

ALGORITMA GENERALIZED SEQUENTIAL PATTERN UNTUK MENGGALI DATA SEKUENSIAL SIRKULASI BUKU PADA PERPUSTAKAAN UK PETRA

Gregorius Satia Budhi¹, Andreas Handoyo², Christine Oktavina Wirawan³

^{1, 2, 3}Jurusan Teknik Informatika, Fakultas Teknologi Industri, Universitas Kristen Petra

Telp. (031)8494830, (031)8439040

E-mail: greg@petra.ac.id, handoyo@petra.ac.id

ABSTRAK

Dengan mengetahui *pattern* sekuensial peminjaman buku pada perpustakaan, banyak putusan/kebijakan strategis yang dapat diambil oleh pimpinan perpustakaan, misalnya: memberi informasi pada customernya tentang buku-buku yang berelasi, menjaga ketersediaan stok buku-buku yang berelasi agar berimbang, pengaturan peletakan buku-buku yang berelasi pada rak-rak buku, dan banyak putusan strategis lain yang bermanfaat. Oleh sebab itu peneliti mencoba memanfaatkan salah satu metode data mining, yaitu *Generalized Sequential Pattern (GSP)* guna menggali informasi dari data transaksi peminjaman buku di perpustakaan UK Petra. Hasil dari penggalian ini adalah informasi tentang buku-buku yang sering dipinjam secara bersamaan (*Association Rules*) dan buku-buku yang sering dipinjam secara berurutan oleh peminjam yang sama (*Sequential Pattern Rules*). Dengan algoritma GSP kedua macam informasi tersebut akan didapat secara bersamaan dalam sekali proses. Dari pengujian disimpulkan bahwa penelitian ini berhasil menggali informasi yang diinginkan dalam waktu singkat. Dari hasil survei pada para pengambil keputusan di perpustakaan UK Petra tentang kelayakan hasil penelitian ini diaplikasikan, didapat nilai rata-rata sebesar 88.34%.

Kata Kunci: Data Mining, Generalized Sequential Pattern, Sirkulasi Buku

1. PENDAHULUAN

Analisa data sederhana menggunakan aplikasi seperti *spreadsheet* dan *query analysis* akan mengalami kesulitan bila data yang akan dianalisa berskala besar. Data sirkulasi buku pada Perpustakaan UK Petra termasuk dalam kategori ini karena setiap bulannya harus menangani ribuan transaksi sirkulasi buku (peminjaman dan pengembalian) dan judul buku yang ada juga sangat banyak. Akibatnya sampai saat ini data sirkulasi buku yang ada tidak pernah dimanfaatkan saat analisa guna pengambilan keputusan di perpustakaan UK Petra, misal: memberi informasi pada customernya tentang buku-buku yang berelasi, menjaga ketersediaan stok buku-buku yang berelasi agar berimbang, pengaturan peletakan buku-buku yang berelasi pada rak-rak buku, dan banyak putusan strategis lain yang bermanfaat.

Pada penelitian ini dicoba untuk melakukan proses *data mining* pada data sirkulasi buku di perpustakaan UK Petra, guna menghasilkan informasi/knowledge yang berguna bagi para pengambil keputusan di dalam perpustakaan UK Petra. Karena data sirkulasi buku ini bersifat sekuensial, maka digunakan metode *sequential pattern mining* menggunakan algoritma *Generalized Sequential Pattern (GSP)*.

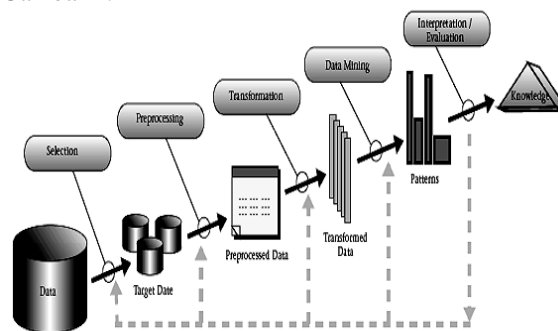
Proses *data mining* terhadap data sirkulasi buku ini akan menghasilkan dua macam informasi/knowledge tentang relasi antar judul buku. Dua macam informasi ini adalah tingkat asosiasi antar buku yang sering dipinjam secara bersamaan (*Association Rules*) dan tingkat asosiasi dari sebuah buku dengan buku-buku lain yang akan dipinjam

secara berurutan dalam kurun waktu yang berdekatan oleh peminjam yang sama (*Sequential Pattern Rules*).

