

IMPLEMENTASI ALGORITMA APRIORI UNTUK MENENTUKAN ITEMSET ANIME YANG JARANG DITONTON PADA MYANIMELIST.NET

IMPLEMENTATION OF APRIORI ALGORITHM TO DETERMINE ANIME ITEMSETS RARE SEEN ON MYANIMELIST.NET

Hilal Abdullah¹⁾, Stefanus Aditya Ferary²⁾, Amalia Anjani Arifiyanti³⁾

E-mail : ¹⁾hilal.spentisa99@gmail.com , ²⁾stefanusaf5@gmail.com , ³⁾amalia_anjani.fik@upnjatim.ac.id

^{1,2,3} Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional
“Veteran” Jawa Timur

Abstrak

MyAnimeList, atau biasa disebut MAL merupakan sebuah situs yang disediakan sebagai katalog berbasis *online* yang berisi tentang anime dan manga. Situs ini juga memberikan layanan sistem yang menampilkan daftar untuk mengatur dan menilai sebuah anime maupun manga, serta sebagai tempat berkumpulnya komunitas para penggemar anime dan manga dari berbagai negara. Para penggemar anime terbantu dengan adanya rekomendasi seri anime yang ada di MyAnimeList. Ada beberapa jenis rekomendasi yang disediakan oleh situs MyAnimeList yaitu, Rekomendasi anime populer, rekomendasi anime yang akan tayang mendatang, dan rekomendasi anime yang sedang tayang. Teknik data mining dapat mengolah data menjadi sebuah itemset, yang mana itemset ini dapat dijadikan sebagai acuan dalam membuat rekomendasi. Tujuan paper ini adalah menentukan itemset dari anime dengan rating tinggi namun jarang dilihat. Dengan menggunakan algoritma apriori didapatkan 8 itemset dari judul anime yang memiliki rating tinggi namun jarang dilihat oleh para pengguna situs MyAnimeList.

Kata kunci: anime, algoritma apriori, aturan asosiasi, itemset, MAL

Abstract

MyAnimeList, or commonly called MAL is a site that is provided as an online-based catalog containing anime and manga. This site also provides a system service that displays a list to organize and rate an anime or manga, as well as a gathering place for a community of anime and manga fans from various countries. Anime fans are helped by the recommendations for anime series on MyAnimeList. There are several types of recommendations provided by the MyAnimeList site, namely, popular anime recommendations, upcoming anime recommendations, and currently airing anime recommendations. Data mining techniques can process data into an itemset, where this itemset can be used as a reference in making recommendations. The purpose of this paper is to determine the itemset from anime with high ratings but rarely seen. By using the apriori algorithm, 8 itemsets are obtained from anime titles that have a high rating but are rarely seen by users of the MyAnimeList site.

Keywords: anime, apriori algorithm, association rule, itemset, MAL

1. PENDAHULUAN

MyAnimeList, atau biasa disebut MAL merupakan sebuah situs yang disediakan sebagai katalog berbasis online yang berisi tentang anime dan manga. Situs ini juga memberikan layanan sistem yang menampilkan daftar untuk mengatur dan menilai sebuah anime maupun manga, serta sebagai tempat berkumpulnya komunitas para penggemar anime dan manga dari berbagai negara. Para penggemar anime terbantu dengan adanya rekomendasi seri anime yang ada di MyAnimeList. Para penggemar tersebut mendiskusikan suatu anime yang sedang hangat dibicarakan dalam sebuah forum yang mana orang tersebut juga bisa memberikan penilaian atau

rating yang bisa menandakan bahwa sebuah anime tersebut bagus atau tidaknya setelah seorang ini menonton anime yang mereka lihat. Semakin tinggi rating maka semakin tinggi juga orang yang menonton anime tersebut. Dari pemberian rating itulah muncul rekomendasi anime yang ditampilkan dalam halaman awal situs. Namun disisi lain, masih ada anime yang jarang ditonton dengan nilai rating yang tinggi.

Rating yang bagus identik dengan jumlah penonton yang banyak. Rating yang tinggi menjadi daya tarik tersendiri bagi penonton anime. Hal ini terkait dengan persepsi yang dimiliki penonton bahwa rating tinggi merupakan representasi dari kualitas yang baik [1]. Rating tinggi juga sangat berpengaruh terhadap minat penonton [2]. Meski tidak semua orang dapat setuju dengan skor rating yang ada, namun mayoritas pengguna adalah sisi yang setuju dengan skor rating. Namun tidak menutup kemungkinan jika anime dengan rating tinggi juga memiliki jumlah penonton yang tidak banyak. Hal ini dikarenakan bahwa rating yang bersifat subjektif ini memiliki tingkat inkonsistensi yang tinggi [3]. Adanya berbagai macam situs yang menyediakan rekomendasi anime underrated (jarang ditonton dengan rating tinggi) menunjukkan bahwa penggemar anime memerlukannya dengan tujuan sebagai referensi mereka untuk menonton anime yang akan mereka saksikan.

