

## ***Sequential Pattern Mining* untuk Data Transaksi Penjualan Supermarket menggunakan Algoritma Generalized Sequential Pattern**

Albert Kurniawan<sup>1</sup>, Ramos Somya<sup>2</sup>

<sup>1,2</sup>Teknik Informatika, Universitas Kristen Satya Wacana, Jl. Dr. O. Notohamidjojo No.1 - 10, Blotongan, Kec. Sidorejo, Kota Salatiga, Jawa Tengah, Indonesia, 50715  
e-mail: <sup>1</sup>672017058@student.uksw.edu, <sup>2</sup>ramos.somya@uksw.edu

Submitted Date: August 05<sup>th</sup>, 2021  
Revised Date: September 06<sup>th</sup>, 2021

Reviewed Date: August 15<sup>th</sup>, 2021  
Accepted Date: MMMM dd, yyyy

### **Abstract**

*Online supermarket sales transaction data is a sequence dataset. This data stores purchase transaction data made by customers, so it can be analyzed using Market Basket Analysis (MBA) approach. The problem that is often experienced by supermarkets is the difficulty of implementing the accurate sales strategy to consumers. Based on these problems, this research will analyze the West Superstore supermarket dataset based on the MBA approach. The algorithm used is the Generalized Sequential Pattern (GSP) algorithm, where this algorithm can generate frequent items and sequence patterns, so that the resulting rules can be more accurate. The GSP algorithm in this study is implemented in the Python programming language. The test results show that the output of Python is in accordance with the output of the GSP algorithm calculation. The time required for rule generation in the GSP algorithm also depends on the number of records being used. The more number of sales transactions to be analyzed, it needs longer time in computation. The analysis conducted on the sales dataset at the West Superstore resulted in 391 rules, where these rules can be used by supermarkets to implement their sales strategies.*

**Keywords:** *Market Basket Analysis; West Superstore; Generalized Sequential Pattern; Python.*

### **Abstrak**

Data transaksi penjualan supermarket online merupakan *sequence dataset*. Data ini menyimpan data transaksi pembelian yang dilakukan oleh pelanggan, sehingga dapat dianalisis menggunakan pendekatan *Market Basket Analysis* (MBA). Masalah yang sering dialami oleh pihak supermarket adalah sulitnya menerapkan strategi penjualan yang akurat kepada para konsumen. Berdasarkan masalah tersebut, pada penelitian ini akan dilakukan analisis terhadap *dataset* supermarket West Superstore berdasarkan pendekatan MBA. Algoritme yang digunakan adalah algoritme Generalized Sequential Pattern (GSP), di mana algoritme ini dapat membangkitkan *frequentitem* dan *sequencepattern*, sehingga aturan yang dihasilkan dapat lebih akurat. Algoritme GSP pada penelitian ini diimplementasikan dalam bahasa pemrograman Python. Hasil pengujian menunjukkan bahwa keluaran dari Python sudah sesuai dengan keluaran dari perhitungan algoritme GSP. Waktu komputasi yang diperlukan untuk pembangkitan aturan pada algoritme GSP juga bergantung pada jumlah *record* yang ada. Semakin banyak jumlah transaksi penjualan yang akan dianalisis, maka waktu komputasinya juga semakin lama. Analisis yang dilakukan pada *dataset* penjualan di West Superstore menghasilkan 391 aturan, di mana aturan tersebut dapat dimanfaatkan oleh pihak supermarket untuk penerapan strategi penjualan.

**Kata kunci:** *MarketBasketAnalysis; West Superstore; Generalized Sequential Pattern; Python*

## 1. Pendahuluan

Data transaksi penjualan merupakan salah satu jenis data yang dihasilkan dari kegiatan penjualan produk ataupun jasa. Data tersebut akan bertambah seiring dengan adanya transaksi penjualan yang terjadi setiap saat. Data transaksi penjualan biasanya memuat informasi produk yang dibeli oleh konsumen, tanggal pembelian dan beberapa informasi lainnya yang biasanya tertera pada struk atau kuitansi pembelian. *Intelligence Enterprise* atau perusahaan yang sudah menerapkan kecerdasan bisnis akan memanfaatkan data tersebut untuk dianalisis dan digunakan dalam penentuan strategi bisnis.

