



IMPLEMENTASI METODE CONTENT BASED FILTERING MENGGUNAKAN SYNOPSIS SIMILARITY UNTUK PEMILIHAN ANIME

Wahyurrahman Ferbiansyah Adha Muhammad Irwan^{1*}, Budi Santoso², Slamet Kacung³

^{1,2,3}Teknik Informatika, Universitas Dr. Soetomo Surabaya

email: wahyurrahmanf@gmail.com^{1*}

Abstrak: Penelitian ini bertujuan mengembangkan sistem rekomendasi anime berbasis Content-Based Filtering dengan memanfaatkan kesamaan sinopsis sebagai parameter utama. Latar belakang penelitian didasari oleh keterbatasan sistem rekomendasi yang hanya mengandalkan genre atau popularitas, tanpa mempertimbangkan kesesuaian naratif secara mendalam. Data sinopsis anime genre fantasi diambil dari dataset MyAnimeList dan diproses melalui tahapan text preprocessing, tokenisasi, dan representasi numerik menggunakan metode TF-IDF. Kemudian, kemiripan antar sinopsis dihitung menggunakan Cosine Similarity untuk menghasilkan rekomendasi judul anime yang memiliki kesamaan konteks cerita dengan preferensi pengguna. Hasil sistem menunjukkan keberhasilan dalam merekomendasikan anime yang relevan, termasuk dalam satu waralaba seperti seri "Bleach". Evaluasi menggunakan metrik Mean Average Precision (MAP) menghasilkan skor 0,416667, yang mengindikasikan bahwa sistem mampu menempatkan rekomendasi relevan pada posisi atas daftar. Temuan ini menunjukkan bahwa pendekatan berbasis sinopsis efektif meningkatkan akurasi dan relevansi sistem rekomendasi, serta memberikan pengalaman personalisasi yang lebih mendalam bagi pengguna.

Kata Kunci: Anime, Content-Based Filtering, Sinopsis, Cosine Similarity, Rekomendasi, Tf-Idf

PENDAHULUAN

Di era digital saat ini, arus informasi berkembang pesat dan tersedia secara instan di berbagai platform. Salah satu bidang yang sangat terpengaruh oleh perkembangan ini adalah industri hiburan, khususnya anime. Anime, sebagai bentuk animasi khas Jepang, telah berkembang menjadi fenomena global dengan jutaan penggemar dari berbagai latar belakang[1]. Berdasarkan data dari MyAnimeList dan platform streaming seperti Crunchyroll, ratusan judul anime baru dirilis setiap tahun dengan beragam genre, gaya visual, serta kompleksitas cerita. Keberagaman ini menghadirkan tantangan bagi penikmat anime dalam menemukan tontonan yang benar-benar sesuai dengan preferensi mereka.

Untuk membantu pengguna dalam menelusuri dan memilih judul yang relevan, banyak platform penyedia konten memanfaatkan sistem rekomendasi. Namun, sistem yang tersedia saat ini umumnya masih mengandalkan pendekatan berbasis popularitas atau genre, yang cenderung merekomendasikan judul-judul populer dan kurang mempertimbangkan kesesuaian naratif secara personal. Akibatnya, pengalaman pengguna menjadi kurang optimal karena tidak semua rekomendasi mencerminkan minat atau preferensi cerita secara spesifik[2].

Dalam pengembangan sistem rekomendasi, telah digunakan berbagai pendekatan seperti Collaborative Filtering, Content-Based Filtering, dan Hybrid Systems[3]. Collaborative Filtering memanfaatkan kesamaan preferensi antar pengguna, tetapi sangat bergantung pada data interaksi yang memadai. Sebaliknya, Content-Based Filtering fokus pada karakteristik item itu sendiri, seperti sinopsis, genre, atau elemen naratif lainnya. Sayangnya, pendekatan berbasis sinopsis sebagai konten naratif masih jarang dikaji secara mendalam, khususnya dalam konteks anime[4].

Beberapa penelitian telah mengaplikasikan metode Content-Based Filtering, namun umumnya hanya menggunakan fitur genre, aktor, atau rating[5][2]. Padahal, sinopsis menyajikan ringkasan naratif yang lebih representatif terhadap isi cerita dan berperan penting dalam membentuk ekspektasi pengguna[6]. Hal ini menandakan adanya celah penelitian (research gap) dalam pemanfaatan sinopsis sebagai dasar utama sistem rekomendasi.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan mengembangkan sistem rekomendasi anime berbasis Content-Based Filtering dengan memanfaatkan kesamaan sinopsis. Analisis dilakukan menggunakan metode pemrosesan teks seperti TF-IDF dan Cosine Similarity[7]. Dataset yang digunakan berasal dari MyAnimeList dan difokuskan pada genre fantasi. Sistem kemudian dievaluasi menggunakan metrik Mean Average Precision (MAP) untuk menilai tingkat relevansi hasil rekomendasi. Diharapkan penelitian ini dapat memberikan kontribusi terhadap pengembangan sistem rekomendasi yang lebih akurat, kontekstual, dan personal di ranah hiburan digital.