Penelitian ini menggunakan teknik asosiasi yang merupakan salah satu dari beberapa teknik data mining. Tujuan utamanya adalah untuk menemukan hubungan antara item [4]. Beberapa penelitian menyebutkan bahwa algoritma apriori dapat menghasilkan pola pada data tertentu berdasarkan parameter yang sudah ditentukan [5]. Selain itu, algoritma ini juga dipilih karena dapat mengolah data dengan cepat dan akurat dalam memberikan rekomendasi [6]. Oleh sebab itu tujuan penelitian ini adalah untuk menentukan itemset seri anime dari situs MyAnimeList.net. Itemset yang didapatkan akan berguna sebagai acuan untuk membuat rekomendasi anime yang memiliki rating tinggi namun dengan jumlah penonton yang sedikit. Hasil penelitian ini diharapkan bisa dijadikan acuan untuk situs MyAnimeList.net untuk membuat sebuah rekomendasi anime bagi para pengguna situs tersebut. Sehingga kebutuhan akan rekomendasi anime yang jarang ditonton tapi memiliki rating tinggi dapat terpenuhi.

2. METODOLOGI

Penelitian ini menggunakan alat yaitu Google Colaboratory untuk melakukan pemrosesan olah data sehingga mampu menghasilkan luaran yang sesuai. Dalam pemrosesan olah data akan dilakukan implementasi algoritma apriori. Berikut merupakan langkah penelitian yang dilakukan.



Gambar 1. Alur Penelitian

2.1 Persiapan Alat

Tahap ini merupakan langkah awal untuk melakukan penelitian. Alat yang dibutuhkan untuk mengelola dataset adalah koneksi internet dan Google Colaboratory. Google Colaboratory merupakan layanan atau tools yang disediakan oleh Google Research dan berbasis cloud. Google Colaboratory merupakan layanan open source dan dapat digunakan oleh siapapun. Google Colaboratory berupa dokumen yang bisa dipakai untuk menyimpan, menulis, serta membagikan program yang telah ditulis melalui Google Drive [7].

2.2 Mencari Dataset

Dataset merupakan kumpulan dari data yang pada umumnya disajikan dalam bentuk tabel berisi baris dan kolom [8]. Baris pada tabel sering disebut sebagai *row/item*, sedangkan kolom pada tabel sering disebut dengan atribut. Tahap pencarian dataset merupakan langkah persiapan bahan penelitian berupa *dataset*. *Dataset* penelitian kami ambil dari sebuah situs *open data* yaitu kaggle.com. *Dataset* ini berisi informasi tentang 17.562 anime dan pilihan dari 325.772 pengguna

MAL yang berbeda [9]. Ada dua dataset dari sumber yang sama dan digunakan pada penelitian ini.

Dataset pertama berisi rincian dari 17.562 anime dengan atribut *MAL_ID*, *Name*, *Score*, *Genres*, *English name*, *Japanese name*, *Type*, *Episodes*, *Aired*, *Premiered*, *Producers*, *Licensors*, *Studios*, *Source*, *Duration*, *Rating*, *Ranked*, *Popularity*, *Members*, *Favorites*, *Watching*, *Completed*, *On-Hold*, *Dropped*, *Plant to Watch*, *Score-10*, *Score-9*, *Score-8*, *Score-7*, *Score-6*, *Score-5*, *Score-4*, *Score-3*, *Score-2*, *Score-1*. Namun tidak semua atribut yang ada dalam dataset akan digunakan dalam penelitian ini. Atribut yang digunakan antara lain: *MAL_ID*, *Name*, *Score*, *English Name*, *Japanese name*, *Watching*.

Dataset kedua berisi rating yang diberikan oleh setiap pengguna situs. Atribut dalam dataset antara lain: *anime_id*, *user_id*, *rating*. Dataset ini adalah dataset utama yang akan digunakan dalam menentukan itemset dari proses implementasi algoritma apriori.