2. TINJAUAN PUSTAKA

2.1 Data Mining

Menurut David Hand (2001), secara sederhana data mining adalah ekstraksi informasi atau pola yang penting atau menarik dari data yang ada di database yang besar. *Data mining* merupakan bagian dari proses *Knowledge Discovery in Databases (KDD)*. Proses dari KDD dapat dilihat pada Gambar 1.



Gambar 1. Tahap-tahap dalam proses *knowledge discovery* (Fayyad, 1996)

2.2 Data Preprocessing

Sebelum data diolah dengan *data mining*, data perlu melalui tahap *preprocessing*. Tahap ini berhubungan dengan pemilihan dan pemindahan data yang tidak berguna (*data cleaning*), penggabungan sumber-sumber data (*data integration*), transformasi data dalam bentuk yang

dapat mempermudah proses (*data transformation*), menampilkan data dalam jumlah yang lebih mudah dibaca (*data reduction*). Semuanya berasal dari data mentah (data transaksi) dan hasilnya akan menjadi data yang nantinya siap untuk diolah dengan data mining (Han, 2003).

2.3 Market Basket Analysis

Market basket analysis adalah suatu metodologi untuk melakukan analisis *buying habit* konsumen dengan menemukan asosiasi antar beberapa *item* yang berbeda, yang diletakkan konsumen dalam *shopping basket* yang dibeli pada suatu transaksi tertentu. Tujuan dari *market basket* adalah untuk mengetahui produk-produk mana yang mungkin akan dibeli secara bersamaan. Analisis data transaksi dapat menghasilkan pola pembelian produk yang sering terjadi. Teknik ini telah banyak digunakan oleh toko grosir maupun retail (Olson, 2007).

2.4 Association Rule

Association rule mining adalah suatu prosedur yang mencari asosiasi antar item dalam suatu data set yang ditentukan. *Interestingness measure* yang dapat digunakan dalam *data mining* adalah (Han, 2003):

- Support, suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu *item* atau *itemset* dari keseluruhan transaksi.
- Confidence, suatu ukuran yang menunjukkan hubungan antar dua *item* secara *conditional*.

2.5 Sequential Pattern Mining

Sebuah record transaksi biasanya berisi tanggal transaksi dan item – item yang dibeli pada transaksi tersebut. Namun sering juga data transaksi berisi identitas dari pelanggannya (CustomerID). Dengan adanya CustomerID ini dapat dilakukan *Mining Sequential Pattern* pada database itu. Contoh dari *pattern* yang dimaksud adalah: bila seorang customer yang menyewa film "Star Wars" berikutnya menyewa film "Empire Strikes Back" kemudian film "Return Of The Jedi". Yang perlu diperhatikan adalah urutannya tidak harus beruntun/*consecutive*, bisa saja customer tersebut menyewa film-film lain bersamaan dengan film dalam contoh. Hal ini tidak menghilangkan pola berurutan (*sequential pattern*) dari penyewaan ketiga film pada contoh (Agrawal, 1995). Agar lebih jelas, dapat dilihat pada ilustrasi pada Tabel 1 dan Tabel 2. Minimum support yang digunakan sebesar 50%.

Informasi dari Tabel 2, dapat digunakan untuk menghasilkan *association rule* maupun *sequential pattern rule*, seperti contoh berikut ini:

- Dari kombinasi (BF), didapat rule-rule:
If Rent(B) Then Rent(F)
If Rent(F) Then Rent(B)
- Dari kombinasi (ABF) didapat rule-rule:

If Rent(A) Then Rent(B and F)
If Rent (B) Then Rent(A and F)
 ... dan seterusnya ...

- Dari kombinasi (BF→A) didapat rule:
If Rent(B and F) Then NextTimeRent(A)
- Dari kombinasi (D→BF→A) didapat rule:
If Rent(D) Then NextTimeRent(B and F)
Then NextTimeRent(A)
 ... dan seterusnya....

