Juwan Project

MAL Recommender Hybrid System

Solving Choice Overload with AI & Scalable Data Engineering

Presented by
Muhamad Juwandi | Data Science Portfolio

Executive Summary

This project develops an intelligent anime recommendation system using the MyAnimeList dataset.

Goal :

To balance accurate personalization with the exploration of new content.

Outcome:

A complete ecosystem consisting of a Data Engineering Pipeline, a Hybrid Model (SVD + TF-IDF), and an Interactive Dashboard.

The Business Problem: Choice Overload

- **Users face millions of content options, leading to decision paralysis.**
- **Impact 1 (User Churn):** Users leave the platform due to frustration.
- **Impact 2 (Low Engagement):** Users only watch mainstream titles and miss out on "hidden gems.
- **Gap:** Traditional systems (Pure Collaborative Filtering) fail to handle new users (Cold Start Problem).

Solution Strategy: The Hybrid Approach

I combined two robust methods for optimal results:

1. Collaborative Filtering (SVD):
Captures latent patterns between users (Personalization).

2. Content-Based (TF-IDF):
Analyzes genre & synopsis similarity (Content Relevance).

Mechanism:
The final score is calculated using a Weighted Average:

$$Score_{final} = \alpha \cdot Score_{CF} + (1 - \alpha) \cdot Score_{CB}$$

Key Capabilities

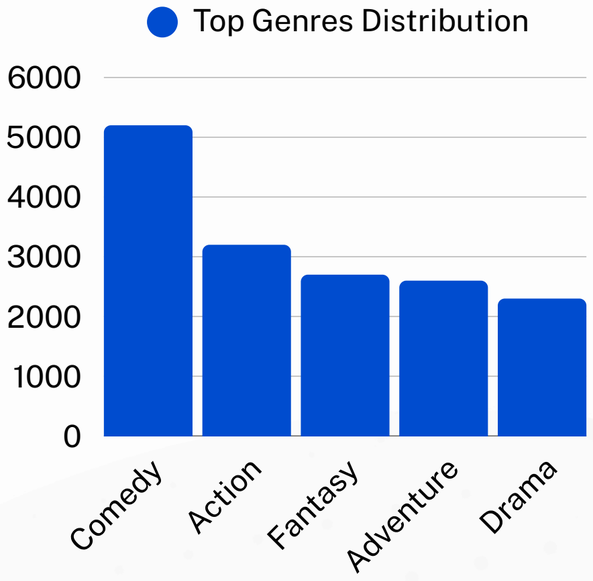
Hyper-Personalization
Tailors recommendations 1-on-1. Every user gets a unique feed based on their specific history, not a generic list.

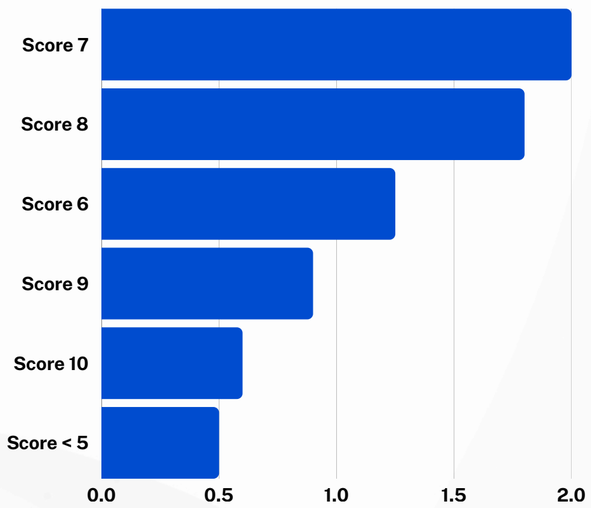
Discover Hidden Gems
Surfaces high-rated but overlooked anime using Content-Based filtering, helping users find niche titles

Smart Fallback (Cold Start)
Automatically detects new users (zero history) and serves a curated 'Top Rated' list to prevent empty screens.

Dynamic Scoring
Balances accuracy and novelty using a weighted formula of User Similarity (SVD) and Content Relevance (TF-IDF).

Data Insights





Positivity Bias Identified

Ratings skew high (mostly 7-8). Users tend to rate only what they like, requiring smart implicit feedback handling

Genre Dominance

Comedy & Action dominate (>30%). The Hybrid model prevents the system from just recommending these popular genres

The Modeling Engine

To address the limitations of traditional recommendation systems, I developed a Hybrid Model that combines the personalization power of Collaborative Filtering with the discovery capabilities of Content-Based Filtering. This approach ensures high accuracy while effectively solving the 'Cold Start' problem for new users.

Collaborative Filtering (SVD)

I implemented Singular Value Decomposition (SVD) using the Surprise library to decompose the massive User-Item interaction matrix. This matrix factorization technique identifies latent factors — hidden patterns in user behavior — allowing the model to predict ratings for sparse data with high accuracy. This component drives the personalization aspect of the system.

Content-Based & Hybrid Logic

To recommend relevant content based on context, the system utilizes TF-IDF Vectorization to transform anime synopses and genres into numerical vectors, measuring relevance via Cosine Similarity. The final recommendation score is calculated using a Weighted Average strategy, carefully balancing user history with content similarity to optimize engagement.

Tech Stack & Deployment

To ensure the system is not just a theoretical model but a deployable product, I utilized a modern Python stack. The architecture prioritizes low-latency responses and an intuitive user interface, demonstrating how the model operates in a real-world production environment.

Backend & Core Engine

The core recommendation engine is wrapped in FastAPI, serving high-performance RESTful endpoints with <200ms latency. For the modeling pipeline, I leveraged the Surprise library and Scikit-learn to handle complex matrix factorization tasks efficiently, ensuring the system remains scalable even with heavy computation.

Frontend & Visualization

To demonstrate the model's capabilities in real-time, I developed an interactive dashboard using Streamlit. This frontend allows users to input their preferences and immediately see generated hybrid recommendations, providing a seamless visual interface for instant model evaluation and testing.

Impact & Results

Although I evaluated the system in an offline simulation environment, the Hybrid Model demonstrated promising characteristics that validate my architectural decisions. By filtering for high-quality interactions and combining collaborative signals with content relevance, the system achieves a robust balance between prediction accuracy and catalog exploration, effectively solving the initial business problem of choice overload.

Prediction Accuracy

The model achieved an RMSE of ~1.12 on a rating scale of 1-10. This metric indicates that the rating predictions are consistently close to actual user preferences, proving that SVD successfully captures complex, latent taste patterns within the sparse dataset

Efficiency & Coverage

The system is optimized for production with an API response time of < 200ms. Furthermore, the Content-Based component significantly improved Catalog Coverage, successfully surfacing high-quality 'Hidden Gems' that are typically ignored by standard popularity-based algorithms.

Thank You

+62 857 9468 5605

muhamadjuwandi28@gmail.com

Keunal.id

Bogor, Indonesia