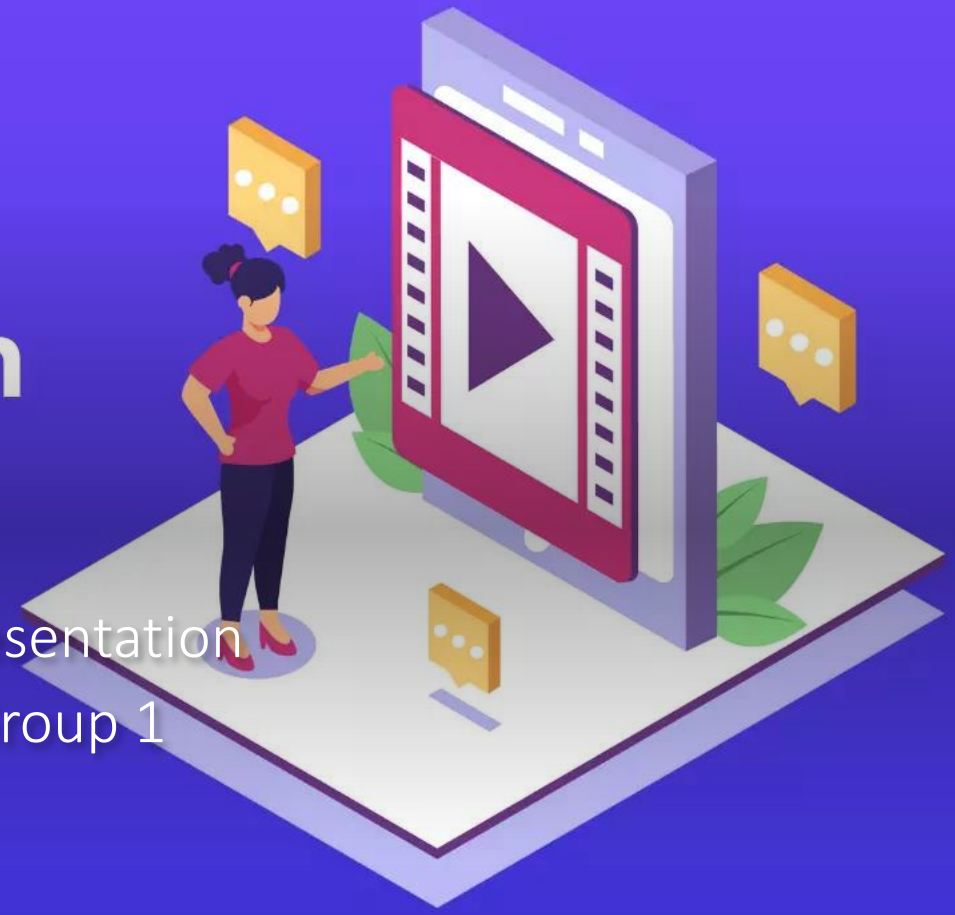
Movie Recommendation System



Non-Technical Presentation

Prepared by: Group 1

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), SVD and a Hybrid 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, movieId, rating, movie title and timestamp.
- Filtered to active users (≥20 ratings) and movies (≥50 ratings).

Data Preparation

- Data cleaning and inspection which involved checking for missing values, duplicate entries and invalid ratings.
- Removed noise and filtered low-activity users/movies.
- Prepared Surprise-compatible format for modeling.

Exploratory Analysis

- Found users favor highly-rated movies in Drama, Comedy, and Thriller genres.
- Active users and popular movies provide strong signals for recommendations.

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

Hybrid Filtering

- Combined SVD with genre-based cosine similarity.
- Addressed cold-start issues for new users.
- Blended scores with alpha=0.7 (70% CF, 30% content-based).

Comparing SVD and Hybrid models Top 5 Recommendations for a User

The hybrid model successfully retained some of the top collaborative picks while introducing new recommendations that align closely in genre

SVD Top 5 for User 1:

	title
0	Shawshank Redemption, The (1994)
1	Casablanca (1942)
2	One Flew Over the Cuckoo's Nest (1975)
3	Chinatown (1974)
4	Guardians of the Galaxy (2014)

Hybrid Top 5 for User 1:

	title
0	Shawshank Redemption, The (1994)
1	North by Northwest (1959)
2	Casablanca (1942)
3	One Flew Over the Cuckoo's Nest (1975)
4	Guardians of the Galaxy (2014)

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.

Contact Information:

Name: James Gatonye

LinkedIn: https://www.linkedin.com/in/james-gatonye-b5144991/