# Building a Movie Recommender System

Phase 4 Machine Learning Project

**Moringa School**

# Group 6 Members

Hildah

Bernice

David

Alice

Erick

Bonface

# Overview

- In the digital age, users face too many choices.
- Recommender systems provide personalized suggestions.
- **Goal:** Recommend the Top 5 movies for each user based on past ratings.
- Dataset: MovieLens – 100,000 ratings, 610 users, 9,724 movies.

# Business and Data Understanding

**Business Problem:**

- Improve engagement with personalized movies.
- Boost user satisfaction and retention.

**Data:**

- Ratings, movie details, tags.
- Cleaned, merged, scaled for stability.

# Modeling Approach

We compared different approaches:

1. Popularity-based – shows trending movies to all.
2. Collaborative Filtering – learns from similar users.
3. Matrix Factorization (SVD) – finds hidden patterns in ratings.

| userId | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| userId | | | | | | | | | | |
| 0 | 0.0 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 1 | 0.0 | 1.000000 | 0.034755 | 0.039183 | 0.165461 | 0.137692 | 0.124770 | 0.143811 | 0.135497 | 0.059023 |
| 2 | 0.0 | 0.034755 | 1.000000 | 0.000000 | 0.000000 | 0.041284 | 0.064996 | 0.068174 | 0.000000 | 0.000000 |
| 3 | 0.0 | 0.039183 | 0.000000 | 1.000000 | 0.002961 | 0.006385 | 0.003619 | 0.000000 | 0.006890 | 0.000000 |
| 4 | 0.0 | 0.165461 | 0.000000 | 0.002961 | 1.000000 | 0.130157 | 0.085226 | 0.125647 | 0.048075 | 0.013898 |

# Evaluation in Plain Language

- **Cross-validation:** Tested model on different groups of users.
- **RMSE (Root Mean Square Error):**
  - Imagine predicting a movie rating out of 5 stars.
  - RMSE tells us how close we were to the actual rating.
  - Lower = better.

# Results Summary

- Popularity model: Simple, not personalized.
- Collaborative filtering: Personalized, but less accurate.
- **SVD + Ridge Regression: Best Performer!**
  - RMSE = **3.29**
  - Balanced accuracy + personalization
  - Generated strong Top-5 recommendations

# Recommendations

- Adopt SVD-based model for recommendations.
- Use popularity-based fallback for new users (**cold start problem**).
- Retrain regularly as more ratings are collected.

# Next Steps

- Test recommender with real users (A/B testing).
- Collect more feedback to fine-tune results.
- Explore hybrid methods combining multiple approaches.

# Thank You

**Questions?**