



## **Analisis Rekomendasi Film dengan Algoritma *Sequential Pattern Discovery Using Equivalence Classes (SPADE)***

*(Film Recommendation Analysis with Sequential Pattern Discovery Algorithm Using Equivalence classes (SPADE))*

Aisyah Amini Nur<sup>1</sup>, Akbar Hidayatullah Harahap<sup>2</sup>, Ihsan Muttaqin Bin Abdul Malik<sup>3</sup>, Muhammad Irfan Nur Imam<sup>4</sup>, Muhammad Thariq Sabiq Bilhaq<sup>5</sup>, Angelyna Angelyna<sup>6</sup>

<sup>1</sup>Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050010@student.uinsgd.ac.id

<sup>2</sup>Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050012@student.uinsgd.ac.id

<sup>3</sup>Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050053@student.uinsgd.ac.id

<sup>4</sup>Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050071@student.uinsgd.ac.id

<sup>5</sup>Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050077@student.uinsgd.ac.id

<sup>6</sup>Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050016@student.uinsgd.ac.id

### **Abstrak**

Saat ini internet merupakan suatu hal yang paling utama di kehidupan manusia. Sehingga adanya perubahan dalam kebutuhan manusia, salah satunya menikmati hiburan. Dengan adanya internet menikmati hiburan bisa dilakukan dimana saja dan kapan saja, salah satu hiburannya yaitu film. Dengan adanya internet banyak penyedia jasa menyediakan situs menonton film secara online. Sehingga masyarakat menjadi tertarik dan banyak menggunakannya untuk menonton film secara online. Dengan meningkatnya penggemar film maka diharuskan adanya sistem rekomendasi film, yaitu membuat pola rekomendasi film berdasarkan tontonan film sebelumnya. Oleh karena itu, studi ini bertujuan untuk memberikan rekomendasi tontonan film menggunakan metode pengenalan pola yang sering muncul. Metode yang digunakan yaitu *sequence pattern mining* menggunakan SPADE untuk menemukan pola dari sekelompok data. Keunggulan dari Spade sendiri adalah kecepatannya dalam menemukan frekuensi *sequence* yang dapat dijadikan sebagai data rekomendasi film dari tontonan sebelumnya.

**Kata kunci:** data mining, sistem rekomendasi film, spade, sequence pattern mining

### **Abstract**

*Currently, the internet is the most essential thing in human life. So that there is a change in human needs, one of which is enjoying entertainment. With the internet, enjoying entertainment can be done anywhere and anytime, one of the entertainment is movies. With the internet, many service providers provide movie watching sites online. So that people become interested and use it a lot to watch movies online. With the increase in film fans, a film recommendation system is required to make a film recommendation pattern based on previous movie viewing. Therefore, this study aims to provide recommendations for viewing films using the pattern recognition method that often occurs. The technique used is sequence pattern mining using SPADE to find patterns from a group of data. SPADE's advantage is its speed in finding sequence frequencies that can be used as movie recommendation data from previous shows.*

**Keywords:** data mining, recommendation system, spade, sequence pattern mining

## 1 Pendahuluan

Internet merupakan sesuatu yang sangat penting di era teknologi saat ini. Internet telah menjadi sebuah kebutuhan pokok di kehidupan manusia, ditandai dengan pengguna internet yang semakin meningkat [1], [2]. Salah satu aktivitas yang menggunakan internet adalah menikmati hiburan seperti menonton film secara online. Banyak sekali media yang menyediakan layanan streaming film secara online. Fitur ini diminati banyak pengguna karena bisa dilakukan kapan saja dan dimana saja. Dengan banyaknya peminat film, maka dibutuhkan sebuah sistem rekomendasi film yang dapat menyesuaikan dengan film sebelumnya. Oleh sebab itu rekomendasi film ini menjadi sebuah masalah yang dapat diselesaikan dengan menggunakan komputer. *Sequence pattern mining* merupakan salah satu metode data mining untuk mendapatkan pola dari sekelompok data [3]. Dengan *sequence pattern mining*, data rekomendasi film ini dapat diolah menjadi *data sequence* dan dapat diproses lebih lanjut lagi.

