

MALハイブリッド型アニメ推薦システム

AIとスケーラブルなデータエンジニアリングによる「選択肢过多」の解決

作成者:

ムハマド・ジュワンディ | データサイエンス・ポートフォリオ



Executive Summary (エグゼクティブ・サマリー)

本プロジェクトでは、MyAnimeListのデータセットを用いた知的アニメ推薦システムを開発しました。

目的：

正確なパーソナライズと、新規コンテンツ探索のバランスを取ること。

成果：

データパイプライン、ハイブリッドモデル (SVD + TF-IDF)、および対話型ダッシュボードからなる完全なエコシステム。





Business Problem (ビジネス課題)

- ✓ ユーザーは数百万ものコンテンツに直面し、どれを見るべきか選べない状況 (Decision Paralysis) に陥っている
- ✓ Impact 1 (User Churn): 選択疲れによるフラストレーションで、プラットフォームから離脱する。
- ✓ Impact 2 (Low Engagement): 有名作品ばかり視聴し、好みに合うはずの「隠れた名作」を見逃している。
- ✓ Gap: 従来のシステム (協調フィルタリングのみ) では、新規ユーザーへの対応 (コールドスタート問題) が困難。

解決策：ハイブリッド・アプローチ

最適な結果を得るため、2つの強力な手法を統合しました：

1. 協調フィルタリング (SVD):

ユーザー間の潜在的なパターンを特定（パーソナライズ）。

2. コンテンツベース (TF-IDF):

ジャンルやあらすじの類似性を分析（コンテンツの関連性）。

メカニズム：

最終スコアは「加重平均」を用いて算出：

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



Key Capabilities (主な機能)



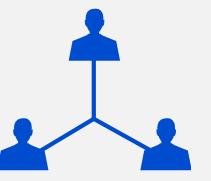
ハイパー・パーソナライゼーション

1対1の最適化。一般的なランキングではなく、各ユーザーの視聴履歴に基づいた独自のフィードを提供。



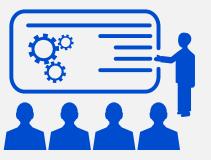
「隠れた名作」の発見

コンテンツベース手法により、高評価ながら見過ごされがちなニッチな作品（Hidden Gems）を提示。



コールドスタート対策

履歴のない新規ユーザーを自動検知し、厳選された「高評価リスト」を表示することで、画面が空白になるのを防ぐ。

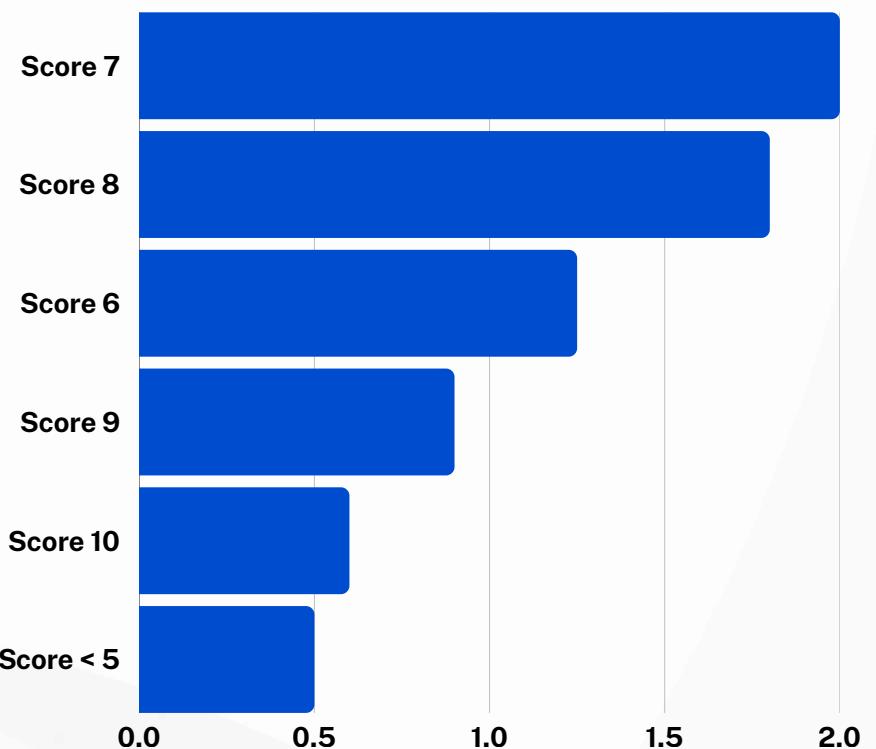
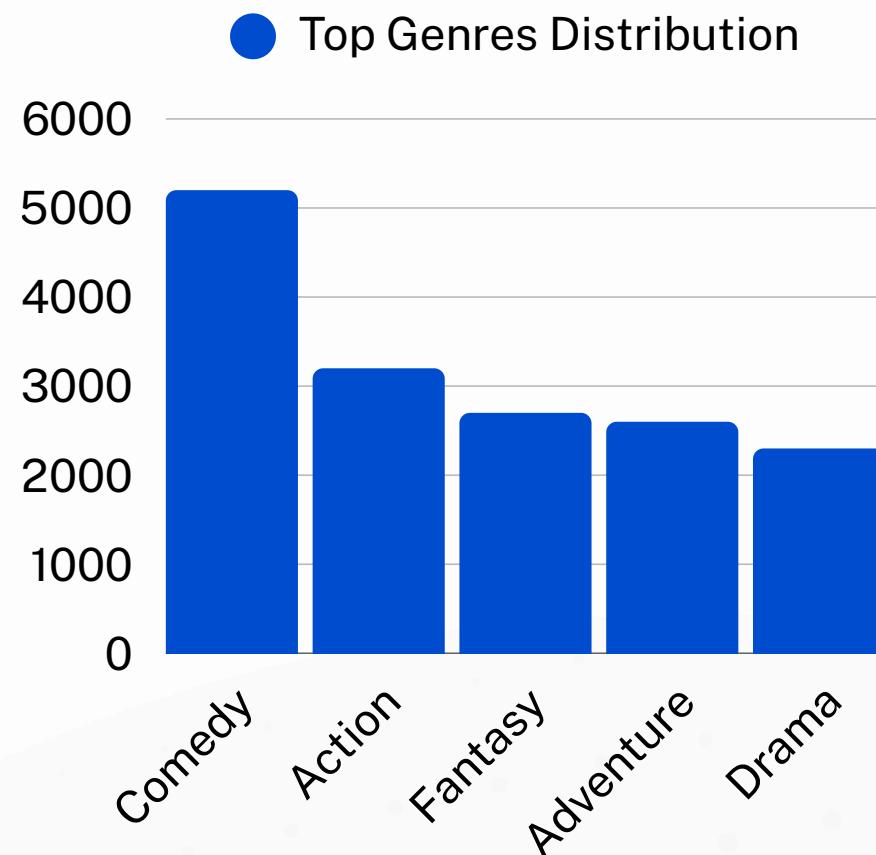


