



MAL Recommender Hybrid System

Solving Choice Overload with AI & Scalable
Data Engineering

Presented by
Muhamad Juwandi | Data Science Portfolio



Executive Summary

Proyek ini mengembangkan sistem rekomendasi anime cerdas menggunakan dataset MyAnimeList.

Goal :

Menyeimbangkan personalisasi akurat dan eksplorasi konten baru.

Outcome:

Ekosistem lengkap yang terdiri dari Pipeline Data Engineering, Model Hybrid (SVD + TF-IDF), dan Dashboard Interaktif.





The Business Problem: Choice Overload

- ✓ Pengguna dihadapkan pada jutaan opsi konten, menyebabkan kebingungan dalam pengambilan keputusan (decision paralysis).
- ✓ Impact 1 (User Churn): Pengguna meninggalkan platform karena frustrasi.
- ✓ Impact 2 (Low Engagement): Pengguna hanya menonton judul mainstream dan melewatkannya "hidden gems".
- ✓ Gap: Sistem tradisional (Collaborative Filtering murni) gagal menangani pengguna baru (Cold Start).

Solution Strategy: The Hybrid Approach

Saya menggabungkan dua metode terbaik untuk hasil optimal:

1. Collaborative Filtering (SVD):

Menangkap pola latent antar pengguna (Personalisasi).

2. Content-Based (TF-IDF):

Menganalisis kemiripan genre & sinopsis (Relevansi Konten).

Mechanism:

Skor akhir dihitung menggunakan Weighted Average:

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



Key Capabilities



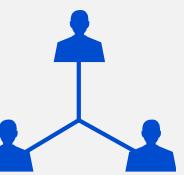
Hyper-Personalization

Rekomendasi disesuaikan secara 1-on-1. Setiap pengguna mendapatkan feed unik berdasarkan riwayat spesifik mereka, bukan daftar umum.



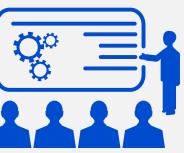
Discover Hidden Gems

Mengangkat anime dengan rating tinggi namun sering terlewatkan menggunakan Content-Based filtering, membantu pengguna menemukan judul niche.



Smart Fallback (Cold Start)

Secara otomatis mendeteksi pengguna baru (tanpa riwayat) dan menyajikan daftar 'Top Rated' terkurasi agar layar tidak kosong.

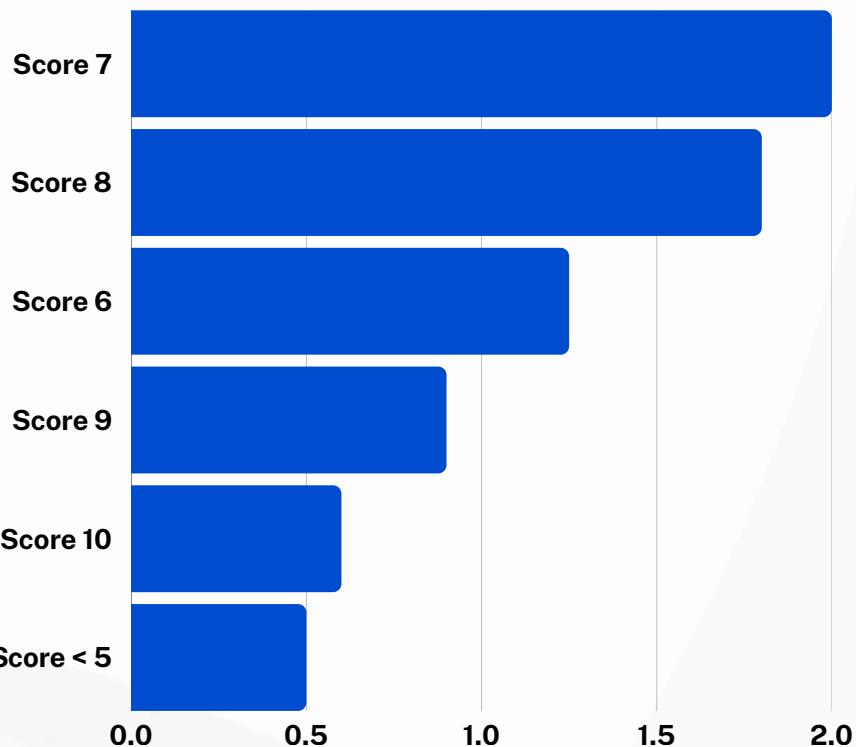
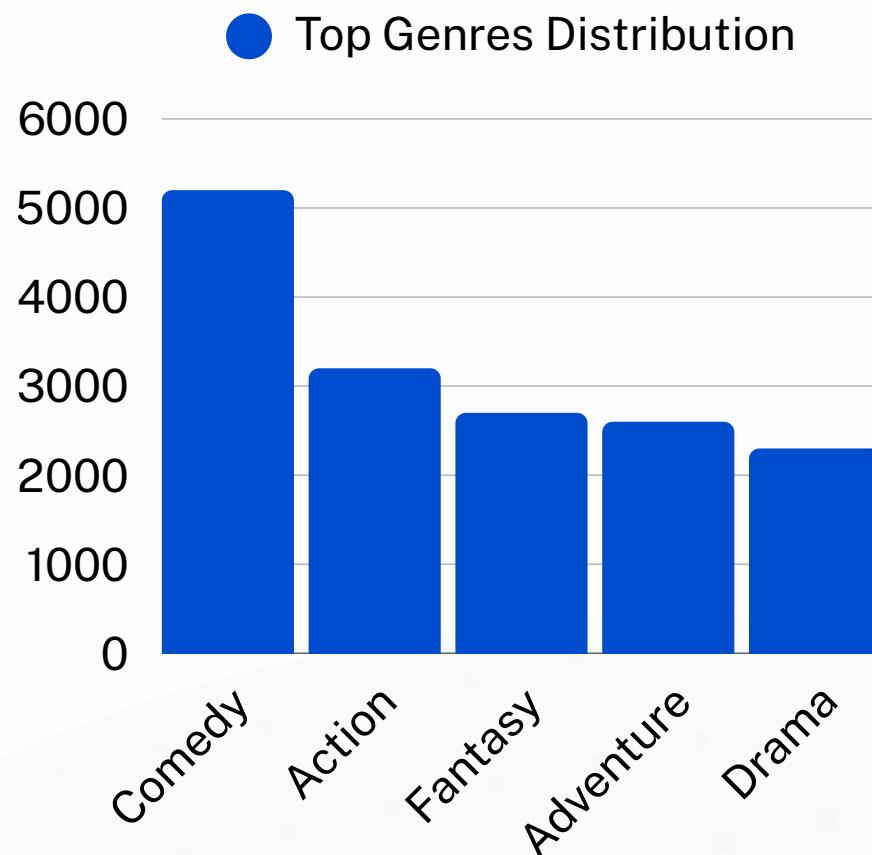