Tabel 1. Transaksi yang diurutkan berdasarkan Customer-Id dan Transaction-Time (Zaki, 1997)

Customer-Id	Transaction-Time	Items
1	10	C D
1	15	A B C
1	20	A B F
1	25	A C D F
2	15	A B F
2	20	E
3	10	A B F
4	10	D G H
4	20	B F
4	25	A G H

Tabel 2. Frequent Sequences dengan perhitungan support-nya (Zaki, 1997)

H-Freq Itemsets	Frequent Sequences with Support
\mathcal{F}_1	{(A)[4], (B)[4], (D)[2], (F)[4]}
\mathcal{F}_2	{(AB)[3], (AF)[3], (B→A)[2], (BF)[4], (D→A)[2], (D→B)[2], (D→F)[2], (F→A)[2]}
\mathcal{F}_3	{(ABF)[3], (BF→A)[2], (D→BF)[2], (D→B→A)[2], (D→F→A)[2]}
\mathcal{F}_4	{(D→BF→A)[2]}

2.6 Algoritma Generalized Sequential Pattern

Menurut Mohammed J. Zaki (1997), Algoritma *generalized sequential patterns* (GSP), atau dengan nama lain *apriori all*, adalah suatu algoritma yang dapat memproses dan menemukan semua pola sekuensial dan non sekuensial yang ada. Algoritma GSP dapat dilihat pada Gambar 2 dan Gambar 3.

```

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 C
      that are contained in c.
    Lk = Candidates in Ck with minimum support.
  end
    
```

Gambar 2. Algoritma GSP (Agrawal, 1995; Zaki, 1997)

```

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;
    
```

Gambar 3. Algoritma untuk meng-generate kandidat dari L_{k-1} (Agrawal, 1995; Zaki, 1997)

2.7 Rule Generation

Algoritma ini digunakan untuk membentuk rule-rule (*Association Rule dan Sequential Pattern Rule*) dari semua *frequent sequence pattern* telah ditemukan. Proses generasi rule dapat dilihat pada Gambar 4.

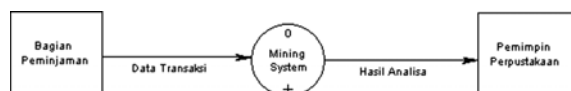
```

RuleGen(F, min_conf):
  for all frequent sequences  $\beta \in F$  do
    for all subsequences  $\alpha \leftarrow \beta$  do
      conf = fr( $\beta$ )/fr( $\alpha$ );
      if (conf  $\geq$  min_conf) then
        output the rule  $\alpha \Rightarrow \beta$ , and conf
    
```

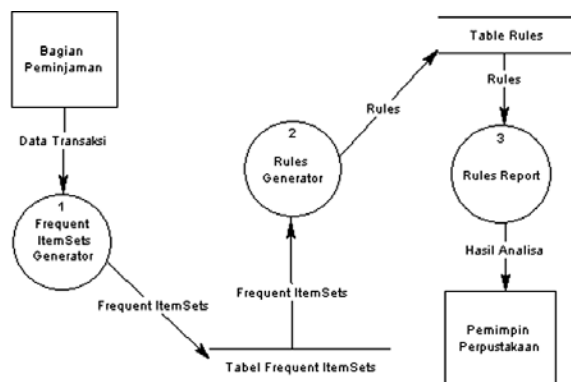
Gambar 4. Algoritma rule generation
(Zaki, 1997)

3. DESAIN APLIKASI

Data Flow Diagram dari aplikasi yang dibuat untuk proses *Sequential Pattern Mining* pada data sirkulasi peminjaman buku di perpustakaan UK Petra dapat dilihat pada Gambar 5 dan Gambar 6. Selain untuk penelitian yang dilakukan, sisten ini sekaligus didesain untuk memenuhi kebutuhan Perpustakaan UK Petra akan adanya sebuah *Decision Support System* yang dapat menunjang pengambilan keputusan.



Gambar 5. Context diagram dari aplikasi



Gambar 6. DFD Level 1 dari aplikasi

Penjelasan dari aplikasi ini adalah sebagai berikut:

- Modul *Frequent ItemSets Generator*, digunakan untuk meng-generate *frequent sequence-frequent sequence* yang tersembunyi dalam data transaksi peminjaman buku perpustakaan UK Petra. Lebih detail langkah-langkahnya yaitu:
 - Memilih range kurun waktu data yang akan diproses (misal: 01/01/05 s/d 31/12/07). Data dalam kurun waktu yang telah ditentukan itu kemudian dikopikan dari data transaksi yang ada ke dalam database lain. Pengkopian ini termasuk juga tabel-tabel

pendukung tabel transaksi, seperti: data master buku, data master customer, dll.

