SVD and KNN are both superior in the e-commerce domain, KNN in the movies, while in the books domain, NCF stands out. The study has created a cross-domain evaluative system that can be reproduced, and for the first time, the potential to gain an interpretable latent representation in a complex system of interactions for user and item relationships. These results form the basis for scalable, multi-domain recommendation systems, for which the only aspect remaining are hybrid deep learning models, systems with transparency, and explainability around recommendations.
Keywords: *Collaborative Filtering; Matrix Factorization, Neural Collaborative Filtering (NCF); E-Commerce Recommendation; Machine Learning; Data Sparsity.*

Date of Submission: 18-12-2025 Date of Acceptance: 28-12-2025

I. Introduction

Recommendation systems give information used by predicting user choices from historical interactions; collaborative filtering is its dominant family of methods for such tasks. The transformation of shopping continues to evolve with e-commerce platforms due to the variety of products and services available. The shopping experience is filtered and expanded with the recommendation system allowing users to see products that match their interest and increasing sales for the platform.

Among the many disciplines like recommendation system, taking agriculture for example, is the use of collaborative filtering to recommend the best crops to plant, the most effective fertilizers to use, and overall, the best decisions for precision farming by analysing the farmer and the environment. In predicting elections, collaborative filtering is applied to predict the votes, perform sentiment and analyse the political content structured as delivery. These are some major applications which highlight the use of collaborative filtering and it strengthens the motivation to evaluate collaborative filtering algorithms among multiple domains.

Research work involves addressing practical problems related to e-commerce services, including selecting an approach that balances predictive accuracy and robustness to sparsity/cold start. For the rating and ranking purposes, the contribution includes a complete literature-driven comparison of traditional machine learning classifiers, SVD, and ALS. An easily understandable and consistent methodology that evaluates the robustness within the creation of synthetic datasets and practical suggestions for the selection of algorithms and execution decisions according to the empirical and algorithmic evidence of knowledge finding. Mobile phones are a highly considered product where attributes like RAM, storage, battery capacity, and pricing strongly influence the impact of buying decisions, which are the specific domain of the study. To start with the unprocessed or unstructured multiple data sources, we proposed and analysed an extensive method to produce and develop a recommendation system. The paper mainly combines a traditional approach to collect the data that brings together and summarizes the various e-commerce information or data. A technique to create a data set of virtual user-item interactions that matches the typical customer behaviour. And also a comparison of several machine learning classifiers to predict if a user is liked or disliked. Two commonly used matrix factorisation methods for rate prediction are put into action and examined. The present study includes a Collaborative -based approach to predict if a weather user will like a product for example a mobile phone.

History of the Research

This section covers the technical development of collaborative filtering, matrix factorisation techniques, different evaluation matrices, traditional machine learning techniques used for recommendation systems, matrices, as well as current development and ongoing issues. In the development of collaborative filtering and taxonomy, the literature survey gives an overview of collaborative filtering types, the advantages and commonly encountered problems, including scarcity, cold start scalability and adversarial or shilling attacks. From memory-based similarity methods, collaborative filtering has developed into model-based and hybrid systems. In different competitions and real-world implementations, matrix factorisation collaborative filtering has gained popularity. [1-3]

The user-item-interaction method is calculated by SVD and related low-rank classifications using structural factors, Pure SVD, and SVD++ significant variations for neighbourhood integration and explicit ratings [11–12]. ALS is a fundamental technique for large-scale implicit-explicit factorisation, where if hidden variables are approached by alternating closed-form least squares, ALS iterates over the minimisation target function. How overfitting is dealt with and implicit weighting are managed with more ALS variants and the standardised ALS-WR7 [12–13]. There are improvements to runtime and scalability to parallelisation and Solvers. Coordinate descent methods outperforms ALS in a handful of sparsely populated cases [5]. Within e-commerce, the use of collaborative filtering for seamless personalization is becoming ubiquitous in numerous applied domains. Similar to this in- the personalization of treatments, patient filtering also suggests adaptive intelligence in healthcare.

When the situation is enriched with extra information, and the difficulty is to formulate it as either ranking or classification as opposed to basic rating prediction the technique of classification, which is of the traditional machine learning kind, it is the supervised classifiers such as decision trees, random forests, support vector machines, and extreme boosts, which is applied to the engineers features of the set, namely user/item features, session/context features, or aggregated interaction signals.

The below table gives us the technical advancement in collaborative filtering.

Table no 1 Literature based collaborative filtering.

