

Penerapan Data Mining untuk Analisa Pemilihan Genre Film menggunakan Algoritma Apriori (Studi Kasus: imdb.com)

Application of Data Mining for Analysis of Film Genre Selection using a Priori Algorithm (Case Study: imdb.com)

Alfina Damayanti^{1*}, Gusmelia Testiana²

^{1,2}Sistem Informasi, Universitas Islam Negeri Raden Fatah Palembang

*damaflaviayanti@gmail.com

Abstrak

Situs imdb.com adalah salah satu sumber data film terbesar dan paling terpercaya yang menyediakan informasi tentang film-film dari berbagai genre. Dari jutaan data film yang pada situs imdb.com, terdapat beberapa genre yang paling sering digunakan, diantaranya action, adventure, animation, comedy, dll. Penerapan data mining bertujuan mempermudah mencapai tujuan yaitu untuk menganalisa pemilihan genre film dengan menggunakan algoritma Apriori. Algoritma Apriori dipilih untuk menentukan aturan asosiasi untuk menganalisis genre film apa yang cocok untuk pengunjung website imdb.com berdasarkan genre film yang pernah ia kunjungi. Metode yang digunakan adalah Pendekatan dimulai dengan pengumpulan data, Algoritma Apriori dan Tahapan data mining yang terdiri dari Pemanaman bisnis, data, persiapan data, pemodelan dan evaluasi. Dengan penelitian ini mendapatkan hasil membantu produsen dan pembuat film dalam membuat strategi pemasaran didunia perfilman dengan efektif dan efisien dalam tingkat keberhasilan pembuatan film dari beberapa genre yang telah ada.

Kata kunci: Algoritma Apriori, Genre, Film, Data Mining.

Abstract

The imdb.com site is one of the largest and most trusted film data sources that provides information about films of various genres. From the millions of film data on the imdb.com site, there are several genres that are most often used, including action, adventure, animation, comedy, etc. The application of data mining aims to make it easier to achieve the goal, namely to analyze the selection of film genres using the Apriori algorithm. The Apriori algorithm was chosen to determine association rules to analyze what film genres are suitable for visitors to the imdb.com website based on the film genres they have visited. The method used is the approach starting with data collection, the Apriori Algorithm and the stages of data mining which consist of business management, data, data preparation, modeling and evaluation. With this research, the results will assist producers and filmmakers in making marketing strategies in the world of cinema effectively and efficiently in terms of the success rate of making films from several existing genres.

Keywords: Apriori Algorithm, Genre, Movie, Data Mining.

PENDAHULUAN

Pada era globalisasi ini, ada banyak kemudahan yang ditawarkan untuk membantu pekerjaan dan memenuhi kebutuhan manusia. Kemudahan ini dapat diperoleh melalui media

cetak maupun elektronik [1]. Dalam perkembangannya, media mampu memberikan hiburan kepada penggunanya, terutama melalui media elektronik seperti televisi dan akses internet. Melalui media ini, berbagai jenis hiburan dapat diakses, termasuk film-film

