



## PEMFLITERAN BERBASIS KONTEN YANG DAPAT MENENTUKAN TOP-N REKOMENDASI ANIME BAGI PENGGUNA

**Hemdani Rahendra Herlianto**

Universitas Pelita Bangsa, Indonesia

[hemdani.rahendra@pelitabangsa.ac.id](mailto:hemdani.rahendra@pelitabangsa.ac.id)

### Abstrak:

Permasalahan dalam memberikan rekomendasi anime yang tepat kepada pengguna, khususnya pengguna baru, menjadi tantangan bagi platform streaming anime karena minimnya informasi historis. Penelitian ini dilakukan untuk mengembangkan sistem rekomendasi anime yang mampu mengatasi kendala tersebut menggunakan pendekatan content-based filtering. Tujuan utama penelitian ini adalah menghasilkan model rekomendasi yang dapat memberikan saran anime secara akurat berdasarkan kemiripan konten tanpa memerlukan masukan awal dari pengguna baru. Penelitian ini menggunakan pendekatan kuantitatif dengan analisis univariat dan pemanfaatan metode TF-IDF dalam ekstraksi fitur genre dari dataset anime dan rating pengguna. Data diolah melalui tahapan praproses, pembuatan model, dan evaluasi rekomendasi. Hasil penelitian menunjukkan bahwa sistem mampu memberikan rekomendasi top-5 anime yang relevan berdasarkan kesamaan genre dengan tingkat kemiripan yang tinggi, seperti yang ditunjukkan pada hasil rekomendasi terhadap anime *Naruto*. Temuan ini menegaskan bahwa content-based filtering efektif digunakan untuk menyusun rekomendasi personalisasi bagi pengguna. Implikasi dari penelitian ini memberikan kontribusi nyata dalam pengembangan sistem rekomendasi pada platform streaming dan dapat diperluas pada domain hiburan digital lainnya.

**Kata Kunci:** Sistem rekomendasi anime, *content-based filtering*, analisis data, pengembangan model, evaluasi model

### Abstract:

*The challenge of providing accurate anime recommendations to users, especially new ones, is a significant issue for anime streaming platforms due to the limited availability of historical data. This research aimed to develop an anime recommendation system that can overcome these*

*obstacles by utilizing a content-based filtering approach. The primary objective of this research is to develop a recommendation model that accurately provides anime suggestions based on content similarities, without requiring initial input from new users. This study employs a quantitative approach, utilizing univariate analysis and the TF-IDF method to extract genre features from anime datasets and user ratings. Data is processed through pre-processing, modeling, and recommendation evaluation. The results of the study show that the system is able to provide recommendations for relevant top-5 anime based on genre similarity, with a high level of similarity, as shown in the results of recommendations for the Naruto anime. These findings confirm that content-based filtering is an effective method for compiling personalized recommendations for users. The implications of this study make a real contribution to the development of recommendation systems on streaming platforms and can be extended to other digital entertainment domains.*

**Keywords:** anime recommendation system, content-based filtering, data analysis, model development, model evaluation

## **Pendahuluan**

Animasi adalah proses menciptakan gerakan dengan mengubah gambar secara berkelanjutan dalam waktu tertentu, disertai elemen suara, ekspresi emosional, dan karakter. Istilah "anime" secara umum di Jepang merujuk pada semua bentuk animasi, tetapi secara global mengacu pada animasi khas Jepang yang memiliki gaya visual dan naratif yang berbeda dari animasi barat seperti Amerika (Pinontoan, 2024; Soni et al., 2023). Anime telah menjadi fenomena global dengan jutaan penggemar yang mengaksesnya melalui platform streaming. Namun, pengguna sering kesulitan menemukan anime yang sesuai dengan preferensi mereka karena keterbatasan informasi, seperti deskripsi yang minim dan ulasan pengguna yang tidak memadai. Masalah ini menjadi signifikan karena berpengaruh langsung pada kepuasan pengguna dan tingkat keterlibatan mereka dalam platform (Caswito et al., 2024; Girsang et al., 2020).