Author(s) & Year	Algorithm / Method	Dataset / Domain	Key Contributions	Limitations / Gaps
[7]Goldberg et al. (1992)	Early Memory-Based Collaborative Filtering	Tapestry Email System	Introduced concept of CF using user opinions and message filtering.	Manual filtering; lacked scalability and automation.
[8]Resnick et al. (1994)	GroupLens (User–User CF)	Netnews Articles	Implemented automatic CF using user similarity and shared ratings.	Limited to small datasets; scalability issues.
[9]Herlocker et al. (2002)	User–User and Item–Item KNN	MovieLens Dataset	Compared similarity metrics (cosine, Pearson) for CF; introduced evaluation standards.	Sensitive to data sparsity and cold start problems.
[10]Sarwar et al. (2001)	Item-Based Collaborative Filtering	MovieLens	Improved scalability and recommendation speed using item similarities.	Lower accuracy on sparse data; limited contextual personalization.
[11]Koren et al. (2009)	Matrix Factorization (SVD)	Netflix Prize Dataset	Introduced latent factor models for CF; achieved breakthrough accuracy.	Struggles with dynamic user preference changes.
[12]Zhou et al. (2008)	Alternating Least Squares (ALS)	Netflix Prize Dataset	Enhanced scalability using parallelized matrix factorization.	Computationally heavy; dependent on hyperparameter tuning.
[13]He et al. (2017)	Neural Collaborative Filtering (NCF)	MovieLens & Pinterest	Combined neural embeddings and MF for non-linear interactions; achieved state-of-the-art accuracy.	Requires large data; higher training cost; limited interpretability.
[14]Harper & Konstan (2016)	GroupLens Dataset Work	MovieLens 1M, 10M, 20M	Provided benchmark datasets for CF research.	No dynamic data updates; limited to explicit feedback.
[15]Ziegler et al. (2005)	Topic Diversification in CF	Book-Crossing Dataset	Proposed diversification in recommendation lists using semantic similarity.	Focused only on books; did not test hybrid CF.
[16]Ricci et al. (2015)	Hybrid Recommendation Models	Multiple Domains	Summarized hybrid CF and content-based techniques in real-world systems.	Theoretical synthesis; lacked empirical comparison.
[17]Su & Khoshgoftaar (2009)	CF Techniques Review	Multiple Datasets	Provided taxonomy of CF algorithms and similarity metrics.	Did not include neural or deep models.

[18]Zhang et al. (2019)	Deep Matrix Factorization	MovieLens, Amazon Reviews	Enhanced MF using multi-layer neural networks for deeper representations.	Overfitting risk; high computational cost.

II. Methodology

This section includes a reproducible pipeline to compare traditional machine learning classifiers and SVD and ALS on our E-Commerce data set as well as synthetic data generation, preprocessing of the data, feature engineering of data, the implementation process and the evaluation framework.

Algorithm	Technique	Library Used	Description
SVD	Matrix Factorization	surprise	Decomposes user–item matrix into latent features.
ALS	Alternating Optimization	implicit	Minimizes squared error via iterative least squares.
KNN	Neighborhood-based	surprise.KNN Basic	Finds similar users/items via cosine or Pearson metrics.
NCF	Deep Neural Network	TensorFlow /PyTorch	Learns non-linear embeddings through multilayer perceptrons.

Dataset and preprocessing:

- The E-Commerce data set (Flipkart mobile data) includes a publicly available data set which is much larger, and then the synthetic user item ratings generated for 50 users with 3900+ products were cleanly applied.
- The movie data set has user ratings on 9000+ movies with the metadata, which includes titles and genres.
- Book data set which includes user ratings on 271000 books with the associated metadata like author and publisher.
- The E-Commerce dataset contents to different csv files containing the data of the products of mobile phones Flipkart Mobile 2.csv (430 records, 16 features) and Flipkart_Mobiles.csv (3114 records, 8 features).

Synthetic data generation:

To train on the data set, the user-item interaction data is essential for collaborative filtering models. The data set we have does not have the user-item interaction, so we created a synthetic data set for 50 users. The generation of the data process was created to simulate realistic reference:

- Every user was allocated a random user_bias.
- Every user was allocated a random preference score for each brand.
- Every user has given a rate to a random sample of 10 to 60 products.
- The small amount of random noise was also added to simulate the natural variance in the user ratings. The final simulator rating was calculated using the below formula: Rating = base_rating + user_bias + brand_preference + noise
- The ratings from users were calculated as a function of the item's base rating, the user's bias and the user's preference for the particular or selected item's brand and a small random noise component. Ratings were then clipped to 1.0 to the 5.0 skill scale. This whole synthetic data generation process produced the synthetic_ratings.csv, which includes user-item-ratings interactions.

Feature engineering and targeting variable

To predict the user preferences a binary target variable, liked, was generated from the synthetic data ratings, the product was considered liked (1) if the rating was 4.0 or higher than that and disliked (0) otherwise.

Classification-based recommendation

To use the binary classification for the recommendation task so that we can predict whether the user will like a product or not we need classification-based recommendation. The label is defined as a light column were liked = 1 only if the rating is higher than or is equal to 4.0 and liked = 0 otherwise. The combination of numeric features such as battery_capacity, display_size, ROM, RAM, sales price and encoded categorical features (user_enc, item_enc, brand_enc).

Collaborative Filtering Recommendation

This paper examines the prediction of the exact rating a user would give to an item using the user-item interaction matrix.

Collaborative filtering is the technique which is used globally over e-commerce with significant impact in different domains. In healthcare, this supports the personalized treatment recommendation with disease risk prediction as well as patient similarity analysis. In healthcare domain, the collaborative filtering helps with personalized treatment recommendations, disease risk predictions, and patient similarity analysis. In agriculture domain, the collaborative filtering enables the crop and fertilizer recommendation, yield prediction, and as well as precision farming decision on the basis of environmental and the soil patterns present in the data. In Politics, Collaborative Filtering helps with model voter behaviour, predict political preferences, and optimise targeted campaigning.

This application gives us that collaborative filtering is a flexible as well as scalable decision support technique over different real-world fields.