TINJAUAN PUSTAKA

Beberapa penelitian sebelumnya telah mengkaji sistem rekomendasi untuk anime dengan berbagai pendekatan yang memberikan wawasan berharga. Penelitian oleh peneliti [5] menerapkan metode Content Based Filtering menggunakan dataset KeywordResto, menunjukkan rata-rata presisi sebesar 0,8915, tetapi akurasi hanya 0,5118, mengindikasikan perlunya perbaikan untuk meningkatkan kualitas rekomendasi. Penelitian lain oleh peneliti [2] menggabungkan Content Based Filtering dengan Monarch Butterfly Optimization (MBO) menggunakan dataset Facebook dan MovieLens 100K, mencapai MAE sebesar 0,716 dan RMSE sebesar 0,915, dengan Precision dan Recall masing-masing 97,35% dan 96,60%. Hasil ini menunjukkan bahwa metode optimasi dapat meningkatkan kinerja sistem rekomendasi, meskipun masih ada ruang untuk eksplorasi lebih lanjut dalam konteks anime.



Selanjutnya, penelitian [8] menggunakan Content Based Filtering dengan Deep Neural Networks (DNNs) pada dataset Fashion Product Images, mencapai akurasi rekomendasi sebesar 73,7%. Meskipun penelitian ini menunjukkan efektivitas DNN, fokusnya pada domain yang berbeda menandakan perlunya penerapan teknik serupa dalam konteks anime.

Dari tinjauan pustaka ini, terlihat bahwa meskipun berbagai pendekatan telah dilakukan, masih ada celah dalam penerapan Content Based Filtering yang berfokus pada kesamaan sinopsis anime. Penelitian ini bertujuan untuk mengisi celah tersebut dengan mengembangkan sistem rekomendasi yang lebih akurat dan relevan, serta memberikan kontribusi positif bagi komunitas penggemar anime. Landasan teori yang mendasari penelitian ini mencakup prinsip-prinsip Content Based Filtering dan teknik analisis konten yang relevan.

Sistem Rekomendasi

Sistem rekomendasi adalah alat yang memberikan saran item yang berguna bagi pengguna, bertujuan untuk mendukung pengambilan keputusan mereka[3]. Konsep dasar sistem ini adalah memanfaatkan berbagai sumber data untuk menyimpulkan preferensi pengguna. Dalam konteks ini, pengguna adalah entitas yang menerima rekomendasi, sedangkan item adalah produk yang direkomendasikan. Rekomendasi sering kali didasarkan pada interaksi sebelumnya antara pengguna dan item, karena minat dan kecenderungan masa lalu dapat menjadi indikator yang baik untuk pilihan di masa depan[9].

Sistem rekomendasi dapat didefinisikan sebagai sistem yang mengumpulkan input dari pengguna dan mengarahkan rekomendasi tersebut kepada penerima yang tepat[3]. Tujuan utama sistem ini adalah memandu pengguna dalam memilih objek yang menarik atau bermanfaat, dan evaluasi sistem dilakukan untuk mengukur sejauh mana tujuan tersebut tercapai.

Tahapan umum dalam sistem rekomendasi terdiri dari lima langkah[3]:

1. Pengumpulan Data: Mengumpulkan data yang relevan, termasuk perilaku pengguna dan konten item.
2. Pengolahan Data: Membersihkan dan mengonversi data ke format yang sesuai, serta melakukan normalisasi.
3. Ekstraksi Fitur: Mengidentifikasi dan mengekstrak fitur dari data yang telah diolah untuk membangun model rekomendasi.
4. Pembangunan Model: Membangun model rekomendasi, yang dapat berupa collaborative filtering, content-based filtering, atau model lainnya.
5. Evaluasi: Mengukur kinerja model dalam menghasilkan rekomendasi yang akurat. Dengan mengikuti tahapan ini, sistem rekomendasi dapat memberikan saran yang lebih relevan dan bermanfaat bagi pengguna.

Text Preprocessing

Text preprocessing adalah tahap penting dalam pengolahan bahasa alami (Natural Language Processing / NLP) yang bertujuan untuk membersihkan dan menyiapkan teks mentah agar dapat dianalisis secara efektif oleh model komputer. Tahap ini krusial karena data teks yang diambil dari dunia nyata umumnya mengandung banyak noise seperti tanda baca, kata-kata umum yang tidak bermakna (stop words), serta variasi bentuk kata (misalnya: “berlari”, “berlari-lari”, “berlarian”) [10].