Analisis terhadap data transaksi penjualan dapat dilakukan menggunakan pendekatan *Market Basket Analysis* (MBA). MBA adalah salah satu metode penggalian data untuk mengetahui barang-barang yang dibeli oleh konsumen secara bersamaan. MBA dapat diterapkan dalam bidang marketing untuk menyediakan informasi kepada perusahaan tentang perilaku konsumen, di mana hal ini dapat membantu perusahaan dalam mengambil keputusan untuk keperluan strategi marketing (Kaur, 2016)

*Association rule mining* merupakan salah satu teknik yang dapat diterapkan dalam MBA, dimana teknik tersebut digunakan untuk menemukan hubungan antar produk item yang tersembunyi dalam dataset transaksi penjualan. Hubungan asosiasi yang sudah ditemukan pada *association rule* direpresentasikan dalam bentuk aturan sebab akibat berdasarkan nilai *support* dan *confidence* (Prasetyo, 2012). Metode kedua yang dapat digunakan untuk MBA adalah *sequential pattern mining*. Metode ini selain dapat menghasilkan *association rule* juga dapat menghasilkan *sequence pattern* atau urutan kombinasi produk yang dibeli oleh pelanggan, sehingga metode ini diharapkan dapat menghasilkan *rule* yang lebih lengkap dalam MBA.

Pada penelitian ini akan dilakukan MBA menggunakan teknik *sequence pattern mining* terhadap dataset penjualan di salah satu supermarket online menggunakan algoritme Generalized Sequential Pattern (GSP). Tujuan dari penelitian ini adalah untuk menemukan aturan asosiasi dan *sequential* berdasarkan dataset transaksi penjualan. Aturan atau *rule* yang dihasilkan akan digunakan untuk menunjang pengambilan keputusan dalam penentuan strategi penjualan.

Terdapat beberapa penelitian terdahulu yang mengkaji tentang *Market Basket Analysis* (MBA). *Association rule mining* pernah dimanfaatkan untuk analisis data transaksi penjualan pada salah satu supermarket di Tamil Nadu, India (Kumar, 2018). Penelitian ini mengkaji tentang masalah *cross selling analysis* yang diperlukan dalam dunia industri retail. Jumlah penjualan pada industri retail dipercaya dapat ditingkatkan melalui penentuan letak produk di toko, dan strategi promosi yang didasarkan pada pergerakan penjualan produk. Pada penelitian ini, Kumar menggunakan data transaksi penjualan selama 5 bulan untuk dicari aturan asosiasinya menggunakan algoritme Apriori. *Software R platform* dimanfaatkan untuk menghasilkan aturan asosiasi, di mana aturan asosiasi tersebut kemudian dipakai untuk membantu dalam pengambilan keputusan terkait dengan strategi pemasaran produk.

Metode MBA pernah diterapkan pada salah satu Usaha Kecil dan Menengah (UKM) minimarket Toko Baru (Novrianto, 2020). Penelitian ini mencoba untuk mengatasi persaingan bisnis menggunakan metode MBA. Algoritme Apriori dimanfaatkan pada MBA tersebut terhadap 100 data transaksi penjualan. Nilai *minimum support* dan *confidence* yang ditetapkan adalah sebesar 10% dan 30%. MBA tersebut menghasilkan keluaran 4 *rule* dengan nilai akurasi sebesar 70% dan nilai presisi sebesar 78%.

