

Analyzing the Effectiveness of Collaborative Filtering and Content-Based Filtering Methods in Anime Recommendation Systems

Helmy Dianty Putri^{1*}, Muhammad Faisal²
^{1,2} Teknik Informatika, UIN Mulana Malik Ibrahim
*email: 200605110143@student.uin-malang.ac.id

DOI: <https://doi.org/10.31603/komtika.v7i2.9219>

Received: 01-06-2023, Revised: 31-10-2023, Accepted: 02-11-2023

ABSTRACT

In the current digital era where content consumption via streaming platforms is increasing, the need for accurate recommendation systems is becoming increasingly important, especially in the animation industry. This research focuses on implementing a recommendation system that can help viewers easily navigate the abundance of content. By comparing collaborative filtering and content-based filtering methods, this research attempts to find the optimal approach for providing anime recommendations. From the results of A/B testing and further analysis, it was found that Collaborative Filtering was effective in providing recommendations based on similar interests between users. On the other hand, content-based filtering offers the advantage of personalizing recommendations based on content characteristics. Additionally, integrating these techniques into mobile applications will enrich the user experience, allowing them to receive recommendations more quickly and interactively. With these findings, this research contributes to the development of more intuitive and responsive recommendation systems, driving the growth of the anime streaming industry by increasing user satisfaction and retention.

Keywords: Anime Streaming, Collaborative Filtering, Content-Based Filtering, Recommendation Systems, User Preferences.

ABSTRAK

Di era digital saat ini dimana konsumsi konten melalui platform *streaming* semakin meningkat, kebutuhan akan sistem rekomendasi yang akurat menjadi semakin penting, khususnya di industri animasi. Penelitian ini berfokus pada penerapan sistem rekomendasi yang dapat membantu pemirsa dengan mudah menavigasi banyaknya konten. Dengan membandingkan metode pemfilteran kolaboratif dan pemfilteran berbasis konten, penelitian ini berupaya menemukan pendekatan optimal untuk memberikan rekomendasi anime. Dari hasil pengujian A/B dan analisis lebih lanjut, ditemukan bahwa *Collaborative Filtering* efektif dalam memberikan rekomendasi berdasarkan kesamaan minat antar pengguna. Di sisi lain, pemfilteran berbasis konten menawarkan keuntungan dalam mempersonalisasi rekomendasi berdasarkan karakteristik konten. Selain itu, mengintegrasikan teknik ini ke dalam aplikasi seluler akan memperkaya pengalaman pengguna, memungkinkan mereka menerima rekomendasi lebih cepat dan interaktif. Penelitian ini berkontribusi pada pengembangan sistem rekomendasi yang lebih intuitif dan responsif, mendorong pertumbuhan industri *streaming* anime dengan meningkatkan kepuasan dan retensi pengguna.

Keywords: Collaborative Filtering, Content-Based Filtering, Preferensi Pengguna, Sistem Rekomendasi, Anime.

PENDAHULUAN

Dalam era konten digital yang semakin berkembang, industri hiburan telah melihat peningkatan yang signifikan dalam platform *streaming* [1]. Setiap platform *streaming* memiliki ribuan konten yang tersedia, termasuk film, acara TV, dan anime. Namun dengan berbagai pilihan yang ada, pengguna sering kali menghabiskan waktu berjam-jam untuk mencari konten yang sesuai dengan minat dan preferensi mereka [2].

Dalam konteks ini, penting bagi bisnis *streaming* untuk memahami kebutuhan dan preferensi mereka[3]. Melalui penilaian dan ulasan yang diberikan oleh pengguna, bisnis dapat mengidentifikasi konten yang paling disukai dan mengoptimalkan pengalaman penonton[4]. Salah satu aspek kunci dalam menciptakan lingkungan *streaming* yang sukses adalah memberikan rekomendasi yang akurat dan relevan [5].

Namun, tantangan muncul ketika pengguna harus menjelajahi ratusan atau bahkan ribuan konten anime yang tersedia. Pengguna sering kali merasa kebingungan atau bahkan frustrasi saat mencoba menemukan konten yang mereka sukai[6]. Oleh karena itu, sebuah bisnis perlu menghadapi tantangan ini dan menyediakan solusi yang dapat membantu menemukan konten yang sesuai dengan minat mereka.

