



Prediksi Film Pilihan Penonton berdasarkan *Genre*, Aktor, dan Sutradara Berbasis Data Mining menggunakan Algoritma Eclat

(Viewer Movie Predictions based on Genres, Actors, and Directors based on Data Mining Using the Eclat Algorithm)

Deden Muhamad Furqon¹, Riki Ahmad Maulana², Ahmad Fauzi³, Nurul Dwi Cahya⁴, Muhammad Nur Sidiq⁵, Tri Kurnia Sandi⁶

¹Teknik Informatika, UIN Sunan Gunung Djati Bandung, furqoncreative24@gmail.com

²Teknik Informatika, UIN Sunan Gunung Djati Bandung, ahmadriki9512@gmail.com

³Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050009@student.uinsgd.ac.id

⁴Teknik Informatika, UIN Sunan Gunung Djati Bandung, nuruldwicahya925@gmail.com

⁵Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1167050104@student.uinsgd.ac.id

⁶Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1167050115@student.uinsgd.ac.id

Abstrak

Film adalah tempat mencari kesenangan dan hiburan setiap orang, industri perfilman menjadi bagian penting dari roda perekonomian di dunia ini. Rata-rata 79% orang di dunia ini menikmati menonton film untuk hiburan mereka. Oleh sebab itu industri film menjadi industri yang sangat besar akan tetapi sulit untuk diprediksi karena keinginan penonton yang sangat bermacam-macam. Oleh karena itu kami membuat sistem prediksi pilihan pengguna berdasarkan rekomendasi yang akan disajikan, kemungkinan besar akan dinikmati oleh penonton berdasarkan genre, aktor, dan sutradara kesukaan mereka. Yang akan berfungsi kepada para produsen untuk menganalisa pasar. Metode yang digunakan adalah data mining menggunakan algoritma Eclat, yang terdapat 5 proses di dalamnya. Hasilnya adalah kita mendapatkan dan menyortir 5043 data dengan 28 kolom dan menghilangkan sebagian dengan support threshold 0,003 dan mengevaluasi hasilnya. Setelah dilakukan evaluasi maka didapatkan data film hasil 1169 baris dengan nilai diatas 0,003 data pendukung untuk memprediksi rekomendasi pilihan pengguna.

Kata kunci: data mining, eclat, sistem rekomendasi

Abstract

Movies are a place for everyone to find pleasure and entertainment. The film industry is an essential part of the economy in this world. On average, 79% of people in the world enjoy watching movies for their entertainment. Therefore, the film industry has become a huge industry, but it is difficult to predict because the audience's desires are very diverse. Therefore, we created a prediction system for user choices based on the recommendations that will be presented, most likely to be enjoyed by audiences based on genre, actor, and their favorite director, which will function for producers to analyze the market. The method used is data mining using the Eclat algorithm, which has five processes in it. The result is that we get and sort 5043 data by 28 columns and omit some with a support threshold of 0.003 and evaluate the results after evaluating the film data obtained from 1169 lines with a value above 0.003 supporting data to predict user choice recommendations.

Keywords: data mining, eclat, recommendation system

1 Pendahuluan

Perkembangan teknologi informasi yang semakin pesat baik pada aspek ekonomi, industry bahkan social, termasuk dalam teknologi audio dan visual salah satunya dalam dunia industry film, dimana teknologi berperan sangat dominan dalam industry tersebut seperti kehadiran Grafika dan Visualisasi Komputer untuk membuat efek film lebih nyata sehingga lahir banyak karya film. Banyaknya produk film yang dihasilkan terkadang membuat bingung dalam menentukan film mana yang akan ditonton atau dalam mengikuti perkembangan industry film untuk mengetahui rating suatu film kita memerlukan data tersebut selain untuk rekomendasi inovasi karena rating nya. Namun berbicara konsumen itu sangat beragam, bahkan menyentuh hal yang sangat abstrak yaitu selera yang sukar untuk diukur berbeda dengan rating yang dapat diukur, hal ini bisa menjadi peluang di pasar industri film.