Salah satu manfaat MBA adalah sebagai acuan dalam pengaturan tata letak produk pada rak supermarket dan untuk pengaturan stok barang pada bagian *inventory*. Hal ini dilakukan dengan menganalisis data transaksi penjualan menggunakan algoritme Apriori (Sunardi, 2020). Penelitian ini menetapkan nilai *minimum support* sebesar 50% dan nilai *minimum confidence* sebesar 60% untuk menghasilkan aturan asosiasi produk. Berdasarkan *rule* yang didapatkan, pihak toko dapat mengatur produk mana saja yang harus dilakukan *restock* pada bagian *inventory* serta bagaimana susunan produk tersebut pada rak supermarket.

Teknik *Multilevel Association Rule* pernah dilakukan untuk proses MBA data pesanan di PT. Unirama (Ramadhani, 2020). Pada penelitian ini, MBA dimanfaatkan untuk menemukan itemset yang dipesan secara bersamaan pada data pemesanan barang untuk pelaksanaan strategi *cross selling*. Pemrosesan *data mining* menghasilkan *rules* yang memenuhi *minimum confidence* dan *minimum support* pada level 1 (Produk) 1 *rule*, pada level 2 (Jenis) 1 *rule*, pada level 3 (Tipe) 12

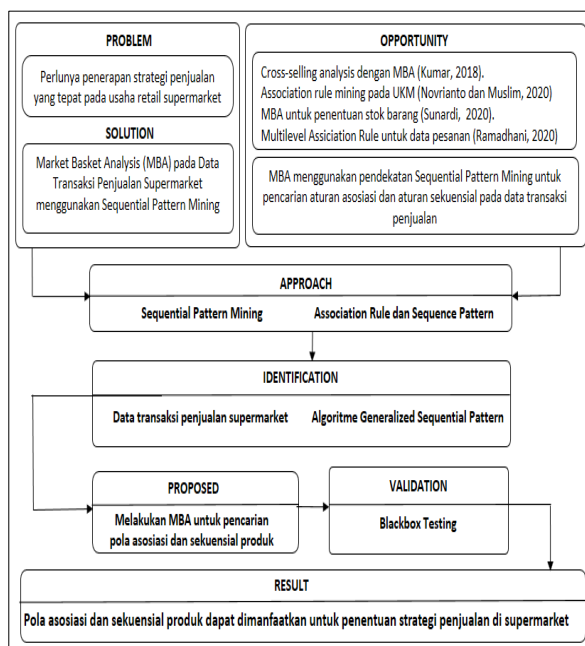
rule, dan pada level 4 (Kemasan) 12 rule. Rule tersebut dipakai sebagai bahan pertimbangan pembuatan strategi distribusi barang yang efektif.

Berdasarkan acuan penelitian terdahulu yang telah dijelaskan, pada penelitian ini akan dilakukan MBA pada dataset transaksi penjualan di West Store Supermarket menggunakan teknik sequential pattern mining, yaitu dengan algoritma Generalized Sequential Pattern (GSP). Algoritma GSP dapat digunakan untuk menghasilkan aturan asosiasi dan aturan sekuensial dari proses ekstraksi terhadap data produk yang dibeli pada data transaksi penjualan. Dengan adanya kedua aturan tersebut, diharapkan dapat menghasilkan strategi penjualan yang lebih baik dibandingkan hanya menggunakan aturan asosiasi saja. Proses pencarian aturan asosiasi dan *sequential* dengan algoritme GSP akan dilakukan menggunakan Python, di mana hasil analisisnya akan divalidasi menggunakan pengujian *Blackbox*.

