

Market Basket Analysis Menggunakan Algoritma Generalized Sequential Pattern

Market Basket Analysis Using Generalized Sequential Pattern Algorithm

¹Yuyun Nisrina, ²Anneke Iswani Achmad

^{1,2}*Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung,*

Jl. Tamansari No.1 Bandung 40116

email: ¹yuyunnisrina20@gmail.com, ²annekeiswani11@gmail.com

Abstract. Data mining is a technique to process data from a large database to obtain information that will help in decision making. One method on data mining is Market Basket Analysis which is used to analyse the patterns of consumer spending behavior to find associations between products in the transaction data. This paper will discuss the Market Basket Analysis using the Generalized Sequential Pattern (GSP) algorithm on the sales transaction data of PT. X in Malang city to find sequential pattern or sequence by involving transaction time measured through support value to show how big the level of dominance of an item of the whole transaction. From the sales transaction data of PT. X in the city of Malang in 2017, the transaction is conducted by as many as 58 consumers consisting of 68 types of products. Result of the data, obtained frequent 3-sequence of extra food (will 1 & 2) and Etta Goat milk (Will 1 & 2) purchased simultaneously at one time transaction, then the next transaction bought Etta Goat milk (Will 1 & 2) with the value of support of 20.69%.

Keywords: Market Basket Analysis, Data Mining, Generalized Sequential Pattern, Support.

Abstrak. *Data mining* adalah suatu teknik mengolah data dari *database* yang besar untuk memperoleh suatu informasi yang akan membantu dalam pengambilan keputusan. Salah satu metode pada *data mining* yaitu *Market Basket Analysis* yang digunakan untuk menganalisis pola perilaku belanja konsumen untuk menemukan asosiasi antar produk dalam data transaksi. Makalah ini akan membahas *Market Basket Analysis* menggunakan algoritma *Generalized Sequential Pattern (GSP)* pada data transaksi penjualan PT.X di Kota Malang untuk menemukan pola sekuensial atau urutan dengan melibatkan waktu transaksi yang diukur melalui nilai *support* untuk menunjukkan seberapa besar tingkat dominasi suatu item dari keseluruhan transaksi. Dari data transaksi penjualan PT.X di Kota Malang tahun 2017, transaksi dilakukan oleh sebanyak 58 konsumen yang terdiri dari 68 jenis produk. Hasil dari data tersebut, diperoleh *frequent 3-sequence* yaitu extra food (will 1 & 2) dan etta goat milk (will 1 & 2) yang dibeli secara bersamaan dalam satu waktu transaksi, kemudian transaksi selanjutnya membeli etta goat milk (will 1 & 2) dengan nilai *support* sebesar 20.69%.

Kata Kunci: Market Basket Analysis, Data Mining, Generalized Sequential Pattern, Support.

A. Pendahuluan

Saat ini teknologi berupa sistem komputerisasi sudah banyak digunakan oleh para pelaku usaha dalam berbagai bidang sehingga menghasilkan data yang berukuran sangat besar. Data yang terkumpul ini merupakan suatu tambang emas yang dapat digunakan sebagai informasi dalam dunia bisnis. *Data mining* merupakan penggalian informasi yang diambil dari kumpulan data besar yang membantu dalam pengambilan keputusan. Dari data yang dihasilkan diperlukan suatu analisis data untuk menggali informasi-informasi yang ada.

Salah satu metode yang digunakan dalam *data mining* adalah *market basket analysis* yaitu metode yang dapat digunakan untuk menganalisis pola perilaku belanja konsumen (Annie dan Kumar, 2012). Tujuan dari analisis ini yaitu untuk menemukan asosiasi antar produk dalam data transaksi. Dalam makalah ini metode *market basket analysis* akan menggunakan bantuan algoritma *Generalized Sequential Pattern (GSP)*.

Algoritma *GSP* atau dengan nama lain *apriori all* adalah suatu algoritma yang dapat memproses dan

menemukan semua pola sekuensial atau urutan yang ada (Zaki,1997). Kelebihan dari algoritma *GSP* ini yaitu melibatkan waktu transaksi sehingga bisa menemukan pola untuk transaksi selanjutnya. Data yang digunakan dalam skripsi ini adalah data transaksi penjualan di PT.X tahun 2017 yang berlokasi di Kota Malang. Dari data transaksi tersebut akan dilihat pola sekuensial atau urutan pembelian produk tertentu.