Studi ini menggunakan algoritma SPADE (Sequence Pattern Discovery using Equivalent class) yang merupakan salah satu dari algoritma pattern mining yang mempunyai keunggulan yaitu dapat menemukan pola frekuensi dengan cepat. Terdapat beberapa penelitian terdahulu yang terkait penelitian ini dalam sebagai sistem rekomendasi menggunakan pattern mining, antara lain: (1) rekomendasi pembelian dan manajemen inventaris produk menggunakan pemfilteran berbasis konten dengan pendekatan penambangan pola sekuensial dengan algoritma SPADE [4]; (2) penelitian yang menganalisis pola konsumsi yang dimiliki pelanggan dan memberikan informasi yang dapat digunakan dalam menentukan tata letak rak toko baru dengan algoritma SPADE [5]; dan (3) pencarian pola perilaku pengguna internet dengan algoritma SPADE [6].

## 2 Metodologi

Sequence pattern mining menggunakan spade untuk menganalisis perilaku dalam memilih rekomendasi film. Data dikumpulkan terlebih kemudian dilakukannya preprocessing pada tahapan ini data di filter, selanjutnya data diproses untuk memperoleh frequent pattern. Data diubah ke dalam bentuk yang sesuai merupakan tahap transformasi data, yaitu sebagai masukan bagi algoritma SPADE. Tahap data mining adalah inti dari analisis data, proses menemukan wawasan, pola menarik, dipahami, dan prediksi model dari data berskala besar.

### 2.1 Sistem Rekomendasi

Sistem rekomendasi merupakan program yang melakukan prediksi sesuatu item [7], seperti rekomendasi baju, elektronik, film, lagu dan lain sebagainya yang menarik pengguna. Sistem ini sumbernya didapat dengan mengumpulkan data dari pengguna baik secara langsung maupun tidak. Hasil dari pengumpulan data tersebut, kemudian dilakukan implementasi algoritma didalamnya dan hasilnya dapat dijadikan sebuah parameter rekomendasi untuk penggunaannya. Sistem rekomendasi juga merupakan salah satu alternatif sebagai mesin pencari suatu item yang dicari oleh pengguna.

### 2.2 Sequential Pattern Mining

Sequential pattern mining digunakan untuk mencari data yang memiliki urutan, data tersebut bisa merupakan urutan transaksi. Sequential pattern mining pertama kali diperkenalkan oleh Agrawal dan Srikant dengan Apriori-nya [8]. Proses sequential pattern mining dapat digambarkan sebagai berikut, diberikan sejumlah sequence, setiap sequence terdiri atas sederetan elemen, dan setiap elemen terdiri atas sejumlah item, serta diberikan nilai minimum- support.

Sequential Pattern Mining digunakan dalam mengekstrak pattern sequential dengan pola analisis sekuensial, klasifikasi & prediksi, dan sistem rekomendasi. Langkah untuk menemukan satu set

urutan data yang lengkap. Tantangan dari *sequential pattern mining* ini ketika sejumlah besar pola berurutan yang mungkin disembunyikan di database, algoritma data mining harus bisa menemukan rangkaian pola yang lengkap dan memenuhi batas minimum.

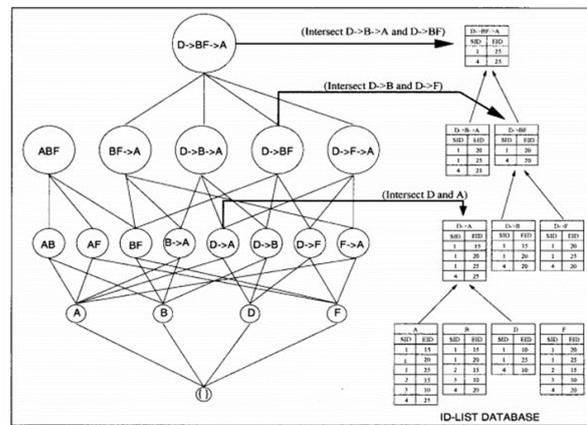
### 2.3 SPADE (Sequential Pattern Discovery Using Equivalence classes)

Sequential Pattern Discovery Using Equivalence Classes (SPADE) merupakan salah satu algoritma dari Sequential Pattern Mining [9]. Algoritma ini digunakan untuk menemukan pola agar mendapatkan informasi yang berguna. Algoritma ini memiliki keunggulan dalam kecepatan menemukan data dibandingkan algoritma-algoritma pencari frequent sequences sebelumnya seperti algoritma Apriori dan GSP (Generalized Sequential Pattern).

