



Implementasi Algoritma Apriori untuk Rekomendasi Film berdasarkan Sutradara dan Durasi Film

(Implementation of the Apriori Algorithm for Film Recommendations based on Director and Movie Duration)

Kamal Zaki Abdurrafi¹, Ryan Reliovani², Nina Nadia Syafitri Husein³, Cecep Rafqi Al Husni⁴, Muhammad Azka Khowarizmi⁵, Karima Marwazia Shaliha⁶

¹Teknik Informatika, UIN Sunan Gunung Djati Bandung, kamalzakiab@gmail.com

²Teknik Informatika, UIN Sunan Gunung Djati Bandung, ryan.reliovani@gmail.com

³Teknik Informatika, UIN Sunan Gunung Djati Bandung, ninanadiasyafitrihusein@gmail.com

⁴Teknik Informatika, UIN Sunan Gunung Djati Bandung, ceceprafqi19@gmail.com

⁵Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050068@student.uinsgd.ac.id

⁶Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050057@student.uinsgd.ac.id

Abstrak

Industri Perfilman merupakan industri yang tidak ada matinya, di Indonesia sendiri selama tiga tahun kebelakang jumlah penonton film terus mengalami peningkatan. Dilansir dari laman indonesia.go.id pada 2018 jumlah produksi film yang diproduksi hampir 200 judul, dari banyaknya jumlah film yang diproduksi tentu para penikmat perfilman memiliki selera film yang berbeda pula, salah satu cara yang dapat digunakan untuk meningkatkan daya tarik dalam film adalah adanya sistem rekomendasi film dengan menganalisis kecenderungan film berdasarkan sutradara dan berapa lama durasi film yang ideal bagi calon penonton. Algoritma apriori dipilih dalam penelitian kali ini untuk mencari dan menentukan pola kecenderungan pemilihan sutradara dan durasi film yang tersedia dalam 1001 data pada data movie, data akan dibagi menjadi himpunan yang terdiri dari 30 item. Sebagai hasil dari penelitian ini adalah rekomendasi film berdasarkan algoritma apriori dengan sutradara dan durasi film sebagai acuan untuk association rule. Hasil yang diperoleh dari penelitian ini adalah algoritma apriori dapat diimplementasikan dalam rekomendasi film berdasarkan sutradara dan durasi film.

Kata kunci: algoritma apriori, association rule, film, sistem rekomendasi

Abstract

The Film Industry is an industry that never dies, in Indonesia itself for the past three years the number of film viewers has continued to increase. Reporting from the indonesia.go.id page in 2018 the number of film productions produced is almost 200 titles, from the large number of films produced, of course film lovers have different tastes for films, one way that can be used to increase attractiveness in films is the existence of film recommendation system based on film trends based on the director and how long the ideal film duration for prospective viewers. The algorithm chosen in this research is to find and determine the pattern of director selection and film duration available in 1001 data on film data, the data will be divided into lists consisting of 30 items. The results of this study are film recommendations based on a priori algorithm with the director and film duration as a reference for association rules. The results obtained from this study are that the apriori algorithm can be implemented in film recommendations based on the director and film duration.

Keywords: apriori algorithm, association rule, film, recommendation system

1 Pendahuluan

Semakin berkembangnya teknologi, manusia semakin mudah untuk memenuhi kebutuhan dan keinginannya salah satu diantaranya hiburan. Dimana orang-orang pada saat ini dapat dengan mudah mengakses berbagai jenis hiburan untuk mengatasi kebosanan mulai dari musik hingga film. Di Indonesia industri film terus mengalami perkembangan yang positif, terbukti dari semakin banyaknya jumlah penonton film dan semakin banyaknya jumlah film yang diproduksi. Baik film dalam negeri maupun negeri memiliki pasar dan tempatnya sendiri di hati para penikmat film Indonesia.

Setiap film yang diproduksi oleh sutradara yang berbeda tentu menghasilkan keunikan dan nilai-nilai tersendiri pada film-film tersebut. Hal ini tentu semakin meningkatkan persaingan di industri film, oleh karenanya pihak yang menayangkan film-film tersebut perlu mengetahui kecenderungan film seperti apa yang disukai oleh para penontonnya siapa sutradara dengan film yang memiliki jumlah penayangan tinggi dan disukai oleh penontonnya, bahkan hingga durasi filmnya. Oleh karenanya diperlukan sebuah rekomendasi yang dapat memberikan informasi saran film yang tepat untuk ditonton. Dalam penelitian ini rekomendasi film dibuat berdasarkan sutradara dan durasi film dengan mengimplementasikan algoritma apriori. Algoritma apriori dipilih karena merupakan salah satu bentuk teknik data mining yaitu analisis asosiasi untuk mencari pola hubungan antar satu item atau lebih dalam suatu dataset