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah dalam penelitian ini sebagai berikut:

1. Bagaimana menemukan pola sekuensial atau urutan dari data transaksi penjualan di PT.X yang berlokasi di Kota Malang dengan Algoritme *Generalized Sequential Pattern*?
2. Berapa besar nilai *support* berdasarkan pola sekuensial atau urutan dari data transaksi penjualan di PT.X yang berlokasi di Kota Malang dengan Algoritme *Generalized Sequential Pattern*?

B. Landasan Teori

Data Mining

1. Pengertian *Data Mining*
Secara sederhana *data mining* adalah ekstraksi informasi atau pola yang penting atau menarik dari data yang ada di *database* yang besar (Hand, 2001). *Data mining* merupakan bidang dari beberapa keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, *database* dan visualisasi untuk penanganan permasalahan pengambilan informasi dari *database* yang besar (Larose, 2005).

2. Tahapan Data Mining

Dalam proses *data mining* terdapat tujuh tahapan (Han dkk, 2012) yaitu pembersihan data (*data cleaning*), integrasi data (*data integration*), seleksi data (*data selection*), transformasi data (*data transformation*), proses *mining*, evaluasi pola (*pattern evaluation*) dan mengemukakan pengetahuan (*knowledge presentation*).

3. Teknik Data Mining

Menurut (Larose, 2005), teknik *data mining* dapat dilakukan dengan menggunakan asosiasi (*association*), klasifikasi (*classification*), pengelompokan (*clustering*), prediksi (*prediction*) dan pola-pola sekuensial (*Sequential Pattern Analysis*).

Market Basket Analysis

Salah satu metode yang ada dalam *data mining* adalah metode *Market Basket Analysis*. Analisis ini merupakan suatu metodologi untuk melakukan analisis *buying habit* konsumen dengan menemukan asosiasi antar beberapa barang yang berbeda, yang diletakkan konsumen dalam *shopping basket* (keranjang belanja) yang dibeli pada suatu transaksi tertentu (Han, 2006).

Algoritma *Generalized Sequential Pattern (GSP)*

Algoritma *GSP* atau dengan nama lain *apriori all*, adalah suatu algoritma yang dapat memproses dan menemukan pola sekuensial yang ada dari data transaksi yang melibatkan waktu transaksi. Algoritma ini digunakan untuk membentuk aturan-aturan (*association rule* dan *sequential pattern rule*) dari semua *frequent sequence pattern* yang telah ditemukan (Zaki, 1997).

Algoritma *GSP* didesain untuk data transaksi, dimana setiap pola merupakan kumpulan dari transaksi berupa *items* (Ahola, 2001). Algoritma ini bekerja menemukan semua pola sekuensial yang sesuai dengan minimum *support* yang ditentukan. Struktur dasar dari algoritma *GSP* sangat mirip dengan algoritma apriori, tetapi *GSP* menggunakan urutan bukan itemset (Srikant, 1996).

Association Rules Mining

Association rules mining adalah metode *data mining* yang menggunakan hubungan satu *item* dengan *item* lainnya yang terdapat di dalam *itemsets* untuk memprediksi pola dari kumpulan data (Olson & Shi, 2008). Ada ukuran kepercayaan (*interestingness measure*) yang digunakan dalam menentukan suatu *association rules*, yaitu *support*. *Support* digunakan untuk mengukur seberapa besar tingkat dominasi suatu item dari keseluruhan transaksi atau bisa diartikan seberapa besar peluang konsumen membeli beberapa item dari keseluruhan transaksi. Berikut rumus untuk menghitung nilai *support* sebuah *item*.

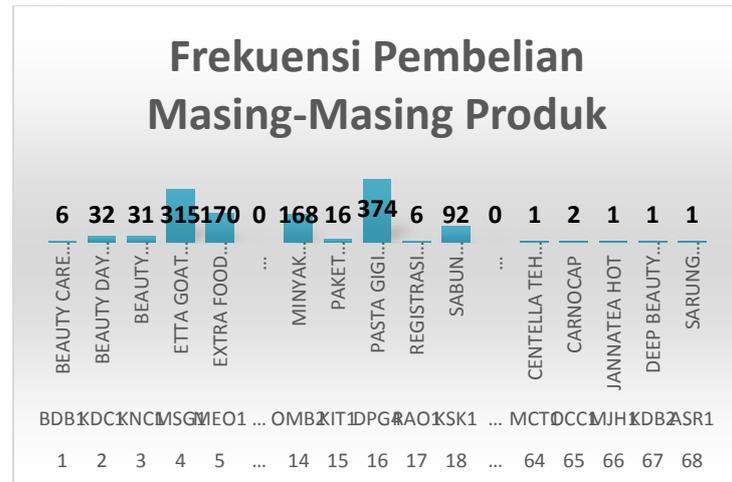
$$\text{Support}(A) = \frac{\text{banyaknya transaksi yang mengandung } A}{\text{total transaksi}} \times 100\% \dots(2.1)$$