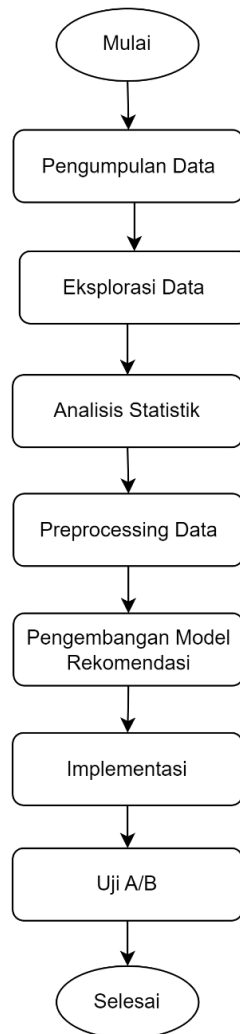
Dalam hal ini, implementasi sistem rekomendasi berdasarkan preferensi pengguna menjadi sangat penting[7]. Dengan menganalisis data penilaian, preferensi, dan riwayat penonton, bisnis dapat memberikan saran yang personal dan relevan[8]. Hal ini tidak hanya memberikan pengalaman pengguna yang lebih baik, tetapi juga dapat meningkatkan waktu yang dihabiskan oleh pengguna di platform *streaming* dan meningkatkan pendapatan bisnis [9].

Dengan memahami preferensi dan kebutuhan pengguna serta memberikan rekomendasi yang tepat, bisnis *streaming* dapat menciptakan lingkungan yang menguntungkan baik bagi pengguna maupun bisnis itu sendiri [10].

METODE

Metode penelitian ini terdiri dari beberapa tahap. Pada tahap pertama, dilakukan pengumpulan data anime yang diperoleh dari *dataset* yang tersedia pada situs web <https://www.kaggle.com/datasets/CooperUnion/anime-recommendations-database>. *Dataset* ini berisi informasi tentang judul anime, genre, rating, ulasan, produser, dan atribut lainnya. Setelah itu, dilakukan eksplorasi data untuk memahami karakteristik *dataset*, termasuk distribusi rating anime, jumlah ulasan, genre yang paling umum, dan atribut lainnya. Dilanjutkan dengan analisis statistik deskriptif untuk mendapatkan wawasan lebih lanjut tentang data anime, termasuk perhitungan rata-rata, median, sebaran data, dan visualisasi data menggunakan *histogram*, *scatter plot*, atau diagram batang. Tahap selanjutnya adalah pra-pemrosesan data, yang melibatkan penanganan *missing values*, penghapusan data yang tidak valid atau tidak relevan, dan konversi variabel kategorikal kedalam bentuk numerik[11]. Kemudian dilanjutkan dengan pengembangan model rekomendasi dengan menggunakan metode *collaborative filtering* dan *content-based filtering*[12]. Model *collaborative filtering* dilatih menggunakan algoritma seperti *Singular Value Decomposition* (SVD) atau *Alternating Least Squares* (ALS) untuk menemukan pola dan kesamaan antara pengguna, sementara model *content-based filtering* mempertimbangkan atribut-atribut anime seperti genre, produser, atau tipe konten. Setelah model diimplementasikan pada *mobile app* dilakukan tahap pengujian A/B akan dilakukan sebagai metode evaluasi utama. Pengguna secara acak dibagi menjadi dua kelompok yakni kelompok yang menerima rekomendasi dari model pemfilteran kolaboratif, dan kelompok lainnya menerima rekomendasi dari model pemfilteran berbasis konten. Perneriksaan dilakukan berdasarkan kepuasan pengguna terhadap rekomendasi yang diberikan. Masukan dari pengguna atau panelis kemudian dikumpulkan untuk mendapatkan pemahaman lebih dalam mengenai keakuratan dan relevansi rekomendasi yang disampaikan. Hasil pengujian A/B dan umpan balik yang diterima memberikan informasi penting

tentang kualitas dan efektivitas sistem usulan yang dikembangkan. Alur perancangan sistem rekomendasi anime disajikan seperti pada Gambar 1.