Dalam penelitian ini, preprocessing dilakukan terhadap data sinopsis anime guna menyiapkan input yang bersih dan terstruktur sebelum dilakukan ekstraksi fitur menggunakan metode TF-IDF. Adapun tahapan-tahapan text preprocessing yang diterapkan antara lain:

1. Case Folding: Mengubah seluruh huruf dalam teks menjadi huruf kecil untuk menghindari perbedaan akibat kapitalisasi (contoh: “Fantasy” dan “fantasy” dianggap sama).
2. Tokenization: Memecah teks menjadi unit-unit kata atau token, yang merupakan elemen dasar untuk analisis lebih lanjut.
3. Stopword Removal: Menghapus kata-kata umum yang tidak memiliki makna signifikan terhadap konteks (contoh: “yang”, “dan”, “dengan”). Ini bertujuan mengurangi kompleksitas teks tanpa kehilangan makna penting.
4. Stemming: Mengembalikan kata ke bentuk dasarnya atau akar katanya. Contohnya, kata “bermain”, “bermainlah”, dan “bermainan” akan dikembalikan ke bentuk dasar “main” [11].
5. Punctuation and Symbol Cleaning: Menghapus tanda baca, angka, serta simbol non-alfabet yang tidak relevan dalam analisis sinopsis.

Tahapan preprocessing ini memastikan bahwa teks sinopsis memiliki bentuk yang konsisten, bersih, dan relevan sehingga proses representasi vektor menggunakan TF-IDF dapat menghasilkan hasil yang optimal dalam sistem rekomendasi. Preprocessing yang efektif juga meningkatkan kualitas perhitungan kemiripan antar sinopsis yang menjadi dasar pemodelan Content-Based Filtering [12].

METODE

1. Data Penelitian

Penelitian ini menggunakan metode Content Based Filtering, yang berfokus pada karakteristik atau fitur dari item yang akan direkomendasikan. Dalam konteks anime, metode ini merekomendasikan anime berdasarkan kesamaan konten,



khususnya sinopsis anime untuk memberikan rekomendasi kepada pengguna. Pendekatan ini berasumsi bahwa jika seorang pengguna menyukai item tertentu, mereka kemungkinan besar akan menyukai item lain yang memiliki karakteristik serupa.

2. Analisis dan penyelesaian Masalah

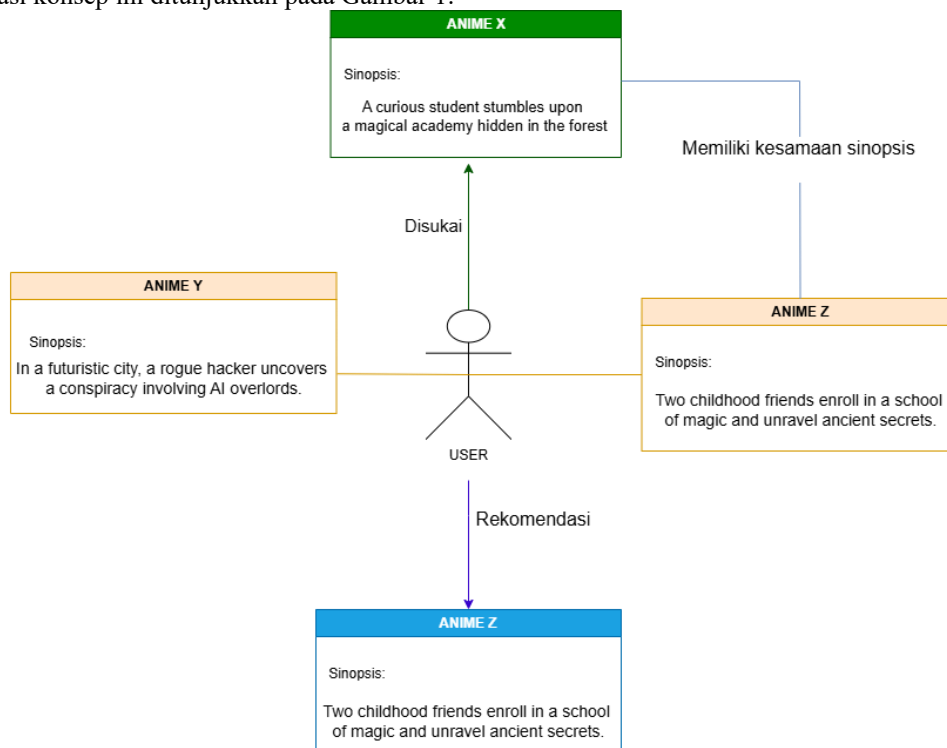
Banyak penikmat anime kesulitan menemukan tontonan yang sesuai karena banyaknya pilihan dengan variasi tema dan cerita. Platform streaming saat ini umumnya memberikan rekomendasi berdasarkan genre atau popularitas, sehingga sering kali kurang relevan dengan preferensi pengguna.

Untuk mengatasi hal ini, digunakan metode rekomendasi berbasis kesamaan konten, dengan fokus pada sinopsis anime. Sistem akan membandingkan sinopsis anime yang dicari pengguna dengan sinopsis dari anime lain dalam dataset. Dengan pendekatan ini, sistem dapat menyarankan judul yang lebih sesuai secara konteks dan isi cerita, sehingga membantu pengguna menemukan tontonan yang lebih relevan dan memuaskan.

