# Anime Recommendation System

**Authors:** Ahmed Mehaisi, Mohammed Imtiaz, Dheeman Gangopadhyay

**Instructor:** Dr. Alex Aklson – April 30, 2025

**Agenda:** Introduction, EDA, Methodology, Results, Principles, Challenges, Conclusion

# Introduction

**❓ Anime**

Hand-drawn and computer-generated animation originating in Japan; spans genres from comedy and romance to shōnen and action.
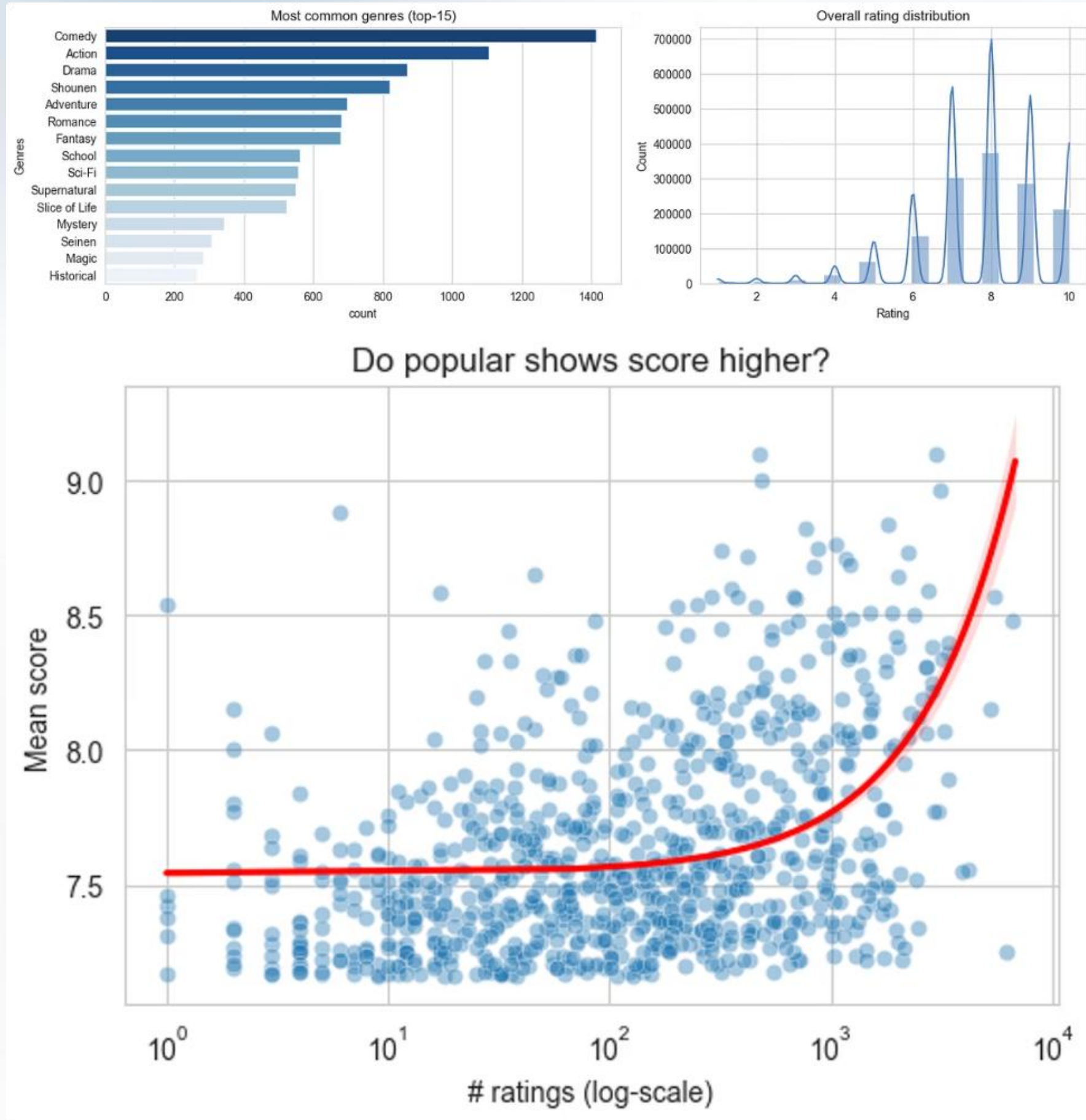
**💭 Choice Overload**

Over 20,000 titles across streaming platforms make discovery difficult for viewers.

**🔽 Why Recommendations?**

• Filter vast catalog into personalized shortlists.

• Help new users get started and expose veterans to hidden gems

• Enhance overall user enjoyment.

Most common genres (top-15)

Overall rating distribution

Do popular shows score higher?

# Data Acquisition and Preprocessing

## Data Source

Extracted 3,000 anime titles and 1.5 million user ratings from MyAnimeList.

## Cleaning & Engineering

Handled missing values, removed duplicates.