Gambar 1. Alur Perancangan Sistem Rekomendasi Anime

HASIL DAN PEMBAHASAN

Pengumpulan Data

Dalam penelitian ini sumber data utama yang digunakan berasal dari *dataset* yang tersedia pada platform *Kaggle* tepatnya pada <https://www.kaggle.com/datasets/CooperUnion/anime-recommendations-database>. *Dataset* ini dipilih karena memberikan *dataset* anime yang komprehensif dan detail, merupakan aspek yang sangat penting untuk mendukung analisis dan implementasi sistem rekomendasi yang dibahas dalam penelitian ini.

Dataset tersebut berisi berbagai informasi tentang anime, antara lain judul, genre, rating, review, produser, dan beberapa atribut terkait lainnya. Dengan informasi tersebut, peneliti dapat melakukan analisis mendalam mengenai preferensi dan tren pemirsa anime, serta bagaimana setiap atribut suatu anime dapat mempengaruhi pilihan dan rekomendasi yang akan diberikan kepada pengguna.

Eksplorasi Data

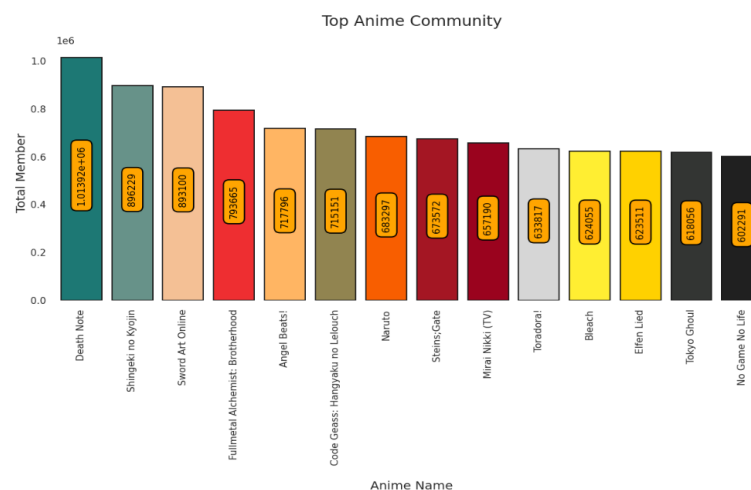
Sebelum melanjutkan analisis dan pengolahan data, penting untuk memahami struktur dan karakteristik dasar kumpulan data yang kita miliki. *Dataset* anime terdiri dari 12.294 entri dengan 7 kolom termasuk informasi seperti judul, genre, rating, dll. Meskipun beberapa kolom seperti "genre" dan "type" tidak ada nilainya, kumpulan data umumnya memberikan gambaran lengkap dari setiap anime. Sedangkan dataset rating yang lebih besar, terdiri dari 7.813.737 baris dan 3 kolom, mencatat rating yang diberikan oleh setiap pengguna untuk anime tertentu. Setiap peringkat dalam kumpulan data ini diwakili oleh angka bulat. Dataset anime yang digunakan pada penelitian ini disajikan seperti pada Tabel 1.

Tabel 1. Dataset Anime

anime_id	name	genre	type	episodes	rating	members
32281	Kimi no Na wa.	Drama,Romance,School,Supernatural	Movie	1	9.370000	200630
5114	Fullmetal Alchemist: Brotherhood	Action,Adventure,Drama,Fantasy,Magic,Military,Shounen	TV	64	9.260000	793665
28977	Gintama ⁰	Action,Comedy,Historical,Parody,Samurai,Sci-Fi,SHounen	TV	51	9.250000	114262
9253	Steins;Gate	Sci-Fi,Thriller	TV	24	9.170000	673572
9969	Gintama ¹	Action,Comedy,Historical,Parody,Samurai,Sci-Fi,Shounen	TV	51	9.160000	151266