2.3 Menyiapkan Data

Dataset yang sudah diperoleh sebelumnya akan dipersiapkan terlebih dahulu agar sesuai dengan keperluan implementasi algoritma. Pada tahap ini dilakukan teknik *data reduction* dan *data cleaning*. *data cleaning* dilakukan untuk menghilangkan data yang tidak relevan, dalam kasus ini ada beberapa value “*Unknown*” yang tidak diperlukan dan bahkan dapat mengganggu hasil penelitian [10]. Sedangkan *data reduction* digunakan untuk mengurangi dimensi data agar proses pengolahan data lebih efisien dan hemat ruang penyimpanan [11].

Data reduction yang dilakukan adalah mengurangi atribut yang tidak dibutuhkan dalam penelitian. Dari 35 atribut yang ada, hanya 6 atribut terpakai dari dataset pertama. selain itu kedua dataset akan disaring sesuai dengan kebutuhan penelitian.

Sedangkan *data cleaning* yang dilakukan adalah menghilangkan data yang memiliki nilai-nilai salah atau anomali, data tersebut disebut sebagai noise [12].

2.4 Implementasi Algoritma Apriori

Data yang sudah disiapkan kemudian diolah menggunakan algoritma Apriori. Algoritma Apriori adalah algoritma yang digunakan untuk mendapatkan itemset yang sering muncul (*frequent*) dalam analisis aturan asosiasi [13]. Aturan asosiasi didapatkan dari pemrosesan data dari data bertipe boolean dan numerik [14].

Aturan yang menyatakan asosiasi antara beberapa atribut sering disebut affinity analysis atau market basket analysis [15]. Analisis asosiasi atau aturan asosiasi adalah sebuah teknik dalam data mining untuk menemukan aturan suatu kombinasi item. Kualitas asosiasi dapat diukur melalui nilai *support* dan *confidence* [16]. Nilai *support* merupakan persentase kombinasi item dalam dataset, sedangkan *confidence* adalah kuatnya hubungan antar item dalam sebuah aturan asosiasi [17].

3. HASIL DAN PEMBAHASAN

Tabel 1. Cuplikan Data Dari Pemrosesan Dataset Pertama

MAL_ID	Name	Score	English Name	Japanese Name	Watching
31812	Kuroshitsuji Movie: Book of...	8.26	Black Butler:Book of the Atlantic	劇場版 黒執事 Book of the Atlantic	3092
572	Kaze no Tani no Nausicaä	8.40	Nausicaä of the Valley of the Wind	風の谷のナウシカ	3082
5205	Kara no Kyoukai 7: Satsujin Kousatsu	8.44	the Garden of sinners Chapter 7:Murder..	劇場版 空の境界 the Garden of sinners 第七 章 『殺人考察 (後) 』	3008
36946	Dragon Ball Super Movie: Br..	8.10	Dragon Ball Super:Broly	ドラゴンボール超 (スーパー) ブロリー	2941
...

Dataset pertama telah disaring sehingga didapatkan sebuah tabel berisi 149 judul anime dengan jumlah ditonton kurang dari rata-rata seluruh judul anime dan memiliki rating ≥ 8 . Cuplikan data dari pemrosesan dataset dapat dilihat pada tabel 1. Sebuah judul anime dengan skor rating 8 telah dianggap hebat. Hal tersebut ditentukan berdasarkan skala yang telah dirumuskan. Berikut merupakan tabel skala nilai rating.

Table 2. Tabel Skala Nilai Rating (1-5)

Skor	Interpretasi
1	Terburuk
2	Sangat Buruk
3	Buruk
4	Dibawah Rata-rata
5	Rata-rata
6	Bagus
7	Sangat Bagus
8	Hebat
9	Luar Biasa
10	Sempurna

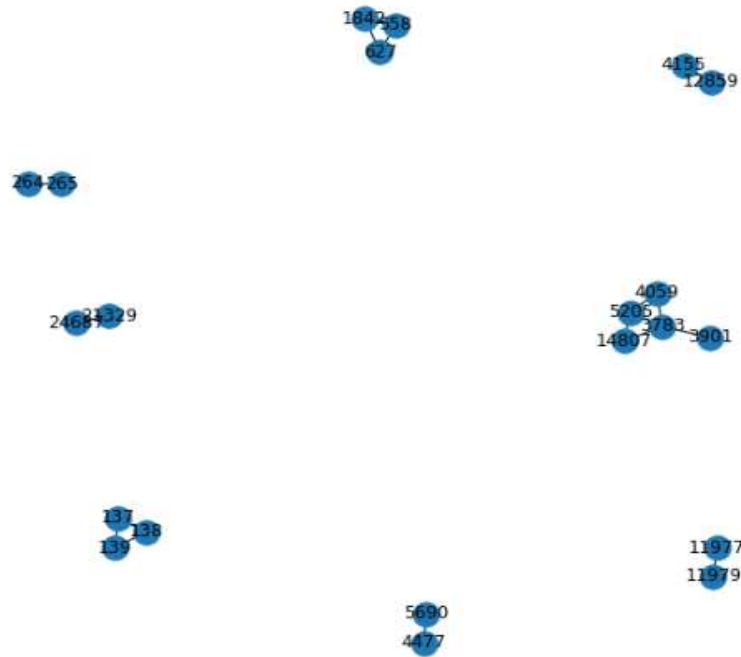
Dataset kedua akan disaring berdasarkan tabel yang didapatkan dari pemrosesan dataset pertama dan minimum skor yang diberikan adalah 8. Sehingga mendapatkan hasil data sebagai berikut.