3. Metode Content Based Filtering

Penelitian ini menerapkan metode Content-Based Filtering untuk menghasilkan rekomendasi anime berdasarkan kemiripan karakteristik kontennya, terutama dari segi sinopsis [5][13]. Berbeda dengan metode berbasis interaksi pengguna, pendekatan ini fokus pada analisis deskriptif dari setiap item yang akan direkomendasikan. Dalam konteks ini, sistem menganalisis sinopsis anime yang telah ditonton atau disukai oleh pengguna, lalu membandingkannya dengan sinopsis dari anime lainnya dalam dataset untuk mengidentifikasi kesamaan dalam konteks cerita.

Pendekatan ini dilandasi asumsi bahwa pengguna cenderung menyukai anime dengan konten yang memiliki narasi atau tema serupa dengan yang telah mereka nikmati sebelumnya. Dengan demikian, bukan hanya genre yang menjadi faktor pertimbangan utama, tetapi isi dan alur cerita yang tercermin dalam sinopsis menjadi penentu relevansi rekomendasi. Ilustrasi konsep ini ditunjukkan pada Gambar 1.



Gambar 1. Konsep cara kerja Content-based Filtering

Proses Content Based Filtering yang menggunakan sinopsis sebagai parameter rekomendasi melibatkan beberapa langkah sebagai berikut[5]:

1. Pengumpulan Data Sinopsis: Data sinopsis dari berbagai anime dikumpulkan. Sinopsis ini berfungsi sebagai deskripsi yang menjelaskan tema dan alur cerita dari anime tersebut.
2. Preprocessing Teks: Sebelum analisis dilakukan, sinopsis yang dikumpulkan perlu diproses untuk menghilangkan noise. Langkah-langkah preprocessing ini meliputi[14]:
Mengubah teks menjadi huruf kecil untuk konsistensi.
Menghapus tanda baca, angka, dan kata-kata yang tidak penting (stopwords) dalam bahasa Inggris.
Menghapus kata-kata yang memiliki dua karakter atau kurang.
Melakukan lemmatization untuk mengubah kata ke bentuk dasarnya.
3. Representasi Teks ke Numerik Menggunakan TF-IDF: Teks sinopsis yang telah dibersihkan diubah menjadi bentuk numerik yang dapat dipahami oleh komputer. Proses ini sangat penting karena model pembelajaran mesin hanya



dapat memproses data numerik[7].

TF-IDF: Kombinasi dari TF dan IDF yang memberikan bobot lebih tinggi pada kata-kata yang penting dan jarang muncul.

Rumus perhitungan TF-IDF:

$$TF-IDF(t, d) = TF(t, d) \times IDF(t) \quad (1)$$

Keterangan :

d = data ke-d

t = teks ke-t dari kata kunci

W = bobot dokumen ke-d terhadap kata ke-t

tf = term frekuensi

idf= inverse document frequency

Setelah bobot (W) diketahui, maka dilakukan proses sorting di mana nilai W semakin besar, semakin besar tingkat kesamaan data tersebut terhadap kata yang dicari, demikian pula sebaliknya.

4. Penghitungan Kemiripan (Cosine Similarity): Setelah setiap sinopsis diubah menjadi vektor numerik, kita dapat menghitung seberapa mirip dua sinopsis dengan menggunakan metrik cosine similarity[15].

Cosine Similarity: Ukuran kemiripan antara dua vektor, dengan nilai berkisar antara -1 hingga 1.

Rumus perhitungan Cosine-similarity:

$$\text{similarity}(x, y) = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2} \cdot \sqrt{\sum y_i^2}} \quad (2)$$

Keterangan:

$\sum x_i \cdot y_i$ adalah produk dot (dot product) dari vektor x dan y .

$\sqrt{\sum x_i^2}$ dan $\sqrt{\sum y_i^2}$ adalah panjang (magnitude) dari vektor x dan y .

Dengan demikian, cosine similarity mengukur seberapa mirip dua vektor dalam ruang vektor.

4. Rekomendasi: Sistem merekomendasikan anime dengan skor kesamaan tertinggi terhadap anime yang disukai pengguna. Misalnya, jika pengguna menyukai anime A, maka sistem akan mencari dan merekomendasikan anime lain dengan sinopsis yang paling mirip.
5. Evaluasi dan Penyesuaian: Rekomendasi dievaluasi menggunakan Evaluasi menggunakan metrik Mean Average Precision (MAP) Jika hasil kurang optimal maka dilakukan penyesuaian preprocessing atau perhitungan kesamaan akan disesuaikan untuk meningkatkan akurasi sistem.

Seluruh proses metode di atas digambarkan secara sistematis pada Gambar 2, yang menunjukkan alur kerja dari mulai pengumpulan data sinopsis, preprocessing teks, representasi teks ke numerik menggunakan TF-IDF, penghitungan similarity, hingga menghasilkan output rekomendasi anime. Framework ini menjadi panduan visual untuk memahami langkah-langkah yang ditempuh dalam pengembangan sistem rekomendasi berbasis konten ini.