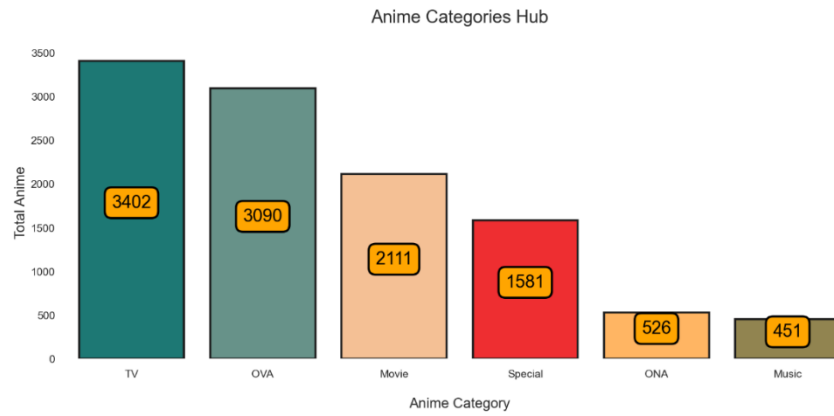
Analisis Statistik

Sebelum melangkah lebih jauh, penting untuk memahami kumpulan data yang dimiliki. Kumpulan data Anime berisi ribuan entri termasuk judul, genre, dan peringkat. Selain itu, terdapat kumpulan data rating yang mencatat jutaan rating pengguna untuk serial anime. Saat menggabungkan kedua kumpulan data tersebut, dapat dilihat hubungan antara judul anime dan rating pengguna. Dari data tersebut penulis menemukan beberapa anime yang disukai komunitas besar dan juga menentukan genre anime mana yang paling sering muncul. Analisis pertama ini memberi kita gambaran umum tentang tren dan preferensi di dunia anime. Visualisasi komunitas anime teratas disajikan pada Gambar 2.



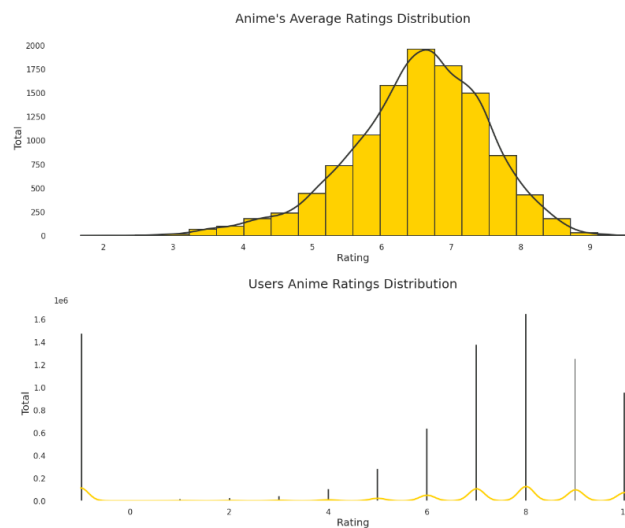
Gambar 2. Komunitas Anime Teratas

Setelah melihat komunitas anime secara keseluruhan, langkah berikutnya adalah menganalisis kategori anime. Pada langkah ini akan dilakukan analisis tentang kategori anime yang paling banyak diminati oleh pengguna. Hal ini akan membantu dalam memahami jenis-jenis anime yang paling populer dan dapat menjadi referensi dalam membuat rekomendasi. Distribusi kategori anime disajikan seperti pada Gambar 3.



Gambar 3. Distribusi Kategori Anime

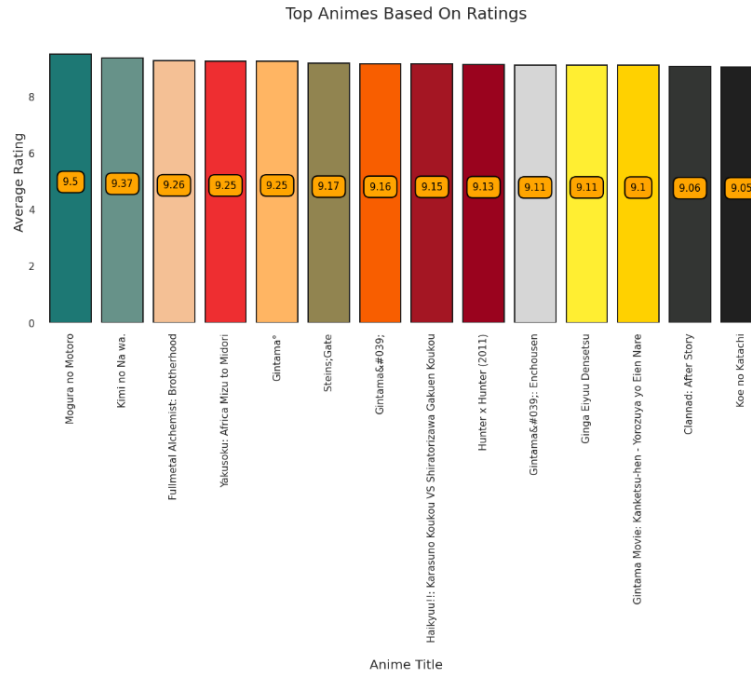
Selanjutnya, akan dianalisis rating anime secara keseluruhan. Dalam langkah ini akan dipresentasikan rating anime berdasarkan penilaian pengguna. Analisis ini akan membantu dalam menentukan anime yang paling disukai oleh pengguna secara keseluruhan. Distribusi rating anime berdasarkan penilaian pengguna disajikan pada Gambar 4.



