Movie Recommendation System – Project Report

# Group Members:

• Aun Ali Kazi (22K-4029)  
• Aizaz Usman (22K-4012)  
• Uzair Ahmed (22K-4070)  
• Fabiha Ayela (22K-8705)

# 1. Introduction

In the era of information overload, recommendation systems play a vital role in helping users discover relevant content. Our project presents a hybrid movie recommendation system that combines collaborative filtering, content-based filtering, and matrix factorization to generate personalized movie suggestions. The goal is to overcome limitations of individual methods and enhance recommendation quality by leveraging their combined strengths.

# 2. Dataset Details

The system uses the MovieLens dataset, a widely-used benchmark dataset for recommender systems. We use the 20M variant, limited to the first 300,000 ratings for computational efficiency. It includes:  
  
- Ratings data: userId, movieId, rating (1–5), timestamp  
- Movie data: movieId, title, genres  
  
This structure enables collaborative and content-based filtering based on user interactions and movie features.

# 3. Explanation of Model Used

Our system is hybrid, combining three methods:  
  
1. Collaborative Filtering:  
 - Constructs a user-item matrix  
 - Calculates cosine similarity between users  
 - Recommends items rated highly by similar users  
  
2. Content-Based Filtering:  
 - Uses TF-IDF on genres and titles  
 - Computes cosine similarity between movies  
 - Recommends similar movies to those liked  
  
3. Matrix Factorization:  
 - Applies Truncated SVD on the user-item matrix  
 - Predicts missing ratings using 50 latent factors  
  
Hybrid Strategy:  
- If input movie is provided: 60% content-based, 20% collaborative, 20% SVD  
- Without movie input: 50% collaborative, 50% SVD

# 4. Details of Experiments

We conducted the following experiments:  
  
• Overall Evaluation:  
 - 80-20 train-test split  
 - Metrics: Precision, Recall, F1 Score, Hit Rate  
  
• Content-Based Evaluation:  
 - Input: Highly-rated user movies  
 - Output: Similar content-based movies  
  
• Parameter Tuning:  
 - Varied k (5–20), rating threshold (3.5–4.5), latent factors (30–70), and hybrid weights

# 5. Results and Discussion

Performance:  
- Precision@10, Recall@10, F1 Score, Hit Rate were evaluated  
  
Comparison:  
- Collaborative: Diverse, cold-start issue  
- Content-Based: Accurate, less diverse  
- Matrix Factorization: Effective, hard to interpret  
- Hybrid: Balanced performance  
  
Limitations:  
- Cold start for new users/movies  
- Computational limits (subset used)  
- No handling of time-based preference shifts  
- Limited diversity and novelty  
  
Future Work:  
- Include movie descriptions, actors, user demographics  
- Use advanced models like BPR, deep learning  
- Improve diversity and personalization balance

# 6. Conclusion

Our hybrid movie recommender system effectively integrates collaborative filtering, content-based filtering, and matrix factorization to produce relevant, personalized movie suggestions. The hybrid approach mitigates the weaknesses of individual methods while leveraging their strengths. The system is explainable, adaptable, and a solid foundation for future improvements.