seperti kartun, layar lebar, animasi, dan drama. Berbagai situs website pun bermunculan untuk membantu penonton menentukan atau memilih film yang sesuai keinginannya, salah satunya adalah situs imdb.com [2]. Situs imdb.com adalah salah satu sumber data film terbesar dan paling terpercaya yang menyediakan informasi tentang film-film dari berbagai genre. Data yang ada di imdb.com meliputi judul film, informasi tentang pemain, sutradara, dan, yang paling penting, genre film [3]. Halaman pada website imdb.com menampilkan daftar film yang sedang rilis, film yang akan rilis, dan film yang menjadi favorit fan. Selain itu website ini juga menampilkan menu explore to streaming, yang berisi daftar film dari berbagai genre secara acak [4]. Situs website imdb.com memiliki jutaan judul film dengan berbagai genre yang berbeda, yaitu action, adventure, animation, comedy, dll. Perkembangan persaingan pada industri film mendorong para kontributor untuk meningkatkan daya tarik film yang mereka produksi agar dapat bersaing dengan lebih baik. Terdapat berbagai cara untuk meningkatkan daya tarik suatu film, di antaranya adalah dengan memilih genre film yang sesuai dan memperhatikan tampilan visual [5]. Salah satu strategi untuk meningkatkan daya tarik film adalah dengan memanfaatkan data preferensi genre yang sering ditonton oleh para penggemar film [6]. Dengan menggunakan data tersebut, situs website dapat menganalisis genre-genre apa yang paling diminati dan cenderung dipilih oleh para penggemar [7]. Pemilihan genre film secara manual bisa menjadi tugas yang rumit dan rentan terhadap kecenderungan subjektivitas. Oleh karena itu, penerapan algoritma Apriori sebagai metode analisis dapat memberikan pendekatan objektif dan otomatis untuk menentukan genre film [8]. Apriori adalah suatu algoritma yang sudah sangat dikenal dalam melakukan pencarian frequent item set dengan menggunakan teknik association rule. Algoritma apriori menggunakan knowledge mengenai frequent itemset yang sebelumnya telah diketahui, untuk memproses informasi selanjutnya. Algoritma ini digunakan dalam data mining dan analisis asosiasi untuk mengidentifikasi pola dan hubungan antara item-item dalam dataset [9]. Metode ini memberikan beberapa manfaat signifikan. Pertama, pemilihan genre

film dapat dilakukan secara otomatis dan objektif, mengurangi ketergantungan pada penilaian subjektif manusia. Kedua, dengan menganalisis dataset dari imdb.com, dapat ditemukan pola dan tren yang mungkin tidak terdeteksi sebelumnya, yang dapat membantu dalam pemahaman pasar dan preferensi pengunjung [10]. Ketiga, hasil dari analisis ini dapat membantu produsen dan pembuat film dalam mengarahkan strategi pemasaran yang efektif dan meningkatkan peluang keberhasilan film. Berdasarkan latar belakang yang telah diuraikan, penelitian ini bertujuan untuk menganalisis pemilihan genre film pada situs imdb.com [11]. Dalam riset ini algoritma yang digunakan dalam pengolahan informasi adalah algoritma apriori. Algoritma tercantum jenis ketentuan asosiasi pada informasi mining, Algoritma Apriori yang bertujuan buat menemukan frequent item sets dijalankan pada sekumpulan informasi [12]. Analisis Apriori didefinisikan sesuatu proses buat menciptakan seluruh aturan apriori yang penuh ketentuan minimum untuk support serta ketentuan minimum buat confidence. algoritma apriori bagus buat digunakan untuk pembuatan itemset, pencarian pola serta lain sebagainya [13].

METODE PENELITIAN

2.1 Metode Pengumpulan Data

Penelitian ini menggunakan data sekunder untuk mendapatkan dataset yang digunakan. Data sekunder adalah data yang didapatkan secara tidak langsung mengacu pada situasi di mana data tidak diberikan langsung kepada pengumpul datanya, melainkan melalui perantara seperti orang lain atau dokumen tertulis. Data penelitian ini diambil dari situs Kaggle.com yang datasetnya diambil langsung dari website resmi imdb.com. Jumlah sampel yang digunakan pada penelitian ini adalah 125 data yang terdiri dari data berbagai genre film.

2.2 Algoritma Apriori

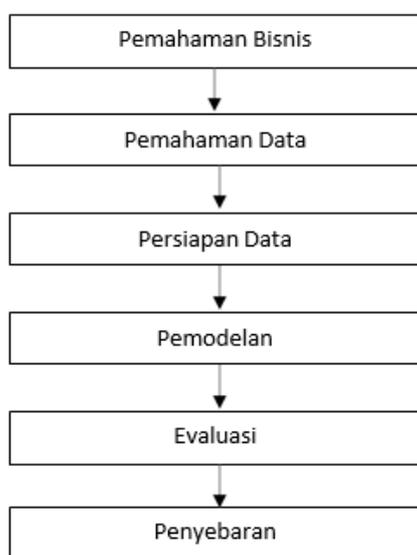
Algoritma Apriori adalah salah satu algoritma pada data mining untuk mencari frequent item/itemset pada transaksional database. Algoritma apriori pertama kali diperkenalkan oleh R. Agarwal dan R. Srikant untuk mencari frequent tertinggi dari suatu database. Untuk menentukan asosiasi rule

mining sebuah transaksi database, diperlukan waktu dalam melakukan proses frequent itemset, menghasilkan kombinasi data yang cukup banyak. Proses ini dilakukan untuk mencari minimum nilai support dan minimum nilai confidence.