Tujuan dari penelitian ini sendiri adalah untuk mengetahui apakah algoritma apriori dapat memberikan informasi berupa rekomendasi film berdasarkan sutradara dan durasi film yang diolah dari data movie yang telah ditonton, dan dapat membantu menyediakan informasi saran film yang sesuai dalam penentuan pemutaran film selanjutnya bagi penonton. Berdasarkan latar belakang diatas maka didapatkan rumusan masalah yang dibahas dalam penelitian ini adalah (1) Bagaimana implementasi algoritma apriori untuk mendapatkan rekomendasi film berdasarkan sutradara dan durasi film? (2) Apakah algoritma apriori dapat membantu menyediakan informasi film yang sesuai dalam penentuan pemutaran film selanjutnya bagi penonton dengan akurasi yang tepat? Terdapat beberapa penelitian terdahulu yang terkait penelitian ini, antara lain: (1) mengenali pola *customer relationship management* dengan menggunakan algoritma Apriori [1]; (2) sistem rekomendasi konten dan produk pada e-commerce dengan menggunakan algoritma Apriori [2], [3]; dan (3) rekomendasi buku dengan menggunakan algoritma apriori [4].

2 Metodologi

2.1 Algoritma Apriori

Algoritma Apriori merupakan algoritma dasar sekaligus algoritma pertama untuk frequent itemset mining (FIM). Algoritma ini mengambil transaksi pada database yang memenuhi dukungan minimum atau nilai ambang batas penggunaan breadth first search untuk mencari semua frequent itemset [5]. Itemset adalah kumpulan item item di dalam sebuah keranjang (support). Algoritma apriori termasuk kedalam jenis aturan asosiasi pada data mining. Adapun yang disebut affinity analysis atau market basket analysis merupakan aturan yang menyatakan asosiasi beberapa atribut. Analisis asosiasi atau association rule mining merupakan suatu teknik dalam data mining untuk menemukan aturan kombinasi item. Analisis yang banyak menarik perhatian peneliti untuk dapat menghasilkan algoritma yang efisien adalah analisis pola frekuensi tinggi (frequent pattern mining). Tolak ukur untuk mengetahui penting atau tidaknya suatu asosiasi, yaitu: support dan confidence. Support (nilai penunjang) adalah persentase dari kombinasi item tersebut di dalam database, sedangkan confidence (nilai kepastian) merupakan kuatnya hubungan antar item dalam aturan asosiasi.

2.2 Association Rule

Analisis asosiasi atau *association rule mining* adalah salah satu teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item [6]–[8]. Contoh dari aturan asosiatif yaitu dari analisa pembelian di suatu pasar swalayan seperti berapa besar kemungkinan seseorang membeli roti bersamaan dengan mentega. Dengan pengetahuan tersebut pemilik pasar swalayan tersebut dapat mengatur posisi barang atau membuat suatu kupon diskon dengan mengkombinasikan barang tertentu [9]. Algoritma aturan asosiasi akan menggunakan data latihan, sesuai dengan pengertian *Data Mining*, untuk menghasilkan pengetahuan. Pengetahuan untuk mengetahui *item-item* belanja yang sering dibeli secara bersamaan dalam suatu waktu. Aturan asosiasi berbentuk “*if....then....*” atau “jika....maka....” merupakan pengetahuan yang dihasilkan dari fungsi aturan asosiasi.

Metodologi dasar analisis asosiasi dibagi menjadi dua tahap, yaitu :

a. Support

Support dari aturan asosiasi adalah presentasi dari kombinasi item dalam *database*, dimana jika memiliki *item A* dan *item B* maka *support* adalah proporsi transaksi di *database* yang berisi A dan B [10], [11]. Nilai *Support* suatu *item* diperoleh dengan rumus (Panjaitan, Amin, Lindawati, Watrianthos, Sihotang, & Sinaga, 2019).

$$Support(A) = \frac{Jumlah\ Transaksi\ Mengandung\ A}{Total\ Transaksi} \quad (1)$$

Setelah nilai *support* dari dua *item* telah didapatkan dari formula dibawah ini:

$$Support(A, B) = P(A \cap B) \quad (2)$$

b. Confidence

Setelah semua pola frekuensi tinggi ditemukan, barulah mencai aturan asosiatif yang memenuhi persyaratan minimum untuk *confidence* dengan menghitung *confidence* dengan menggunakan rumus berikut:

$$Confidence = P(B|A) = \frac{Jumlah\ Transaksi\ Mengandung\ A\ dan\ B}{Jumlah\ Transaksi\ Mengandung\ A} \quad (3)$$

3 Hasil dan Pembahasan

3.1 Data awal

Data yang berbentuk .csv dimasukkan ke dalam dataset lalu mengganti nilai nilai sel yang null menjadi nol (0). Terdapat beberapa variasi tipe data dari data yang didapat.

