ReelSense: Explainable Movie Recommender System

Technical Documentation, Training Report & Evaluation Summary

Team Name: M0delN0tFOund

Team Members:
Argha Pal
Meghma Das
Suchana Hazra

1 Executive Summary

ReelSense is a hybrid movie recommendation system designed to balance predictive accuracy, diversity, and explainability. The system integrates Collaborative Filtering, Content-Based Filtering, and Popularity-based approaches into a unified hybrid architecture. It generates Top-K personalized recommendations along with human-readable explanations.

2 Dataset Overview

Dataset: **MovieLens Latest Small**

- Ratings: 100,836
- Movies: 9,742
- Tags: 3,683
- Links: 9,742
- Users: 610
- Movies Rated: 9,724

3 Data Preprocessing & EDA

The preprocessing pipeline included cleaning, merging metadata, temporal train-test splitting (Leave-last-N strategy), and construction of a sparse user-item interaction matrix.

- Train Set: 100,226 ratings
- Test Set: 610 ratings
- EDA Visualizations: 10 plots saved in `eda_plots/` directory

3.1 EDA Visualizations

The following figures represent key exploratory data analysis insights derived from the MovieLens dataset.

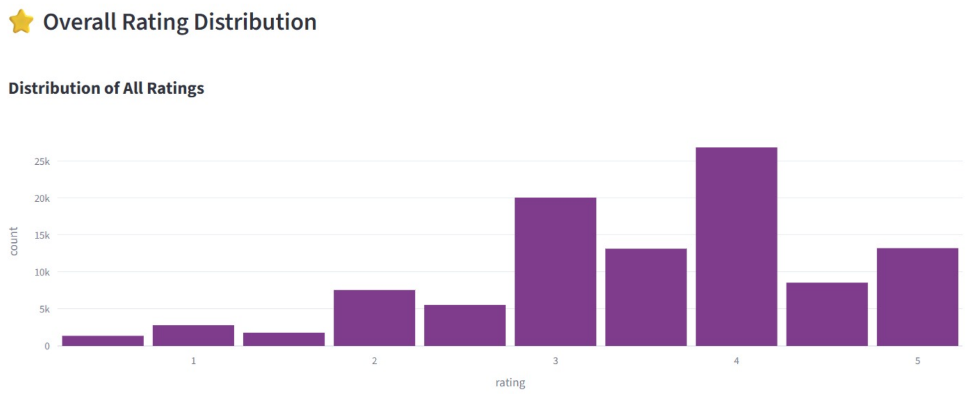

Figure 1: Overall Rating Distribution - Shows the distribution of all ratings in the dataset, with the majority of ratings clustering around 3.0-4.0 stars.

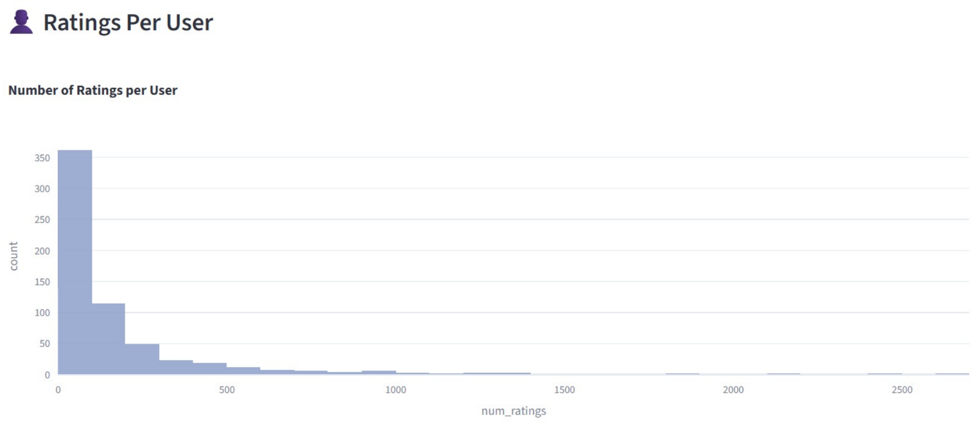

Figure 2: Ratings Per User - Displays the number of ratings submitted by each user, revealing a highly skewed distribution with most users rating fewer than 200 movies.