SPADE adalah algoritma yang digunakan untuk penemuan pola sekuensial dengan cepat. Sebagian besar algoritma penambangan pola sekuensial mengasumsikan tata letak basis data horizontal. Spade menggunakan format basis data vertikal, tempat kami mempertahankan daftar id berbasis disk untuk setiap item, ditampilkan pada Gambar 1.

A		B		D		F	
SID	EID	SID	EID	SID	EID	SID	EID
1	15	1	15	1	10	1	20
1	20	1	20	1	25	1	25
1	25	2	15	4	10	2	15
2	15	3	10			3	10
3	10	4	20			4	20
4	25						

Gambar 1 Contoh Id List



Gambar 2 Menghitung Support

Gambar 2 menunjukkan contoh bagaimana kita membuat ID-List dari suatu urutan id-list ( $D \rightarrow B$ ) dan ( $D \rightarrow F$ ). Untuk ( $D \rightarrow B \rightarrow A$ ) kita harus melakukan gabungan sementara pada  $L(D \rightarrow B)$  dan  $L(D \rightarrow A)$ . Akhirnya, urutan-2 diperoleh dengan menggabungkan atom secara langsung. Pada gambar menunjukkan proses lengkap, dimulai dengan database vertikal awal dari id-list untuk setiap atom. Seperti yang bisa kita lihat, di setiap titik hanya pasangan (sid, id) yang disimpan dalam list-id (yaitu, hanya eid untuk item terakhir dari suatu urutan yang disimpan).

## 3 Hasil dan Pembahasan

### 3.1 Hasil Eksperimen

Data yang digunakan untuk eksperimen yaitu data yang didapat dari movielens. Eksperimen dilakukan dengan menggunakan Bahasa pemrograman Python. Algoritma SPADE memiliki:

- *Item*, yang dalam kasus ini yaitu ID film atau movie\_id, sehingga perlu ditambahkan kolom baru
- Sequences ID (SID), yaitu nomor unik dari item set film yang ditonton oleh user dalam waktu yang ditentukan
- Event ID (EID), yaitu event atau waktu saat user menonton film SID dan EID akan dibuat pada saat mengasumsikan pola tontonan User

```
dataFrame = pd.DataFrame(dataset[['movie_title', 'director_name', 'genres', 'actor_1_name', 'movie_id']])
dataFrame
```

	movie_title	director_name	genres	actor_1_name	movie_id
0	Avatar	James Cameron	Action Adventure Fantasy Sci-Fi	CCH Pounder	10001
1	Pirates of the Caribbean: At World's End	Gore Verbinski	Action Adventure Fantasy	Johnny Depp	10002
2	Spectre	Sam Mendes	Action Adventure Thriller	Christoph Waltz	10003
3	The Dark Knight Rises	Christopher Nolan	Action Thriller	Tom Hardy	10004
4	Star Wars: Episode VII - The Force Awakens	Doug Walker	Documentary	Doug Walker	10005
...	...	...	...	...	...
5038	Signed Sealed Delivered	Scott Smith	Comedy Drama	Eric Mabius	15039
5039	The Following	0	Crime Drama Mystery Thriller	Natalie Zea	15040
5040	A Plague So Pleasant	Benjamin Roberds	Drama Horror Thriller	Eva Boehnke	15041
5041	Shanghai Calling	Daniel Hsia	Comedy Drama Romance	Alan Ruck	15042
5042	My Date with Drew	Jon Gunn	Documentary	John August	15043

5043 rows × 5 columns

Gambar 3 Implementasi olah data

### 3.2 Pre-Processing Data

Pada pre-processing terdapat 2 tahap, yang pertama filtering dan kedua formatting. Membuat sampel transaksi kebiasaan user menontoh film dalam 30 hari, serta mengasumsikan bahwa film selanjutnya yang ditonton oleh user adalah:

- Memiliki aktor yang sama dari film sebelumnya; atau
- Memiliki sutradara yang sama dari film sebelumnya; atau
- Memiliki genre yang sama dari film sebelumnya.