2.3 Tahapan Data Mining

Dalam melakukan analisa dan mencari pola data untuk dijadikan sebuah dataset dalam memudahkan penelitian dan dapat berjalan dengan sistematis dan memenuhi tujuan yang diinginkan maka dibuat alur dalam tahapan penelitian yang akan dilakukan berikut: Data mining merupakan proses ekstraksi informasi yang berharga dari kumpulan data besar dan kompleks dengan tujuan menghasilkan informasi baru yang dapat digunakan dalam pengambilan keputusan.

Data mining melibatkan analisis data yang beragam dan menarik kesimpulan yang menjadi pengetahuan atau pola penting untuk meningkatkan keuntungan, mengurangi biaya pengeluaran, atau bahkan mencapai kedua hal tersebut. Berikut adalah tahapan data mining.



Gambar 1 Tahapan Data Mining Pada Pemrosesan Pemilihan Genre Film

- 1) Tahap Pemahaman Bisnis
 - a) Mendefinisikan dengan rinci tujuan proyek dan kebutuhan dalam konteks bisnis atau unit penelitian secara menyeluruh.
 - b) Mengubah tujuan dan batasan menjadi formula yang mewakili

permasalahan dalam data mining.

- c) Merancang strategi awal untuk mencapai tujuan yang telah ditetapkan.
- 2) Tahap Pemahaman Data
 - a) Melakukan pengumpulan data.
 - b) Menerapkan analisis investigasi data untuk lebih memahami data dan mencari pemahaman awal.
 - c) Menilai kualitas data.
- 3) Tahap Persiapan Data
 - a) Mempersiapkan data awal dengan mengumpulkan data yang akan digunakan dalam semua fase berikutnya.
 - b) Memilih kasus dan variabel yang ingin dianalisis sesuai dengan jenis analisis yang akan dilakukan.
 - c) Melakukan perubahan pada beberapa variabel jika diperlukan.
 - d) Menyiapkan data awal agar siap digunakan dalam proses pemodelan.
- 4) Tahap Pemodelan
 - a) Memilih dan menerapkan teknik pemodelan yang tepat.
 - b) Jika diperlukan, proses dapat kembali ke tahap pengolahan data untuk mengubah data menjadi format yang sesuai dengan spesifikasi teknik data mining yang dibutuhkan.
- 5) Tahap Evaluasi
 - a) Melakukan evaluasi terhadap satu atau lebih model yang digunakan dalam tahap pemodelan untuk mengukur kualitas dan efektivitas sebelum diterapkan secara luas.
 - b) Menentukan apakah terdapat model yang memenuhi tujuan yang ditetapkan pada tahap awal.
 - c) Mengidentifikasi apakah ada permasalahan penting dalam bisnis atau penelitian yang tidak tertangani dengan baik oleh model yang telah dibangun.
 - d) Mengambil keputusan terkait penggunaan hasil dari data mining, mempertimbangkan implementasinya dalam konteks yang relevan.

- 6) Tahap Penyebaran
 - a) Menerapkan model yang telah dihasilkan. Namun, pembentukan model itu sendiri tidak menandakan bahwa proyek telah selesai.

HASIL DAN PEMBAHASAN

3.1 Tahap Pemahaman Bisnis

Tahap pemahaman bisnis, yang juga dikenal sebagai tahap pemahaman penelitian, bertujuan untuk memahami permasalahan yang ada dan menetapkan tujuan dalam melakukan data mining. Seperti yang telah dijelaskan dalam latar belakang permasalahan, penelitian ini bertujuan untuk menganalisis pemilihan genre film pada situs imdb.com.

3.2 Tahap Pemahaman Data

Tahap kedua adalah pemahaman data. Tahap ini bertujuan untuk memahami karakteristik data yang akan diolah. Adapun data yang akan digunakan dalam proses data mining adalah sampel data dari situs Kaggle.