## 2. Metode Penelitian

### 2.1 Kerangka Penelitian

Kerangka pemikiran dalam penelitian ini merupakan kerangka konsep pemecahan suatu masalah yang telah diidentifikasi atau dirumuskan, yang melingkupi Permasalahan (*Problem*), Kesempatan (*Opportunity*), Solusi (*Solution*), Pendekatan (*Approach*), Usulan (*Proposed*), Validasi (*Validation*) dan Hasil (*Result*). Kerangka pemikiran penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Kerangka penelitian

Kerangka penelitian ini ditunjukkan pada Gambar 1. Identifikasi masalah dilakukan pada West Store Supermarket, di mana difokuskan pada penentuan strategi penjualan yang lebih baik berdasarkan data penjualan yang dimiliki. Pendekatan MBA dengan teknik *Sequential Pattern Mining* akan dimanfaatkan untuk menghasilkan keluaran berupa pola asosiasi dan sekuensial produk. Pola tersebut dapat dimanfaatkan untuk penentuan strategi penjualan di supermarket.

### 2.2 Data Mining

*Data mining* merupakan pendekatan *machine learning* untuk menggali pengetahuan dari sekumpulan *dataset* menggunakan pendekatan kecerdasan buatan, statistika dan matematika. *Data Mining* biasanya dioperasikan terhadap *dataset* dalam jumlah besar yang tersimpan dalam basis data, gudang data atau media penyimpanan data lainnya (Turban, 2007). *Data mining* juga merupakan salah satu bagian dari konsep *Knowledge Discovery in Databases* (Han, 2000), melalui proses eksplorasi dan analisis secara otomatis maupun semi otomatis terhadap data dalam jumlah besar (Larose, 2005). Fungsi *datamining* dibagi berdasarkan dua jenis pembelajaran, yaitu *supervised* dan *unsupervised*, di mana kedua fungsi tersebut kemudian dibagi lagi menjadi lebih spesifik. Membagi fungsi *data mining* menjadi beberapa fungsi, yaitu fungsi *association*, *sequence*, *clustering*, *classification*, *regretion*, *forecasting* dan *solution* Haskett (2000).

### 2.3 Market Basket Analysis (MBA)

Istilah *Market Basket Analysis* (MBA) muncul dari kebiasaan konsumen yang memasukkan barang ke dalam keranjang belanja. Informasi mengenai produk-produk yang dibeli oleh para konsumen dalam satu waktu dapat memberikan informasi bagi pengelola toko atau swalayan untuk menambah keuntungan bisnisnya (Gunawan, 2015).

### 2.4 Sequential Pattern Mining

*Dataset* transaksi penjualan biasanya memuat informasi tentang ID transaksi, tanggal transaksi, produk yang dibeli beserta jumlah yang harus dibayar oleh konsumen. Namun, sering kali data transaksi juga berisi identitas dari pelanggannya (*Customer ID*). Dengan adanya informasi tentang ID konsumen, pihak toko dapat melakukan *Sequential Pattern Mining* terhadap dataset transaksi penjualan (Budhi, 2009). Selain

dapat menghasilkan pola asosiasi, teknik *Sequential Pattern Mining* juga dapat menghasilkan pola sekuensial produk yang dibeli oleh konsumen. Misalnya, jika konsumen membeli produk A pada transaksi pertama, maka di transaksi berikutnya akan membeli produk A dan produk B (Srikant, 1995). *Sequential Pattern Mining* adalah proses analisa suatu pola pembelian oleh konsumen untuk menemukan pola sering atau berkali-kali pada sebuah urutan database yang dipandang penting pada dalam *data mining* (Han, 2012).

## 2.5 Algoritme Generalized Sequential Pattern (GSP)

Algoritme GSP atau yang disebut dengan Apriori All adalah suatu algoritme untuk memproses dan menemukan pola sekuensial dan non sekuensial yang ada. Algoritme GSP digunakan untuk memecahkan masalah *mining sequence* yang didasarkan pada konsep algoritme Apriori. Selain untuk menemukan aturan asosiasi, algoritme GSP digunakan untuk menemukan pola sekuensial atau urutan (Zaki, 1997).