Untuk menjawab tantangan tersebut, sistem rekomendasi berbasis konten (content-based filtering) menjadi salah satu solusi yang dapat digunakan. Sistem ini menyajikan rekomendasi berdasarkan kesamaan fitur dari konten yang telah disukai pengguna sebelumnya. Meski menjanjikan, implementasi pendekatan ini dalam domain anime memiliki tantangan tersendiri, seperti kompleksitas genre, variasi preferensi individu, serta masalah cold-start untuk pengguna baru. Studi sebelumnya telah menerapkan pendekatan serupa pada rekomendasi film dan musik, namun sedikit yang fokus secara khusus pada anime dengan pendekatan TF-IDF untuk genre (Prakash et al., 2022; Roziqin, 2024)

Literatur yang ada menunjukkan bahwa sistem rekomendasi terbagi dalam tiga pendekatan utama: collaborative filtering, content-based filtering, dan hybrid.

Collaborative filtering mengandalkan data historis antar pengguna, namun tidak efektif untuk pengguna baru. Hybrid menggabungkan dua pendekatan namun lebih kompleks dan membutuhkan lebih banyak data. Oleh karena itu, content-based filtering dipilih dalam studi ini karena mampu memberikan hasil personalisasi meskipun informasi pengguna terbatas (Ramu, 2022).

Penelitian ini memiliki kebaruan dalam penerapan teknik TF-IDF pada dataset anime untuk menghasilkan sistem rekomendasi yang relevan, akurat, dan efisien. Tidak seperti studi sebelumnya yang lebih banyak fokus pada penggabungan metode, penelitian ini menitikberatkan pada pengolahan genre sebagai fitur utama melalui teknik pemodelan konten murni.

Dengan demikian, tujuan dari penelitian ini adalah merancang dan menguji sistem rekomendasi anime berbasis content-based filtering yang mampu memberikan rekomendasi top-N secara akurat, bahkan kepada pengguna baru tanpa riwayat preferensi. Kontribusi utama dari penelitian ini adalah penyederhanaan proses rekomendasi berbasis konten dalam domain hiburan digital dengan tingkat akurasi tinggi namun tetap efisien untuk diterapkan di dunia nyata.

## Metode

### A. Data Understanding & Preprocessing

Terdapat dua dataset, yaitu dataset anime yang berisi setiap judul anime beserta dengan genre, tipe, episode, rata-rata rating, dan jumlah anggota komunitas anime tersebut dan dataset anime\_rating yang berisi pemberian rating anime dari setiap user.

#### 1. Dataset anime

- Dataset memiliki *format CSV*.
- Dataset memiliki 12294 sample dengan 7 fitur.
- Dataset memiliki 1 fitur bertipe *float(64)*, 2 fitur bertipe *int(64)*, dan 4 fitur bertipe object.
- Terdapat missing value pada dataset.

#### 2. Dataset anime\_rating

- Dataset memiliki *format CSV*.
- Dataset memiliki 7813737 sample dengan 3 fitur.
- Dataset memiliki 1 fitur bertipe *int(64)*.
- Tidak ada missing value pada dataset.

### B. Variable - Variable Pada Dataset

#### 1. Dataset anime

- anime\_id = ID unik dari setiap anime.
- name = Judul anime.
- genre = genre anime.
- type = Tipe tayang anime, seperti TV, OVA, etc.

- e. episodes = Banyaknya episode setiap anime.
- f. rating = Rata-rata rating setiap anime terhadap jumlah user yang memberi rating.
- g. members = Jumlah anggota komunitas setiap anime.

## 2. Dataset anime\_rating

- a. user\_id = ID unik dari setiap user
- b. anime\_id = ID dari anime yang diberi peringkat oleh user
- c. rating = Rating yang diberikan oleh user.

## C. Data preprocessing

- a. Menghapus missing value pada dari dataset anime.
- b. Menghapus duplikat sample dari rating anime dataset.
- c. Menghapus symbol pada judul anime.