Singular Value Decomposition SVD: SVD is a matrix factorisation technique where it decomposes that user-item rating matrix into your dimensional user-latent and item-latent factor matrices. The first implementation populates the sparse training park; the training park populates the past training; the past training populates the sparse training matrix by filling the missing values with specific mean ratings. To produce the predictions, the SVD is applied to the degrading matrix.

$$\hat{r}ui = \mu + b_u + b_i + p_u^T q_i \quad (1)$$

Where:

- $\hat{r}_{ui}$= Predicted rating
- μ= Global average rating
- b_u, b_i= User and item biases
- p_u, q_i= Latent feature vectors of user and item

Alternating Least Squares ALS: ALS is the iterative optimisation algorithm, which is particularly effective for sparse matrices. It also fixes the user-latent factors while solving for the item-latent practice factors. This approach escapes the need for dense matrix imputation to make it computationally well organised. For the implementation of the method, we have set the number of latent factors to k = 20 and successfully calculated for 15 iterations.

$$min \sum_{(u,i)\in R} (r_{ui} - p_u^T q_u)^2 \ + \lambda(\| \, pu \, \|^2 + \| \, qi \, \|^2) \quad (2)$$

Where λ is a regularization constant.

Neural Collaborative Filtering (NCF): The framework of Neural Collaborative Filtering was proposed by He et al. (2017) to generalize the solution of the restrictive nature of the inner product in Matrix Factorization. NCF is the recommendation system that makes use of a neural network for the prediction of user preferences on the basis of their interaction with items.

$$\hat{y}ui = \sigma\{f_\theta\{p_u \oplus q_i\}\} \quad (3)$$

Where:

- $p_u \oplus q_i$= Concatenation of embeddings
- f_θ= Multilayer neural network
- σ= Output activation (sigmoid)

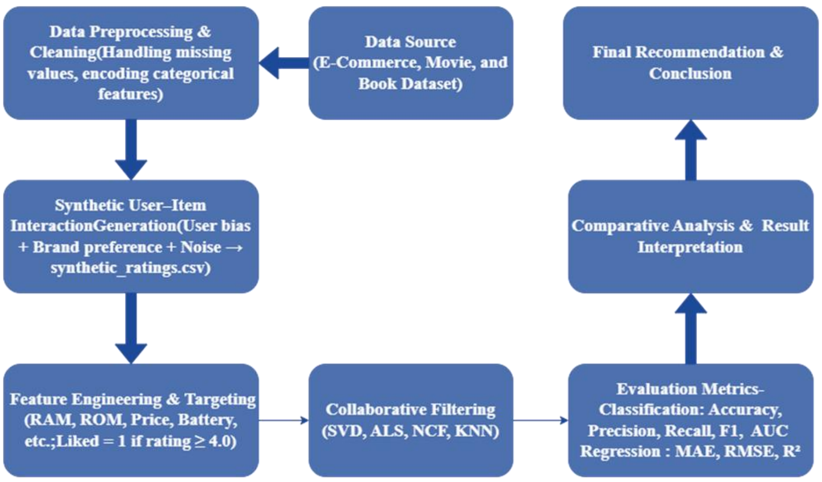

Fig. 1. Methodology of Research

Fig. 1. gives a complete research methodology workflow. It starts with data acquisition technique, pre-processing of the data, and synthetic data generation, which is then followed by feature engineering, model implementation, and the evaluation of model. This given pipeline gives a consistency across all the three dataset domains, i.e. e-commerce, movies, and books, which allows a fair comparison between all the traditional machine

learning classifiers as well as collaborative filtering algorithms. All the stages included in the workflow contributes to the reproducibility of the recommendation system and ensures the sparse and dense interaction patterns which are represented during the evaluation of the data.

Evaluation Matrix

Evaluating the performance of all methods used in the research are done using different evaluation matrices as they perform the evaluation for each task. Accuracy, precision, recall, F1 score, and area under the ROC curve are used for the evaluation of classification algorithms. Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE), and R-square (R2) are used for the regression algorithm for the collaborative filtering.

III. Results And Discussions

This section of result presents a complete comparative evaluation of four collaborative filtering algorithms which we have used in our research that is Singular Value Decomposition SVD, Alternating Least Squares ALS, K-Nearest Neighbours, (KNN), and Neural Collaborative Filtering, (NCF) on three datasets. The evaluation process includes regression-based accuracy metrics, classification-based decision reliability measures, and clustering-based interpretability analysis to give an accurate assessment of the performance and latent representation quality of algorithms.

Quantitative Performance Evaluation

The models evaluated using the standard statistical metrics, Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2) also with the classification metrics along with Accuracy, Precision, Recall, F1-Score, and ROC-AUC.