	user_id	anime_id	rating
102	1	4053	9
425	3	14807	9
426	3	5205	10
455	3	3783	8
480	3	22507	8
...
57632388	353398	1842	8
57632402	353398	3226	8
57632414	353398	5028	8
57632415	353398	7655	8
57632725	353398	38088	8

[874655 rows x 3 columns]

Gambar 2. Data yang disiapkan dari dataset kedua

Data pada Gambar 2 akan diolah lagi menggunakan algoritma apriori. Dengan menambahkan parameter yakni minimum *support* sebesar 0.035 dan minimum *lift ratio* adalah 2. Kedua nilai tersebut telah ditentukan dalam penelitian. Sehingga mendapatkan hasil berupa beberapa itemset yang disajikan seperti pada gambar 3.



Gambar 3. Itemset yang didapat

Dari Gambar 3 dapat diketahui bahwa terdapat 8 itemset yang didapatkan dari hasil implementasi algoritma apriori. Angka-angka yang menjadi kombinasi adalah id dari judul anime. Maka dari itu kombinasi-kombinasi yang terdapat pada Gambar 3 akan disajikan dalam tabel 4.

Tabel 4. Kombinasi Judul Anime

Kombinasi	MAL_ID	Judul Anime
1	264	Hajime no Ippo: Champion Road
	265	Hajime no Ippo: Mashiba vs. Kimura
2	4477	Nodame Cantabile: Paris-hen
	5690	Nodame Cantabile: Finale
3	21329	Mushishi: Hihamukage
	24687	Mushishi Zoku Shou: Odoro no Michi
4	137	Hunter x Hunter: Original Video Animation
	138	Hunter x Hunter: Greed Island
	139	Hunter x Hunter: Greed Island Final
5	11977	Mahou Shoujo Madoka★Magica Movie 1: Hajimari n...
	11979	Mahou Shoujo Madoka★Magica Movie 2: Eien no Mo...
6	1842	Major S3
	558	Major S2
	627	Major S1
7	4115	One Piece Film: Strong World
	12859	One Piece Film: Z
8	3738	Usavich II
	3901	Baccano! Specials
	4059	Clannad: Mou Hitotsu no Sekai, Tomoyo-hen

Kombinasi	MAL_ID	Judul Anime
	5205	Kara no Kyoukai 7: Satsujin Kousatsu (Go)
	14807	Kara no Kyoukai: Mirai Fukuin

Delapan kombinasi yang ada dalam tabel 4 merepresentasikan aturan asosiasi. Sebagai contoh kombinasi 1 dapat diartikan bahwa Hajime no Ippo: Champion Road dan Hajime no Ippo: Mashiba vs. Kimura saling berkaitan dan dapat menjadi sebuah item dalam rekomendasi yang diberikan kepada pengguna situs myanimelist.net.

4. KESIMPULAN DAN SARAN

Delapan itemset yang didapatkan bisa menjadi acuan untuk membuat sebuah rekomendasi judul anime pada situs MyAnimeList.net. Itemset tersebut merupakan kombinasi dari judul anime yang sudah sesuai dengan kondisi yang ditentukan, yakni judul anime dengan jumlah ditonton rendah dengan rating tinggi. Kombinasi yang dihasilkan dapat menjadi acuan untuk memberikan rekomendasi anime yang jarang ditonton namun memiliki rating tinggi kepada pengguna situs myanimelist.net.

Penelitian ini masih terbatas oleh data yang digunakan. Agar hasil penelitian lebih maksimal, diharapkan data yang digunakan harus merupakan data terbaru sehingga hasil yang didapatkan dapat relevan dengan situasi sebenarnya.