- Hasil copy data transaksi yang ada selanjutnya ditransformasikan menjadi bentuk lain seperti tampak pada contoh di Tabel 1 (Disorting berdasarkan *Customer-Id* dan *Transaction-Time*)
 - Dari data yang telah ditransformasi itu selanjutnya di-generate *frequent items* yang didalamnya terdapat *frequent sequences*. Proses ini dilakukan dengan menggunakan algoritma GSP yang ada pada Gambar 2 dan 3.
 - Simpan semua *frequent items* yang dihasilkan ke dalam *Tabel Frequent Itemsets*. Setiap kali modul *Frequent Itemsets Generator* dijalankan, hasilnya akan disimpan pada sebuah *Tabel Frequent Itemsets* yang berbeda.
- Modul *Rules Generator*, dibuat untuk meng-generate *association rules* dan *sequential pattern rules* dari *Tabel Frequent Itemsets* yang dipilih. Proses generate rule ini menggunakan algoritma yang ada pada Gambar 4. Hasil dari modul ini disimpan pada tabel *rules*. Setiap kali proses pada modul ini, sebuah tabel *rules* baru akan di-create.
 - Modul *Rules Report*, digunakan untuk menampilkan *association rule* dan *sequential pattern rule* yang dihasilkan kepada pengguna dalam bentuk tabel, grafik dan *tree view*.

4. PENGUJIAN

Pengujian dilakukan pada sebuah PC dengan spesifikasi sebagai berikut:

- CPU: Intel Pentium 4 1.73 GHz.
- RAM: 512 Mb.
- OS: Ms. Windows XP
- Harddisk: 80 Gb.
- Database: Ms. Access 2003
- Programming Tool: Delphi 7

IDRULE	KOMBINASI	SUPPORT	CONFIDENCE	RULE
1	2	2	2	100 01403/03=>01749/02
2	2	2	2	100 01749/02=>01403/03
3	2	2	2	100 01403/03=>04404/96
4	2	2	2	100 04404/96=>01403/03
5	2	2	2	100 01431/00=>02937/93
6	2	2	2	100 02937/93=>01431/00

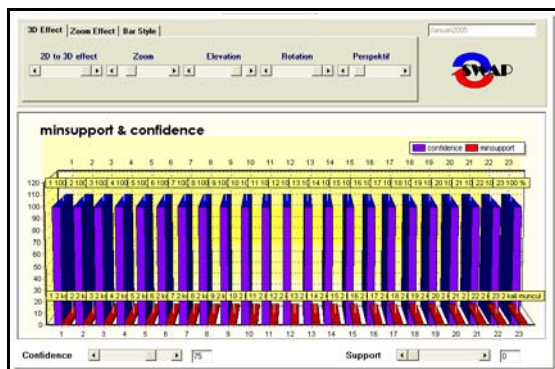
Book Id	Book Title
01431/00	Pengantar akuntansi I
02937/93	Statistics for business and economics

Gambar 7. Tampilan *association rules* hasil aplikasi dalam bentuk tabel.

Ada beberapa macam pengujian yang dilakukan untuk penelitian ini, yaitu:

- a. Pengujian aplikasi yang dibuat secara keseluruhan: Data sirkulasi yang digunakan dalam pengujian ini adalah data sirkulasi peminjaman buku perpustakaan UK Petra tahun 2005. Tampilan dari *rule-rule* yang dihasilkan dapat dilihat pada Gambar 7 sampai dengan Gambar 10. Pada pengujian dapat dilihat bahwa *association rules* dan *sequential pattern rules* dapat di-generate dengan benar dan ditampilkan pada antar muka yang mudah dipahami.

Gambar 8. Tampilan *sequential pattern rules* hasil aplikasi dalam bentuk tabel.



Gambar 9. Tampilan *sequential pattern rules* hasil aplikasi dalam bentuk grafik.

Gambar 10. Tampilan *association rules* dan *sequential pattern rules* hasil aplikasi dalam bentuk tree view.

- b. Pengujian kecepatan proses aplikasi yang dibuat dalam meng-generate rule dari beberapa macam data transaksi berbeda. Hasil dapat dilihat pada Tabel 3.

Tabel 3. Hasil Pengujian kecepatan proses

Jml Transaksi	Min. Support	ItemSet	Waktu Proses
Transaksi 1903	2	Candidate 34900 Frequent: 44	6m 6s
Jumlah Item 1749	3	Candidate 809 Frequent: 5	33s
Transaksi 1097	2	Candidate 4036 Frequent: 76	17s
Jumlah Item 1039	3	Candidate 219 Frequent: 15	7s
Transaksi 965	2	Candidate 1618 Frequent: 37	16s
Jumlah Item 917	3	Candidate 219 Frequent: 15	7s

Dari pengujian kecepatan proses ini dapat disimpulkan bahwa semakin banyak jumlah transaksi dan jumlah item yang diproses, waktu proses yang dibutuhkan semakin lama. Lama waktu proses ini masih kurang dari 10 menit untuk memproses transaksi yang cukup banyak, sehingga aplikasi ini masih wajar untuk dipakai sebagai *decision support sistem* di perpustakaan UK Petra.