C. Hasil Penelitian dan Pembahasan

Bab ini berisikan tentang hasil dan pembahasan dari penerapan algoritma *Generalized Sequential Pattern (GSP)* untuk data transaksi penjualan PT.X yang berlokasi di Kota Malang tahun 2017. Data transaksi ini berukuran 58 x 68. Kemudian algoritma *GSP* dilakukan dengan bantuan *software Phyton 3*. Algoritma *GSP* digunakan untuk menemukan pola sekuensial dan menghitung nilai *support* berdasarkan *frequent sequence pattern* yang telah ditemukan.

Deskripsi Data

Dalam bagian ini akan mendeskripsikan data transaksi penjualan PT.X di Kota Malang tahun 2017. Dari data transaksi terlebih dahulu akan dicari frekuensi dari semua jenis produk yang dibeli oleh konsumen. Dengan bantuan *software Microsoft Excel*, hasilnya disajikan dalam Gambar 4.1.



Gambar 4.1 Grafik Frekuensi Pembelian untuk Masing-Masing Produk

Hasil Aplikasi Algoritma *GSP*

Setelah data dideskripsikan, selanjutnya akan diaplikasikan algoritma *GSP* pada data transaksi penjualan PT.X di Kota Malang tahun 2017. Langkah selanjutnya adalah membuat tabel *frequent 1-sequence* yang transaksinya memenuhi syarat *minimum support count*. Dalam pembahasan ini akan digunakan *minimum support count* 20% dari 58 transaksi yang terjadi, sehingga didapat *minimum support count* nya sebesar 12 transaksi. Dimana hasilnya terdapat 10 jenis item dengan bantuan *software Phyton 3* sebagaimana disajikan pada Tabel 4.1.

Tabel 4.1 *Frequent 1-Sequence*

Item	Banyaknya Transaksi
DPG4	30
KIT1	12
KSK1	16
MEO1	22
MHC1	15
MKE1	16
MSG1	26
MSK1	17
OMB2	26
OZC1	15

Berdasarkan tabel 4.1 ada 10 jenis item yang banyak transaksinya lebih dari atau sama dengan *minimum support count*. Dari tabel 4.1 terlihat bahwa item yang jumlah transaksinya paling banyak yaitu DPG4 sebanyak 30 transaksi, MSG1 sebanyak 26 transaksi, OMB2 sebanyak 26 transaksi, dan seterusnya. Informasi yang didapat pada Tabel 4.1 akan digunakan untuk mencari *frequent 2-sequence*. Hasil *frequent 2-sequence* akan disajikan pada Tabel 4.2 dengan bantuan *software Python 3* sebagai berikut.

Tabel 4.2 *Frequent 2-Sequence*

Item	Banyaknya Transaksi
DPG4,KSK1	14
DPG4,MEO1	16
DPG4,MSG1	17
DPG4,OMB2	15
DPG4,OZC1	13
MEO1,MSG1	15
MEO1,OMB2	13
MHC1,MSG1	12
MSG1,OMB2	16
DPG4→DPG4	15
DPG4→MSG1	13
MEO1→MEO1	14
MEO1→MSG1	13
MEO1→MSK1	12
MEO1→OMB2	12
MSG1→DPG4	12
MSG1→MSG1	13
OMB2→DPG4	12

OMB2→MEO1	15
OMB2→OMB2	12

Berdasarkan Tabel 4.2 ada 20 jenis kombinasi yang memuat *frequent 2-sequence* yang banyak transaksinya lebih dari atau sama dengan *minimum support count*. Dari tabel 4.3 terlihat bahwa ada 14 transaksi yang memuat 2 item DPG4 dan KSK1 yang dibeli secara bersama-sama dalam satu waktu transaksi, ada 16 transaksi yang memuat 2 item DPG4 dan MEO1 yang dibeli secara bersama-sama dalam satu waktu transaksi, dan seterusnya.

Langkah selanjutnya adalah membuat tabel *frequent 3-sequence* yang banyak transaksinya lebih dari atau sama dengan *minimum support count*. Hasilnya disajikan dalam tabel 4.3 sebagai berikut.

