



## Penentuan Rekomendasi Film dengan Metode *Asosiasi Generalized Sequence Pattern (GSP)*

(*Determination of Film Recommendations using the Generalized Sequence Pattern (GSP) Association Method*)

Acep Razif Andriyan<sup>1</sup>, Dinda Meysya Rochma<sup>2</sup>, Melani Nur Mudyawati<sup>3</sup>, Miftahul Jannah<sup>4</sup>, Siti Lufia Dwi Agustini<sup>5</sup>, Arham Aulia Nugraha<sup>6</sup>

<sup>1</sup>Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050002@student.uinsgd.ac.id

<sup>2</sup>Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050032@student.uinsgd.ac.id

<sup>3</sup>Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050060@student.uinsgd.ac.id

<sup>4</sup>Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050061@student.uinsgd.ac.id

<sup>5</sup>Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050007@student.uinsgd.ac.id

<sup>6</sup>Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050017@student.uinsgd.ac.id

### Abstrak

Seiring perkembangan zaman film tidak hanya dijadikan sebagai media komunikasi saja, melainkan juga sebagai media hiburan yang dapat ditonton oleh penggemar dimana saja. Tingginya tingkat peminat film membuat pola tontonan yang dapat dijadikan data rekomendasi. Selain itu sistem tersebut dapat memberikan keuntungan pada perusahaan terkait. Pada penelitian ini, kami menggunakan Algoritma Generalized Sequential Pattern (GSP) dengan metode Association Mining yang fungsi utamanya menemukan urutan yang atau sequensial dari kumpulan atribut yang sering muncul bersamaan sehingga menemukan hasilnya untuk mengetahui informasi data rekomendasi setelah film sebelumnya habis.

**Kata kunci:** Data Mining, Generalized Sequential Pattern, Association Mining, Rekomendasi Film

### Abstract

*Along with the development of the times, films are used as a medium of communication and as a medium of entertainment that can be watched by fans everywhere. The high level of film enthusiasts makes viewing patterns that can be used as recommendation data. Besides, this system can provide benefits to the company concerned. This study used the Generalized Sequential Pattern (GSP) Algorithm with the Association Mining method. The primary function is to find the sequence or sequential of a set of attributes that often appear together to see the results to find out the recommendation data information after the previous film runs out.*

**Keywords:** association mining, data mining, Generalized Sequential Pattern, recommendation system

## 1 Pendahuluan

Film merupakan media penyampaian pesan melalui bentuk video yang sudah melalui berbagai proses editing agar dapat disajikan dengan kesan yang estetik dan menjiwai peminatnya. Pesan dalam film merupakan komunikasi massa dapat berbentuk apa saja tergantung misi film tersebut, dapat berupa pesan pendidikan, informasi, kekeluargaan, dan banyak lainnya. Saat ini film banyak digandrungi oleh berbagai kalangan karena film dianggap dapat mempengaruhi pikiran penonton dengan cerita-

cerita yang disediakan mengenai fakta kehidupan, sehingga lebih menarik perhatian penonton. Setiap film yang diproduksi pasti memiliki ciri khas masing - masing dari setiap genre, sehingga tidak heran industri film masa kini semakin menarik perhatian penonton. Banyaknya film yang ditonton oleh penggemar akan menghasilkan suatu informasi menarik yang dapat dijadikan data, dimana data tersebut dapat menguntungkan bagi suatu perusahaan dalam membuat film selanjutnya. Data tersebut berasal dari tayangan film yang ditonton penggemar sehingga menjadi pola tontonan yang dapat dijadikan rekomendasi film. Masalahnya adalah menemukan semua pola berurutan dengan dukungan minimum yang ditentukan pengguna, dengan dukungan pola sekuensial adalah persentase urutan data yang mengandung pola tersebut. Berkaitan dengan rekomendasi film maka dalam penelitian ini menggunakan metode data mining

Studi ini menggunakan Teknik *Data Mining* dengan pendekatan pengenalan pola menggunakan algoritma *Generalized Sequential Pattern* (GSP). Terdapat beberapa penelitian terdahulu yang terkait penelitian ini, antara lain: (1) Pengembangan Aplikasi Market Basket Analysis Menggunakan Algoritma Generalized Sequential Pattern pada Supermarket [1]; (2) penelitian yang membahas tentang rekomendasi produk dengan menggunakan metode Generalized Sequence Pattern (GSP) untuk mengetahui produk yang paling diminati konsumen (produk laku) dan metode Apriori agar dapat memberikan rekomendasi produk yang benar kepada konsumen [2]; dan (3) aplikasi data mining menggunakan algoritma *generalized sequential pattern* untuk menentukan aturan asosiasi pada transaksi peminjaman buku [3].