# Rating	# Runtime	A_Censor	A_Total_Gross	A_main_genre	A_side_genre
9.3	148	UA	Gross Unknown	Action	Adventure, Drama
9.8	152	UA	\$334.86M	Action	Crime, Drama
9.8	201	U	\$377.85M	Action	Adventure, Drama
8.6	148	UA	\$192.58M	Action	Adventure, Sci-Fi
8.8	179	UA	\$342.55M	Action	Adventure, Drama
8.8	178	U	\$115.54M	Action	Adventure, Drama

Gambar 2 Dataset Awal Pada Pemrosesan Pemilihan Genre Film

3.3 Tahap Persiapan Data

Dataset IMDB Movies Genre memiliki beberapa atribut yaitu Movie_Title, Year, Director, Actors, Rating, Main_Genre, Side_Genre, Runtimes, Cencor, dan Total_Gross.

Dalam penerapan algoritma Apriori untuk menentukan *association rule* genre film situs

website imdb.com, hanya digunakan atribut Main_Genre dan Side_Genre saja. Selanjutnya kedua atribut tersebut digabung hingga menjadi atribut tunggal yang dapat digunakan untuk menentukan *association rule*.

Tabel 1 Daftar Genre Pada Pemrosesan Pemilihan Genre Film

No.	Genre
1	Action
2	Adventure
3	Animation
4	Comedy
5	Crime
6	Drama
7	Family
8	Horror
9	Mystery
10	Romance
11	Thriller

Jumlah sampel yang digunakan pada penelitian ini adalah 125 data.

Tabel 2 Daftar Genres

No.	Genres
1	Comedy, Romance
2	Action, Drama, Thriller
3	Adventure, Comedy, Family
4	Comedy, Romance
5	Crime, Drama, Thriller
6	Comedy, Romance
7	Action, Adventure, Crime
8	Drama, Romance
9	Comedy, Romance
10	Action, Adventure, Crime
11	Drama, Horror, Mystery
12	Adventure, Comedy, Family
13	Action, Crime, Drama
14	Crime, Drama, Thriller
15	Action, Adventure, Thriller
16	Horror, Thriller
17	Drama, Romance, Thriller
18	Comedy, Romance
19	Comedy, Family, Romance
20	Animation, Drama, Family
21	Comedy, Drama, Romance
22	Action, Adventure, Crime
23	Comedy, Drama, Romance
24	Drama, Horror, Mystery
25	Action, Adventure, Crime

26	Comedy, Romance	76	Crime, Drama
27	Action, Drama, Thriller	77	Comedy, Romance
28	Adventure, Comedy, Family	78	Horror, Thriller
29	Comedy, Romance	79	Crime, Drama
30	Comedy, Drama, Romance	80	Adventure, Comedy, Family
31	Animation, Drama, Family	81	Drama, Horror, Mystery
32	Crime, Drama, Thriller	82	Adventure, Comedy, Family
33	Action, Comedy, Crime	83	Action, Crime, Drama
34	Animation, Drama, Family	84	Crime, Drama, Thriller
35	Crime, Drama, Thriller	85	Action, Adventure, Thriller
36	Action, Comedy, Crime	86	Horror, Thriller
37	Crime, Drama, Thriller	87	Drama, Romance, Thriller
38	Action, Adventure, Thriller	88	Comedy, Romance
39	Comedy, Romance	89	Comedy, Family, Romance
40	Crime, Drama, Thriller	90	Animation, Drama, Family
41	Action, Adventure, Thriller	91	Comedy, Romance
42	Horror, Thriller	92	Drama, Horror, Mystery
43	Drama, Thriller	93	Crime, Drama, Thriller
44	Animation, Drama, Family	94	Horror, Thriller
45	Crime, Drama	95	Animation, Drama, Family
46	Adventure, Comedy, Family	96	Crime, Drama, Thriller
47	Drama, Horror, Mystery	97	Action, Adventure, Crime
48	Comedy, Drama, Romance	98	Crime, Drama, Thriller
49	Comedy, Romance	99	Action, Adventure, Thriller
50	Crime, Drama, Thriller	100	Comedy, Romance
51	Horror, Thriller	101	Horror, Thriller
52	Adventure, Comedy, Family	102	Crime, Drama, Thriller
53	Crime, Drama, Thriller	103	Comedy, Romance
54	Drama, Horror, Mystery	104	Action, Adventure, Crime
55	Comedy, Drama, Romance	105	Comedy, Drama, Romance
56	Action, Adventure, Crime	106	Crime, Drama, Thriller
57	Action, Comedy, Crime	107	Drama, Horror, Thriller
58	Comedy, Drama, Romance	108	Animation, Drama, Family
59	Drama, Horror, Mystery	109	Crime, Drama
60	Drama, Thriller	110	Action, Adventure, Crime
61	Drama, Romance	111	Crime, Drama, Thriller
61	Animation, Drama, Thriller	112	Action, Adventure, Thriller
63	Drama, Horror, Mystery	113	Comedy, Romance
64	Adventure, Comedy, Family	114	Action, Drama, Thriller
65	Crime, Drama	115	Adventure, Comedy, Family
66	Horror, Thriller	116	Action, Adventure, Crime
67	Animation, Drama, Family	117	Crime, Drama
68	Crime, Drama	118	Crime, Drama, Thriller
69	Action, Comedy, Crime	119	Action, Adventure, Thriller
70	Crime, Drama, Thriller	120	Comedy, Romance
71	Action, Adventure, Thriller	121	Crime, Drama
72	Comedy, Drama, Romance	122	Drama, Romance
73	Action, Drama, Thriller	123	Action, Adventure, Crime
74	Adventure, Comedy, Family	124	Action, Adventure, Crime
75	Action, Adventure, Crime	125	Adventure, Comedy, Family