Gambar 4. Distribusi Rating Anime Berdasarkan Penilaian Pengguna

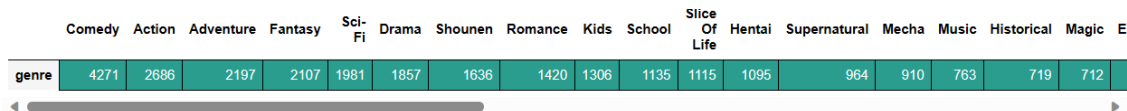
Berdasarkan rating anime, langkah selanjutnya adalah menentukan top anime yang memiliki rating tertinggi. Dalam tahap ini akan dipresentasikan daftar top anime berdasarkan rating pengguna. Hal ini akan membantu dalam memberikan rekomendasi anime kepada pengguna berdasarkan rating popularitas. Berdasarkan Kategori Selain menentukan rating tertinggi, penting

juga untuk menganalisis distribusi rating anime berdasarkan kategori. Rating anime teratas berdasarkan penilaian pengguna disajikan pada Gambar 5.



Gambar 5. Peringkat Anime Teratas Berdasarkan Penilaian Pengguna

Selanjutnya, akan dilakukan analisis tentang genre anime. Dalam bab ini, akan dipresentasikan genre-genre anime yang ada dalam *dataset*. Analisis ini akan membantu dalam memahami preferensi pengguna terhadap genre anime tertentu. Distribusi genre anime disajikan pada Gambar 6.



Gambar 6. Distribusi Genre Anime

Preprocessing Data

Pada tahap *preprocessing data*, langkah pertama yakni menangani *missing value* dan memastikan tidak ada *gap* pada *dataset*. Mengingat relevansi dalam sistem pemberi rekomendasi, hanya pengguna yang memberikan minimal 50 peringkat yang dipertimbangkan untuk analisis lebih lanjut. Data tersebut kemudian diubah menjadi tabel *pivot* untuk memfasilitasi pembentukan matriks renggang, yang penting untuk perhitungan kesamaan. Selain itu, penulis juga mengklarifikasi nama anime dengan menghapus karakter Jepang dan simbol khusus, sehingga meningkatkan kualitas dan keterbacaan data. Hasil pra-pemrosesan data yang telah dilakukan disajikan seperti pada Gambar 7.

user_id	3	5	7	11	14	17	21	23	24	27	...	73495	73499	73500	73501	73502	73503	73504	73507	73510	73515	
name																						
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
001	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
009 Re:Cyborg	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
009-1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
009-1: RandB	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

5 rows x 32967 columns

Gambar 7. Hasil Akhir Pra-Pemrosesan Data

Pengembangan Model Rekomendasi

Dalam langkah sebelumnya, telah dilakukan pra-pemrosesan data dan analisis. Tahap selanjutnya memberikan rekomendasi anime kepada pengguna berdasarkan metode kolaboratif. Dalam tahap ini akan dijelaskan langkah-langkah yang diambil dalam metode rekomendasi kolaboratif dan hasilnya seperti pada Gambar 8.