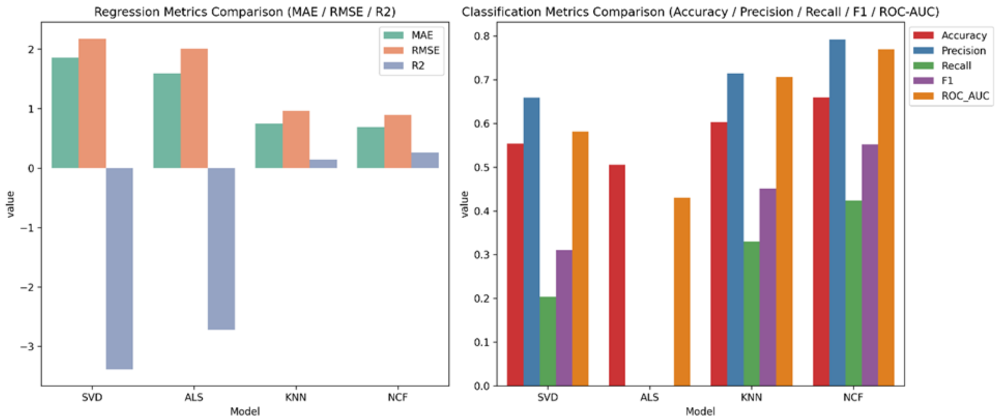

Fig. 2. Regression and Classification Metrics for Movie Dataset

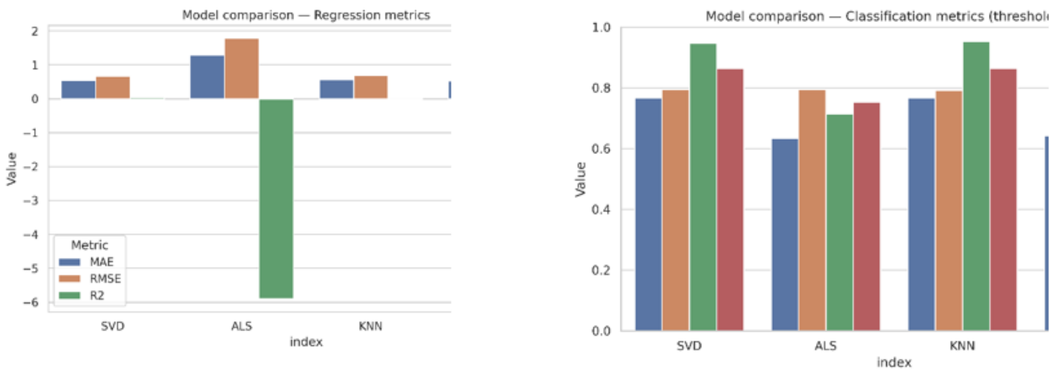

Fig. 3. Regression and Classification Metrics for E-Commerce Dataset

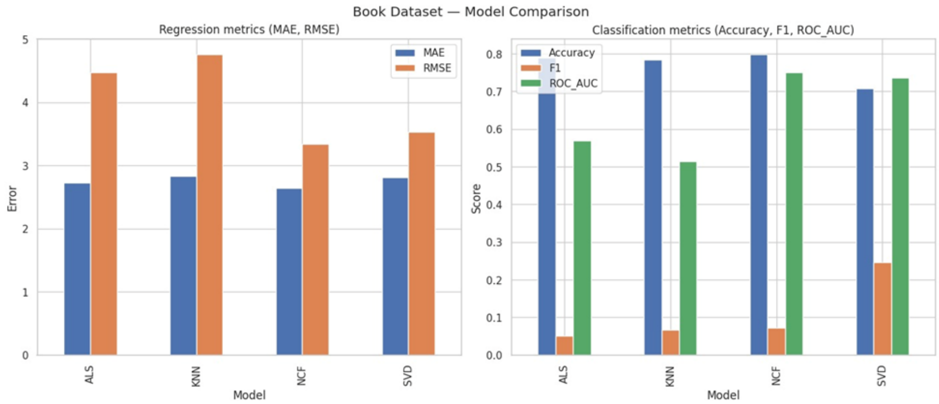

Fig. 4. Regression and Classification Metrics for Book Dataset

On the book dataset gives overall performance which is different because of the extreme sparse user-item interaction. The algorithms SVD, ALS, and KNN model showed the higher RMSE values, which reflects the difficulty in learning using minimal explicit ratings, where MCF significantly outperformed the traditional ML models, which achieved the lowest RMSE, 3.38, but a positive R-square score, i.e., 0.21, which indicates that neural architectures can capture the non-linear relationships, even though user behaviour is shown inconsistent or sparse in the dataset. The classification metrics also showed the improvement when threshold tuning was applied during the performance.

Analysis of Regression Metrics

In the MovieLens dataset, NCF model achieved low RMSE 0.8953, and KNN achieved 1.02, and ALS achieved 2.93. All this evaluation demonstrates that the Deep Neural Approach formed non-linear dependencies in between user and items. NCF's R-square value, 0.26, gives a proper proportion of variance in weightings, which reflects better model generalization. Whereas, the ALS shows highly negative value for R^2, that is, -6.91, which highlights the overfitting in sparse interaction regions, that is a very common limitation of matrix factorization under the low feedback density. KNN performed well in terms of RMS 1.02, which is a stable result, which confirms its reliability and interpretability in a moderate-sized dataset. For Book dataset, all the traditional SVD, ALS, and KNN) models' performance was low. Where, The NCF achieved the lowest RMSE (3.38) and positive R^2 (0.21), shows the ability to learn from sparse implicit feedback. This model improvement shows that the neural architectures take the latent reading preferences more successfully than linear models. The following Figure 4. Shows that Scatter plots of True vs Predicted Ratings for best model (NCF) in each dataset.