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name	actor_1_facebook_likes	gross	
0	Color	James Cameron	723.0	178.0	0.0	855.0	Joel David Moore	1000.0	760505847.0	Action
1	Color	Gore Verbinski	302.0	169.0	563.0	1000.0	Orlando Bloom	40000.0	309404152.0	Ac
2	Color	Sam Mendes	602.0	148.0	0.0	161.0	Rory Kinnear	11000.0	200074175.0	A
3	Color	Christopher Nolan	813.0	164.0	22000.0	23000.0	Christian Bale	27000.0	448130642.0	
4	0	Doug Walker	0.0	0.0	131.0	0.0	Rob Walker	131.0	0.0	

Gambar 1. Dataset yang didapatkan dari file .csv

3.2 Pre-processing

Dari dataset yang sudah didapat kami mengambil 1001 sampel data dimana kolom *duration* sudah diberi diklasifikasikan dan diberi label.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
0	Color	James Cameron	723.0	Medium	0.0	855.0	Joel David Moore	1000.0	760505847.0	Action Adventure Fantasy Sci-Fi	CCH Pounder	Avatar	886204	4834	Wes Studi	0.0	avatar future
1	Color	Gore Verbinski	302.0	Medium	563.0	1000.0	Orlando Bloom	40000.0	309404152.0	Action Adventure Fantasy	Johnny Depp	Pirates of the Caribbean: At World's End	471220	48350	Jack Davenport	0.0	goddess m
2	Color	Sam Mendes	602.0	Medium	0.0	161.0	Rory Kinnear	11000.0	200074175.0	Action Adventure Thriller	Christoph Waltz	Spectre	275868	11700	Stephanie Sigman	1.0	bomb esp
3	Color	Christopher Nolan	813.0	Medium	22000.0	23000.0	Christian Bale	27000.0	448130642.0	Action Thriller	Tom Hardy	The Dark Knight Rises	1144337	106759	Joseph Gordon-Levitt	0.0	deception impris
4	Color	Andrew Stanton	462.0	Medium	475.0	530.0	Samantha Morton	640.0	73058679.0	Action Adventure Sci-Fi	Daryl Sabara	John Carter	212204	1873	Polly Walker	1.0	all

Gambar 2. Dataset dengan kolom *duration* (3) yang sudah diberi label

Dataset berisi 1001 sampel akan dibuat menjadi banyak list dengan masing masing terdiri atas 30 item yang di filter berdasarkan sutradara dan durasi film.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
0	Gone Girl	Funny People	The Amazing Spider-Man 2	White House Down	Ghost Rider	Safe House	Tangled	Get Smart	A Christmas Carol	Resident Evil: Retribution	The Adventures of Tintin	Quantum of Solace	The Haunting	In the Name of the King: A Dungeon Siege Tale	Me, Myself & Irene	Nutty Professor II: The Klumps	The Matrix Revolutions
1	Avatar	Pirates of the Caribbean: At World's End	Spectre	The Dark Knight Rises	John Carter	Spider-Man 3	xxx: State of the Union	Avengers: Age of Ultron	Harry Potter and the Half-Blood Prince	I Spy	Superman Returns	The Dictator	Pirates of the Caribbean: Dead Man's Chest	The Lone Ranger	Man of Steel	The Chronicles of Narnia: Prince Caspian	The Avengers
2	Avatar	Pirates of the Caribbean: At World's End	Spectre	The Dark Knight Rises	John Carter	Spider-Man 3	Inkheart	Avengers: Age of Ultron	Harry Potter and the Half-Blood Prince	Town & Country	Superman Returns	Mona Lisa Smile	Pirates of the Caribbean: Dead Man's Chest	The Lone Ranger	Man of Steel	The Chronicles of Narnia: Prince Caspian	The Avengers
3	Avatar	Pirates of the Caribbean: At World's End	Spectre	The Dark Knight Rises	John Carter	Spider-Man 3	Captain America: The Winter Soldier	Avengers: Age of Ultron	Harry Potter and the Half-Blood Prince	Wall Street: Money Never Sleeps	Superman Returns	Town & Country	Pirates of the Caribbean: Dead Man's Chest	The Lone Ranger	Man of Steel	The Chronicles of Narnia: Prince Caspian	The Avengers
4	Blackhat	Click	The Core	Fury	The Girl with the Dragon Tattoo	Fool's Gold	Tangled	Arthur and the Invisibles	Resident Evil: Retribution	X-Men	Casino	Quantum of Solace	Australia	The Revenant	The Boxtrolls	Resident Evil: Retribution	Eat Pray Love