## D. Univariate Analysis

*Univariate Analysis* adalah menganalisis setiap fitur secara terpisah.

**Tabel 1.** Analisis setiap atribut dataset anime

	anime_id	rating	members
count	12017.00	12017.00	12017.00
mean	13638.00	6.48	18348.88
std	11231.08	1.02	55372.50
min	1.00	1.67	12.00
25%	3391.00	5.89	225.00
50%	9959.00	6.57	1552.00
75%	23729.00	7.18	9588.00
max	34519.00	10.00	1013917.00

Dataset anime memiliki rating anime terendah 1.67 dan rating tertinggi 10 dengan rata-rata 6.48. Dalam dataset juga memiliki jumlah anggota komunitas anime terendah 12 dan terbanyak 1013917 dengan rata-rata 18348. Perbedaan nilai min dan max dari jumlah anggota komunitas anime cukup jauh dan hal ini wajar dikarenakan beberapa anime memang sangat populer dan beberapa tidak.

**Tabel 2.** Analisis setiap atribut numerik dataset anime\_rating

	user_id	anime_id	rating
count	7813736.00	7813736.00	7813736.00
mean	36727.96	8909.07	6.14
std	20997.95	8883.95	3.73
min	1.00	1.00	-1.00
25%	18974.00	1240.00	6.00
50%	36791.00	6213.00	1552.00
75%	54757.00	14093.00	9.00
max	73516.00	34519.00	10.00

Dataset rating anime memiliki rating terendah yang diberikan user pada suatu anime adalah -1 dan rating tertinggi adalah 10. Rating -1 menandakan bahwa user menonton anime, namun tidak memberikan rating. Sample user yang tidak memberikan rating tidak akan digunakan sehingga dihapus.

## Hasil dan Pembahasan

### 1. Content Based Filtering

Sistem rekomendasi berbasis konten adalah sistem yang merekomendasikan konten yang mirip dengan konten yang disukai pengguna sebelumnya. Apabila suatu konten memiliki karakteristik yang sama atau hampir sama dengan konten lainnya, maka kedua konten tersebut dapat dikatakan mirip (Fajriansyah et al., 2021; Tyas, 2017). Misalkan dalam sistem rekomendasi anime, jika pengguna menyukai anime Jujutsu Kaisen, sistem dapat merekomendasikan anime dengan genre action lainnya.



**Gambar 1.** Sistem *Content Based Filtering*

## 2. *TF-IDF*

*TF-IDF* atau *Term Frequency - Inverse Document Frequency*, Bahasa adalah ukuran statistik yang menggambarkan pentingnya istilah dalam dokumen dalam kumpulan atau korpus. Metrik ini biasanya digunakan sebagai faktor pembobotan untuk pencarian informasi, penambahan teks, dan pemodelan pengguna. Nilai *TF-IDF* meningkat secara linier dengan jumlah kemunculan suatu term dan bergantung pada jumlah dokumen dalam korpus yang memuat term tersebut (Fajriansyah et al., 2021; Sani et al., 2022). *TF-IDF* digunakan pada sistem rekomendasi anime untuk menentukan representasi fitur penting dari setiap genre anime. Untuk menjalankan *TF-IDF* digunakan fungsi `tfidfvectorizer()` dari library *sklearn* (Fahrezi, 2024; Feurer et al., 2020).

Setelah menghitung nilai *TF-IDF* untuk setiap kata kunci dalam dokumen, langkah selanjutnya adalah mentransformasikan hasil tersebut ke dalam bentuk matriks menggunakan fungsi `to_dense()`. Proses transformasi ini penting karena mengubah representasi data dari format yang berbasis teks menjadi format matriks yang lebih mudah diproses oleh model pemodelan atau analisis selanjutnya. Dengan matriks *TF-IDF* yang dihasilkan, informasi penting tentang kepentingan relatif setiap kata kunci dalam dokumen dapat diakses dan dimanfaatkan dalam berbagai aplikasi, seperti klasifikasi teks, kategorisasi dokumen, atau pencarian informasi. Dengan demikian, proses transformasi matriks *TF-IDF* menjadi langkah penting dalam analisis teks yang holistik dan informatif (Rofiqi et al., 2019; Yuana et al., 2018).

