



# MOVIELENS RECOMMENDATION SYSTEM

Leveraging Collaborative and Content-Based Filtering

By Nelima Wanyama



# WHY DO WE NEED RECOMMENDATION SYSTEMS?

- Streaming platforms have thousands of movies, making selection difficult.
- Users benefit from personalized recommendations based on preferences.
- Our system utilizes MovieLens data to generate personalized movie recommendations.



# PROBLEM STATEMENT

- Users struggle to find movies matching their interests.
- Manual searching is time-consuming.
- Existing recommendation systems have limitations (e.g., cold-start problem).



# MOVIELENS DATASET

- Movies Dataset: 9,742 movies with titles & genres.
- Ratings Dataset: 100,836 ratings from users.
- Tags Dataset: 3,683 tagged movies.



# DATA SAMPLE

- Movie example: Toy Story (1995),
- Genres: Animation, Comedy, Fantasy
- Rating example: User 234 rated Movie 567 with 4.5 stars



# DATA PREPROCESSING

- 1. Handling Missing Values: Checked for null entries and removed them.
- 2. Data Cleaning: Processed timestamps, standardized genres.
- 3. Data Transformation: Converted user ratings into a matrix format for collaborative filtering.

# METHODOLOGY

- **Collaborative Filtering**

- Concept: Users with similar preferences rate movies similarly.
- Technique Used: Singular Value Decomposition (SVD) to factorize the rating matrix.
- Predict missing ratings based on learned latent factors.

- **Content-Based Filtering**

- Concept: Recommend movies similar to those a user liked, based on metadata.
- Technique Used: TF-IDF Vectorization\* on movie descriptions. Cosine Similarity\* to find similar movies.





# METHODOLOGY

- **Hybrid Recommendation Approach**
- Combining Both Methods:
- Collaborative filtering predicts ratings from similar users.
- Content-based filtering suggests similar movies.
- Weighted Hybrid: Balancing both approaches for better accuracy.





# MODEL EVALUATION

Metrics Used:

- 1. Root Mean Squared Error (RMSE)
- 2. Mean Absolute Error (MAE)

Results:

- SVD Model RMSE: 0.89, MAE: 0.70
- Interpretation:
- Lower RMSE and MAE indicate better accuracy in prediction.



# RESULTS & INTERPRETATION

- Top recommended movies align with user preferences.
- Hybrid approach performs better than individual models.
- Limitations exist in handling new users (cold start issue).



# CHALLENGES

Cold-start problem:

- New users lack historical data.
- Scalability issues
- Large datasets increase computation time.

Future Improvements:

- Implement Deep Learning-based recommendations
- Deploy as a web-based API for real-time recommendations.



# CONCLUSION

## Summary:

- Built a MovieLens-based recommendation system.
- Used collaborative filtering, content-based filtering, and a hybrid approach
- Achieved good prediction accuracy using RMSE & MAE

## Future Scope:

- Improve scalability, personalization, and integration.



*Thank- You*