Genre merupakan representasi dari selera sehingga dapat menganalisa peluang pasar dalam industry film, dengan pengklasifikasian genre dapat memudahkan system dalam mencari rekomendasi berdasarkan tipe tertentu dan juga kita bisa mengetahui genre mana yang sangat banyak diminati dan dijadikan peluang, oleh karena faktor-faktor di atas, terutama pada genre yang selalu berkembang seiring dengan waktu sesuai dengan pola tertentu dan membaca peluang pasar maka perlu sebuah rekomendasi dalam menunjang hal tersebut.

Pada tahun 2016 dilakukan penelitian mengenai system rekomendasi film dengan menggunakan algoritma Simple Additive Weighting (SAW) oleh dimana focus objeknya adalah rekomendasi film berdasarkan genre serta preferensinya sehingga dapat dianalisa kebutuhan pasar berdasarkan selera konsumen [1]. Kemudian dilakukan juga penelitian oleh Syafina Dwi Arinda mengenai Implementasi data mining menggunakan algoritma Eclat dimana mencari kombinasi layanan yang sering dilakukan pada toko kendaraan bermotor sehingga memperoleh data berupa kombinasi layanan yang sering dilakukan kepada konsumen [2], [3]. Algoritma Eclat juga digunakan pada penelitian pengenalan pola peminjaman buku perpustakaan [4].

Algoritma Eclat (Equivalence Class Transformation) merupakan algoritma dengan metode depth first search yang kami gunakan untuk mendeteksi rekomendasi film apa yang kira-kira akan muncul dan ditonton untuk selanjutnya berdasarkan parameter, actor yang sama di dalam film/ sutradara yang sama/genre yang sama. Yang akan diurutkan berdasarkan nilai support. Masalah ini dikaji untuk bisa memprediksi rekomendasi yang tepat yang akan disediakan kepada penonton sehingga menghasilkan keuntungan untuk industry perfilman. Dengan memprediksi genre kita juga bisa memprediksi selera penonton dan menampilkan sajian yang tepat. Minimal nilai support yang digunakan adalah 0,003.

Berdasarkan uraian latar belakang masalah di atas maka perumusan masalah dalam penelitian ini adalah bagaimana menerapkan Data Mining menggunakan metode Association Rules dengan algoritma Eclat [5], untuk mengetahui pola rekomendasi film. Sehingga dapat diketahui informasi rekomendasi film yang akan muncul ketika telah menonton sebuah film.

2 Metodologi

Metodologi yang digunakan adalah uji coba langsung pada data movie yang memiliki 5043 baris dan 28 kolom yang berisi judul film, genre, nama-nama actor, sutradara, rating, jumlah like dll. Setelah menyiapkan data- data tersebut, barulah data tersebut dilakukan mining menggunakan algoritma Eclat. Tabel 1 menyajikan contoh data yang akan digunakan dalam eksperimen.

Tabel 1 Contoh Data Eksperimen

Director	Critic	Durati on	Director Facebook	Actor 2 Name	Gross	Genres	Actor 1 Name
James Cameron	723	178	0	Joel David Moore	760505847	Action Adventure Fantasy Sci- Fi	CCH Pounder
Gore Verbinski	302	169	563	Orlando Bloom	309404152	Action Adventure Fantasy	Johnny Depp
Sam Mendes	602	148	0	Rory Kinnear	200074175	Action Adventure Thriller	Christoph Waltz
Christopher Nolan	813	164	22000	Christia n Bale	448130642	Action Thriller	Tom Hardy

Untuk mencapai hasil yang diinginkan terdapat 5 tahapan yang harus dilalui pada metode ini, yaitu :