Pada tahap ini, SID dan EID akan dicantumkan, dimana Urutan data untuk algoritma SPADE [SID, EID, [sequence]].

```
[ ] SID = 1
    EID = 0
    item = 0
    lastDay = 30
    lastCellIndex = 100 * lastDay
    userTransactions=[]

    for i in range(0,lastCellIndex+1):
        EID = EID + 1
        userTransactions.append([])
        userTransactions[i].append(SID)
        userTransactions[i].append(EID)
        userTransactions[i].append([int(movieTransactions[SID][EID-1][4]))

    if(EID == lastDay):
        SID = SID + 1
        EID = 0
        item = 0
```

userTransactions		
[1, 1, [10041]]		
[1, 2, [10521]]		
[1, 3, [10001]]		
[1, 4, [10296]]		
[1, 5, [10558]]		
[1, 6, [10248]]		
[1, 7, [10651]]		
[1, 8, [10949]]		
[1, 9, [10800]]		
[1, 10, [10213]]		
[1, 11, [10834]]		
[1, 12, [10142]]		
[1, 13, [10486]]		
[1, 14, [10737]]		
[1, 15, [10436]]		
[1, 16, [10261]]		
[1, 17, [10137]]		
[1, 18, [10682]]		
[1, 19, [10348]]		
[1, 20, [10786]]		
[1, 21, [10752]]		
[1, 22, [10651]]		
[1, 23, [10161]]		
[1, 24, [10650]]		
[1, 25, [10993]]		
[1, 26, [10120]]		
[1, 27, [10399]]		
[1, 28, [10100]]		

Gambar 4 Pre-processing

### 3.3 Implementasi Algoritma

Setelah pre-processing selesai, kemudian masukkan algoritma didalamnya dan mulai untuk pengoperasiannya. Gambar 5 menunjukkan prosen implementasi algoritma SPADE dan Gambar 6 menunjukkan pola yang dihasilkan dari algoritma SPADE.

### ▼ Implementasi Algoritma SPADE

```

from pycspade.helpers import spade

result = spade(data=userTransactions, support=0.02, maxlen=2)

[ ] def print_spade(result):
    nseqs = result['nsequences']
    print('{{0:>9s} {1:>9s} {2:>9s} {3:>9s} {4:>9s} {5:>80s}'.format('Occurs', 'Accum', 'Support', 'Confid', 'Lift',
                                                              'Sequence'))
    for mined_object in result['mined_objects']:
        conf = 'N/A'
        lift = 'N/A'
        name = ''
        if mined_object.confidence:
            conf = '{:0.7f}'.format(mined_object.confidence)
        if mined_object.lift:
            lift = '{:0.7f}'.format(mined_object.lift)
        if (conf != 'N/A'):
            print('{{0:>9d} {1:>9d} {2:>0.7f} {3:>9s} {4:>9s} {5:>80s}'.format(
                mined_object.noccurs,
                mined_object.accum_occurs,
                mined_object.noccurs / nseqs,
                conf,
                lift,
                '->'.join(list(map(str, mined_object.items)))))

```

Gambar 6 Implementasi Algoritma SPADE

Occurs	Accum	Support	Confid	Lift	Sequence
4	0.0396040	0.1538462	3.1076923		(10001)->(10010)
3	0.0297030	0.1153846	2.3307692		(10001)->(10054)
3	0.0297030	0.1153846	2.9134615		(10001)->(10136)
4	0.0396040	0.1538462	2.2197802		(10001)->(10248)
3	0.0297030	0.1153846	1.9423077		(10001)->(10253)
3	0.0297030	0.1153846	2.9134615		(10001)->(10261)
3	0.0297030	0.1153846	3.8846154		(10001)->(10445)
3	0.0297030	0.1153846	3.8846154		(10001)->(10657)
3	0.0297030	0.1153846	1.9423077		(10001)->(10728)
3	0.0297030	0.1153846	1.9423077		(10001)->(10756)
4	0.0396040	0.1538462	2.2197802		(10001)->(10842)
3	0.0297030	0.1153846	2.3307692		(10001)->(10918)
3	0.0297030	1.0000000	3.8846154		(10016)->(10001)
3	0.0297030	0.7500000	12.6250000		(10077)->(10745)
3	0.0297030	0.5000000	7.2142857		(10084)->(10584)
3	0.0297030	0.6000000	7.5750000		(10189)->(10031)
4	0.0396040	1.0000000	3.8846154		(10241)->(10001)
3	0.0297030	0.3333333	6.7333333		(10299)->(10055)
3	0.0297030	1.0000000	3.8846154		(10307)->(10001)
3	0.0297030	0.3000000	1.1653846		(10309)->(10001)
3	0.0297030	0.3000000	3.3666667		(10309)->(10063)
3	0.0297030	0.3000000	6.0600000		(10309)->(10587)
3	0.0297030	0.3750000	9.4687500		(10473)->(10326)
3	0.0297030	1.0000000	3.8846154		(10521)->(10001)
3	0.0297030	1.0000000	33.6666667		(10547)->(10725)
3	0.0297030	0.4285714	10.8214286		(10558)->(10183)
3	0.0297030	0.4285714	3.6071429		(10561)->(10591)
3	0.0297030	0.7500000	2.9134615		(10578)->(10001)
3	0.0297030	1.0000000	3.8846154		(10607)->(10001)
3	0.0297030	0.6000000	2.3307692		(10639)->(10001)
3	0.0297030	0.7500000	12.6250000		(10693)->(10210)
3	0.0297030	0.4285714	5.4107143		(10710)->(10501)
3	0.0297030	0.7500000	2.9134615		(10735)->(10001)
3	0.0297030	0.5000000	1.9423077		(10756)->(10001)
3	0.0297030	0.7500000	9.4687500		(10863)->(10684)
3	0.0297030	0.7500000	18.9375000		(10931)->(10183)