Tabel 1. *Sequence* dataset transaksi

ID Konsumen	Time	Items
1	10,15,20,25	<{cd}{abc}{abf}{acdf}>
2	15,20	<{abf}{e}>
3	10	<{abf}>
4	10,20,25	<{dgh}{bf}{agh}>

Tabel 1 merupakan *dataset* transaksi penjualan yang memuat data *sequence*. Misalnya ID konsumen 1 melakukan transaksi sebanyak 4 kali selama periode waktu tertentu. Pada transaksi pertama, produk yang dibeli adalah {cd}, transaksi kedua {abc} dan seterusnya. *Dataset* transaksi penjualan yang memuat data *sequence* tersebut dapat dicari pola sekuensialnya menggunakan algoritme GSP. Terdapat 2 langkah utama dalam algoritme ini, yaitu *candidate generation* dan *support counting* (Agrawal, 1996). Tahap *candidate generation* terdiri 2 langkah, yaitu *join phase* dan *prune phase*. Sedangkan *support counting* merupakan tahapan *counting candidates* atau menemukan semua kandidat *sequence*. Secara umum, algoritme GSP ditunjukkan pada Gambar 2.

```

F1 = the set of frequent 1-sequence
k=2,
do while Fk-1 != Null;
    Generate candidate sets Ck (set of candidate k-sequences);
    For all input sequences s in the database D
    do
        Increment count of all a in Ck if s supports a
    End do
    Fk = {a ∈ Ck such that its frequency exceeds the threshold}
    k = k+1;
End do
Result = Set of all frequent sequences is the union of all Fk's
    
```

Gambar 2. Algoritme generalized sequential pattern

Berdasarkan Gambar 2, maka tahapan pada algoritme GSP tersebut adalah 1) menentukan nilai *minimum support*, 2) untuk  $k = 1$ , tentukan semua *frequent sequence* untuk *frequent k-sequence*, 3) pilih *frequent sequence* yang memenuhi nilai *minimum support*, dan 4) ulangi langkah 2 dan 3 untuk  $k = k + 1$  sampai tidak ada *frequent sequence* atau kandidat yang ditemukan.

## 3. Hasil dan Pembahasan

Algoritme GSP pada penelitian ini diimplementasikan dalam bahasa pemrograman Python sesuai dengan aturan atau langkah-langkah yang ada pada algoritme GSP. Setelah kode program Python selesai dibuat, dilakukan pengujian *Blackbox* terlebih dahulu sebelum digunakan untuk menganalisis *dataset* West Store Supermarket. Pengujian ini dilakukan terhadap *dataset* sederhana supaya dapat dilihat alur perhitungan dan hasil akhirnya, sehingga dapat dicocokkan dengan keluaran dari Python. Adapun pengujian ini bertujuan untuk memastikan bahwa *rule* yang dihasilkan oleh Python sudah sesuai dengan perhitungan dari algoritme GSP. Tabel 1 merupakan *dataset* yang digunakan untuk pengujian *Blackbox*.

Tabel 2. *Datasetsequence* transaksi

SID	Sequence
1	({A,B}, {C}, {F,G}, {G}, {E})
2	({A,D}, {C}, {B}, {A,B,E,F})
3	({A}, {B}, {F,G}, {E})
4	({B}, {F,G})

*Datasetsequence* pada Tabel 2 terdiri dari 4 konsumen, di mana masing-masing konsumen memiliki beberapa sekuen atau urutan transaksinya. Misalnya untuk konsumen pertama, pada transaksi pertama membeli produk A dan B,

di transaksi kedua membeli produk C dan seterusnya. Dataset ini akan dicari pola sekuennya menggunakan algoritme GSP dengan nilai *minimum support* sebesar 2. Langkah pertama pada algoritme GSP adalah mencari *support count* atau frekuensi kemunculan dari masing-masing produk yang muncul di *dataset* seperti terlihat pada Tabel 3.