- c. Pengujian oleh calon pengguna. Pengujian ini dalam bentuk kuisioner yang diisi oleh calon pengguna aplikasi, yaitu karyawan UK Petra yang bekerja pada perpustakaan dan berhak/berkewajiban/berhubungan dengan pengambilan keputusan strategis. Hasil Kuisioner dapat dilihat pada Tabel 4.

Tabel 4. Hasil Kuisioner

No	Jabatan Responden	Poin Kuisioner			
		1	2	3	4
1	Kepala Perpustakaan	4	4	5	4
2	KaSubBid. TU Perpustakaan	4	4	5	5
3	Staf Ref. & Kerjasama	4	5	5	4
4	Staf Pengolahan Koleksi	4	4	5	5
5	Staf Sirkulasi	4	5	5	4
6	Staf Sis. Inform. Perpustakaan	4	4	4	5
Total:		4	4.3	4.8	4.5
Total dalam prosentase:		80	86.7	96.7	90

Keterangan:

Nilai: 1. Sangat kurang s/d 5. Sangat baik

Point Kuisioner:

No. 1: Keakuratan informasi yang dihasilkan

No. 2: Desain antar muka

No. 3: Kemudahan penggunaan aplikasi

No. 4: Nilai tambah aplikasi untuk proses pengambilann keputusan di perpustakaan UK Petra

Hasil dari pengujian 4.c. menyatakan bahwa para pengambil keputusan di perpustakaan UK Petra menganggap aplikasi ini telah memenuhi harapan mereka dan dapat digunakan sebagai *decision support sistem* bagi perpustakaan UK Petra.

5. KESIMPULAN

Dari hasil pengujian aplikasi dapat disimpulkan bahwa algoritma *Generalized Sequential Patterns* (GSP) dapat digunakan untuk menggali *association rule* dan *sequential pattern rule* dari data transaksi sirkulasi buku perpustakaan UK Petra. Kecepatan proses dari aplikasi yang dibuat juga cukup cepat sehingga masih wajar bila diaplikasikan menjadi sebuah aplikasi *decision support system*. Dari hasil kuisioner yang didapat, yaitu rata-rata 88.34%, dapat pula disimpulkan bahwa aplikasi yang dibuat telah dapat memenuhi harapan para pengambil keputusan di perpustakaan UK Petra.

PUSTAKA

- Agrawal, Rakesh, dan Srikant, Ramakrishnan (1995). Mining Sequential Patterns. *International Conference on Data Engineering*.
- Agrawal, Rakesh, Imielinsky, Tomasz, dan Swami, Arun (1993). Mining Association Rules between Sets Of Items In Large Databases. *Proceeding Of The 1993 ACM SIGMOD Conference, Washington DC, USA*.
- Agrawal, Rakesh, dan Srikant, Ramakrishnan (1994). Fast algorithms for mining association rules. *Proceeding Of The 1994 International Conference Very Large Data Bases, Santiago, Chile*.
- Fayyad, U., Shapiro, G. Piatetsky, dan Smyth, Padhraic (1996). From data mining to knowledge discovery in databases. *AI Magazine*, pp. 37-54.
- Han, Jiawei, dan Kamber, Micheline (2003). *Data Mining: Concepts and Techniques*. 2nd Edition. Morgan Kaufmann.
- Hand, David, Mannila, Heikki, dan Smyth, Padhraic (2001). *Principles Of Data Mining*. The MIT Press.
- Olson, David dan Yong, Shi (2007). *Introduction To Business Data Mining*. New York: McGraw-Hill.
- Srikant, Ramakrishnan, dan Agrawal, Rakesh (1996). Mining Sequential Patterns: Generalizations and performance improvements. *5th International Conference Extending Database Technology*.
- Zaki, Mohammed J. (1997). Fast Mining of Sequential Patterns in Very Large Databases. *The University of Rochester Computer Science Department Rochester, New York 14627. Technical report 668*.
- Zaki, Mohammed J., Parthasarathy, S., Ogihara, M., dan Li, W. (1997). New Parallel Algorithms for Fast Discovery of Association Rules. *Data Mining and Knowledge Discovery: An International Journal*.