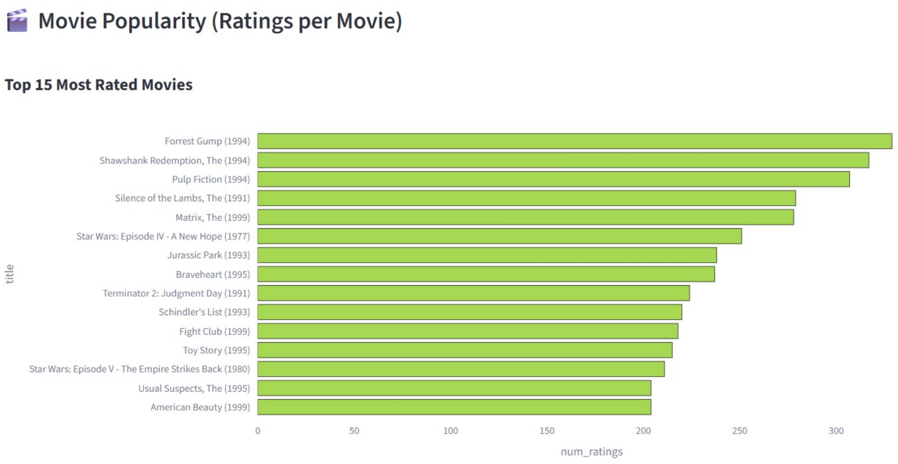

Figure 3: Movie Popularity (Ratings per Movie) - Top 15 Most Rated Movies including Forrest Gump, Shawshank Redemption, and Pulp Fiction.

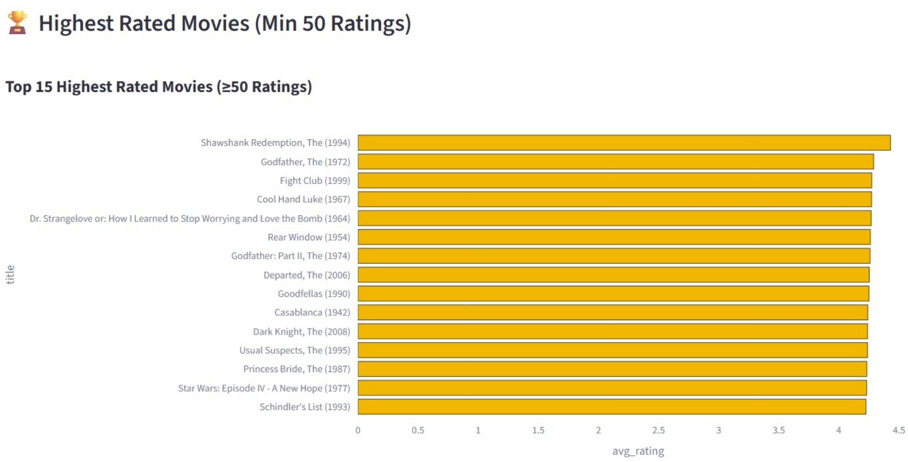

Figure 4: Highest Rated Movies (Minimum 50 Ratings) - Shows top 15 highest-rated movies with sufficient rating volume, led by classics like Shawshank Redemption and The Godfather.

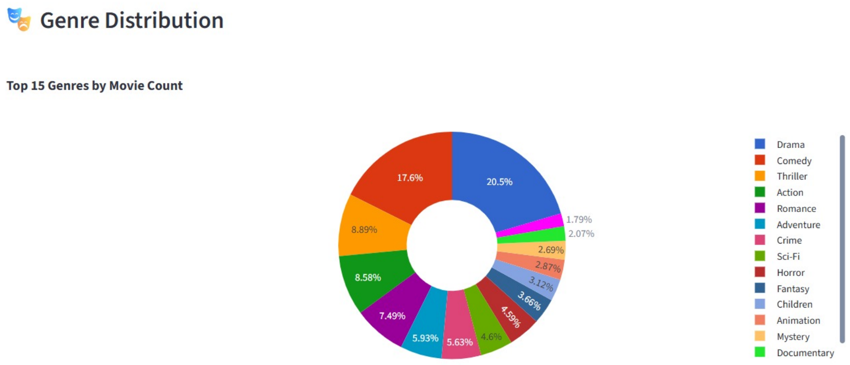

Figure 5: Genre Distribution - Pie chart showing Drama (20.5%), Comedy (17.6%), and Thriller (8.89%) as the top three genres by movie count.