- Business Understanding, tujuan bisnis untuk memahami kebutuhan dan keperluan yang harus ditempuh dalam proses pengambilan hasil.
- Data Preprocessing, pada tahap ini yaitu proses menyiapkan dan menyusun data yang relevan untuk di mining agar mendapatkan hasil yang diinginkan.
- Modeling, untuk tahapan pemodelan ini akan digunakan teknik data mining dengan algoritma Eclat menggunakan Bahasa python dan juga tools google collabs sebagai media code.
- Evaluation, pada tahapan ini akan menguji dan membahas apakah hasil sudah sesuai yang diinginkan ataukah terjadi kecacatan dalam proses uji coba.
- Result, menentukan hasil yang sesuai dari proses sebelumnya

Algoritma Eclat

Algoritma Eclat sendiri menggunakan metode dept first pada kisi bagian dan menentukan dukungan set item dengan memotongkan daftar transaksi [6]–[8]. Pada saat ini algoritma Eclat hanya mendukung diffset dan beberapa varian algoritma lainnya. Algoritma ini melintasi pohon prefix di urutan pertama dan memperluas itemset barang hingga mencapai batas item set yang sering dan jarang terjadi, kemudian kembali ke awal untuk mengerjakan awalan berikutnya. Eclat menentukan support dari item set yang ditetapkan dengan membangun daftar pengenalan transaksi yang mengandung item set dengan cara memotong dua daftar pengenalan transaksi didalam dua set item yang berbeda hanya dengan satu item dan bersama-sama membentuk item yang ditetapkan saat diproses.

Cara yang digunakan untuk mempresentasikan transaksi untuk algoritma ini menggunakan matriks bit, dimana setiap baris sesuai dengan item masing-masing kolom untuk transaksi. Bit adalah set didalam matriks jika item yang sesuai dengan baris terkandung didalam transaksi yang sesuai dengan kolom, selain itu akan dihapus. Pada dasarnya terdapat dua cara untuk merepresentasikan matriks bit baik sebagai yang benar dengan satu bit memori untuk setiap item dan transaksi, atau menggunakan daftar untuk setiap baris di dalam kolom bit yang ditetapkan dimana cara kedua ini setara dengan menggunakan daftar pengenalan transaksi untuk setiap item dan cara merepresentasikan matrik bit ini terga tung pada kepadatan dataset.

Nilai support sebuah item diperoleh dengan rumus berikut:

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

Sedangkan untuk dua item diperoleh dengan rumus berikut:

$$Support(A \cup B) = \frac{Jumlah\ transaksi\ mengandung\ (A\ dan\ B)}{Total\ Transaksi} \quad (2)$$

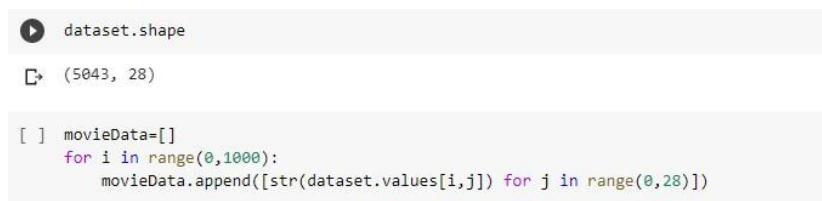
Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiatif yang memenuhi syarat minimum untuk confidence dengan menghitung confidence aturan asosiatif $A \rightarrow B$. Nilai confidence dari aturan $A \rightarrow B$ diperoleh dari rumus berikut:

$$Confidence = P(A|B) = \frac{\text{Jumlah transaksi mengandung (A dan B)}}{\text{Jumlah transaksi mengandung (A)}} \quad (3)$$