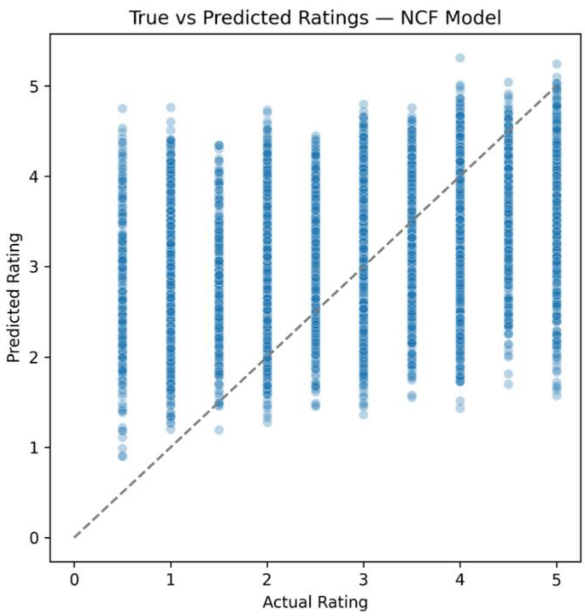

Fig. 5. Scatter plots of True vs Predicted Ratings for best model (NCF) in movie dataset.

Analysis of Classification Metrics

Binary classification problem that is light is equal to rating greater than or equal to 4.0 for movie data set and greater than or equal to 8.0 for book data set. NCF consistently received the highest ROC-AUC values that is 0.7699 for movie data set and 0.7479 for book data set which indicates the superior discriminative ability of the algorithm. The ALS and KNN showed a slightly higher precision in some class with low recall value (0.007 -0.10) showing the models favours the majority /non-liked classes and fails to detect the positive preferences. For Book dataset, Threshold tuning in the NCF model increased the Recall from 0.026 to 0.436 and F1-score from 0.05 to 0.43.

Visual Analysis and Interpretability

The NCF model have given the best result among all, with highest Silhouette 0.06 and lowest DBI 4.4, which shows that it is very meaningful and well-structured groups, as well as SVD the moderate clustering with Silhouette 0.67 and DBI 1.6, whereas ALS performed very poor with weak separation. And for the book dataset, the algorithm trend was also similar. NCF gave clear and smoother clusters, while ALS and SVD showed scattered clusters. So, overall, the visual analysis shows that the deep learning models like NCF can predict better but also create more interpretable and meaningful feature spaces.

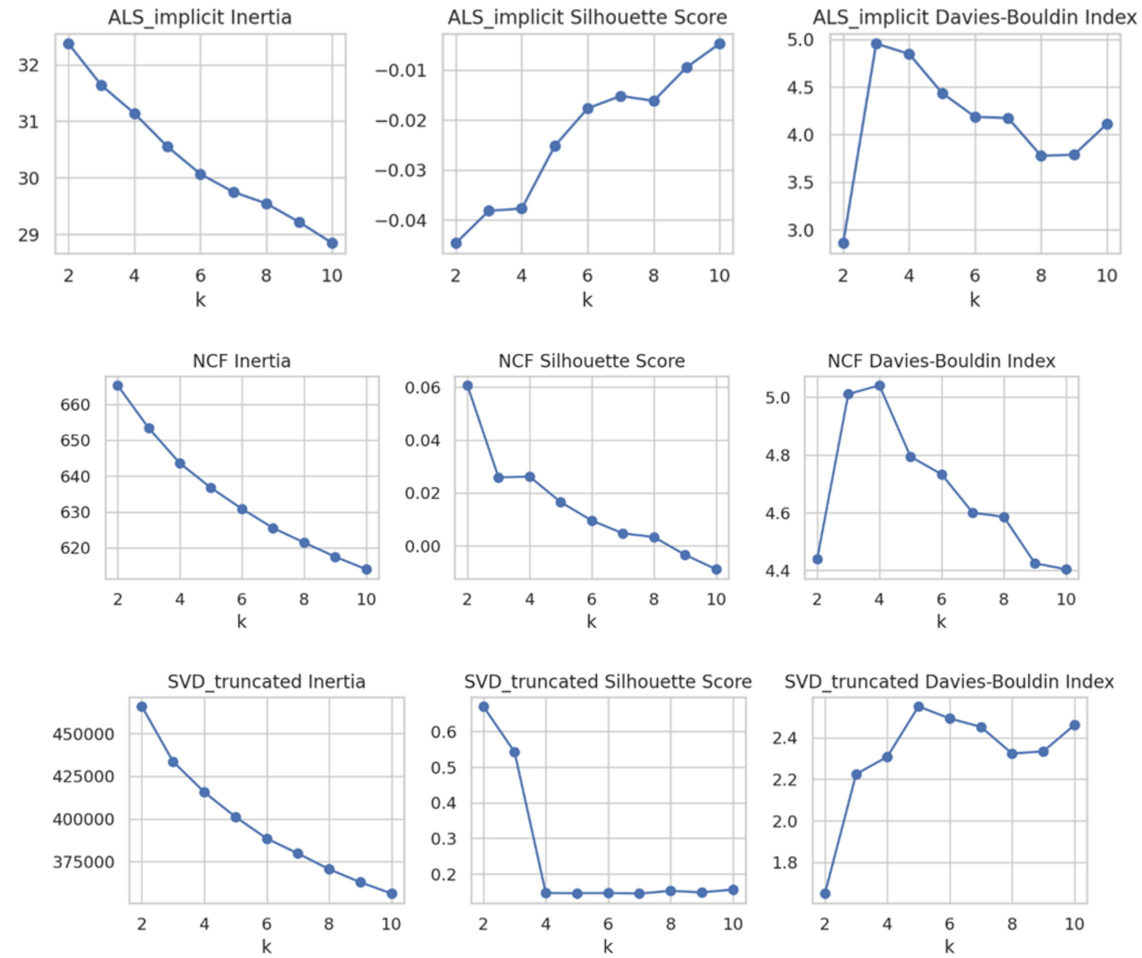

Fig. 6. K-Means cluster visualizations and metric plots (Silhouette, DBI, Inertia)

