Sinema Pamoja: Personalized Movie Recommendation System

Using Data-Driven Personalization to Enhance Viewer Engagement and Retention

Team Members:

- Beatrice Ndanu
- Mercy Barminga
- Sharon Chebet
- Shem Nderitu
- Valentine Kweyu
- Nelson Muia

Business Problem & Project Goal

The Kenyan movie industry is experiencing growth, with more viewers accessing films through digital platforms

The Challenge

Viewers face choice overload and may miss movies they'd enjoy → disengagement and reduced loyalty

Proposed Solution

Build a system to recommend the Top-5 movies each user is most likely to watch and enjoy

Business Impact:

Boosts viewership and revenue.

Strengthens brand loyalty.

Helps audiences discover hidden gems.

Differentiates Sinema Pamoja from competitors.

Data Understanding

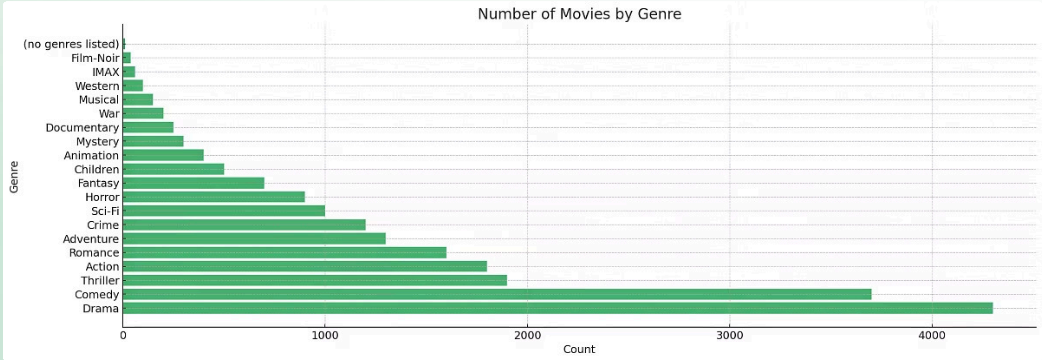
We load the core MovieLens datasets — movies, ratings, links, and tags — into a single comprehensive dataframe

The merging process ensures that all ratings are retained, even if some movies lack tags.

Dataset includes:

- **User ratings** - how much users liked a movie (1–5 stars)
- **Movie details** - title and genre
- **User-generated tags** - words users add to describe movies (e.g. *funny, action-packed, romantic*)
- **Links** - connects movies to external sites like IMDb or TMDb for extra details

After merging, the dataset provides a rich foundation for building personalized recommendations



Exploratory Data Analysis (EDA)

- Drama and Comedy dominate the catalog with the largest number of movies
- Action and Thriller also have strong representation among popular genres
- Niche genres like Film-Noir, IMAX, and Western are rare, reflecting the long-tail of movie availability

Content Balance

The dataset is heavily skewed toward mainstream genres, highlighting the need for personalization to help users explore beyond the most common categories

Genre Popularity

Romance, Adventure, Crime and Sci-fi genres also have strong representation beyond Drama and Comedy

Movie Popularity

Musicals, War, and Documentaries remain limited

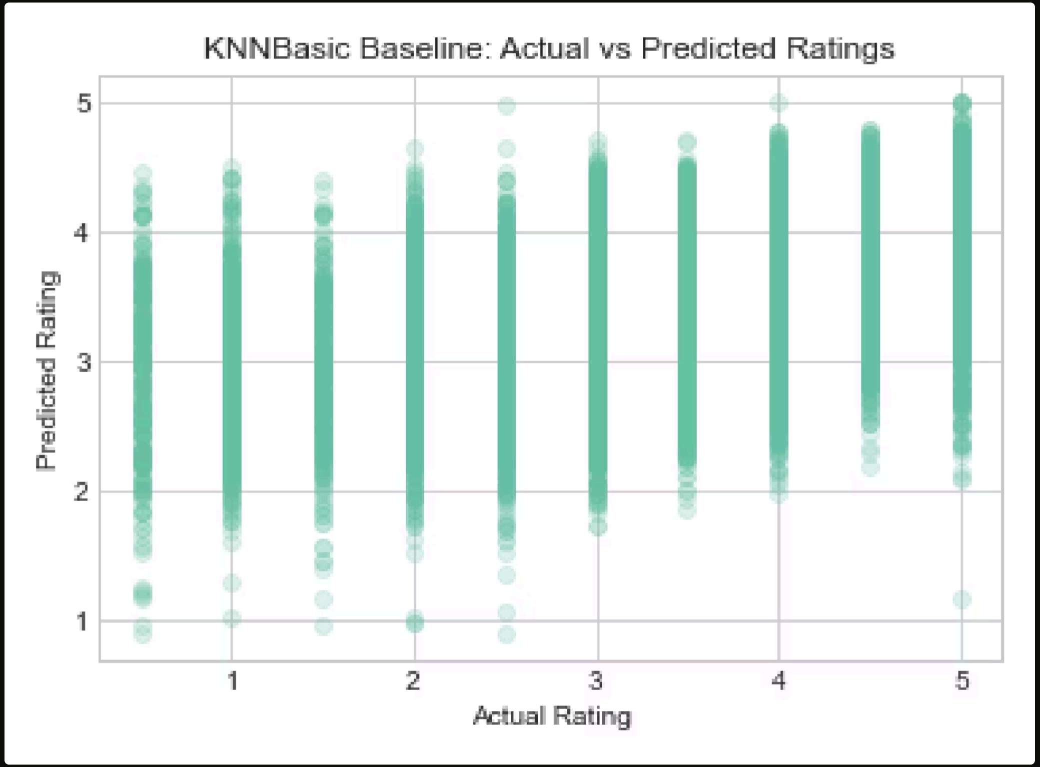
Baseline Modeling – KNN

KNN is intuitive — the model predicts ratings by looking at the k nearest users or items and averaging their ratings