Dynamic Scoring

Menyeimbangkan akurasi dan kebaruan menggunakan rumus terbobot dari Kemiripan Pengguna (SVD) dan Relevansi Konten (TF-IDF).



Data Insights



Positivity Bias Identified

Distribusi rating sangat condong ke kiri (rata-rata 7-8). Pengguna cenderung hanya menilai apa yang mereka sukai, menuntut model untuk menangani umpan balik implisit dengan cerdas.

Genre Dominance

Judul Komedi & Aksi mendominasi dataset (>30%). Strategi Hybrid sangat krusial di sini untuk mencegah sistem hanya merekomendasikan genre populer ini kepada pengguna dengan selera unik.

The Modeling Engine

Untuk mengatasi keterbatasan sistem rekomendasi tradisional, saya mengembangkan Model Hybrid yang menggabungkan kekuatan personalisasi Collaborative Filtering dengan kemampuan penemuan Content-Based Filtering. Pendekatan ini memastikan akurasi tinggi sekaligus memecahkan masalah 'Cold Start'.

Collaborative Filtering (SVD)

Saya mengimplementasikan Singular Value Decomposition (SVD) menggunakan pustaka Surprise untuk memecah matriks interaksi User-Item yang masif. Teknik faktorisasi matriks ini mengidentifikasi faktor laten—pola tersembunyi dalam perilaku pengguna—yang memungkinkan model memprediksi rating untuk data yang jarang (sparse) dengan akurasi tinggi.

Content-Based & Hybrid Logic

Untuk merekomendasikan konten yang relevan secara konteks, sistem menggunakan Vektorisasi TF-IDF untuk mengubah sinopsis dan genre menjadi vektor numerik. Skor rekomendasi akhir dihitung menggunakan strategi Rata-rata Terbobot, menyeimbangkan riwayat pengguna dengan kemiripan konten.



Tech Stack & Deployment

Agar sistem ini bukan sekadar model teoritis melainkan produk yang siap pakai (deployable), saya menggunakan stack Python modern. Arsitekturnya memprioritaskan latensi rendah dan antarmuka pengguna yang intuitif.

Backend & Core Engine

Mesin rekomendasi inti dibungkus dalam FastAPI, menyediakan endpoint RESTful berkinerja tinggi dengan latensi <200ms. Untuk pipeline pemodelan, saya memanfaatkan pustaka Surprise dan Scikit-learn untuk menangani komputasi matriks yang kompleks secara efisien.

Frontend & Visualization

Untuk mendemonstrasikan kemampuan model secara real-time, saya mengembangkan dashboard interaktif menggunakan Streamlit. Frontend ini memungkinkan pengguna memasukkan preferensi dan langsung melihat rekomendasi hybrid yang dihasilkan secara visual.

Impact & Results

Meskipun evaluasi dilakukan dalam lingkungan simulasi offline, Model Hybrid menunjukkan karakteristik menjanjikan yang memvalidasi keputusan arsitektur saya. Sistem mencapai keseimbangan yang kuat antara akurasi prediksi dan eksplorasi katalog.



Prediction Accuracy

Model mencapai RMSE ~1.12 pada skala rating 1-10. Metrik ini menunjukkan bahwa prediksi rating secara konsisten mendekati preferensi pengguna yang sebenarnya, membuktikan SVD berhasil menangkap pola selera yang kompleks.

Efficiency & Coverage

Sistem dioptimalkan untuk produksi dengan waktu respon API < 200ms. Selain itu, komponen Content-Based secara signifikan meningkatkan Cakupan Katalog (Catalog Coverage), berhasil memunculkan 'Hidden Gems' berkualitas tinggi yang biasanya diabaikan oleh algoritma populer.





Thank You

+62 857 9468 5605

muhamadjuwandi28@gmail.com

Keunal.id

Bogor, Indonesia