HASIL DAN PEMBAHASAN

Preprocessing Dataset

Data yang digunakan dalam penelitian ini berasal dari dataset MyAnimeList yang diperoleh melalui situs Kaggle.com. Penelitian ini menggunakan satu dataset utama, yaitu anime_dataset.csv, yang berisi informasi lengkap mengenai anime, termasuk judul, sinopsis, genre, serta penilaian atau rating yang diberikan oleh pengguna. Dari dataset tersebut, akan dilakukan proses penyaringan untuk hanya menyertakan anime dengan genre fantasy sebagai fokus utama dalam penelitian ini.

Table 1. Detail-anime.csv

anime_id	Name	English name	Other name	Score	Genres	Synopsis
1	Cowboy Bebop	Cowboy Bebop	カウボーイビバップ	8.75	Action, Award Winning, Sci-Fi	rime is timeless. By the year 2071, humanity has expanded across the galaxy, filling the surface of other planets with settlements like those on Earth.
6	Trigun	Trigun	トライガン	8.22	Action, Adventure, Sci-Fi	Vash the Stampede is the man with a



						\$60,000,000,000 bounty on his head.
--	--	--	--	--	--	--

Ada beberapa tahap untuk membersihkan data diantaranya

1. Cleaning data

Tahap cleaning merupakan proses pembersihan data yang bertujuan meningkatkan kualitas data sebelum dianalisis. Dalam penelitian ini, tahap cleaning mencakup beberapa langkah, antara lain: memeriksa dan menangani missing values, menghapus data yang duplikat, serta membersihkan kolom teks seperti Genres dan Synopsis dengan menghilangkan karakter yang tidak perlu dan spasi berlebih. Selain itu, dilakukan juga standarisasi teks, seperti mengubah seluruh teks menjadi huruf kecil, terutama jika data akan digunakan untuk analisis berbasis teks.

Table 2. Data setelah dilakukan proses cleaning

anime_id	title	score	genres	synopsis	type	episodes
1	cowboy bebop	8.75	action, award winning, sci-fi	crime is timeless. by the year 2071, humanity has expanded across the galaxy, filling the surface of other planets with settlements like those on earth.	tv	26
5	cowboy bebop: tengoku no tobira	8.38	action, sci-fi	nother day, another bounty—such is the life of the often unlucky crew of the bebop.	movie	1

Table 3. lanjutan

aired	status	studio	rating	popularity	favorites	image url
1998	finished airing	sunrise	r - 17+ (violence & profanity)	43	78525	https://cdn.myanimelist.net/images/anime/4/19644.jpg
2001	finished airing	bones	r - 17+ (violence & profanity)	602	1448	https://cdn.myanimelist.net/images/anime/1439/93480.jpg

2. Tokenisasi

Pada tahap tokenisasi, bagian synopsis dari setiap anime dalam dataset diproses dengan menggunakan library spaCy yang sekaligus mencakup lemmatization dan penghapusan stopword. Tokenisasi bertujuan untuk memecah teks, seperti synopsis anime, menjadi unit-unit kata (token) yang lebih kecil dan bermakna. Dalam proses ini, setiap token akan diubah ke bentuk dasarnya melalui lemmatization untuk menyederhanakan variasi kata, serta dibersihkan dari kata-kata umum yang tidak memiliki makna penting dalam analisis (stopwords), seperti “dan”, “yang”, atau “adalah”. Langkah ini penting untuk menyiapkan data teks agar lebih efisien dan relevan dalam proses analisis selanjutnya.

Table 4. Sinopsis sebelum proses tokenisasi

title	synopsis
cowboy bebop	"crime is timeless. by the year 2071, humanity has expanded across the galaxy, filling the surface of other planets with settlements like those on earth. these new societies are plagued by murder, drug use, and theft, and intergalactic outlaws are hunted by a growing number of tough bounty hunters.
cowboy bebop: tengoku no tobira	another day, another bounty—such is the life of the often unlucky crew of the bebop.

Dalam proses ini, setiap token akan diubah ke bentuk dasarnya melalui lemmatization untuk menyederhanakan variasi kata, serta dibersihkan dari kata-kata umum yang tidak memiliki makna penting dalam analisis (stopwords), seperti “dan”, “yang”, atau “adalah”. Langkah ini penting untuk menyiapkan data teks agar lebih efisien dan relevan dalam proses analisis selanjutnya.

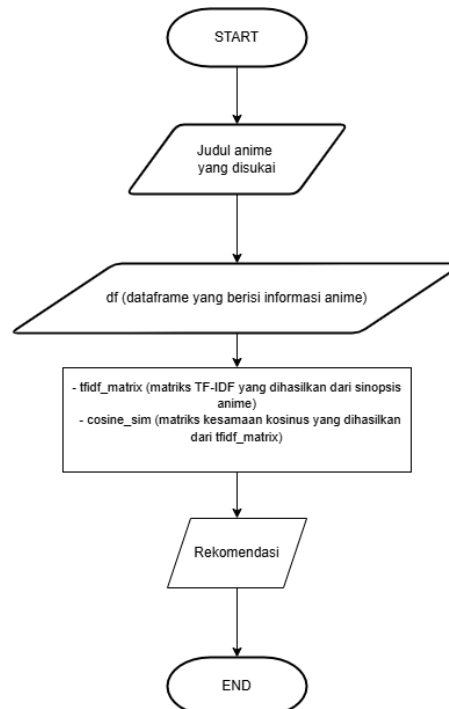


Table 4. Sinopsis setelah proses tokenisasi

title	synopsis
cowboy bebop	crime timeless year humanity expand galaxy fill surface planet settlement like earth new society plague murder drug use theft intergalactic outlaw hunt grow number tough bounty hunter
cowboy bebop: tengoku no tobira	day bountysuch life unlucky crew bebop