Tabel 4.3 *Frequent 3-Sequence*

Item	Banyaknya Transaksi
MEO1,MSG1→MSG1	12

Berdasarkan tabel 4.3 terlihat bahwa ada 1 pola terdiri dari 3 item yang banyak transaksinya lebih dari atau sama dengan 20% dari keseluruhan transaksi atau 12 transaksi. Pada tabel 4.3 terlihat bahwa dengan *minimum support count* 20% dari keseluruhan transaksi atau 12 transaksi, *Frequent Sequence* yang dihasilkan hanya bisa sampai *Frequent 3-Sequence* tidak bisa lebih. Oleh karena itu, algoritma *GSP* dihentikan.

Hasil Perhitungan Support

Perhitungan ini akan diuraikan perhitungan nilai *support*, yaitu ukuran yang menunjukkan seberapa besar tingkat dukungan data terhadap validitas aturan yang dikembangkan. **Tabel 4.4** dibawah ini berisikan nilai *support* untuk seluruh *frequent 1-sequence* dalam bentuk yang sudah diurutkan dari nilai *support* terbesar sampai terkecil.

Item	Banyaknya Transaksi	Support
DPG4,MSG1	17	29.31%
DPG4,MEO1	16	27.59%
MSG1,OMB2	16	27.59%
DPG4,OMB2	15	25.86%
MEO1,MSG1	15	25.86%
DPG4→DPG4	15	25.86%
OMB2→MEO1	15	25.86%
DPG4,KSK1	14	24.14%
MEO1→MEO1	14	24.14%
DPG4,OZC1	13	22.41%

Tabel 4.4 Nilai *Support Frequent 1-Sequence*

Item	Banyaknya Transaksi	Support
DPG4	30	51.72%
MSG1	26	44.83%
OMB2	26	44.83%
MEO1	22	37.93%
MSK1	17	29.31%
KSK1	16	27.59%
MKE1	16	27.59%
MHC1	15	25.86%
KIT1	12	20.69%
OZC1	15	25.86%

Berdasarkan tabel 4.4 nilai *support* yang ada menjadi informasi bagi PT.X Kota Malang dalam hal manajemen stok penjualannya. Sebagai contoh, item DPG4 harus terus ditambah stoknya, karena paling sering dibeli oleh konsumen dengan peluang sebesar 51.72% agar tidak terjadi kekurangan stok. Begitu seterusnya untuk *frequent 1-sequence* yang ada pada tabel 4.4. Selanjutnya akan dihitung nilai *support* untuk *frequent 2-sequence* dengan menggunakan persamaan (2.1).

Tabel 4.5 Nilai *Support Frequent 2-Sequence*

Item	Banyaknya Transaksi	Support
MEO1,OMB2	13	22.41%
DPG4→MSG1	13	22.41%
MEO1→MSG1	13	22.41%
MSG1→MSG1	13	22.41%
MHC1,MSG1	12	20.69%
MEO1→MSK1	12	20.69%
MEO1→OMB2	12	20.69%
MSG1→DPG4	12	20.69%
OMB2→DPG4	12	20.69%
OMB2→OMB2	12	20.69%

Berdasarkan tabel 4.5 nilai *support* yang ada menjadi informasi bagi PT.X Kota Malang dalam hal menyusun tata letak rak penjualan. Sebagai contoh, item DPG4 dan MSG1 harus diletakkan berdampingan di rak-rak penjualan, karena paling sering dibeli secara bersamaan dalam satu waktu transaksi oleh konsumen, dengan peluang sebesar 29,31%. Item DPG4 dan MEO1 harus diletakkan berdampingan di rak-rak penjualan, karena berada di urutan ke-2 yang sering dibeli secara bersamaan dalam satu waktu transaksi oleh konsumen, dengan peluang sebesar 27,59%. Begitu seterusnya untuk *frequent 2-sequence* yang ada pada tabel 4.5.

Tabel 4.6 Nilai *Support Frequent 3-Sequence*

Item	Banyaknya Transaksi	Support
MEO1,MSG1→MSG1	12	20.69%

Nilai ini menunjukkan bahwa ada sekitar 20.69% dari 58 transaksi

yang memuat item MEO1, MSG1 → MSG1 yang dibeli secara bersamaan dalam satu waktu transaksi oleh konsumen. Dimana nilai *support* yang ada menjadi informasi bagi PT.X Kota Malang dalam hal menyusun tata letak rak penjualan. Item MEO1, MSG1 → MSG1 harus diletakkan berdampingan di rak-rak penjualan, karena paling sering dibeli secara bersamaan dalam satu waktu transaksi oleh konsumen, dengan peluang sebesar 20.69%.