Tabel 3. Support count produk

Items	Frekuensi Kemunculan
{A}	3
{B}	4
{C}	2
{D}	1
{E}	3
{F}	4
{G}	3

Berdasarkan Tabel 3, item yang memenuhi nilai *minimum support* adalah A, B, C, E, F, G sehingga, A, B, C, E, F, G merupakan *frequent 1-Sequence* ( $L_1$ ) seperti ditunjukkan pada Tabel 4. Item yang tidak memenuhi nilai *minimum support* tidak dipakai pada tahap berikutnya.

Tabel 4. Frequent 1-sequence ( $L_1$ )

Items	Frekuensi Kemunculan
{A}	3
{B}	4
{C}	2
{E}	3
{F}	4
{G}	3

Langkah berikutnya adalah melakukan *Join Phase* dengan menggabungkan  $L_1$  dan  $L_1$  untuk membangkitkan *frequent sequence 2-itemset* ( $C_2$ ). Data  $C_2$  yang memenuhi nilai *minimum support* ditunjukkan pada Tabel 5. Sebagai catatan, {A, B} adalah sama dengan {B, A}, sedangkan {AB} tidak sama dengan {BA}. Artinya, {A, B} merupakan item-item yang dibeli sekaligus dalam satu kali transaksi, sedangkan {AB} merupakan item-item pada transaksi yang berbeda.

Tabel 5. Frequent 2-sequence ( $L_2$ )

Items	Frekuensi Kemunculan
{AB}	2
{AC}	2
{AE}	3
{AF}	2
{AG}	2
{(A,B)}	3

{BE}	3
{BF}	4
{BG}	3
{CE}	2
{CF}	2
{FE}	2
{(F,G)}	3
{GE}	2

Langkah berikutnya adalah melakukan *Join Phase* kembali yaitu kali ini dengan menggabungkan  $L_2$  dan  $L_2$  untuk membangkitkan *frequent sequence 3-itemset* ( $C_3$ ). Sebagai contoh, sequence {AB} dan {BE} dapat digabungkan menjadi {ABE}, {AB} dan {BF} dapat digabungkan menjadi {ABF}, sedangkan {AB} dan {AC} tidak dapat digabungkan. Hasil penggabungan  $L_2$  dengan  $L_2$  dan yang memenuhi nilai *minimum support* ditunjukkan pada Tabel 6.

Tabel 6. Frequent 3-sequence ( $L_3$ )

Items	Frekuensi Kemunculan
{ABE}	2
{ABF}	2
{ACE}	2
{ACF}	2
{AFE}	2
{A(F,G)}	2
{AGE}	2
{BFE}	2
{B(F,G)}	3
{BGE}	2
{(F,G)E}	2

Hasil pembangkitan sekuen pada Tabel 6 akan diproses lagi dengan cara menggabungkan  $L_3$  dengan  $L_3$  untuk mendapatkan  $L_4$ . Hasil penggabungan ini ditampilkan pada Tabel 7.

Tabel 7. Frequent 4-sequence ( $L_4$ )

Items	Frekuensi Kemunculan
{A(F,G)E}	2
{B(F,G)E}	2

Hasil pada  $L_4$  merupakan hasil akhir, di mana sekuen pada  $L_4$  tersebut sudah tidak bisa lagi untuk digabungkan, karena akan menghasilkan sekuen yang sama. Adapun hasil akhir dari pembangkitan *sequence* ini hanya ada 2 sequence yaitu {A (F, G) E} dan {B (F, G) E} dengan frekuensi kemunculan sebanyak 2 kali. Hasil lengkap untuk pembangkitan sekuen yang sudah dilakukan ditunjukkan pada Tabel 8.