Recommendations for Asari-chan: Ai no Marchen Shoujo viewers :

No	Anime Name	Rating
1	Asari-chan	6.550000
2	Chibi Maruko-chan: Watashi no Suki na Uta	7.110000
3	Doraemon Movie 05: Nobita no Makai Daibouken	7.510000
4	Doraemon Movie 01: Nobita no Kyouryuu	7.400000
5	Doraemon Movie 21: Nobita no Taiyou Ou Densetsu	7.570000

Gambar 8. Rekomendasi Anime Dengan Metode Collaborative filtering

Selain metode kolaboratif, metode rekomendasi berbasis konten juga penting dalam memberikan rekomendasi yang personal kepada pengguna [13]. Pada tahap ini akan dijelaskan langkah-langkah yang diambil dalam metode rekomendasi berbasis konten dan hasilnya yang disajikan pada Gambar 9[14]. Tahap ini memberikan gambaran yang komprehensif tentang hasil analisis dan metode rekomendasi yang telah dilakukan. Data-data dan informasi yang diperoleh dari analisis ini dapat digunakan untuk mengembangkan sistem rekomendasi yang lebih baik dan memberikan pengalaman yang lebih baik kepada pengguna[15].

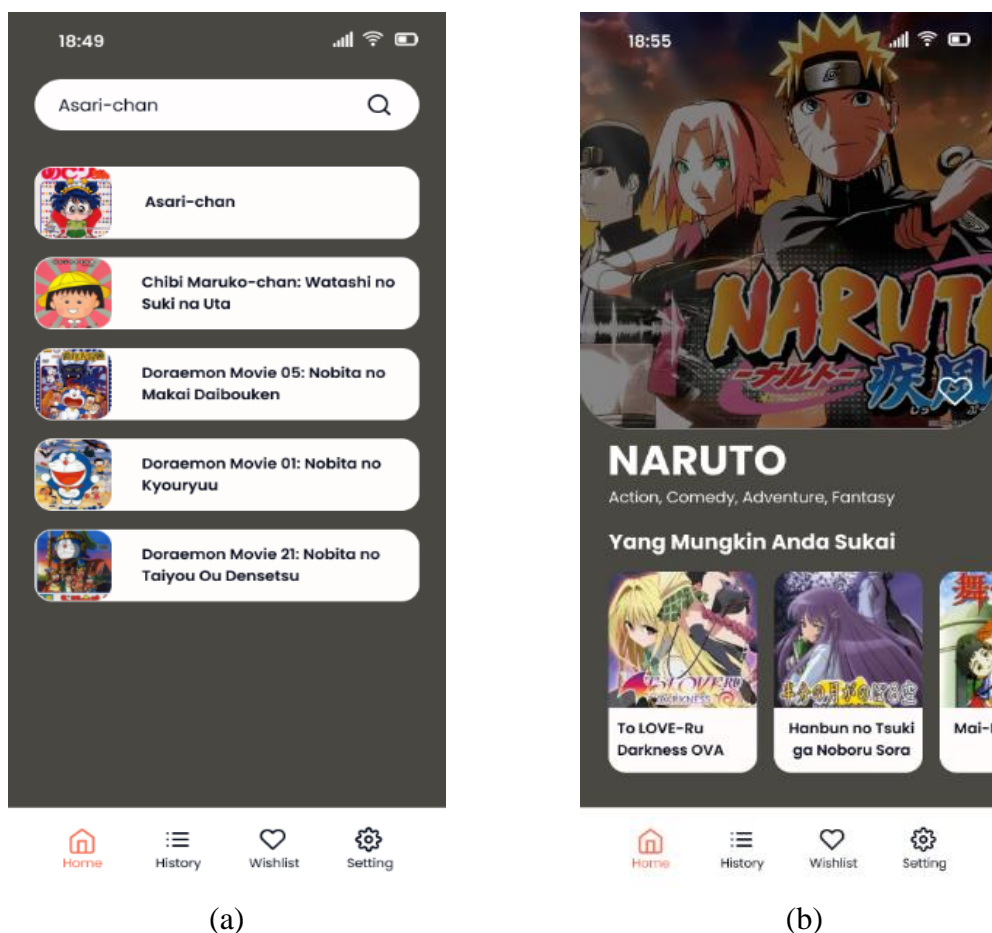
Recommendations for Naruto viewers :