Here, in the above figure, the x-axis represents the latent embedding dimensions which are extracted from each algorithm, whereas the y-axis represents the cluster assignment generated using the k-means model. The metric plots the existing measures of the cluster include:

- Silhouette Score (higher = better separation),
- Davies–Bouldin Index (DBI) (lower = better cohesion),
- Inertia (lower = tighter clusters).

This annotated axis and matrix also helps to interpret which algorithm has done better and organizes user-item relationships in the latent space.

Overall Summary

Neural Collaborative Filtering has achieved 30-40% higher record as compared to the other traditional CF algorithms which highlights the potential for scalable and data-rich recommendation scenarios.

Matrix factorization models (SVD/ALS) have evaluated very well. KNN has continued to offer utility as a predictor and for attaining consistency. As for this analysis, the embedding was completed via the visual and quantitative mechanism available to the Latin architecture and the algorithm to set a threshold and hybridize to realize the recall improvement and the F1 score to a level of satisfaction for the actual tools to the dataset in question and the level of sparsity of the dataset.

The analysis of the datasets in the domain of e-commerce, movie, and book exhibits the same pattern in terms of consistency in the performance evaluation. Among all of them, NCF was able to achieve the best performance and demonstrated improvement in all metrics and was able to achieve a lower RMSE, ROC, and AOC and demonstrated improvement in the overall quality of clusters. In this case, KNN showed a lower performance as a consistency compared to the NCF but demonstrated a more stable performance Mid-range performance. In contrast to other models, ALS here struggled a little in a sparse dataset. A combination of charts summarizes the variation and the demonstration of how sparsity in a dataset affects the behaviour of a model. The findings of the evaluation showed great importance in the selection of the algorithm for the dataset. The findings of the evaluation showed great importance in the selection of the algorithm for the dataset.

Table no 2 For E-commerce dataset.

Model	MAE	MSE	RMSE	R2
SVD	0.5388	0.4409	0.6640	0.0386
ALS	1.2864	3.1627	1.7784	-5.8956
KNN	0.5540	0.4655	0.6823	-0.0150
NCF	0.5221	0.4270	0.6535	0.0688

Regression Matrics for E-commerce dataset

Model	Accuracy	Precision	Recall	F1	ROC_AUC
SVD	0.7676	0.7937	0.9461	0.8632	0.5555
ALS	0.6344	0.7940	0.7138	0.7517	0.5496
KNN	0.7676	0.7905	0.9528	0.8641	0.5506
NCF	0.6422	0.8773	0.6262	0.7308	0.7235

Classification Matrics for E-commerce dataset

Table no 3 For Movie Dataset.

Model	MAE	MSE	RMSE	R2
SVD	1.8594	4.7474	2.1788	-3.3841
ALS	1.5901	4.0266	2.0066	-2.7185
KNN_user	0.7470	0.9255	0.9620	0.1453
KNN_item	0.6741	0.7820	0.8843	0.2778
NCF	0.6677	0.7443	0.8627	0.3125
NCF_tuned	0.6743	0.7668	0.8756	0.2918

Regression Matrics for Movie Dataset

Model	Accuracy	Precision	Recall	F1	ROC_AUC
SVD	0.5539	0.6586	0.2033	0.3107	0.5809
ALS	0.5054	0.8229	0.0076	0.0150	0.4303
KNN_user	0.6033	0.7144	0.3298	0.4513	0.7064
KNN_item	0.6705	0.8219	0.4262	0.5613	0.7822
NCF	0.6457	0.8357	0.3531	0.4964	0.7913
NCF_tuned	0.6569	0.8115	0.3989	0.5349	0.7814

Classification Matrics for Movie Dataset

Table no 4 For Book Dataset.

Model	MAE	MSE	RMSE	R^2
ALS	2.7238	19.991	4.4711	-0.3761
KNN	2.8299	22.713	4.7658	-0.5220

NCF	2.6457	11.195	3.3459	0.2279
SVD	2.8107	12.501	3.5357	0.1579

Regression Matrics for Book Dataset

Model	Accuracy	Precision	Recall	F1	ROC_AUC
ALS	0.7898	0.4181	0.0265	0.0499	0.5683
KNN	0.7832	0.6111	0.0350	0.0663	0.5145
NCF	0.7990	0.8683	0.0375	0.0720	0.7499
SVD	0.7084	0.6717	0.1506	0.2461	0.7363

Classification Matrics for Book Dataset

The Book dataset includes a separate analysis because of extreme sparsity.

IV. Novelty And Contributions

The novelty of this research lies in developing a unified multi-domain evaluation and interpretability framework for Collaborative Filtering algorithms. Unlike most prior studies that focus on a single dataset or a single algorithm, this work integrates four distinct CF techniques—SVD, ALS, KNN, and Neural Collaborative Filtering (NCF)—and evaluates them across three heterogeneous domains (E-Commerce, Movies, and Books).