Dataframe baru dibuat untuk menunjukkan matriks *TF-IDF* untuk beberapa anime dan genre. Semakin tinggi nilai matriks menunjukkan semakin erat

hubungan antara anime dengan genre tersebut. Misalkan anime *Asa da yo! Kaishain* merupakan *genre comedy* terlihat dari nilai matriks 1.0 yang didapat anime tersebut pada genre comedy.

**Tabel 3.** *Dataframe* baru matriks *TF-IDF*

	school	slice	sci	vampire	power	shounen	military	life	samurai	parody	...	ecchi	horror	comedy	fa
name															
Bucket no Ana	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.0	...	0.000000	0.0	0.000000	0.0
Tonagural	0.479050	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.0	...	0.569369	0.0	0.283324	0.0
Tenkai Knights	0.000000	0.000000	0.278399	0.0	0.429446	0.293865	0.000000	0.000000	0.0	0.0	...	0.000000	0.0	0.197200	0.0
Toshokan Sensou: Koi no Shougai	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.638911	0.000000	0.0	0.0	...	0.000000	0.0	0.287964	0.0
Noisy Birth	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.0	...	0.000000	0.0	0.000000	0.0
Kagaku na Yatsura	0.390037	0.000000	0.325664	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.0	...	0.463574	0.0	0.230680	0.0
Shin Maple Town Monogatari: Palm Town-hen	0.000000	0.522789	0.000000	0.0	0.000000	0.000000	0.000000	0.522789	0.0	0.0	...	0.000000	0.0	0.000000	0.4
Super Robot Taisen OG: Divine Wars Special	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.0	...	0.000000	0.0	0.413276	0.0
Asa da yo! Kaishain	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.0	...	0.000000	0.0	1.000000	0.0
Summer	0.000000	0.577350	0.000000	0.0	0.000000	0.000000	0.000000	0.577350	0.0	0.0	...	0.000000	0.0	0.000000	0.0

Fungsi *anime\_recommendations* dibuat untuk menemukan rekomendasi anime menggunakan *similarity* yang telah didefinisikan sebelumnya. Fungsi ini bekerja dengan cara mengambil anime dengan *similarity* terbesar dari index yang ada (Sitorus, 2023; Yuana et al., 2018)Selanjutnya adalah menemukan rekomendasi yang mirip dengan anime Naruto :

**Tabel 4.** Hasil rekomendasi anime Naruto

anime_id	name	genre	type	episodes	rating	members
20	Naruto	Action, Comedy, Martial Arts, Shounen, Super Power	TV	220	7.81	683297

**Tabel 5.** Hasil top 5 rekomendasi film naruto

name	genre
Naruto: Shippuuden Movie 4 - The Lost Tower	Action, Comedy, Martial Arts, Shounen, Super Power
Naruto Shippuuden: Sunny Side Battle	Action, Comedy, Martial Arts, Shounen, Super Power
Boruto: Naruto the Movie - Naruto ga Hokage ni Natta Hi	Action, Comedy, Martial Arts, Shounen, Super Power
Naruto x UT	Action, Comedy, Martial Arts, Shounen, Super Power
Naruto: Shippuuden	Action, Comedy, Martial Arts, Shounen, Super Power

Sistem telah berhasil merekomendasikan top 5 persen anime yang mirip dengan naruto, yaitu beberapa film dan seri dari naruto itu sendiri. Jadi, jika pengguna menyukai naruto, maka sistem dapat merekomendasikan seri atau movie naruto lainnya.