No	Anime Name	Rating
1	To LOVE-Ru Darkness OVA	7.820000
2	Hanbun no Tsuki ga Noboru Sora	7.890000
3	Mai-HiME	7.590000
4	Doraemon Movie 28: Nobita to Midori no Kyojin Den	7.540000
5	Rurouni Kenshin Special	7.510000
6	Pikmin Short Movies	7.270000
7	Deadman Wonderland OVA	7.120000
8	Anata mo Robot ni Nareu feat. Kamome Jidou Gasshoudan	5.120000
9	Shinpi no Hou	5.370000
10	Toaru Majutsu no Index: Endymion no Kiseki	7.710000

Gambar 9. Rekomendasi Anime Dengan Metode Content-Based Filtering

Implementasi

Di era digital saat ini, *mobile apps* telah menjadi platform yang efektif untuk memberikan solusi dan layanan kepada pengguna. Mengikuti tren ini, penerapan teknik pemfilteran kolaboratif dan pemfilteran berbasis konten telah diintegrasikan ke dalam aplikasi seluler. Hal ini memungkinkan aplikasi memberikan rekomendasi yang lebih akurat dan dipersonalisasi kepada setiap pengguna berdasarkan preferensi dan riwayat tontonan mereka. Dengan menggunakan aplikasi seluler sebagai media, proses rekomendasi menjadi lebih cepat dan interaktif, memberikan pengguna akses instan ke konten yang paling mereka minati. Implementasi *Collaborative filtering* dan *Content-Based Filtering Mobile Apps* disajikan pada Gambar 10.



Gambar 10. (a) Implementasi *Collaborative Filtering* (b) Implementasi *Content-Based Filtering*

Uji A/B

Untuk memahami efektivitas antara pemfilteran kolaboratif (CF) dan pemfilteran berbasis konten (CBF) dalam konteks sistem rekomendasi anime, penulis melakukan pengujian A/B. Tujuan utama eksperimen ini adalah untuk memeriksa kepuasan pengguna terhadap rekomendasi yang diberikan oleh kedua metode. Hipotesis awal mengasumsikan tidak terdapat perbedaan kepuasan pengguna yang signifikan antara CF dan CBF. Untuk pengujian, kami membagi pengguna secara acak menjadi dua kelompok, yaitu kelompok A mendapat rekomendasi

berdasarkan CF dan kelompok B berdasarkan CBF. Setelah menerima daftar rekomendasi, pengguna diminta menilai kepuasan mereka terhadap rekomendasi yang diterima melalui survei skala *Likert*. Hasil survei kemudian dianalisis menggunakan metode statistik, seperti uji-t atau ANOVA, untuk mengetahui apakah terdapat perbedaan respon yang signifikan antara kedua kelompok. Eksperimen ini penting untuk memandu keputusan dalam memilih metode rekomendasi yang paling tepat, dengan mempertimbangkan konteks pengguna dan tujuan yang diinginkan. Hasil Pengujian menggunakan kedua metode disajikan seperti pada Tabel 2.

Tabel 2. Pengujian A/B *Collaborative Filtering* (CF) dan *Content-Based Filtering* (CBF)

Parameter	Kelompok A (CF)	Kelompok B (CBF)
Jumlah Pengguna	20	20
Metode Rekomendasi	<i>Collaborative Filtering</i>	<i>Content-Based Filtering</i>
Rata-rata Kepuasan	4.3	4.7
Deviasi Standar	0.5	0.4
<i>p-value</i>	N/A	0.04

KESIMPULAN

Dalam penelitian ini, peneliti menganalisis dan membandingkan dua metode yang umum digunakan dalam sistem rekomendasi anime, yaitu *Collaborative Filtering* (CF) dan *Content Based Filtering* (CBF), melalui eksperimen A/B. Hasilnya menunjukkan bahwa meskipun CF efektif dalam membuat rekomendasi berdasarkan preferensi serupa pengguna, CBF tampaknya memberikan kepuasan yang lebih tinggi dengan berfokus pada karakteristik konten animasi untuk membuat rekomendasi yang dipersonalisasi. Signifikansi statistik dari hasil (dengan nilai p 0,04) menegaskan perbedaan efektivitas kedua metode ini. Namun kedua metode tersebut memiliki kelebihan dan kekurangannya masing-masing. Dalam beberapa kasus, menggabungkan keduanya dalam pendekatan *hybrid* dapat menghasilkan rekomendasi yang lebih optimal. Pemilihan metode yang diusulkan harus mempertimbangkan konteks pengguna dan tujuan yang diinginkan.