3.4 Tahap Pemodelan

Tahap ini merupakan proses menggalidata dari sebuah dataset atau kumpulan data untuk mendapatkan informasi tersembunyi dari data yang diolah. *Tool Microsoft Excel* digunakan untuk melakukan perhitungan manual menggunakan, sedangkan untuk pengujiannya menggunakan *Software Rapid Miner*.

Pada penelitian ini diterapkan asosiasi dengan menggunakan algoritma yang telah dipilih yaitu Apriori. 125 data dijadikan sampel untuk proses perhitungan algoritma Apriori untuk menganalisis pemilihan genre film pada imdb.com. Tahap-tahap pemodelan data menggunakan algoritma Apriori adalah sebagai berikut.

1) Pembentukan *Items*

Langkah pertama pada tahap ini adalah menentukan itemset. Dalam penelitian ini tahap pembentukan items juga digunakan untuk mengetahui genre yang paling sering dilihat oleh para pengunjung situs imdb.com Berikut adalah penyelesaian pembentukan 1 itemset dengan jumlah minimum support 20%. Untuk menghitung nilai support dari masing-masing genre digunakan rumus sebagai berikut :

$$= \frac{\sum \text{Pemilihan Mengandung A}}{\sum \text{Pemilihan}} * 100\% \quad [1]$$

Rumus diatas menjelaskan bahwa pembentukan item didapat dengan cara membagi jumlah pemilihan mengandung item A dibagi dengan jumlah total pemilihan yang diambil, hasilnya kemudian dikalikan 100%. Dengan menggunakan rumus tersebut, berikut hasil nilai *support* untuk tiap item yang ada.

a) Menghitung Item Action

$$= \frac{30}{125} * 100\% \\ = 24\%$$

b) Menghitung Item Adventure

$$= \frac{31}{125} * 100\% \\ = 25\%$$

c) Menghitung Item Animation

$$= \frac{9}{125} * 100\% \\ = 7\%$$

d) Menghitung Item Comedy

$$= \frac{42}{125} * 100\% \\ = 34\%$$

e) Menghitung Item Crime

$$= \frac{43}{125} * 100\% \\ = 34\%$$

f) Menghitung Item Drama

$$= \frac{64}{125} * 100\% \\ = 51\%$$

g) Menghitung Item Family

$$= \frac{21}{125} * 100\% \\ = 17\%$$

h) Menghitung Item Horror

$$= \frac{17}{125} * 100\% \\ = 14\%$$

i) Menghitung Item Mystery

$$= \frac{8}{125} * 100\% \\ = 6\%$$

j) Menghitung Item Romance

$$= \frac{31}{125} * 100\% \\ = 25\%$$

k) Menghitung Item Thriller

$$= \frac{42}{125} * 100\% \\ = 34\%$$

Tabel 3 Calon Pola 1 Itemset

No.	Item	Support
1	Action	24%
2	Adventure	25%

3	Animation	7%
4	Comedy	34%
5	Crime	34%
6	Drama	51%
7	Family	17%
8	Horror	14%
9	Mystery	6%
10	Romance	25%
11	Thriller	34%

