International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

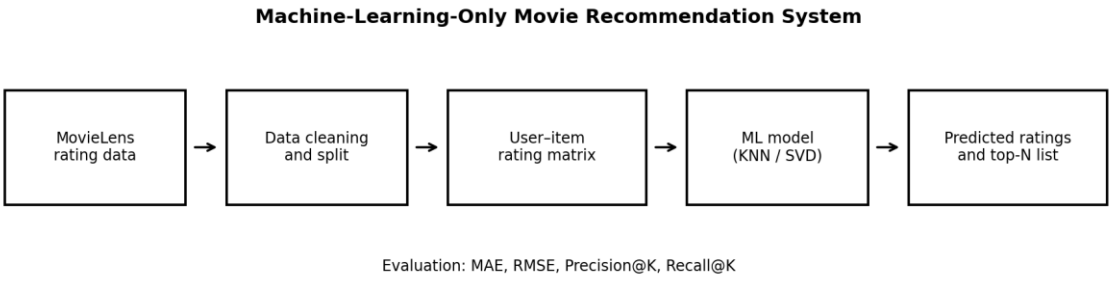
A Movie Recommendation System Using Machine Learning-Based Collaborative Filtering

Atharv Mane, Audamber Karad, Amar Tamboli, and Rushikesh Kulkarni

Department of Computer Science and Engineering
Bhagwant Institute of Technology, Barshi, Maharashtra, India

Abstract - Movie platforms contain large catalogs, making it difficult for users to select films that match their interests. This paper proposes a movie recommendation system that uses machine-learning-based collaborative filtering only. The system learns from historical user–movie ratings rather than using content-based similarity, genre matching, or a hybrid content model. After cleaning the rating data, it builds a user–item interaction matrix and trains collaborative models such as K-nearest neighbors (KNN) and matrix factorization using singular value decomposition (SVD). The trained model estimates ratings for movies a user has not rated and ranks candidates to produce a top-N recommendation list. MovieLens is proposed as a public dataset for implementation and reproducible evaluation. The system can be evaluated using mean absolute error (MAE), root mean squared error (RMSE), Precision@K, and Recall@K. No numerical experimental results are claimed because the models have not been run on a dataset as part of preparing this document. The proposed design provides a clear ML-only workflow and can be extended through parameter tuning, cold-start strategies, and user-interface development.

Keywords - Machine learning, movie recommendation, collaborative filtering, K-nearest neighbors, matrix factorization, singular value decomposition, MovieLens



I. INTRODUCTION

Online movie platforms provide many choices, and users may spend significant time finding films they are likely to enjoy. A recommender system addresses this problem by predicting which unseen movies a user may rate positively.

This paper focuses exclusively on machine-learning-based collaborative filtering. The models learn from user–movie rating interactions. Content-based filtering, genre/text similarity, and hybrid combinations with content features are outside the scope of this design.

The proposed workflow prepares rating data, trains collaborative ML models, predicts ratings for unseen user–movie pairs, and ranks the candidate movies. The paper presents the related work, methodology, evaluation plan, comparison, and future improvements.

II. LITERATURE REVIEW

Collaborative filtering uses patterns in user-item interactions to generate recommendations. Sarwar et al. presented item-based collaborative filtering, in which item similarities are derived from users' rating behavior [1]. This provides a useful neighborhood-based baseline, although performance depends on sufficient overlapping ratings.

Koren, Bell, and Volinsky discussed matrix-factorization methods for recommender systems [2]. Such methods learn latent user and item factors from observed ratings and can estimate preferences even when users have not rated every item. Their performance depends on data quality, regularization, and parameter selection.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

GroupLens distributes MovieLens datasets containing user ratings and movie identifiers, enabling repeatable recommender-system experiments [3]. This paper proposes evaluating KNN collaborative filtering and SVD-based matrix factorization using the same data split. The document does not claim that either method is superior before experiments are run.

III. PROPOSED MACHINE LEARNING METHODOLOGY

A. Dataset: Use a specified MovieLens release with user IDs, movie IDs, and rating values. The main model input is the rating interaction; movie descriptions, genre text, image features, and content similarity are not used.