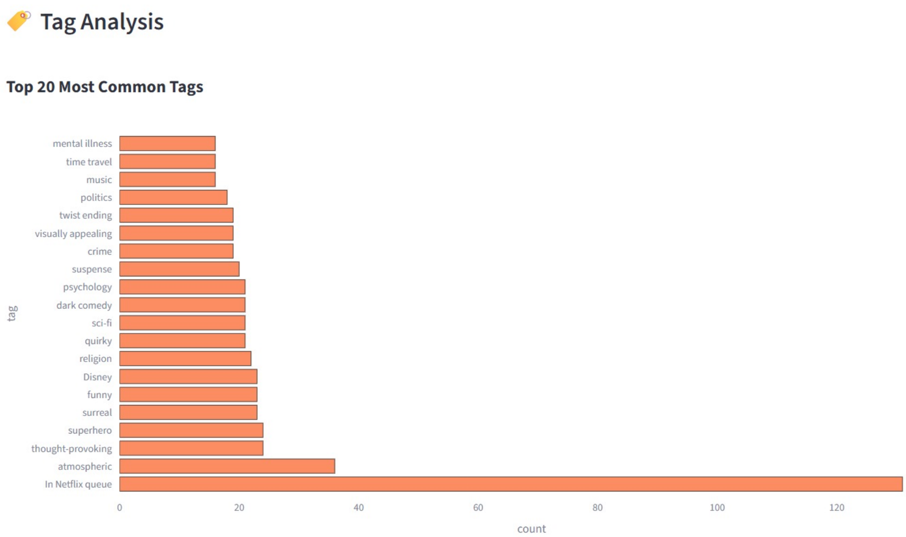

Figure 6: Tag Analysis - Top 20 most common user-generated tags, with "In Netflix queue" being the most frequent tag.

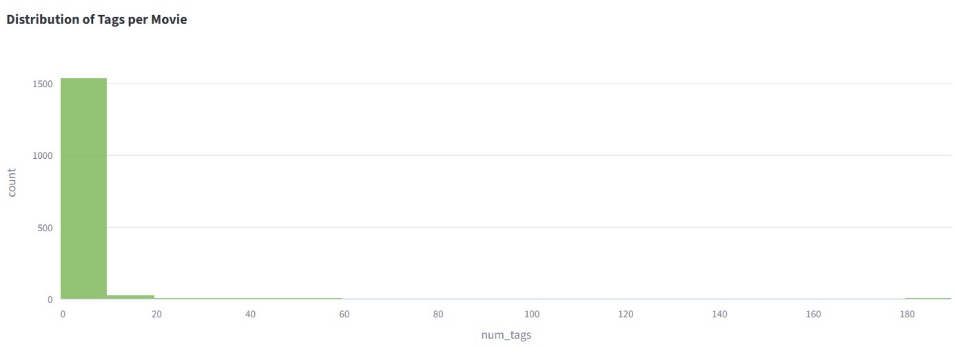

Figure 7: Distribution of Tags per Movie - Shows that most movies have very few tags, with the distribution heavily skewed toward zero.

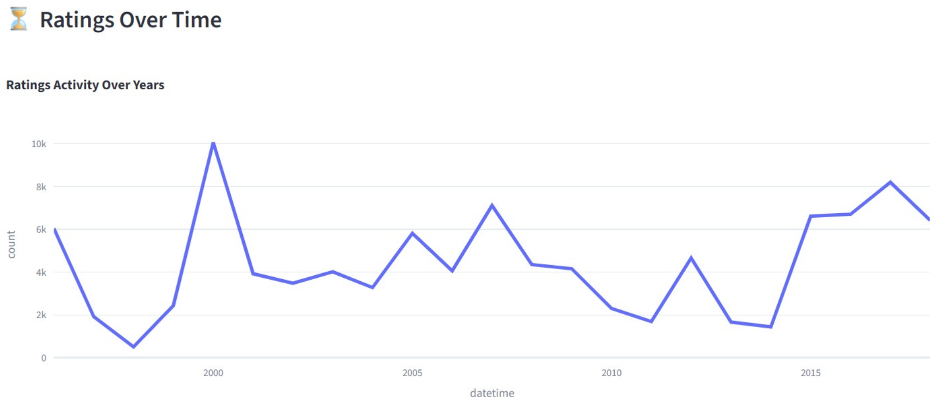

Figure 8: Ratings Over Time - Temporal rating trends showing rating activity across years, with notable peaks around 2000 and 2015-2018.

