

# Movie Recommendation System

---

Non-Technical Presentation  
Prepared by: Group 4



# Project Overview

- ❖ Developed a recommendation system using the MovieLens 32M dataset.
- ❖ Goal: Suggest top 5 movies to users based on past ratings.
- ❖ Models used: KNNBasic (baseline) and SVD (final model).

# Business Understanding

- ❖ Users face decision fatigue due to too many content choices.
- ❖ Streaming platforms benefit from recommending relevant content.
- ❖ Improves user satisfaction, engagement, and retention.

# Data Overview

- ❖ MovieLens 32M dataset: 200K users, 87K movies, 32M ratings.
- ❖ Features used: UserId, movied, rating, movie title.
- ❖ Filtered to active users ( $\geq 20$  ratings) and movies ( $\geq 50$  ratings).

# Data Preparation

- ❖ Merged ratings with movie metadata.
- ❖ Removed noise and filtered low-activity users/movies.
- ❖ Prepared Surprise-compatible format for modeling.

# Modeling Approach

- ❖ KNNBasic: item-item collaborative filtering using cosine similarity.
- ❖ SVD: matrix factorization for latent preferences.
- ❖ Evaluation: 80/20 train-test split with RMSE and MAE metrics.

# Model Performance

- ❖ KNNBasic RMSE: 0.8844 | MAE: 0.6753
- ❖ SVD RMSE: 0.8368 | MAE: 0.6395 (better performance)
- ❖ SVD chosen for deployment due to improved accuracy.

# Top Recommendations (User 555)

- Alien (1979) — Predicted Rating: 4.55
- Raiders of the Lost Ark (1981) — Predicted Rating: 4.53
- A Clockwork Orange (1971) — Predicted Rating: 4.52
- Star Wars: Episode V (1980) — Predicted Rating: 4.50
- Terminator 2: Judgment Day (1991) — Predicted Rating: 4.46



# Business Impact & Insights

- ❖ Deploy SVD model into a web or mobile app backend.
- ❖ Gather user feedback and retrain as data grows.
- ❖ Explore hybrid methods using genres and tags.

# Next Steps

- Deploy model to flag high-risk customers weekly.
- Integrate predictions into CRM for personalized retention actions.
- Continue refining the model with new data.

# Appreciation

We would like to thank MovieLens team for dataset access which made these analysis possible.

Questions?

We would love to hear your take on this! Please feel free to get in touch.