## 2 Metodologi

### 2.1 Data Mining

Data Mining merupakan proses untuk mendapatkan informasi yang berguna dari basis data yang besar [4]. Salah satu teknik yang dibuat dalam data mining adalah bagaimana menelusuri data yang ada untuk membangun sebuah model, kemudian menggunakan model tersebut agar dapat mengenali pola data yang lain yang tidak berada dalam basis data yang tersimpan. Data mining merupakan bagian dari proses Knowledge Discovery in Database (KDD) [5]. Dalam penelitian kali ini, metode Data Mining yang digunakan adalah metode Asosiasi. Metode Asosiasi adalah suatu metode yang digunakan untuk menemukan suatu atribut yang muncul dalam satu waktu [6]. Metode Asosiasi yang digunakan dalam penelitian ini adalah Algoritma Generalized Sequential Pattern.

### 2.2 Generalized Sequential Pattern (GSP)

Algoritma Generalized Sequential Pattern untuk mencari informasi apakah algoritma ini dapat memberikan informasi berupa rekomendasi film dari film - film yang telah ditonton secara acak, dan merekomendasikan kembali kepada penonton yang lain pola yang telah tersimpan berupa film selanjutnya. Generalized Sequential Pattern (GSP) memiliki nama lain Apriori All. GSP adalah algoritma yang dapat memproses dan menemukan semua pola sekuensial dan non sekuensial [7]. Berikut GSP dalam bentuk Pseudocode berikut:

```
L1 = {large 1-sequences};  
for (k = 2; Lk ; k++) do  
begin  
Ck = New candidates generated from Lk-1  
foreach customer-sequences in the database do  
Increment the count of all candidates in Ck  
that are contained in c.
```

```
Lk = Candidates in Ck with minimum support.  
end  
insert into Ck  
select p.litemset1, ..., p.litemsetk-1, q.litemsetk-1  
from Lk-1 p, Lk-1 q  
where p.litemset1 = q.litemset1, ...,  
p.litemsetk-2 = q.litemsetk-2;
```

Algoritma GSP ini digunakan pada mining sequence dan sangat baik untuk memecahkan masalah mining sequence yang banyak didasarkan pada sebuah algoritma Apriori. Fungsi utama dari algoritma GSP yaitu menemukan pola sekuensial atau urutan [2]. Sedangkan metode Apriori digunakan untuk menemukan jumlah itemset frequent. Algoritma ini merupakan pelopor algoritma analisis asosiasi dalam mencari juga menemukan kandidat itemset frequent dan pembangkit aturan asosiasi yang bisa dibentuk.

Penerapan Algoritma GSP :

1. Hitung jumlah item, lalu tentukan minimum support
2. Hapus item yang memiliki jumlah dibawah minimum support
3. Kombinasikan item - item kedalam 2 bagian, temukan jumlah gabungan item tertinggi
4. Hapus item gabungan tersebut yang jumlahnya dibawah minimum support

### 3 Hasil dan Pembahasan

#### 3.1 Hasil Eksperimen

Data yang kami pakai berisi kumpulan Judul film, sutradara, genre, dan lain-lain. Data ini merupakan data mentah yang harus melalui tahap preprocessing data terlebih dahulu. Data mentah perlu dibersihkan dan dikodekan menjadi nilai numerik sebelum diberikan ke model pembelajaran mesin, proses pembersihan dan pengkodean ini disebut sebagai teks preprocessing. Oleh karena itu, diperlukan proses perubahan bentuk menjadi data yang terstruktur untuk kebutuhan lebih lanjut (sentiment analysis, topic modelling, dan sebagainya).

Pada penelitian kali ini disini kita menggunakan salah satu metode dari data mining, yaitu menggunakan metode association mining dengan menggunakan algoritma GSP (Generalized Sequential Pattern), algoritma GSP ini fungsi utamanya untuk menemukan sequential atau urutan sehingga pada artikel ini guna untuk mengetahui rekomendasi film. Pada tahapan dalam coding, ada preprocessing, data training dan visualisasi data. Kami memulainya dengan preprocessing yaitu awalnya data itu berupa data mentah selanjutnya kita bersihkan dan nanti hasilnya akan menjadi data yang siap untuk diolah dengan data mining. Selanjutnya data training pada tahap ini digunakan untuk melatih algoritma mencari model yang cocok yang menghasilkan suatu model machine learning. Dan yang terakhir tahap visualisasi data pada proses ini menemukan kandungan yang tersembunyi sehingga mendapatkan hasil rekomendasi film sesuai yang diharapkan menyajikan hasil visual dari data. Gambar 1-3 menunjukkan implementasi penelitian yang menggunakan Bahasa pemrograman Python. Mulai dari tahap pre-processing data, proses algoritma GSP, hingga pola yang dihasilkan oleh algoritma GSP.

```
[ ] dataset = pd.read_csv('movie_metadata.csv')
dataset.fillna(value=0,axis=1,inplace=True)
dataset.head()
```