### **Kesimpulan**

Sistem rekomendasi anime yang optimal diarahkan untuk didasarkan secara eksklusif pada sejarah pemutaran pengguna. Secara konseptual, sistem ini mampu mengevaluasi kesamaan antara profil pengguna dan menyediakan prediksi atas kemungkinan kepuasan pengguna terhadap anime tertentu. Kendati demikian, karena batasan perhitungan yang berkaitan dengan volume data yang besar, hanya 100 ribu sampel data yang terpilih untuk dijadikan dasar analisis. Dengan adopsi mesin yang lebih canggih, eksperimen dengan sampel data hingga 1 juta dapat dilakukan. Makalah ini mengilustrasikan pendekatan yang simpel namun efektif dalam pengembangan sistem rekomendasi yang relevan.

### **Daftar Pustaka**

- Caswito, A., Sayidah, N., & Adi, S. U. (2024). Faktor Penggunaan E-Commerce Terhadap Pengambilan Keputusan melalui Perilaku Konsumen (Kajian Studi Literatur Review). *Jurnal Minfo Polgan*, 13(1), 1–8.
- Fahrezi, I. A. (2024). *Analisis Sentimen Twitter terhadap Opini Publik atas Isu Hak Angket Menggunakan Ekstraksi Fitur TF-IDF dan Algoritma Support Vector Machine*.
- Fajriansyah, M., Adikara, P. P., & Widodo, A. W. (2021). Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 5(6), 2188–2199.

- Feurer, M., Eggenesperger, K., Falkner, S., Lindauer, M., & Hutter, F. (2020). Auto-sklearn 2.0: The next generation. *ArXiv Preprint ArXiv:2007.04074*, 24(8).
- Girsang, A. S., Al Faruq, B., Herlianto, H. R., & Simbolon, S. (2020). Collaborative recommendation system in users of anime films. *Journal of Physics: Conference Series*, 1566(1), 12057.
- Pinontoan, S. Z. R. (2024). *Pengaruh Globalisasi Terhadap Penerimaan Fujoshi Dalam Komunitas Lgbtq+ Di Indonesia Tahun 2020-2024*. Universitas Nasional.
- Prakash, V., Raghav, S., Sood, S., Pandey, M., & Arora, M. (2022). Deep Anime Recommendation System: Recommending Anime Using Collaborative and Content-based Filtering. *2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, 718–723.
- Ramu, P. (2022). *Deep Learning-Based Anime and Movie Recommendation System*. Dublin, National College of Ireland.
- Rofiqi, M. A., Fauzan, A. C., Agustin, A. P., & Saputra, A. A. (2019). Implementasi Term-Frequency Inverse Document Frequency (TF-IDF) Untuk Mencari Relevansi Dokumen Berdasarkan Query. *ILKOMNIKA*, 1(2), 58–64.
- Roziqin, N. M. (2024). *Sistem rekomendasi pemilihan film animasi menggunakan metode Content Based Filtering*. Universitas Islam Negeri Maulana Malik Ibrahim.
- Sani, R. R., Pratiwi, Y. A., Winarno, S., Udayanti, E. D., & Alzami, F. (2022). Analisis Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Berita Hoax pada Berita Online Indonesia. *Jurnal Masyarakat Informatika*, 13(2), 85–98.
- Sitorus, M. R. (2023). *Implementasi Model-Based Collaborative Filtering pada Sistem Rekomendasi Film Menggunakan Algoritma SVD (Singular Value Decomposition)*.
- Soni, B., Thakuria, D., Nath, N., Das, N., & Boro, B. (2023). RikoNet: A Novel Anime Recommendation Engine. *Multimedia Tools and Applications*, 82(21), 32329–32348.
- Tyas, A. S. P. (2017). Identifikasi kuliner lokal Indonesia dalam

pembelajaran bahasa Inggris. *Jurnal Pariwisata Terapan*, 1(1), 38–51.

Yuana, R. A., Harjunowibowo, D., Karyanta, N. A., & Budiyanoto, C. W. (2018). Data similarity filtering of Wartegg Personality Test result using cosine-similarity. *International Journal of Recent Contributions from Engineering, Science & IT (IJES)*, 6(3), 19–28.