Key Contributions

1. Cross-Domain Framework:

A single experimental pipeline designed to analyze algorithmic adaptability under varying data sparsity, user–item interaction density, and domain complexity.

2. Integrated Evaluation Metrics:

Combination of statistical regression, classification, and clustering-based metrics (Silhouette, DBI, Inertia) to assess both performance and interpretability.

3. Neural Collaborative Filtering Extension:

Implementation of a deep neural architecture for learning non-linear user–item relationships and evaluating its generalization across datasets.

4. Threshold-Tuned Recommendation Model:

Dynamic threshold tuning to optimize precision–recall balance for sparse datasets such as Book-Crossing.

5. Latent Space Interpretability:

Visualization and clustering of learned item embeddings to understand how algorithms capture user preference structures.

6. Benchmark Dataset Comparison:

Empirical comparison showing how each algorithm performs optimally under specific domain and data conditions. This multi-dimensional benchmarking approach provides a novel foundation for developing hybrid, interpretable, and domain-adaptive recommender systems.

V. Conclusion

This research signifies an evaluation of collaborative filtering algorithms on our e-commerce, movie, and book dataset domains. The integration of pre-processing, synthetic interaction modelling, feature engineering, and multi-metric evaluation in this research shows the strengths and limitations of traditional and neural techniques. On the basis of evaluation, among the evaluated models, neural collaborative filtering is consistently given the superior accuracy, generalization, and cluster interpretability on the given datasets. The matrix factorization method, such as SVD and ALS, has performed reliably in denser datasets but has struggled a little with sparse interactions. Overall, the work here establishes a unified reproducible evaluation framework, which provides practical guidelines to select the algorithm. On the basis of domain characteristics and data sparsity, it is easy to select the algorithm. Neural Collaborative Filtering showed the highest performance over all three datasets we used in the research e-commerce, movies, books specially excelled in sparse environment. In denser data sets, both ALS and SVD did greatly, while KNN had an okay, robust performance that definitely struggles with very sparse data. Finally for e-commerce data KNN and SVD had the highest accuracy being 76% and ALS was the lowest with 63%. For the movie data, KNN had the highest accuracy being 67% and ALS was the lowest with 50%. While with the book dataset NCF had 79 % being the highest and SVD 70% being the lowest. For the future, there could be some research into hybrid deep learning methods, and also with explainable recommendation systems, which could be combined into scalable systems that could greatly improve multi-domain recommendations.

References

[1]. Ibrahim, O. A. S., Younis, E. M. G., Mohamed, E. A., & Ismail, W. N., (2025). Revisiting Recommender Systems: An Investigative Survey. Neural Computing And Applications, 37(4), 2145–2173. Doi:10.1007/S00521-024-10828-5

[2]. Andika, H. G., Hadinata, M. T., Huang, W., Anderies, & Iswanto, I. A., (2022). Systematic Literature Review: Comparison On Collaborative Filtering Algorithms For Recommendation Systems. In Proceedings Of The IEEE International Conference On Communication, Networks And Satellite (COMNETSAT), 3–4 November (Pp. 56–61). Solo, Indonesia. Doi:10.1109/COMNETSAT56033.2022.9994316

[3]. Wu, X., (2024). Review Of Collaborative Filtering Recommendation Systems. Advances In Computer Engineering, 43(1), 76–82. Doi:10.54254/2755-2721/43/20230811

[4]. Martins, G. B., Papa, J. P., & Adeli, H., (2020). Deep Learning Techniques For Recommender Systems Based On Collaborative Filtering. Expert Systems, 37(6), E12647. Doi:10.1111/Exsy.12647

[5]. Li, D., (2024). On The Theoretical Foundation Of Model Comparison And Evaluation For Recommender Systems. Arxiv Preprint. Doi:10.48550/Arxiv.2411.01843

[6]. Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S., (2007). Collaborative Filtering Recommender Systems. In Brusilovsky, P., Kobsa, A., & Nejdl, W. (Eds.), The Adaptive Web (Pp. 291–324). Springer. Doi:10.1007/978-3-540-72079-9_9

[7]. Goldberg, D., Nichols, D., Oki, B. M., & Terry, D., (1992). Using Collaborative Filtering To Weave An Information Tapestry. Communications Of The ACM, 35(12), 61–70. Doi:10.1145/138859.138867

[8]. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J., (1994). Grouplens: An Open Architecture For Collaborative Filtering Of Netnews. In Proceedings Of The ACM Conference On Computer Supported Cooperative Work (CSCW), 22–26 October (Pp. 175–186). Chapel Hill, NC.