3. Proses Content-based Filtering

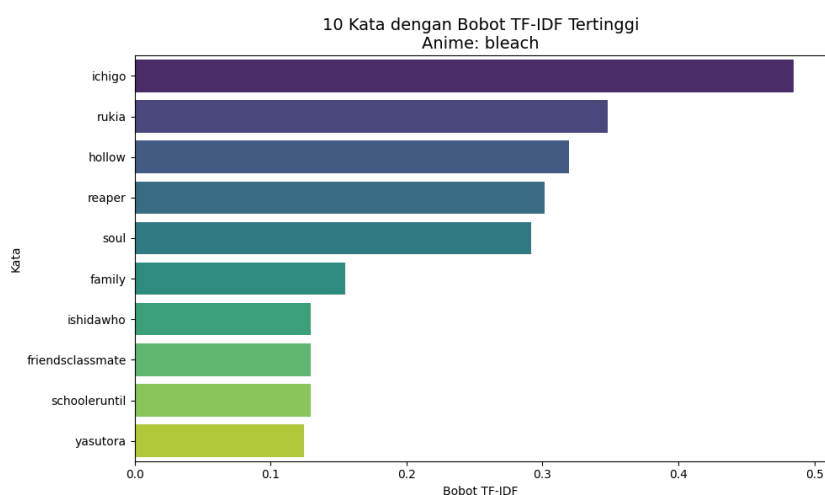
Proses Content-based Filtering digunakan untuk mengembangkan rekomendasi anime yang akurat. Teknik ini bekerja dengan membandingkan vektor sinopsis anime yang dipilih pengguna dengan vektor sinopsis anime lainnya. Vektor sinopsis anime yang dipilih pengguna diambil dan digunakan sebagai acuan untuk membandingkan dengan vektor sinopsis anime lainnya menggunakan metode Cosine Similarity. Hasil perbandingan diurutkan berdasarkan skor similarity tertinggi, dan 5 anime dengan skor similarity tertinggi diambil sebagai rekomendasi anime yang paling sesuai dengan anime yang dipilih pengguna. Proses ini dapat dilihat lebih jelas pada flowchart di bawah ini, yang menunjukkan langkah-langkah yang dilakukan dalam proses Content-based Filtering.



Gambar 2. flowchart sistem

1. Ekstraksi fitur dengan TF-IDF

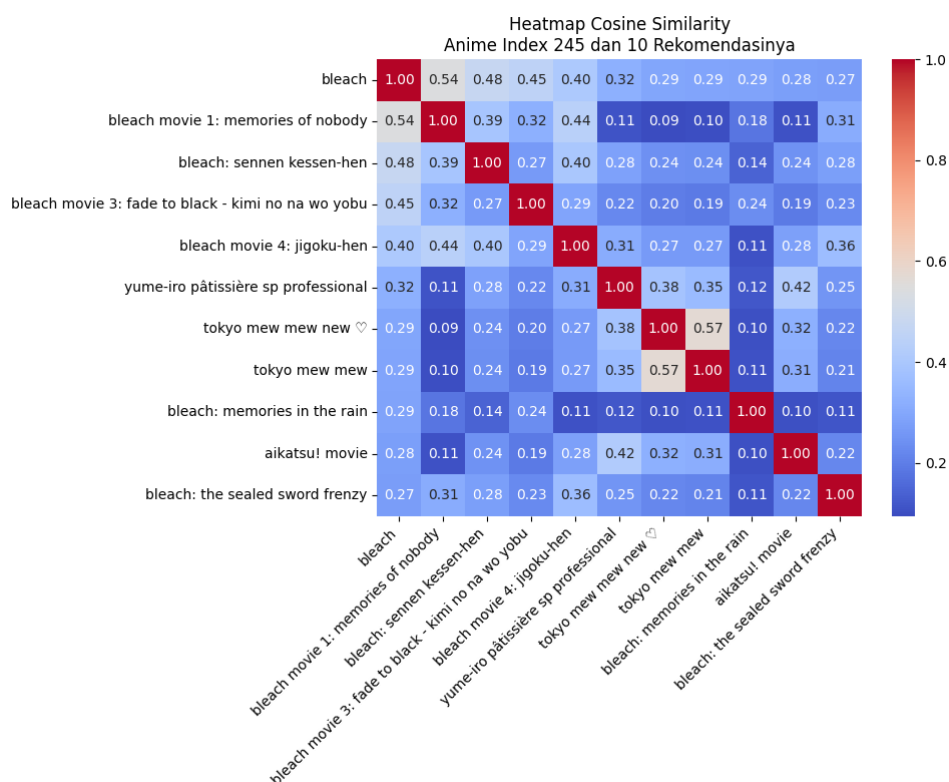
Pada tahap ini, sinopsis anime direpresentasikan ke dalam bentuk vektor numerik menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency) untuk menekankan kata-kata yang unik dan relevan dalam setiap dokumen. Proses ini dilakukan dengan menggunakan kelas TfidfVectorizer dari pustaka scikit-learn, disertai penghapusan stopwords dalam bahasa Inggris agar hanya kata penting yang diperhitungkan. Sebagai ilustrasi, Gambar 4 menampilkan grafik batang 10 kata dengan bobot TF-IDF tertinggi dari sinopsis anime Naruto, yang menunjukkan kata-kata kunci paling mewakili isi sinopsis tersebut..