Tabel 8. Hasil pembangkitan *frequent sequence* dari  $L_1$  hingga  $L_4$

$L_1$		$L_2$		$L_3$		$L_4$	
{A}	3	{AB}	2	{ABE}	2	{A(F,G)E}	2
{B}	4	{AC}	2	{ABF}	2	{B(F,G)E}	2
{C}	2	{AE}	3	{ACE}	2		
{E}	3	{AF}	2	{ACF}	2		
{F}	4	{AG}	2	{AFE}	2		
{G}	3	{(A,B)}	3	{A(F,G)}	2		
		{BE}	3	{AGE}	2		
		{BF}	4	{BFE}	2		
		{BG}	3	{B(F,G)}	3		
		{CE}	2	{BGE}	2		
		{CF}	2	{(F,G)E}	2		
		{FE}	2				
		{(F,G)}	3				
		{GE}	2				
		{AB}	2				

Setiap *sequence* yang diperoleh dapat dihitung nilai *support* dan *confidence*. Nilai *support* atau nilai penunjang menunjukkan persentase keberadaan *sequence* tersebut dalam *dataset*, sedangkan nilai *confidence* atau nilai kepastian akan menunjukkan seberapa kuat kombinasi item dalam *sequence* tersebut. *Sequence* pada  $L_4$  yaitu {A (F, G) E} memiliki nilai *supportcount* sejumlah 2 dari jumlah semua konsumen (yaitu 4 konsumen), sehingga nilai *support* adalah  $2 / 4 * 100\% = 50\%$  atau 0.500. Sedangkan nilai *confidence* untuk *sequence* tersebut adalah sebesar nilai *support count* dibagi jumlah konsumen yang membeli item pada bagian *antecedent* (A), sehingga nilai *confidence* adalah  $2 / 3 * 100\% = 66.67\%$  atau 0.667.

Hasil perhitungan yang sudah dilakukan sebelumnya kemudian akan dicocokkan dengan hasil yang dikeluarkan oleh kode program Python. Pada kode program tersebut, pengguna akan diminta untuk memasukkan nilai *minimum support* seperti terlihat pada Gambar 3.

OPSI

- MS dengan frekuensi
  - MS dengan nilai pecahan
- Pilihan : 1

Masukkan Minimal support =

Gambar 3. Inputan nilai *minimum support*

Setelah nilai *minimum support* dimasukkan, maka program Python akan melakukan pembangkitan *sequence pattern* menggunakan algoritme GSP. Hasil pola sekuen yang berhasil

dibangkitkan akan disimpan dalam *file .txt*. Tampilan hasil ini ditunjukkan pada Gambar 4.

```

26 No of length :3 Frequent sequences: 11
27 <{'A'}{'B'}{'E'}> count: 2 support: 0.5 confidence: 0.6666666666666666
28 <{'A'}{'B'}{'F'}> count: 2 support: 0.5 confidence: 0.6666666666666666
29 <{'A'}{'C'}{'E'}> count: 2 support: 0.5 confidence: 0.6666666666666666
30 <{'A'}{'C'}{'F'}> count: 2 support: 0.5 confidence: 0.6666666666666666
31 <{'A'}{'F'}{'E'}> count: 2 support: 0.5 confidence: 0.6666666666666666
32 <{'B'}{'F'}{'E'}> count: 2 support: 0.5 confidence: 0.6666666666666666
33 <{'A'}{'F', 'G'}> count: 2 support: 0.5 confidence: 0.6666666666666666
34 <{'B'}{'F', 'G'}> count: 3 support: 0.75 confidence: 0.75
35 <{'A'}{'G'}{'E'}> count: 2 support: 0.5 confidence: 0.6666666666666666
36 <{'B'}{'G'}{'E'}> count: 2 support: 0.5 confidence: 0.5
37 <{'F', 'G'}{'E'}> count: 2 support: 0.5 confidence: 0.5
38
39 No of length :4 Frequent sequences: 2
40 <{'A'}{'F', 'G'}{'E'}> count: 2 support: 0.5 confidence: 0.6666666666666666
41 <{'B'}{'F', 'G'}{'E'}> count: 2 support: 0.5 confidence: 0.5
42
43 No of length :5 Frequent sequences: 0
    
```