Gambar 7 Hasil Implementasi Algoritma SPADE.

## 4 Simpulan

Berdasarkan hasil pengujian algoritma, dapat diambil beberapa kesimpulan diantaranya: (1) data yang digunakan untuk dilakukan pattern mining merupakan data yang valid dan merupakan transaksi yang dilakukan oleh pengguna; (2) berdasarkan analisis yang dilakukan penggunaan algoritma SPADE dalam menemukan hasil sangat cepat; (3) penggunaan algoritma SPADE terhadap data rekomendasi film berhasil dilakukan; dan (4) rekomendasi film ini sangat membantu dalam pencarian film selanjutnya. Saran untuk penelitian selanjutnya agar lebih sempurna untuk penggunaan data cari yang sedikit nilai kosongnya, kemudian variasi film nya bisa dikurangi dan menggunakan data film yang sedang tren dimasa nya.

## Referensi

- [1] Anonymous, “Ketika Pengguna Internet dan Smartphone Terus Meningkat, Android Dominasi Pasar Indonesia,” *TribunJabar.id*, 2019. <http://jabar.tribunnews.com/2019/01/24/ketika-pengguna-internet-dan-smartphone-terus-meningkat-android-dominasi-pasar-indonesia-dan-dunia> (accessed Jan. 26, 2019).
- [2] Y. Pratomo, “APJII: Jumlah Pengguna Internet di Indonesia Tembus 171 Juta Jiwa,” *Kompas.com*, 2019. <https://tekno.kompas.com/read/2019/05/16/03260037/apjii-jumlah-pengguna-internet-di-indonesia-tembus-171-juta-jiwa> (accessed Jul. 14, 2019).
- [3] 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.
- [4] A. C. Raharja, I. S. Sitanggang, and A. Buono, “Purchase Recommendation and Product Inventory Management using Content Based Filtering with Sequential Pattern Mining Approach,” *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, pp. 287–294, 2018, doi: 10.22219/kinetik.v3i4.663.
- [5] D. A. Saputra, E. T. Tosida, and F. D. W, “PENENTUAN POLA SEKUENSIAL DATA TRANSAKSI PENJUALAN MENGGUNAKAN ALGORITMA SEQUENTIAL PATTREN DISCOVERY USING EQUIVALENT CLASSES (SPADE),” *Komputasi J. Ilm. Ilmu Komput. dan Mat.*, vol. 16, no. 2, pp. 271–282, 2019, doi: 10.33751/komputasi.v16i2.1621.
- [6] D. Yunianto, C. Dewi, S. Kom, and N. Yudistira, “Sequential Pattern Mining Pada Pencarian Pola Perilaku Penggunaan Internet Menggunakan Algoritma SPADE.” Universitas Brawijaya, 2014.
- [7] F. Ricci, B. Shapira, and L. Rokach, “Recommender systems: Introduction and challenges,” in *Recommender Systems Handbook, Second Edition*, 2015, pp. 1–34.
- [8] 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.
- [9] M. J. Zaki, “SPADE: An efficient algorithm for mining frequent sequences,” *Mach. Learn.*, vol. 42, no. 1–2, pp. 31–60, 2001, doi: 10.1023/A:1007652502315.