Gambar 3. Grafik chart 10 kata dengan bobot TF-IDF tertinggi

2. Perhitungan Kemiripan Antar Sinopsis dengan Cosine Similarity

Setelah sinopsis direpresentasikan dalam bentuk vektor TF-IDF, kemiripan antar sinopsis dihitung menggunakan Cosine Similarity, yang mengukur kesamaan arah antara dua vektor dengan nilai antara 0 hingga 1. Metode ini umum dipakai dalam sistem rekomendasi berbasis konten karena fokus pada arah vektor. Sebagai contoh, Gambar 5 menampilkan heatmap Cosine Similarity untuk 10 sinopsis anime pertama, di mana warna semakin merah menunjukkan kemiripan yang lebih tinggi, sehingga memudahkan pemahaman pola kemiripan sebelum rekomendasi diberikan.



Gambar 4. Heatmap Cosine Similarity antar sinopsis

3. Hasil Rekomendasi

Hasil rekomendasi anime yang diperoleh dari proses Content-based Filtering dapat dilihat pada tabel di bawah ini. Tabel 5 menunjukkan 5 anime yang memiliki skor similarity tertinggi dengan anime yang dipilih pengguna.

Table 5. Hasil rekomendasi						
index	title		genres	score	synopsis	skor_similarity
1533	bleach	movie 1:	action,	7.43	when a life ends, its soul departs to its final	0.543393



	memories of nobody	adventure, fantasy		resting place known as the soul society.	
16617	bleach: sennen kessen-hen	action, adventure, fantasy	9.07	substitute soul reaper ichigo kurosaki spends his days fighting against hollows, dangerous evil spirits that threaten karakura town.	0.475907
3840	bleach movie 3: fade to black - kimi no na wo yobu	action, adventure, fantasy	7.5	after a mysterious pair attack rukia kuchiki and erase her memories while she is in seireitei, ichigo kurosaki, a substitute soul reaper, briefly forgets rukia, until he is reminded of her by kon, an underpod mod-soul.	0.450370
5290	bleach movie 4: jigoku-hen	action, adventure, fantasy	7.61	when a soul has committed irredeemable sins, it is removed from the cycle of reincarnation monitored by the soul society, and the gates of hell open to condemn the soul to eternal damnation.	0.404246
5755	yume-iro pâtissière sp professional	gourmet, slice of life	7.58	upon her return to japan after a two year study break in paris, ichigo amano is met with news of team ichigo's separation, as each of the sweets princes takes a different path toward their dream.	0.324264

Dari hasil rekomendasi yang disajikan, Sistem berhasil merekomendasikan "Bleach Movie 1: Memories of Nobody" sebagai anime teratas dengan skor kemiripan sinopsis 0.543393 terhadap anime input "Bleach", menunjukkan kemampuan sistem mengidentifikasi konten serupa. Bahkan, rekomendasi kedua pun masih dalam franchise yang sama. Untuk mengukur kualitas urutan rekomendasi, kami menggunakan Mean Average Precision (MAP). Sistem mencapai nilai MAP 0.416667, mengindikasikan bahwa anime relevan (berdasarkan ambang batas cosine similarity) cenderung ditempatkan di posisi atas daftar rekomendasi.

KESIMPULAN DAN SARAN

Penelitian ini berhasil mengembangkan sistem rekomendasi anime berbasis Content-Based Filtering dengan memanfaatkan kemiripan sinopsis sebagai parameter utama. Dengan menggunakan metode TF-IDF untuk representasi teks dan Cosine Similarity untuk pengukuran kemiripan antar sinopsis, sistem mampu memberikan rekomendasi anime yang relevan terhadap preferensi pengguna. Hasil pengujian menunjukkan bahwa sistem dapat mengidentifikasi dan merekomendasikan anime yang berada dalam waralaba atau memiliki tema cerita yang serupa, seperti ditunjukkan dalam kasus rekomendasi terhadap anime Bleach.

Evaluasi menggunakan metrik Mean Average Precision (MAP) menghasilkan skor 0.416667, yang menunjukkan bahwa sistem mampu menempatkan anime yang relevan pada urutan atas dalam daftar rekomendasi. Hal ini menandakan bahwa pendekatan berbasis kesamaan sinopsis memiliki potensi yang baik dalam memberikan pengalaman personalisasi yang lebih mendalam bagi pengguna, khususnya dalam domain hiburan seperti anime.