ダイナミック・スコアリング

ユーザー類似度(SVD)とコンテンツ関連度(TF-IDF)のバランスを動的に調整し、精度と新規性の両立を図る。



Data Insights (データ分析の洞察)



ポジティビティ・バイアス

高評価への偏り：評価分布は7～8点に偏っている。ユーザーは好みの作品しか評価しない傾向があるため、暗黙のフィードバック処理が重要となる。

ジャンルの偏り

特定ジャンルの支配：コメディとアクションが全体の30%以上を占める。ハイブリッド戦略は、人気ジャンルばかりが推薦されるのを防ぐために不可欠。

The Modeling Engine (モデリング・エンジン)

従来の推薦システムの限界を克服するため、協調フィルタリングの「パーソナライズ力」とコンテンツベースの「発見力」を融合したハイブリッドモデルを開発しました。これにより、高精度な予測とコールドスタート問題の解決を両立しています。

Collaborative Filtering (SVD)

ライブラリ Surprise を用い、**特異値分解 (SVD) を実装。大規模なユーザー/アイテム行列を分解し、ユーザー行動の潜在因子 (Latent Factors) **を特定することで、疎 (スペース) なデータセットからも高精度な予測を実現しました。

Content-Based & Hybrid Logic

文脈に基づいた推薦を行うため、TF-IDFベクトル化を用いてあらすじやジャンルを数値化し、コサイン類似度で関連性を測定。最終スコアは加重平均戦略で算出され、ユーザー履歴とコンテンツ類似度のバランスを最適化しています。

Tech Stack (使用技術)

単なる理論モデルではなく、実運用可能なプロダクトとして構築するため、モダンなPythonスタックを採用。低レイテンシと直感的なUIを重視しました。

Backend & Core Engine

推論エンジンにはFastAPIを採用し、200ms未満の高速レスポンスを実現。モデリングにはSurpriseとScikit-learnを活用し、複雑な行列計算を効率的に処理しています。

Frontend & Visualization

リアルタイムでのデモンストレーション用に、Streamlitを用いた対話型ダッシュボードを開発。ユーザーが好みを入力すると、即座に推薦結果が可視化されるインターフェースを構築しました。

Impact & Results (成果とインパクト)

オフライン・シミュレーション環境での評価において、本ハイブリッドモデルは有望な結果を示し、アーキテクチャの有効性が実証されました。予測精度とカタログ探索性のバランス確保に成功しています。



Prediction Accuracy(予測精度)

RMSE（二乗平均平方根誤差）約1.12（10段階評価）を達成。この数値は、モデルの予測が実際のユーザーの好みに極めて近いことを示しており、SVDが複雑な好みのパターンを正確に捉えていることを証明しています。

Efficiency & Coverage (効率性とカバレッジ)

API応答速度は200ms未満で、プロダクション環境に最適化されています。また、コンテンツベース要素の導入によりカタログカバレッジが大幅に向上し、既存の人気ランキングでは埋もれてしまう良作の提示に成功しました。





Thank You

+62 857 9468 5605

muhamadjuwandi28@gmail.com

Keunal.id

Bogor, Indonesia