Pada penelitian ini, minimum support yang ditentukan adalah 20%, sehingga Item yang memiliki nilai support di bawah 21% dihapuskan dari daftar Itemset. Berikut adalah data itemset yang telah valid.

Tabel 4 Pola 1 Itemset Valid

No.	Item	Support
1	Action	24%
2	Adventure	25%
3	Comedy	34%
4	Crime	34%
5	Drama	51%
6	Romance	25%
7	Thriller	34%

2) Kombinasi 2 Itemset

Pada proses penyusunan 2 Itemset, persentase *minimal support* masih tetap pada 20%. Untuk menghitung nilai *support 2* Itemset, digunakan rumus berikut:

$$= \frac{\sum \text{Pemilihan Mengandung } A \cap B}{\sum \text{Pemilihan}} * 100\% \quad [2]$$

Yaitu dengan cara membagi jumlah pemilihan yang mencakup tipe genre A dan tipe genre B dengan total jumlah pemilihan yang dilakukan, lalu hasilnya dikalikan dengan 100%. Sebelum menghitung *support 2* itemset, terlebih dahulu dicari kombinasi antara banyak itemset valid (*n*) dengan banyak kombinasi yang akan digunakan (*r*), dengan rumus:

$$C(n, r) = \frac{n!}{r!(n-r)!} \quad [3]$$

Maka hasilnya:

$$C(7,2) = \frac{7!}{2!(7-2)!} = 21 \text{ item}$$

Berikut merupakan tabel 2 itemset yang dihitung berdasarkan rumus.

Tabel 5 Pola 2 Itemset

No.	Itemset	Support
1	Action, Adventure	16%
2	Action, Comedy	3%
3	Action, Crime	14%
4	Action, Drama	4%
5	Action, Romance	0%
6	Action, Thriller	10%
7	Adventure, Comedy	9%
8	Adventure, Crime	10%
9	Adventure, Drama	0%
10	Adventure, Romance	0%
11	Adventure, Thriller	6%
12	Comedy, Crime	3%
13	Comedy, Drama	6%
14	Comedy, Romance	21%
15	Comedy, Thriller	0%
16	Crime, Drama	22%
17	Crime, Romance	0%
18	Crime, Thriller	13%
19	Drama, Romance	10%
20	Drama, Thriller	21%
21	Romance, Thriller	2%

Berdasarkan nilai minimum yang telah ditentukan sebelumnya, maka nilai *support* di bawah 20% akan dihapuskan sehingga terbentuklah 2itemset valid sebagai berikut.

Tabel 6 Pola 2 Itemset Valid

No.	Itemset	Support
1	Comedy, Romance	21%
2	Crime, Drama	22%
3	Drama, Thriller	21%

3) Kombinasi 3 Itemset

Dalam pembentukan kombinasi 3 itemset masih menggunakan nilai *support* terendah 20%. Dengan kombinasi:

$$C(3,3) = \frac{3!}{3!(3-3)!} = 1 \text{ item}$$

Kombinasi 3 itemset hanya menghasilkan satu pola, yaitu sebagai berikut.

Tabel 7 Pola 3 Itemset

Itemset	Support
Crime, Drama, Thriller	14%

Tabel tersebut menampilkan nilai *support* dari 3 itemset adalah di bawah 20%, artinya bisa dipastikan bahwa penelitian ini tidak menghasilkan kombinasi 3 itemset.