Untuk pengembangan lebih lanjut, disarankan agar sistem rekomendasi menggunakan dataset yang lebih besar dan mencakup beragam genre anime guna meningkatkan cakupan dan akurasi hasil rekomendasi. Selain itu, mempertimbangkan penggunaan sinopsis dalam bahasa asli pengguna, seperti bahasa Jepang atau Indonesia, dapat meningkatkan relevansi konten. Integrasi dengan metode lain seperti Collaborative Filtering atau algoritma pembelajaran mesin juga dapat dipertimbangkan untuk meningkatkan kinerja sistem. Evaluasi sebaiknya tidak hanya terbatas pada metrik MAP, tetapi juga mencakup metrik lain seperti Recall dan F1-Score, serta uji coba langsung kepada pengguna untuk menilai kepuasan mereka. Akhirnya, pengembangan sistem dalam bentuk aplikasi web atau mobile akan memungkinkan implementasi nyata dan memberikan manfaat langsung bagi pengguna.

DAFTAR PUSTAKA

- [1] N. Bastin, "pengertian anime," 2022, *nahason bastin, sidoarjo*. [Online]. Available: https://www.google.co.id/books/edition/Apakah_Itu_Anime/3m1wEAAQBAJ?hl=id&gbpv=1&dq=anime&pg=PA141&printsec=frontcover
- [2] S. Sridhar, D. Dhanasekaran, and G. C. P. Latha, "Content-Based Movie Recommendation System Using MBO with DBN," *Intelligent Automation and Soft Computing*, vol. 35, no. 3, pp. 3241–3257, 2023, doi: 10.32604/iasc.2023.030361.
- [3] P. B. K. Francesco Ricci, Lior Rokach, Bracha Shapira (auth.), Francesco Ricci, Lior Rokach, Bracha Shapira, *Handbook Recommender System*, 1st ed. New York: Springer US, 2011. doi: 10.1007/978-0-387-85820-3.
- [4] A. Said and A. Bellogin, "Comparative recommender system evaluation: Benchmarking recommendation frameworks," in *RecSys 2014 - Proceedings of the 8th ACM Conference on Recommender Systems*, Association for Computing Machinery, Oct. 2014, pp. 129–136. doi: 10.1145/2645710.2645746.



- [5] E. Salim, J. Pragantha, and M. D. Lauro, "Perancangan Sistem Rekomendasi Film menggunakan metode Content-based Filtering."
- [6] Y. C. Chang and C. M. Hung, "Improving content-based movie recommendation using narrative similarity from plot summaries," *Expert Syst. Appl.*, vol. 198, May 2022.
- [7] M. Nurjannah, I. Fitri Astuti, and D. Program Studi, "PENERAPAN ALGORITMA TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF) UNTUK TEXT MINING Mahasiswa S1 Program Studi Ilmu Komputer FMIPA Universitas Mulawarman 2,3)," 2013.
- [8] N. Yarahmadi Gharaei, C. Dadkhah, and L. Daryoush, "Content-based Clothing Recommender System using Deep Neural Network," in *26th International Computer Conference, Computer Society of Iran, CSICC 2021*, Institute of Electrical and Electronics Engineers Inc., Mar. 2021. doi: 10.1109/CSICC52343.2021.9420544.
- [9] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowl Based Syst.*, vol. 46, pp. 109–132, Jul. 2013, doi: 10.1016/j.knosys.2013.03.012.
- [10] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze, "irbookonlinereading," Apr. 2009, Accessed: Jun. 03, 2025. [Online]. Available: Website: <http://www.informationretrieval.org/>
- [11] Steven Bird, Ewan Klein, and Edward Loper, *Natural Language Processing with Python*, 1st ed. O'Reilly Media, 2009. Accessed: Jun. 03, 2025. [Online]. Available: <http://my.safaribooksonline.com>
- [12] A. A. Nafea *et al.*, "A Brief Review on Preprocessing Text in Arabic Language Dataset: Techniques and Challenges," *Babylonian Journal of Artificial Intelligence*, vol. 2024, pp. 46–53, May 2024, doi: 10.58496/bjai/2024/007.
- [13] M. J. Pazzani and D. Billsus, "Content-Based Recommendation Systems," in *The Adaptive Web*, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 325–341. doi: 10.1007/978-3-540-72079-9_10.
- [14] E. Kumar, *Natural Language Processing*. New Delhi: I.K. International Publishing House Pvt. Ltd., 2013. [Online]. Available: <https://books.google.co.id/books?id=FpUBFNfuKWgC&lpg=PP1&hl=id&pg=PP5#v=onepage&q&f=false>
- [15] P. Xia, L. Zhang, and F. Li, "Learning similarity with cosine similarity ensemble," *Inf Sci (N Y)*, vol. 307, pp. 39–52, Jun. 2015, doi: 10.1016/j.ins.2015.02.024.