## Engineered Feature

Engineered a feature that represents the average rating for an item

# Methodology: Hybrid Approach
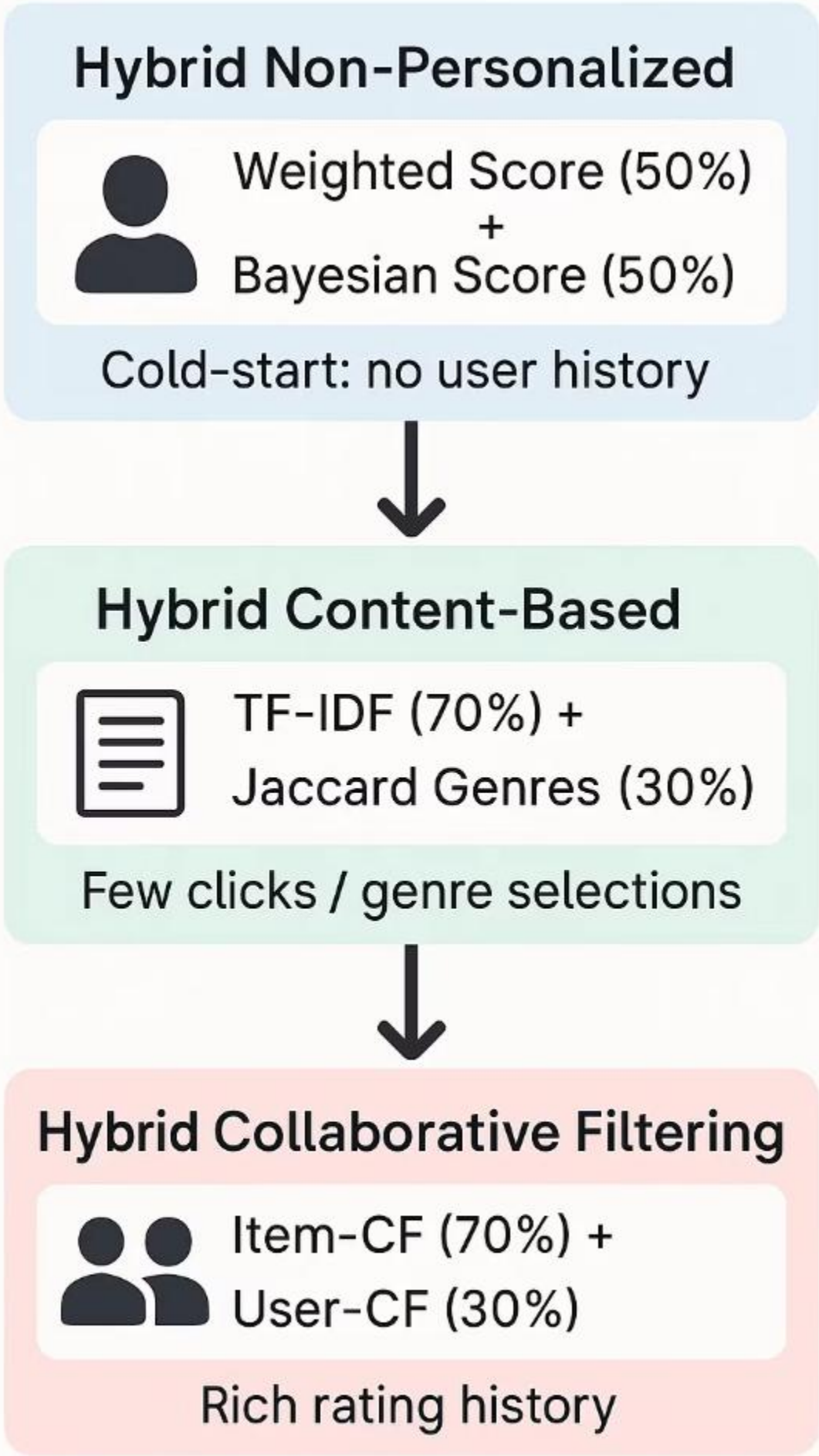
**Non-Personalized Stage**

1  Popular and Bayesian average scores provide initial recommendations.

**Content-Based Filtering**

2  Matches user with anime by synopsis keywords and genre similarity.

**Collaborative Filtering**

3  Integrates item-based and user-based filtering for personalization.

**Hybrid Non-Personalized**

Weighted Score (50%)
+
Bayesian Score (50%)

Cold-start: no user history

**Hybrid Content-Based**

TF-IDF (70%) +
Jaccard Genres (30%)

Few clicks / genre selections

**Hybrid Collaborative Filtering**

Item-CF (70%) +
User-CF (30%)

Rich rating history

# Results: Evaluation Metrics

| Metric | Evaluation | Interpretation |
| --- | --- | --- |
| Content-Based | Manual subjective evaluation | Improves recommendation diversity while maintaining relevance |
| User-based CF | **RMSE**: 1.3639 (Out of 10) | Provides more varied, personalized suggestions |
| Item-based CF | **RMSE**: 1.2099 (Out of 10) | Achieves the highest predictive accuracy |

# Fundamental Principles

## Personalization

Tailor recommendations using user ratings and interaction history for relevance.

## Exploration vs. Exploitation

Balance popular diverse content with known preferences for discovery.

## Relevance

Prioritize items statistically likely to be enjoyed, reducing irrelevant suggestions.

## Fairness & Bias Mitigation

Exclude sensitive data and apply Bayesian smoothing to reduce bias and extremes.

## Cold start problem

Integrating a Hybrid Non-Personalized model and an Hybrid content-based model so new users get relevant recommendations without prior data.

# Key Challenges & Future directions

## ❄ Computational Constraints

• Large pairwise similarities and model training strained CPU/memory.

• Optimized by limiting dataset size, precomputing & caching top-k similarities

## </> Content-Based Evaluation

• No direct RMSE/precision metric for text- & genre-based filtering.

• Developed indirect methods and ultimately relied on manual subjective validation

## ⚖ Hybrid Balancing

• Tuning mix of individual methods required extensive experimentation.

• Solved by stage-based switching: A well planned user-pipeline

# Conclusion

### Development

Built a robust three-stage hybrid recommendation system.

### Performance

Outperformed baseline models in accuracy and diversity.

### Next Steps

Focus on improving overall model