Gambar 3. Format dataset sampel yang telah disaring berdasarkan sutradara dan durasi film

3.3 Implementasi Algoritma

Dataset yang telah melewati tahap *Pre-processing* maka selanjutnya akan melewati proses pengolahan data menggunakan algoritma Apriori dengan variabel variabel berikut :

```
min_support=0.003
min_confidence=0.2
min_lift=3
```

Kami menggunakan library Apriori yang sudah disediakan python. Berikut hasil dari algoritma tersebut yang menghasilkan data berupa film yang terpilih beserta rekomendasinya.

	Film yang dipilih	Rekomendasi	Support	Confidence	Lift
0	Unforgotten	102 Dalmatians	0.004024	0.20	6.855172
1	2 Fast 2 Furious	Constantine	0.004024	0.20	4.733333
2	2 Fast 2 Furious	Everest	0.004024	0.20	5.372973
3	2 Fast 2 Furious	The Avengers	0.004024	0.20	3.550000
4	2 Fast 2 Furious	U-571	0.004024	0.20	6.212500
...
733	The Hobbit: The Desolation of Smaug	X-Men: Days of Future Past	0.003018	1.00	24.850000
734	The Scorch Trials	You Don't Mess with the Zohan	0.003018	0.75	25.706897
735	Tomorrowland	Two Weeks Notice	0.003018	1.00	32.064516
736	V for Vendetta	What Lies Beneath	0.003018	0.75	24.048387
737	Waterworld	Yes Man	0.003018	0.75	19.618421

738 rows x 5 columns

Gambar 4. Dataset yang dihasilkan dari implementasi algoritma Apriori

Kami juga menyajikan data berdasarkan *Support*, *Confidence*, dan *Lift* dimulai dari variabel dengan nilai tertinggi.

	Film yang dipilih	Rekomendasi	Support	Confidence	Lift
148	Men in Black 3	Quantum of Solace	0.032193	0.571429	8.352941
142	Man of Steel	Skyfall	0.015091	0.306122	6.915584
213	The Hobbit: The Battle of the Five Armies	The Hobbit: The Desolation of Smaug	0.012072	0.300000	6.934884
134	King Kong	The Hobbit: The Desolation of Smaug	0.012072	0.266667	6.164341
188	Spider-Man 3	Robin Hood	0.010060	0.256410	5.097436
...
391	The Man from U.N.C.L.E.	Walking Tall	0.003018	0.600000	15.694737
390	The Bourne Supremacy	The Man from U.N.C.L.E.	0.003018	0.750000	21.300000
389	Ladder 49	Mad Max: Fury Road	0.003018	0.750000	23.296875
388	Cats & Dogs	The Lord of the Rings: The Two Towers	0.003018	0.750000	23.296875
737	Waterworld	Yes Man	0.003018	0.750000	19.618421

738 rows x 5 columns

Gambar 5. Dataset berdasarkan nilai *Support*

	Film yang dipilih	Rekomendasi	Support	Confidence	Lift
283	Analyze That	The Love Guru	0.003018	1.0	34.275862
633	Men in Black 3	Quantum of Solace	0.003018	1.0	14.617647
735	Tomorrowland	Two Weeks Notice	0.003018	1.0	32.064516
229	2 Guns	Man of Steel	0.003018	1.0	20.285714
230	2 Guns	Tangled	0.003018	1.0	20.708333
...
157	Monster House	Nutty Professor II: The Klumps	0.004024	0.2	6.024242
103	Harry Potter and the Goblet of Fire	Jurassic World	0.004024	0.2	3.106250
104	Harry Potter and the Goblet of Fire	Star Trek Beyond	0.004024	0.2	6.855172
195	Tomorrow Never Dies	Something's Gotta Give	0.004024	0.2	5.231579
93	Gladiator	Tangled	0.004024	0.2	4.141667