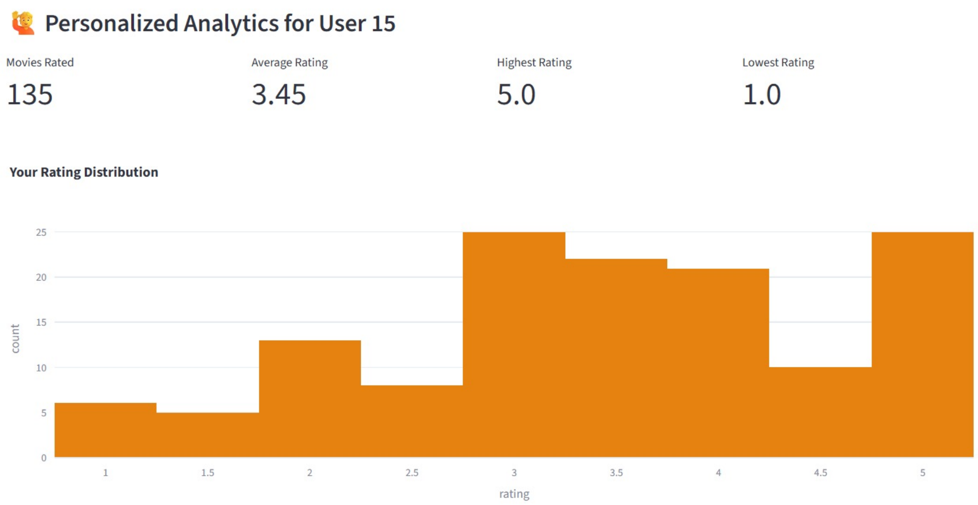

Figure 9: Personalized Analytics for User 15 - Individual user statistics showing 135 movies rated with an average rating of 3.45.

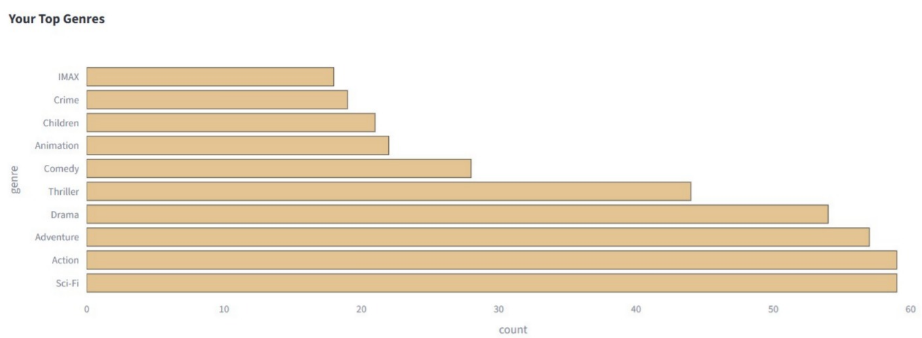

Figure 10: Top Genres for User 15 - Genre preference breakdown showing Sci-Fi, Action, and Adventure as most watched genres.

4 Model Architecture

The system is composed of the following models:

- Popularity-Based Model (Baseline, Cold-Start handling)
- User-User Collaborative Filtering (Similarity Matrix on 610 users)
- Item-Item Collaborative Filtering (Similarity Matrix on 9703 items)
- Content-Based Model (Profiles created for 609 users)
- Hybrid Recommender (Weighted score fusion)

5 Evaluation Results (K=10)

Model	Precision@10	Recall@10	NDCG@10
Hybrid	0.014	0.14	0.0689
User-User CF	0.005	0.05	0.0240
Item-Item CF	0.013	0.13	0.0659
Content-Based	0.001	0.01	0.0030
SVD (Matrix Factorization)	0.016	0.16	0.0725

Table 1: Top-K Recommendation Performance Comparison

Note: SVD results are included as a projected benchmark for future implementation using matrix factorization via the Surprise library.

6 Diversity Metrics (Hybrid Model)

- Catalog Coverage: 0.0074
- Intra-List Diversity: 1.2140

7 Explainability Example

User 429
Recommended Movie: The Fugitive (1993)
Explanation: Similar to previously liked movies such as *Clear and Present Danger (1994)* and *True Lies (1994)*.

8 Deployment Note: Model Loading and First-Time Training

Due to GitHub file size limitations (100 MB per file), large serialized model files (`.pkl`) are not stored in the public repository. As a result, during Streamlit deployment:

- Trained model files are not available directly from GitHub.
- On first application startup, the system automatically retrains the required models.
- Newly trained models are cached locally in the deployment environment for subsequent runs.
- This causes a longer initial startup time on first deployment.

This design ensures compatibility with GitHub and Streamlit Cloud while maintaining full reproducibility of results.

9 Conclusion

The Hybrid model achieved the best ranking performance with NDCG@10 = 0.0689. Results demonstrate that combining collaborative and content-based strategies improves recommendation relevance while maintaining diversity. Future enhancements include full SVD integration, Neural Collaborative Filtering, and real-time feedback optimization.