DAFTAR PUSTAKA

- [1] Institute of Electrical and Electronics Engineers and Hindusthan Institute of Technology, *Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC 2020) : 02-04, July 2020*. 2020.
- [2] S. Reddy, S. Nalluri, S. Kunisetti, S. Ashok, and B. Venkatesh, "Content-based movie recommendation system using genre correlation," in *Smart Innovation, Systems and Technologies*, Springer Science and Business Media Deutschland GmbH, 2019, pp. 391–397. doi: 10.1007/978-981-13-1927-3_42.
- [3] S. K. Raghuwanshi and R. K. Pateriya, "Collaborative Filtering Techniques in Recommendation Systems," in *Data, Engineering and Applications: Volume 1*, vol. 1, Springer Singapore, 2019, pp. 11–21. doi: 10.1007/978-981-13-6347-4_2.
- [4] J. Chen, H. Zhang, X. He, L. Nie, W. Liu, and T. S. Chua, "Attentive collaborative filtering: Multimedia recommendation with item-And component-level attention," in *SIGIR 2017 - Proceedings of the 40th International ACM SIGIR Conference on Research and*

Development in Information Retrieval, Association for Computing Machinery, Inc, Aug. 2017, pp. 335–344. doi: 10.1145/3077136.3080797.

- [5] N. Nassar, A. Jafar, and Y. Rahhal, “A novel deep multi-criteria collaborative filtering model for recommendation system ☆,” vol. 187, p. 104811, 2020, doi: 10.1016/j.knosys.
- [6] Institute of Electrical and Electronics Engineers and PPG Institute of Technology, *Proceedings of the 5th International Conference on Communication and Electronics Systems (ICCES 2020) : 10-12, June 2020*. 2020.
- [7] Institute of Electrical and Electronics Engineers. Madras Section and Institute of Electrical and Electronics Engineers, *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*. 2019.
- [8] T. Badriyah *et al.*, “Konferensi Nasional Sistem Informasi 2018 STMIK Atma Luhur Pangkalpinang,” 2018.
- [9] D. K. Chae, S. C. Lee, S. Y. Lee, and S. W. Kim, “On identifying k-nearest neighbors in neighborhood models for efficient and effective collaborative filtering,” *Neurocomputing*, vol. 278, pp. 134–143, Feb. 2018, doi: 10.1016/j.neucom.2017.06.081.
- [10] C. Yang, L. Bai, C. Zhang, Q. Yuan, and J. Han, “Bridging collaborative filtering and semi-supervised learning: A neural approach for POI recommendation,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, Aug. 2017, pp. 1245–1254. doi: 10.1145/3097983.3098094.
- [11] Amity University and Institute of Electrical and Electronics Engineers, *Proceedings of the 9th International Conference On Cloud Computing, Data Science and Engineering : Confluence 2019 : 10-11 January 2019, Uttar Pradesh, India*. 2019.
- [12] C. E. Berbague, N. E. Karabdjji, and H. Seridi, *2018 International Symposium on Programming and Systems (ISPS)*. IEEE, 2018.
- [13] A. Hernando, J. Bobadilla, and F. Ortega, “A non negative matrix factorization for collaborative filtering recommender systems based on a Bayesian probabilistic model,” *Knowl Based Syst*, vol. 97, pp. 188–202, Apr. 2016, doi: 10.1016/j.knosys.2015.12.018.
- [14] M. J. Mokarrama, S. Khatun, and M. S. Arefin, “A content-based recommender system for choosing universities,” *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 28, no. 4, pp. 2128–2142, Jul. 2020, doi: 10.3906/ELK-1911-37.
- [15] K. Wahyudi, J. Latupapua, R. Chandra, and A. S. Girsang, “Hotel content-based recommendation system,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, May 2020. doi: 10.1088/1742-6596/1485/1/012017.