3 Hasil dan Pembahasan

3.1 Pengolahan Data Eksperimen

Pada tahap ini dilakukan proses perhitungan data untuk menentukan panjang data yang akan di mining sehingga dapat ditentukan perkiraan waktu nya dan terlihat lebih tersusun. Gambar 1 menunjukkan hasil ditemukan 5043 baris data dengan 28 kolom tersedia.



```
dataset.shape
(5043, 28)

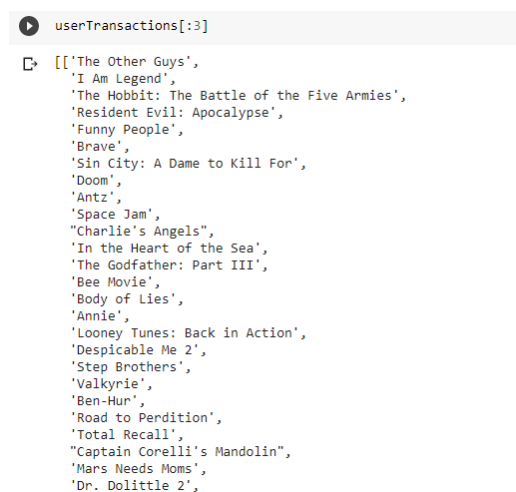
[ ] movieData=[]
for i in range(0,1000):
    movieData.append([str(dataset.values[i,j]) for j in range(0,28)])
```

Gambar 1 Ketersediaan Data

Berikutnya adalah proses membuat sample transaksi user/ simulasi jika user melakukan pilihan pilihan tertentu untuk mendapatkan sebuah hasil yang sesuai dengan sample data yang tersedia, berikut adalah hasil proses transaction sample dengan menggunakan kode program:

```
movieTransactions=[]
userTransactions=[]
for i in range(0,1000):
    firtsMovie = (random.choice(movieData))
    userTransactions.append([])
    movieTransactions.append([]) movieTransactions[i].append(getNextMovie(firtsMovie))
for j in range(0,30):
    movieTransactions[i].append(getNextMovie(movieTransactions[i][j]))
    userTransactions[i].append(movieTransactions[i][j][11].replace(u'\xa0', u''))
```

Contoh hasil yang diperoleh terdapat pada Gambar 2.



```
userTransactions[:3]
[['The Other Guys',
'I Am Legend',
'The Hobbit: The Battle of the Five Armies',
'Resident Evil: Apocalypse',
'Funny People',
'Brave',
'Sin City: A Dame to Kill For',
'Doom',
'Antz',
'Space Jam',
'Charlie's Angels',
'In the Heart of the Sea',
'The Godfather: Part III',
'Bee Movie',
'Body of Lies',
'Annie',
'Looney Tunes: Back in Action',
'Despicable Me 2',
'Step Brothers',
'Valkyrie',
'Ben-Hur',
'Road to Perdition',
'Total Recall',
'Captain Corelli's Mandolin',
'Mars Needs Moms',
'Dr. Dolittle 2',
']]
```

Gambar 2 Contoh hasil pengolahan data

3.2 Hasil Eksperimen

Gambar 3 menunjukkan hasil pengelompokan itemset data yang telah di olah menggunakan Eclat model, dengan menentukan relasi pada setiap film berdasarkan genre, sutradara yang sama, dan actor yang sama. Maka penentuan nilai minimum support adalah:

$\text{rules}=\text{apriori}(\text{userTransactions}, \text{min_support}=0.003)$ yang artinya $\text{min_support} = \text{film ditonton } 1x \text{ sehari} * 30 \text{ hari} / 1000 \text{ user} = 0.003$

```
results=list(rules)
results
[RelationRecord(items=frozenset({'13 Hours', 'Now You See Me 2'}), support=0.006, ordered_statistics=[OrderedStatistic(items_base=frozenset({'13 Hours'}), item_
RelationRecord(items=frozenset({'13 Hours', 'S.W.A.T.'}), support=0.005, ordered_statistics=[OrderedStatistic(items_base=frozenset({'S.W.A.T.'}), items_add=fro
RelationRecord(items=frozenset({'The SpongeBob Movie: Sponge Out of Water', '2 Guns'}), support=0.007, ordered_statistics=[OrderedStatistic(items_base=frozens
RelationRecord(items=frozenset({'50 First Dates', 'Catwoman'}), support=0.006, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Catwoman'}), items_
RelationRecord(items=frozenset({'50 First Dates', 'Salt'}), support=0.004, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Salt'}), items_add=froze
RelationRecord(items=frozenset({'A Civil Action', 'WALL-E'}), support=0.008, ordered_statistics=[OrderedStatistic(items_base=frozenset({'WALL-E'}), items_add=froze
RelationRecord(items=frozenset({'Mission: Impossible', 'A.I. Artificial Intelligence'}), support=0.004, ordered_statistics=[OrderedStatistic(items_base=frozens
RelationRecord(items=frozenset({'Absolute Power', 'White House Down'}), support=0.006, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Absolute Pou
RelationRecord(items=frozenset({'After Earth', 'Dinosaur'}), support=0.005, ordered_statistics=[OrderedStatistic(items_base=frozenset({'After Earth'}), items_
RelationRecord(items=frozenset({'After Earth', 'Total Recall'}), support=0.006, ordered_statistics=[OrderedStatistic(items_base=frozenset({'After Earth'}), ite
RelationRecord(items=frozenset({'Air Force One', 'We Are Marshall'}), support=0.004, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Air Force One
RelationRecord(items=frozenset({'The Thin Red Line', 'Alice Through the Looking Glass'}), support=0.004, ordered_statistics=[OrderedStatistic(items_base=froze
RelationRecord(items=frozenset({'Allegiant', 'Men in Black 3'}), support=0.006, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Allegiant'}), item_
RelationRecord(items=frozenset({'The Chronicles of Riddick', 'Almost Famous'}), support=0.006, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Almo
RelationRecord(items=frozenset({'The Patriot', 'Alvin and the Chipmunks'}), support=0.004, ordered_statistics=[OrderedStatistic(items_base=frozenset({'The Patri
RelationRecord(items=frozenset({'Perfume: The Story of a Murderer', 'Anger Management'}), support=0.006, ordered_statistics=[OrderedStatistic(items_base=froze
RelationRecord(items=frozenset({'Armageddon', 'Dante's Peak'}), support=0.006, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Dante's Peak'}), ite
RelationRecord(items=frozenset({'Armageddon', 'Red Planet'}), support=0.006, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Red Planet'}), items_
RelationRecord(items=frozenset({'Armageddon', 'Rise of the Planet of the Apes'}), support=0.005, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Ri
RelationRecord(items=frozenset({'Around the World in 80 Days', 'Resident Evil: Retribution'}), support=0.005, ordered_statistics=[OrderedStatistic(items_base=froze
RelationRecord(items=frozenset({'Asterix at the Olympic Games', 'Tangled'}), support=0.005, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Asteri
```

Gambar 3 Hasil pengelompokan itemset data

	Film yang dipilih	Rekomendasi	Support
26	Avatar	Star Wars: Episode III - Revenge of the Sith	0.047
23	Man of Steel	Avatar	0.028
25	Star Wars: Episode II - Attack of the Clones	Avatar	0.028
22	End of Days	Avatar	0.025
28	The League of Extraordinary Gentlemen	Avatar	0.023
...
552	Hotel Transylvania 2	Star Wars: Episode II - Attack of the Clones	0.003
553	Hotel Transylvania 2	The Matrix Reloaded	0.003
554	Hotel Transylvania 2	The Perfect Storm	0.003
555	How the Grinch Stole Christmas	Man of Steel	0.003
1168	The Twilight Saga: Breaking Dawn - Part 2	You Don't Mess with the Zohan	0.003

1169 rows x 3 columns

Gambar 3 Hasil Rekomendasi Film

Setelah melakukan training model data pada algoritma Eclat, selanjutnya adalah proses untuk melakukan evaluasi dan result untuk melihat hasil yang terjadi. Code yang digunakan adalah :

```
movie,r1,support=[],[],[]
for result in results:
    movie.append(tuple(result[2][0][0])[0])
    r1.append(tuple(result[2][0][1])[0])
    support.append(result[1])
columns=list(zip(movie,r1,support))
result_df=pd.DataFrame(columns)
result_df.columns=['Film yang dipilih','Rekomendasi','Support']
result_df
```

Setelah baris code tersebut akan menghasilkan keluaran seperti pada Gambar 4 yang merupakan rekomendasi film yang dihasilkan dengan minimum support yang terpenuhi.