B. Preprocessing: Remove invalid or duplicate rating rows where appropriate, check rating ranges, and split the interaction records into training, validation, and test sets. A fixed random seed or a time-based split should be recorded. The test set must not be used to tune model parameters.

C. User–Item Matrix: Represent users as rows and movies as columns, with observed ratings as matrix entries. Missing entries represent unrated movies and must not be treated as explicit zero ratings.

D. Model 1 — K-Nearest Neighbors Collaborative Filtering: Train a neighborhood model using user-rating patterns. An item-based variant identifies movies with similar rating histories; a user-based variant identifies users with similar rating patterns. Predict a target rating from relevant neighbors, with safeguards for sparse neighborhoods.

E. Model 2 — Matrix Factorization (SVD): Learn latent vectors for users and movies from the training ratings. Estimate a rating using the user and item latent factors, optionally with user/item biases. Tune the number of factors and regularization on validation data.

F. Recommendation Generation: For a selected user, estimate ratings for movies not present in that user's training history, sort candidates by predicted rating, and return the top N. Already-rated movies should be excluded from the final list.

G. Algorithm: (1) Load user–movie ratings; (2) validate and clean the data; (3) create training, validation, and test sets; (4) build the user–item matrix; (5) train KNN and SVD models; (6) tune parameters on validation data; (7) predict ratings for unseen movies; (8) rank candidates; (9) evaluate the final model on the untouched test set.

IV. SYSTEM ARCHITECTURE

The architecture contains rating-data input, data cleaning and splitting, user–item matrix construction, collaborative ML model training, rating prediction, and top-N output. KNN and SVD should first be evaluated independently. The model selected for deployment should be based on measured validation and test results rather than assumptions.

V. RESULTS AND DISCUSSION

No model execution or measured dataset output was supplied for this paper. Therefore, this section does not invent accuracy or error values. The proposed comparison is an experimental plan rather than a claim of completed results.

KNN collaborative filtering is relatively interpretable because recommendations can be linked to neighboring users or movies, but it may be affected by sparse rating overlap. SVD-based matrix factorization learns compact latent factors and can model broader rating patterns, but it requires training and parameter tuning. Both approaches rely on historical interactions and can struggle with users or movies that have no ratings.

Evaluate rating prediction using MAE and RMSE. For top-N recommendations, report Precision@K and Recall@K, and optionally NDCG@K. Compare KNN and SVD on the same test split, report parameter settings and dataset version, and discuss results only after running the experiment.

VI. COMPARATIVE ANALYSIS

ML method	Strength	Limitation
KNN collaborative	Neighborhood-based and interpretable	Needs rating overlap
SVD matrix factorization	Learns latent user/item patterns	Needs training and tuning

VII. CONCLUSION AND FUTURE WORK

This paper described a movie recommendation system based only on machine-learning collaborative filtering. It uses user–movie rating interactions to train KNN and SVD models, predict ratings for unseen movies, and rank a personalized recommendation list. It intentionally excludes content-based filtering, metadata similarity, and hybrid content-based scoring.

Future work includes implementation on a fixed MovieLens dataset, fair comparison and tuning of KNN and SVD, handling cold-start users and movies, improving scalability, and building a web interface. Any claim about which model performs best should be supported by reproducible test results.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

REFERENCES

[1] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl, “Item-Based Collaborative Filtering Recommendation Algorithms,” in Proceedings of the 10th International Conference on World Wide Web, 2001, pp. 285–295, doi: 10.1145/371920.372071.

[2] Y. Koren, R. Bell, and C. Volinsky, “Matrix Factorization Techniques for Recommender Systems,” Computer, vol. 42, no. 8, pp. 30–37, Aug. 2009, doi: 10.1109/MC.2009.263.

[3] GroupLens Research, “MovieLens Datasets.” [Online]. Available: https://grouplens.org/datasets/movielens/. Accessed: Oct. 30, 2026.

ACKNOWLEDGMENT

The authors acknowledge the CSE department and the developers of the publicly available MovieLens datasets. Project guide details may be added before submission.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)