[9]. Herlocker, J. L., Konstan, J. A., Borchers, A., & Riedl, J., (2004). Evaluating Collaborative Filtering Recommender Systems. ACM Transactions On Information Systems, 22(1), 5–53. Doi:10.1145/963770.963772

[10]. Sarwar, B., Karypis, G., Konstan, J., & Riedl, J., (2001). Item-Based Collaborative Filtering Recommendation Algorithms. In Proceedings Of The 10th International Conference On World Wide Web, 1–5 May (Pp. 285–295). Hong Kong, China. Doi:10.1145/371920.372071

[11]. Koren, Y., Bell, R., & Volinsky, C., (2009). Matrix Factorization Techniques For Recommender Systems. Computer, 42(8), 30–37. Doi:10.1109/MC.2009.263

[12]. Zhou, Y., Wilkinson, D., Schreiber, R., & Pan, R., (2008). Large-Scale Parallel Collaborative Filtering For The Netflix Prize. In Fleischer, R., & Xu, J. (Eds.), Algorithmic Aspects In Information And Management (Pp. 337–348). Springer, Berlin, Heidelberg. Doi:10.1007/978-3-540-68880-8_32

[13]. He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T.-S., (2017). Neural Collaborative Filtering. In Proceedings Of The 26th International Conference On World Wide Web (WWW), 3–7 April (Pp. 173–182). Perth, Australia. Doi:10.1145/3038912.3052569

[14]. Harper, F. M., & Konstan, J. A., (2016). The Movielens Datasets: History And Context. ACM Transactions On Interactive Intelligent Systems, 5(4), 1–19. Doi:10.1145/2827872

[15]. Ziegler, C. N., Mcnee, S. M., Konstan, J. A., & Lausen, G., (2005). Improving Recommendation Lists Through Topic Diversification. In Proceedings Of The 14th International Conference On World Wide Web (WWW), 10–14 May (Pp. 22–32). Chiba, Japan. Doi:10.1145/1060745.1060754

[16]. Ricci, F., Rokach, L., & Shapira, B., (Eds.), (2015). Recommender Systems Handbook (2nd Ed.). Springer, New York. Doi:10.1007/978-1-4899-7637-6

[17]. Su, X., & Khoshgoftaar, T. M., (2009). A Survey Of Collaborative Filtering Techniques. Advances In Artificial Intelligence, 2009, 421425. Doi:10.1155/2009/421425

[18]. Zhang, S., Yao, L., Sun, A., & Tay, Y., (2019). Deep-Learning Based Recommender System: A Survey And New Perspectives. ACM Computing Surveys, 52(1), 1–38. Doi:10.1145/3285029

[19]. Aggarwal, Charu C. Recommender Systems: The Textbook. Springer, 2016. K10plus ISBN.

[20]. Bobadilla, J., Et Al. “Recommender Systems Survey.” Knowledge-Based Systems, Vol. 46, July 2013, Pp. 109–32. DOI.Org (Crossref), Https://Doi.Org/10.1016/J.Knosys.2013.03.012.

[21]. Fajie, U, Et Al. “A Simple But Hard-To-Beat Baseline For Session-Based Recommendations.” Arxiv,2018.DOI.Org(Datacite), Https://Doi.Org/%20https://Doi.Org/10.48550/Arxiv.1808.05163.

[22]. Gantner, Zeno, Et Al. “Mymedialite: A Free Recommender System Library.” Proceedings Of The Fifth ACM Conference On Recommender Systems [Chicago Illinois USA], 2011, Pp. 305–08. DOI.Org (Crossref), Https://Doi.Org/10.1145/2043932.2043989.

[23]. Wang, Hao, Et Al. “Collaborative Deep Learning For Recommender Systems.” Proceedings Of The 21th ACM SIGKDD International Conference On Knowledge Discovery And Data Mining [Sydney NSW Australia], 2015, Pp. 1235–44. DOI.Org (Crossref), Https://Doi.Org/10.1145/2783258.2783273.

[24]. Wu, Yao, Et Al. “Collaborative Denoising Auto-Encoders For Top-N Recommender Systems.” Proceedings Of The Ninth ACM International Conference On Web Search And Data Mining [San Francisco California USA], 2016, Pp. 153–62. DOI.Org (Crossref), Https://Doi.Org/10.1145/2835776.2835837.

[25]. Zhang, Yongfeng, And Xu Chen. “Explainable Recommendation: A Survey And New Perspectives.” Foundations And Trends® In Information Retrieval, Vol. 14, No. 1, 2020, Pp. 1–101. DOI.Org (Crossref), Https://Doi.Org/10.1561/1500000066.

[26]. “Explainable Recommendation: A Survey And New Perspectives.” Foundations And Trends® In Information Retrieval, Vol. 14, No. 1, Mar. 2020, Pp. 1–101. Www.Nowpublishers.Com, Https://Doi.Org/10.1561/1500000066.

[27]. Koren, Yehuda. “Factorization Meets The Neighborhood: A Multifaceted Collaborative Filtering Model.” Proceedings Of The 14th ACM SIGKDD International Conference On Knowledge Discovery And Data Mining [Las Vegas Nevada USA], 2008, Pp. 426–34. DOI.Org (Crossref), Https://Doi.Org/10.1145/1401890.1401944.

[28]. He, Xiangnan, Et Al. “Fast Matrix Factorization For Online Recommendation With Implicit Feedback.” Proceedings Of The 39th International ACM SIGIR Conference On Research And Development In Information Retrieval [Pisa Italy], 2016, Pp. 549–58. DOI.Org (Crossref), Https://Doi.Org/10.1145/2911451.2911489.

[29]. Ferrari Dacrema, Maurizio, Et Al. “Are We Really Making Much Progress? A Worrying Analysis Of Recent Neural Recommendation Approaches.” Proceedings Of The 13th ACM Conference On Recommender Systems [Copenhagen Denmark], 2019, Pp. 101–09. DOI.Org (Crossref), Https://Doi.Org/10.1145/3298689.3347058.