4) Pembentukan Aturan Asosiasi

Setelah pola frekuensi tinggi ditemukan menggunakan langkah- langkah di atas, langkah selanjutnya adalah mencari aturan asosiasi. Pada proses ini ditemukan aturan asosiasi yang melewati syarat minimum untuk *confidence* berdasarkan perhitungan *confidence* aturan asosiatif $A \rightarrow B$, persamaan ditunjukkan padapersamaan (4).

$$= \frac{\sum \text{Pemilihan Mengandung } A \cap B}{\sum \text{Pemilihan Mengandung } A} * 100\% \quad [4]$$

Untuk menentukan nilai *confidence* dari $A \cap B$ adalah dengan membagi jumlah pemilihan yang mengandung item A dan mengandung item B dengan jumlah pemilihan item yang mengandung item A, lalu dikalikan dengan 100%. Pada penelitian ini minimum nilai *confidence* yang digunakan adalah 60%.

Tabel 8 Pembentukan Aturan Asosiasi

Aturan Asosiasi	Confidence
{Comedy, Romance} → Comedy	62%
{Comedy, Romance} → Romance	84%
{Crime, Drama} → Crime	63%
{Crime, Drama} → Drama	42%

{Drama, Thriller} → Drama	41%
{Drama, Thriller} → Thriller	62%

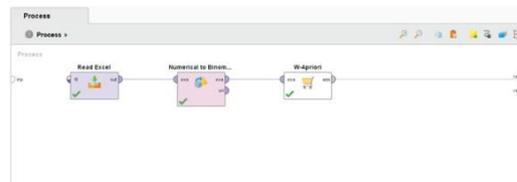
Dari tabel di atas, dapat ditentukan aturan asosiasi berdasarkan nilai minimum *confidence* 60%, yaitu sebagai berikut.

Tabel 9 Aturan Asosiasi

Aturan Asosiasi	Confidence
Jika menonton Comedy, maka menonton Romance	62%
Jika menonton Romance, maka menonton Comedy	84%
Jika menonton Crime maka, menonton Drama	63%
Jika menonton Thriller, maka menonton Drama	62%

3.5 Tahap Evaluasi

Tahap evaluasi bertujuan untuk mengetahui performa akurasi dari model yang telah dibangun. Pada penelitian ini, evaluasi dilakukan pada *software* Rapid Miner.



Gambar 3 Pemodelan di Rapid Miner

Hasil dari pemodelan tersebut adalah sebagai berikut:

W-Apriori

```

Apriori
=====
Minimum support: 0.2
(25instances)
Minimum metric <confidence>:
0.6Number of cycles performed:
7

Generated sets of large itemsets:

Size of set of large
itemsetsL(1): 7
    
```

```
Large Itemsets
L(1):Action=true
30
Adventure=true 31
Comedy=true 42
Crime=true 43
Drama=false 61
Romance=true 31
Thriller=true 43

Size of set of large
itemsetsL(2): 10
```

```
Large Itemsets L(2):
Action=true Adventure=true
20
Action=true Crime=true 18
Action=true Drama=false 24
Adventure=true Drama=false 31
Comedy=true Drama=false 34
Comedy=true Romance=true 26
Crime=true Drama=true 27
Crime=true Thriller=true 17
Drama=false Romance=true 18
Drama=true Thriller=true 26
```

```
Size of set of large
itemsetsL(3): 4
```

```
Large Itemsets L(3):
Action=true
Adventure=true
Drama=false 20
Action=true Crime=true
Drama=false16
Comedy=true Drama=false
Romance=true 18
Crime=true
Drama=true
Thriller=false 17
```

Best rules found:

1. Adventure=true 31 ==>
Drama=false 31
conf:(1)
2. Action=true Adventure=true 20
=> Drama=false 20 conf:(1)
3. Drama=false Romance=true 18
=>Comedy=true 18 conf:(1)
4. Crime=true Drama=false 16
=>Action=true 16 conf:(1)
5. Action=true Crime=true 18
=>Drama=false 16 conf:(0.89)
6. Romance=true 31 ==>
Comedy=true26 [conf:\(0.84\)](#)

3.6 Tahapan Penyebaran

Tahapan penyebaran dapat dilakukan dengan menyusun laporan, yaitu mendokumentasikan apa yang relevan dengan pekerjaan penelitian yang dilakukan. Menuliskan temuan dari penelitian dan mengubahnya menjadi argumen yang disajikan dalam bentuk laporan agar dapat digunakan sebagai referensi.