5. DAFTAR RUJUKAN

- [1] A. R. Yosafat and Y. Kurnia, "Aplikasi Prediksi Rating Film dengan Perbandingan Metode Naïve Bayes dan KNN Berbasis Website Menggunakan Framework Codeigniter," *Jurnal ALGOR*, vol. 1, no. 1, pp. 16-26, 2019.
- [2] Z. F. Auliya, M. R. K. Umam and S. K. Prastiwi, "Online Customer Reviews (OTRs) dan Rating: Kekuatan Baru pada Pemasaran Online di Indonesia," *Jurnal Ilmiah Bidang Ekonomi Bisnis dan Perbankan*, pp. 89-98, 2017.
- [3] D. Kotkov, A. Medlar, U. R. Satyal and A. Maslov, "Rating consistency is consistently underrated: An exploratory analysis of movie-tag rating inconsistency," in *ACM/SIGAPP Symposium on Applied Computing*, Virtual, 2022.
- [4] M. Kaur and S. Kang, "Market Basket Analysis: Identify the changing trends of market data using association rule mining," in *International Conference on Computational Modeling and Security*, Bengaluru, 2016.
- [5] K. Z. Abdurrafi, R. Reliovani, N. N. S. Husein, C. R. A. Husni, M. A. Khowarizmi and K. M. Shaliha, "Implementasi Algoritma Apriori untuk Rekomendasi Film berdasarkan Sutradara dan Durasi Film," *Gunung Djati Conference Series*, vol. 3, pp. 12-18, 2021.
- [6] I. P. Y. D. Jaya, B. Dirgantoro and I. , "IMPLEMENTASI ALGORITMA A-PRIORI UNTUK SISTEM REKOMENDASI PRODUK MITRA PERUSAHAAN
- [7] E-COMMERCE EATAJA," *e-Proceeding of Engineering*, vol. 8, no. 5, pp. 6222-6229, 2021.
- [8] N. Hidayah, "Mengenal Google Colab, Aplikasi Pengolahan Data Open Source," *Anak Teknik Indonesia*, 20 November 2021. [Online]. Available: <https://www.anateknik.co.id/111620868375095332510/articles/mengenal-google-colab-aplikasi-pengolahan-data-open-source>. [Accessed 16 Juni 2022].
- [9] M. 2, "Apa itu Dataset? Berikut Pengertian, Contoh, serta Jenis-jenisnya!," *Bitlabs Academy*, 12 Februari 2022. [Online]. Available: https://bitlabs.id/blog/dataset-adalah/#Apa_itu_Dataset. [Accessed 17 Juni 2022].
- [10] H. Valdivieso, "Anime Recommendation Database 2020 | Kaggle," *Kaggle*, 2021. [Online]. [Accessed 16 Juni 2022].

-
- [11] DQLab, "Langkah Awal dalam Pemrosesan Data: Data Preprocessing dalam Data Mining," DQLab, 21 Mei 2021. [Online]. Available: <https://www.dqlab.id/langkah-awal-dalam-pemrosesan-data-dalam-data-mining>. [Accessed 16 Juni 2022].
- [12] D. Uhm, S. Jun and S.-J. Lee, "A Classification Method Using Data Reduction," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 12, pp. 1-5, 2012.
- [13] T. Setyorini and R. S. Wahono, "Penerapan Metode Bagging untuk Mengurangi Data Noise pada Neural Network untuk Estimasi Kuat Tekan Beton," *Journal of Intelligent Systems*, vol. 1, pp. 37-42, 2015.
- [14] M. A. Azis, N. Hadianto, J. Mhiharja and S. Rifai, "IMPLEMENTASI ALGORITMA APRIORI UNTUK ANALISA PEMILIHAN TIPE GENRE FILM ANIME (STUDI KASUS :MYANIMELIST.NET)," *Jurnal PILAR Nusa Mandiri* , vol. 15, pp. 209-216, 2018.
- [15] F. Zhan, X. Zhu, L. Zhang, X. Wang, L. Wang and C. Liu, "Summary of Association Rules," in *IOP Conf. Series: Earth and Environmental Science*, 2018.
- [16] M. R. Rahman, S. M. Arefin, S. Rahman, A. Ahmed, T. Islam, P. K. Dhar and O.-j. Kwon, "A Comprehensive Survey on Affinity Analysis, Bibliomining, and Technology Mining: Past, Present, and Future Research," *Applied Sciences*, vol. 12, no. 10, p. Mei, 2022.
- [17] R. Karthiyayini and R. Balasubramanian, "Affinity Analysis and Association Rule Mining using Apriori Algorithm in Market Basket Analysis," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 6, no. 10, pp. 241-246, 2016.
- [18] F. A. Sianturi, "PENERAPAN ALGORITMA APRIORI UNTUK PENENTUAN TINGKAT PESANAN," *Jurnal Mantik Penusa*, vol. 2, no. 1, pp. 50-57, 2018.