D. Kesimpulan

Berdasarkan pembahasan dalam penelitian ini, peneliti menyimpulkan beberapa hasil penelitian sebagai berikut:

1. Pola sekuensial atau urutan algoritma *GSP* dengan *minimum support count* 20% atau 12 transaksi dihentikan di *frequent 3-sequence* karena sudah tidak ada lagi *frequent sequence* yang memenuhi *minimum support count*. Hasilnya didapatkan 10 *frequent 1-sequence*, 20 *frequent 2-sequence* dan 1 *frequent 3-sequence* yaitu extra food (wil 1 & 2) dan etta goat milk (wil 1 & 2) yang dibeli secara bersamaan dalam satu transaksi, kemudian pada transaksi selanjutnya membeli etta goat milk (wil 1 & 2).
2. Nilai *support* terbesar untuk *frequent 1-sequence* yaitu pasta gigi hebal HPAI new sebesar 51.72%, artinya ada sekitar 51.72% dari 58 transaksi memuat item pasta gigi hebal HPAI new. *Support* terbesar untuk *frequent 2-sequence* yaitu pasta gigi hebal HPAI new dan etta goat milk (wil 1 & 2) sebesar 29.31% , artinya peluang konsumen membeli pasta gigi hebal HPAI new dan etta goat milk (wil 1 & 2) sebesar 29.31%.

Nilai *support* untuk *frequent 3-sequence* yaitu extra food (wil 1 & 2), etta goat milk (wil 1 & 2) dan etta goat milk (wil 1 & 2) sebesar 20.69% artinya peluang konsumen membeli extra food (wil 1 & 2), etta goat milk (wil 1 & 2) dan etta goat milk (wil 1 & 2) sebesar 20.69%.

E. Saran

1. Disarankan kepada PT.X di Kota Malang untuk menata produk yang dijual pada etalase/rak penjualan dimana menyesuaikan dengan hasil analisis yang peneliti lakukan yang berkaitan dengan produk-produk yang dibeli secara bersamaan dalam satu waktu transaksi agar lebih mengefektifkan waktu dan tenaga.
2. Disarankan kepada PT.X di Kota Malang untuk memperhatikan stok produk yang sering atau yang jarang dibeli agar tidak terjadi penumpukkan atau kekurangan stok.
3. Disarankan kepada peneliti lain untuk meneliti *market basket analysis* dengan bantuan algoritma yang dapat memperlihatkan berapa banyaknya jumlah penjualan dari *frequent sequence* yang dihasilkan sehingga perusahaan dapat mengetahui stok produk dari setiap *frequent sequence*.
4. Disarankan kepada peneliti lain untuk meneliti *market basket analysis* dengan bantuan algoritma yang berbeda seperti algoritma *Prefix SPAN (Prefix-projected Sequential Pattern Mining)*, *SPADE (An Efficient Algorithm for Mining Frequent Sequences)*, dan lainnya.

Daftar Pustaka

- Ahola, Jussi. (2001). *Mining Sequential Patterns*. Finland: VTT Information Technology.
- Annie, M. C., L. C., dan Kumar, D., A. (2012). *Market Basket Analysis for a Supermarket Based on Frequent Itemset Mining*. International Journal of Computer Science Issues, 257-264.
- Han, J., dan Kamber, M. (2006). *Data Mining: Concept amd Techniques*. Waltham: Morgan Kaufmann Publishers.
- Han, J., Kamber, M., dan Pei, J. (2012). *Data Mining: Concept amd Techniques, Third Edition*. Waltham: Morgan Kaufmann Publishers.
- Hand, David, Mannila, Heikki, dan Smyth, Padhraic (2001). *Priciples of Data Mining*. The MIT Press.
- Larose, Daniel T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. Canada: John Willey and Sons. Inc.
- Olson, D., dan Shi, Y. (2008). *Pengantar Ilmu Penggalian Data Bisnis*. Salemba Empat, Jakarta.
- Srikant, Ramakrishnan. (1996). *Fast Algorithm for Mining Association Rules and Sequential Patterns*. Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy at the University of Wisconsin – Madison.
- Zaki, Mohammed J. (1997). *Fast Mining of Sequential Patterns in Very Large Databases*. The University of Rochester Computer Science Departement Rochester, New York 14627. Technical Report 668.