KESIMPULAN

Kesimpulan pada penelitian ini dapat diperoleh dari proses perhitungan yang telah dilakukan. Penerapan data mining pada penelitian ini bertujuan mempermudah mencapai tujuan. Berdasarkan perhitungan yang telah dilakukan, didapatkan genre yang paling banyak digunakan adalah genre Drama. Dengan data asosiasi yang telah didapatkan dapat membantu memilih tampilan pada beranda website imdb.com, yaitu pada menu explore to streaming, untuk menentukan genre apa saja yang cocok untuk ditampilkan dan tidak. Selain itu hasil dari penelitian ini juga dapat membantu pengunjung menentukan lebih cepat film yang ingin dilihatnya. Hasil implementasi telah dilakukan dengan menghasilkan nilai strong confidence paling tinggi 84%. Jika diterapkan secara efektif, pemanfaatan dan analisis data Asosiasi dapat memberikan kontribusi yang signifikan dalam perencanaan manajemen perfilman. Untuk penelitian berikut dapat menggunakan data yang lebih banyak dengan algoritma lain yang berbeda.

REFERENSI

- [1] R. N. Dasmien, S. Elfahmi, and W. D. Septiani, "Analisa Jaringan Local Area Network (LAN) dengan Aplikasi Cisco Packet Tracer," vol. 2, no. 2, pp. 45–49, 2022.
- [2] R. N. Dasmien, F. Fatoni, A. Wijaya, B. Tujni, and S. Nabila, "Pelatihan uji kegunaan website menggunakan System Usability Scale (SUS)," *ABSYARA J. Pengabd. Pada Masy.*, vol. 2, no. 2, pp. 146–158, 2021, doi: 10.29408/ab.v2i2.4031.
- [3] R. N. Dasmien and F. Kurniawan,

- “Digital Forensik Deleted Cyber Crime Evidence pada Pesan Instan Media Sosial,” *Techno.Com*, vol. 20, no. 4, pp. 527–539, 2021, doi: 10.33633/tc.v20i4.5170.
- [4] F. Azhiman and R. N. Dasmien, “Sosialisasi Performansi Antena Omnidirectional Palembang PT . LEN REKAPRIMA SEMESTA LRT Kota,” vol. 5, no. 2, pp. 359–371, 2022.
- [5] F. Azhiman, R. N. Dasmien, and T. D. Purwanto, “Pelatihan Pemanfaatan Online Storage pada SMK Nurul Huda Pemulutan Barat dengan Live Zoom Meeting,” vol. 1, pp. 21–26, 2023.
- [6] M. Alwi and N. Anggraini, “Analisis Data Mining Pada Pemilihan Jenis Game,” vol. 11, no. 1, pp. 9–15, 2023.
- [7] F. Azhiman, R. N. Dasmien, A. Putra, and W. Agustian, “Collecting Data Desa di Kecamatan Rambutan Kabupaten Banyuasin untuk Pengimplementasian Sistem Digital Desa,” vol. 2, no. 6, pp. 1733–1742, 2022.
- [8] H. Wintolo and A. Farhati, “Pembagian jaringan komputer menggunakan virtual local area network guna mendukung perpustakaan digital,” vol. 8, no. 2, pp. 133–150, 2020.
- [9] A. T. Laksono and M. A. H. Nasution, “Implementasi Keamanan Jaringan Komputer Local Area Network Menggunakan Access Control List pada Perusahaan X,” vol. 1, pp. 83–88, 2020, doi: 10.30865/json.v1i2.1920.
- [10] S. Hidayatulloh and M. Mastur, “Penerapan Simple Queue Dalam Pengelolaan Bandwidth Local Area Network (Studi Kasus : PT Sumber Berkah Niaga),” vol. 2, no. 2, 2020.
- [11] A. Gunawan, S. Ningsih, and D. A. Lantana, “Pelatihan Local Area Network (LAN) di MTS Asyafi ’ Iyah 04 Jakarta,” vol. 05, no. 05, pp. 545–551, 2022.
- [12] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, “Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner,” *JBASE - J. Bus. Audit Inf. Syst.*, vol. 4, no. 1, pp. 13–21, 2021, doi: 10.30813/jbase.v4i1.2729.
- [13] M. A. Azis, N. Hadiano, 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.