Why baseline? It's simple, interpretable, and sets the benchmark floor

Results:

- **RMSE** ≈ 0.9 → On average, predictions differ from actual ratings by just under one star
- **MAE** ≈ 0.7 → Most predictions are within about three-quarters of a star of the true rating
- There is visible spread, meaning the model sometimes over- or under-predicts by 1–2 stars
- Overall, KNN captures general trends but lacks precision compared to more advanced models like SVD



Advanced Models – SVD & NMF

Matrix factorization (SVD, NMF) reduces high-dimensional rating data into latent features that capture hidden taste factors

SVD

Captures patterns like *Action vs Romance* preferences

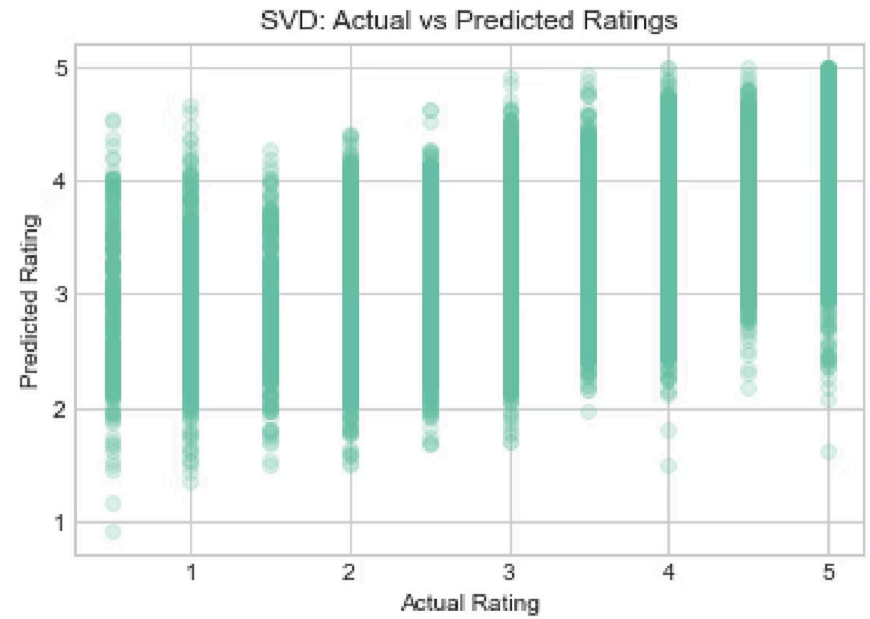
NMF

Ensures interpretability while improving prediction accuracy

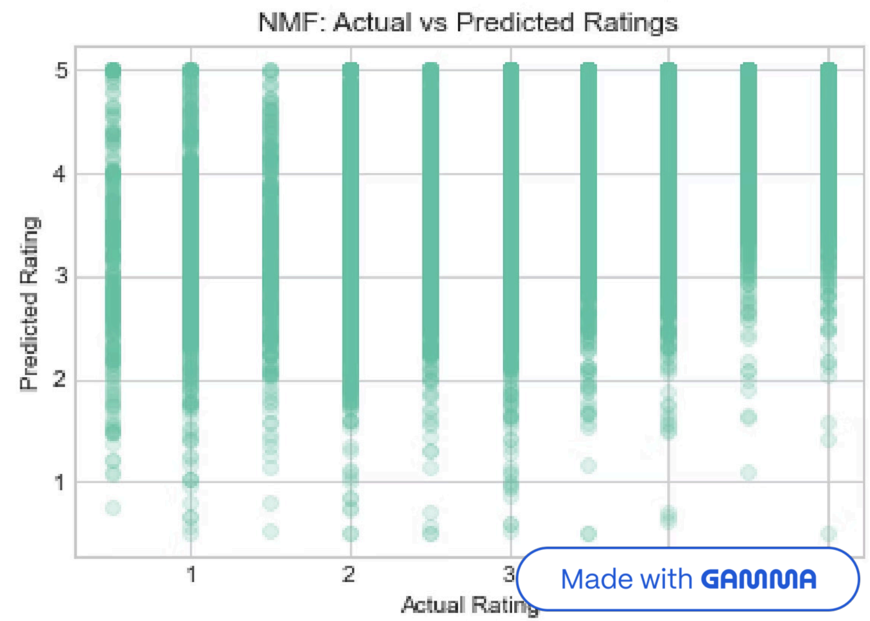
Results

Both outperform KNN in accuracy and Top-5 ranking quality

SVD: Actual vs Predicted Ratings



NMF: Actual vs Predicted Ratings



Model Comparison

- **KNN (Baseline):** RMSE ≈ 0.9, MAE ≈ 0.7 — predictions usually off by less than 1 star
- **NMF (Advanced):** Higher RMSE and MAE than KNN and SVD — less effective in capturing user preferences
- **SVD (Advanced):** Lowest RMSE (≈ 0.87) and MAE (≈ 0.67) — most accurate and consistent predictions

SVD clearly outperforms the baseline KNN model and NMF, reducing errors and making more reliable predictions. This demonstrates the business value of moving from a simple baseline to an advanced model

Business Value & Conclusion

Personalized recommendations: keep users engaged longer

Higher satisfaction → stronger loyalty & repeat usage

Helps **Sinema Pamoja** stand out with data-driven personalization

Next Steps:

Deploy

Deploy recommender prototype

Enhance

Add hybrid (content + collaborative) for new users

Integrate

Integrate into **Sinema Pamoja** platform

Thank You

We appreciate your time and attention

Open for questions and feedback