Didapatkan total data 1169 baris yang telah memenuhi nilai support minimal yaitu 0,003. Berdasarkan hasil tersebut sisa data lainnya tereliminasi dikarenakan tidak memenuhi nilai support. Jika melihat data diatas maka dapat dilihat bahwa hasil tertinggi adalah jika user memilih film Avatar maka rekomendasi selanjutnya yang akan muncul agar film ditonton adalah film “ Star Wars : Episode III – Revenge Of The Sith “ dengan nilai support tertinggi yaitu 0,047.

4 Simpulan

Berdasarkan hasil analisis yang telah dilakukan pada data movie yang memiliki 5043 baris dan 28 kolom yang berisi judul film, genre, nama-nama actor, sutradara, rating, jumlah like dll. Pada data mining dengan menggunakan algoritma Eclat maka dapat diambil kesimpulan didapatkan total data 1169 baris yang telah memenuhi nilai support minimal yaitu 0,003. Berdasarkan hasil tersebut sisa data lainnya tereliminasi dikarenakan tidak memenuhi nilai support. hasil tertinggi jika user memilih film Avatar maka rekomendasi selanjutnya yang akan muncul agar film ditonton adalah film “ Star Wars : Episode III – Revenge Of The Sith “ dengan nilai support tertinggi yaitu 0,047. Dengan terbentuknya pola rekomendasi ini berfungsi untuk mengetahui hubungan antar film dengan tujuan meningkatkan minat penonton dalam menonton film sesuai perilaku si penonton sendiri.

Referensi

- [1] E. W. Sumarlin, S. Hansun, and Y. W. Wiratama, “RANCANG BANGUN APLIKASI REKOMENDASI FILM DENGAN MENGGUNAKAN ALGORITMA SIMPLE ADDITIVE WEIGHTING,” *J. Inform.*, vol. 10, no. 2, 2016, doi: 10.26555/jifo.v10i2.a5066.
- [2] S. D. Arinda and S. Sulastrri, “IMPLEMENTASI DATA MINING MENGGUNAKAN ALGORITMA ECLAT,” in *Prosiding SINTAK 2017*, 2017, [Online]. Available: <https://www.unisbank.ac.id/ojs/index.php/sintak/article/view/5549/1674>.
- [3] S. Sulastrri, E. Zuliarso, and Y. Anis, “IMPLEMENTASI ALGORITMA APRIORI DAN ALGORITMA ECLAT PADA AHASS AKMAL JAYA PURWODADI,” *Dinamik*, vol. 22, no. 1, pp. 50–56, 2017, doi: 10.35315/dinamik.v22i1.7105.
- [4] M. Subianto, F. AR, and M. Hijriyana P., “Pola peminjaman buku di perpustakaan Universitas Syiah Kuala menggunakan Algoritma Eclat,” *Berk. Ilmu Perpust. dan Inf.*, vol. 14, no. 1, p. 35, 2018, doi: 10.22146/bip.32089.
- [5] 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.
- [6] M. J. Zaki, “Scalable algorithms for association mining,” *IEEE Trans. Knowl. Data Eng.*, vol. 12, no. 3, pp. 372–390, 2000, doi: 10.1109/69.846291.
- [7] D. S. Maylawati, “The Concept of Frequent Itemset Mining for Text,” in *AASEC 2018*, 2018.
- [8] U. Grag and M. Kaur, “ECLAT Algorithm for Frequent Itemsets Generation,” *Int. J. Comput. Syst.*, vol. 01, no. 03, pp. 82–84, 2014, [Online]. Available: <http://www.ijcsonline.com/>.