	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
1	Color	Gore Verbinski	302.0	169.0	563.0	1000.0	Orlando Bloom	40000.0	309404152.0
2	Color	Sam Mendes	602.0	148.0	0.0	161.0	Rory Kinnear	11000.0	200074175.0
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 Tahap Pra Proses Data

	Avatar	Pirates of the Caribbean: At World's End	Spectre	Star Wars: Episode VII - The Force Awakens	The Dark Knight Rises
0	True		False	True	False
1	False		False	True	True
2	True		True	False	True
3	True		False	True	False
4	False		True	False	True
...	...		...	...	...
95	True		True	False	False
96	False		True	False	False
97	True		False	True	True
98	False		True	True	False
99	False		True	False	True

100 rows x 5 columns

Gambar 2 Tahap Proses Pencarian Pola dengan GSP

	Seq 1	Seq 2	Support
0	Avatar	Avatar	10
1	Avatar	Star Wars: Episode VII - The Force Awakens ...	11
2	Pirates of the Caribbean: At World's End	Pirates of the Caribbean: At World's End	11
3	Pirates of the Caribbean: At World's End	Star Wars: Episode VII - The Force Awakens ...	13
4	Star Wars: Episode VII - The Force Awakens ...	Pirates of the Caribbean: At World's End	11
5	Star Wars: Episode VII - The Force Awakens ...	The Dark Knight Rises	10

Gambar 3 Contoh Hasil Pola dengan GSP

#### 4 Simpulan

Berdasarkan penelitian kami mengenai implementasi algoritma Generalized Sequential Pattern, maka dapat disimpulkan bahwa algoritma Generalized Sequential Pattern Mining berhasil digunakan untuk menggali associate rule dan sequential pattern rule dari hasil transaksi film dalam merekomendasikan film yang dimaksud untuk meningkatkan minat menonton film sesuai perilaku penonton. Pada Visualisasi Data menunjukkan rekomendasi film berdasarkan genre atau kategori filmnya sesuai dengan nilai support yang dimiliki masing-masing film. Untuk percobaan selanjutnya diharapkan lebih teliti lagi dalam menganalisis pada saat preprocessing agar data yang dihasilkan lebih akurat dan tepat.

## Referensi

- [1] F. H. Gunawan, Alex Xandra Albert Sim, M Hawari Simanullang, M Firkhan Siregar, “Pengembangan Aplikasi Market Basket Analysis Menggunakan Algoritma Generalized Sequential Pattern pada Supermarket,” in *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, 2015, pp. 1–6, [Online]. Available: <https://journal.uui.ac.id/Snati/article/view/3527>.
- [2] E. Muningsih, “PENERAPAN METODE ASOSIASI GSP DAN APRIORI UNTUK STOK DAN REKOMENDASI PRODUK,” *Konf. Nas. Ilmu Sos. Teknol.*, pp. 412–416, 2017.
- [3] A. Malau, E. Buulolo, and K. Ulfa, “APLIKASI DATA MINING MENGGUNAKAN ALGORITMA GENERALIZED SEQUENTIAL PATTERN (GSP) UNTUK MENENTUKAN ATURAN ASSOSIASI PADA TRANSAKSI PEMINJAMAN BUKU DI PERPUSTAKAAN STMIK BUDIDARMA MEDAN,” *Pelita Inform. Inf. dan Inform.*, vol. 6, no. 4, pp. 206–209, 2018.
- [4] J. Han, J. Pei, Y. Yin, and R. Mao, “Mining frequent patterns without candidate generation: A frequent-pattern tree approach,” *Data Min. Knowl. Discov.*, vol. 8, no. 1, pp. 53–87, 2004, doi: 10.1023/B:DAMI.0000005258.31418.83.
- [5] F. Gorunescu, *Data Mining: Concepts, models and techniques*. 2011.
- [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] R. Srikant and R. Agrawal, “Mining sequential patterns: Generalizations and performance improvements,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1996, vol. 1057 LNCS, pp. 3–17, doi: 10.1007/bfb0014140.