738 rows x 5 columns

Gambar 6. Dataset berdasarkan nilai *Confidence*

	Film yang dipilih	Rekomendasi	Support	Confidence	Lift
365	Burlesque	The SpongeBob Movie: Sponge Out of Water	0.003018	1.000000	66.266667
513	Madagascar: Escape 2 Africa	The Dark Knight	0.003018	1.000000	47.333333
361	Bridget Jones: The Edge of Reason	Warcraft	0.003018	1.000000	45.181818
275	Robots	War of the Worlds	0.003018	1.000000	45.181818
246	A Beautiful Mind	Men in Black	0.003018	1.000000	45.181818
...
80	Escape Plan	Jurassic World	0.005030	0.208333	3.235677
127	Marley & Me	Jurassic World	0.006036	0.206897	3.213362
47	Brave	Jurassic World	0.006036	0.206897	3.213362
128	Star Trek Beyond	Jurassic World	0.006036	0.206897	3.213362
103	Harry Potter and the Goblet of Fire	Jurassic World	0.004024	0.200000	3.106250

738 rows × 5 columns

Gambar 7. Dataset berdasarkan nilai *Lift*

4 Simpulan

Berdasarkan percobaan kami mengenai implementasi algoritma priori. Dapat dilihat bahwa Algoritma apriori memiliki kemampuan untuk mencari pola pada data tertentu berdasarkan parameter parameter yang ditentukan. Juga melihat dari hasil data yang didapatkan terdapat banyak sekali data dengan nilai Confidence = 1.0 yang berarti jumlah data masih belum cukup untuk mendapatkan hasil yang akurat dan lebih memuaskan. Selain itu juga terdapat faktor kurangnya analisis yang mendalam pada saat Pre-Processing yang dapat mengakibatkan kesalahan interpretasi data sehingga data yang dihasilkan tidak sesuai dengan apa yang diharapkan.

Referensi

- [1] R. R. Ramadina, T. H. Pudjiantoro, and I. Santikarama, "Pembangunan Sistem Customer Relationship Management (CRM) Menggunakan Metode Asosiasi Algoritma Apriori," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 2, no. 2, p. 78, 2020, doi: 10.36499/jinrpl.v2i2.3539.
- [2] T. Badriyah, R. Fernando, and I. Syarif, "Sistem Rekomendasi Content Based Filtering Menggunakan Algoritma Apriori," *Konf. Nas. Sist. Inf.*, vol. 1, no. 1, pp. 554–559, 2018.
- [3] C. S. Fatoni, E. Utami, and F. W. Wibowo, "Sistem Rekomendasi Produk Online Store Menggunakan Metode Apriori," *J. Inf. Politek. Indonusa Surakarta*, vol. 4, no. 2, pp. 20–27, 2018.
- [4] S. Saefudin and D. Fernando, "PENERAPAN DATA MINING REKOMENDASI BUKU MENGGUNAKAN ALGORITMA APRIORI," *JSii (Jurnal Sist. Informasi)*, vol. 7, no. 1, p. 50, 2020, doi: 10.30656/jsii.v7i1.1899.
- [5] D. S. Maylawati, "The Concept of Frequent Itemset Mining for Text," in *AASEC 2018*, 2018.
- [6] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," *J. Comput. Sci. Technol.*, vol. 15, no. 6, pp. 487–499, 1994, doi: 10.1007/BF02948845.
- [7] J. Han and M. Kamber, "Data Mining: Concepts and Techniques," *Ann. Phys. (N. Y.)*, vol. 54,

- p. 770, 2006, doi: 10.5860/CHOICE.49-3305.
- [8] M. A. Azis, N. Hadianto, J. Miharja, and S. Rifai, “IMPLEMENTASI ALGORITMA APRIORI UNTUK ANALISA PEMILIHAN TIPE GENRE FILM ANIME (STUDI KASUS : MYANIMELIST.NET),” *J. Pilar Nusa Mandiri*, vol. 14, no. 2, p. 209, 2018, doi: 10.33480/pilar.v14i2.932.
- [9] K. Tampubolon, H. Saragih, B. Reza, K. Epicentrum, and A. Asosiasi, “Implementasi Data Mining Algoritma Apriori pada sistem persediaan alat-alat kesehatan,” *Maj. Ilm. Inf. dan Teknol. Ilm.*, vol. 1, no. 1, pp. 93–106, 2013.
- [10] S. Dutt, N. Choudhary, and D. Singh, “An Improved Apriori Algorithm based on Matrix Data Structure,” *Glob. J. Comput. Sci. Technol. C Softw. Data Eng.*, vol. 14, no. 5, pp. 1–5, 2014.
- [11] S. Liu and X. Jiyi, “An improved apriori algorithm based on matrix,” in *Proceedings - 2020 12th International Conference on Measuring Technology and Mechatronics Automation, ICMTMA 2020*, 2020, pp. 488–491, doi: 10.1109/ICMTMA50254.2020.00111.