Gambar 4. Tampilan hasil pembangkitan *sequence pattern*

Gambar 4 merupakan hasil pembangkitan pola sekuen dari  $L_1$  sampai  $L_4$ . Pada gambar tersebut ditampilkan pola sekuen khusus pada  $L_3$  dan  $L_4$ . Berdasarkan keluaran pada kode program Python tersebut dapat disimpulkan bahwa perhitungan pada algoritme GSP telah berhasil diimplementasikan. Selanjutnya kode program Python ini akan digunakan untuk melakukan pembangkitan *sequence pattern* pada *dataset* West Superstore. *Dataset* ini ditunjukkan pada Gambar 5.

Row ID	Order ID	Order Date	Ship Date	Ship Mod	Customer ID	Customer Name	Segment	Country	City	State	Postal Co	Region	Product ID	Category	Sub-Category	Product Name
5328	CA-2011-130813	06/01/2013	08/01/2013	Standard	CI-LS-17230	Lycoris Saunders	Consumer	United States	Los Angeles	California	90049	West	OFF-PA-10	Office Sup	Paper	Merax 225
4938	CA-2011-157147	13/01/2013	18/01/2013	Standard	BD-11605	Brian Dahlen	Consumer	United States	San Francisco	California	94109	West	OFF-SF-100	Office Sup	Storage	Tenzo 6- and 18-Comp
4939	CA-2011-157147	13/01/2013	18/01/2013	Standard	BD-11605	Brian Dahlen	Consumer	United States	San Francisco	California	94109	West	FUR-BO-10	Furniture	Bookcases	O'Sullivan Elevations Bo
4940	CA-2011-157147	13/01/2013	18/01/2013	Standard	BD-11605	Brian Dahlen	Consumer	United States	San Francisco	California	94109	West	OFF-AR-10	Office Sup	Art	8009 Highlighters by San
5365	CA-2011-129477	10/01/2013	21/01/2013	Standard	CI-DHW-13295	David Wiener	Corporate	United States	Springfield	Oregon	97477	West	OFF-AP-10	Office Sup	Appliance	Fellowes Mighty 8 Comp
5463	CA-2011-146591	19/01/2013	20/01/2013	First Class	TS-21340	Toby Swindell	Consumer	United States	Scottsdale	Arizona	85254	West	OFF-BI-100	Office Sup	Binders	6BC Standard Recycled F
5464	CA-2011-146591	19/01/2013	20/01/2013	First Class	TS-21340	Toby Swindell	Consumer	United States	Scottsdale	Arizona	85254	West	OFF-PA-100	Office Sup	Paper	TOPS Carbonless Receipt
5465	CA-2011-146591	19/01/2013	20/01/2013	First Class	TS-21340	Toby Swindell	Consumer	United States	Scottsdale	Arizona	85254	West	OFF-BN-10	Office Sup	Envelope	Typelink Top-Opening Pee
5466	CA-2011-146591	19/01/2013	20/01/2013	First Class	TS-21340	Toby Swindell	Consumer	United States	Scottsdale	Arizona	85254	West	FUR-BO-10	Furniture	Bookcases	O'Sullivan 4-Shell Bookc
5737	CA-2011-148614	20/01/2013	25/01/2013	Standard	MMV-17465	Mark Van Huff	Consumer	United States	Los Angeles	California	90049	West	OFF-PA-10	Office Sup	Paper	Wirebound Service Call
5738	CA-2011-148614	20/01/2013	25/01/2013	Standard	MMV-17465	Mark Van Huff	Consumer	United States	Los Angeles	California	90049	West	FUR-FU-100	Furniture	Furniture	Eldon Expressions Desk.
9156	CA-2011-102645	23/01/2013	28/01/2013	Standard	MM-15055	Ionia McGrath	Consumer	United States	Las Vegas	Nevada	89115	West	OFF-PA-10	Office Sup	Paper	Merax 195

Gambar 5. Tampilan *dataset* West Superstore

*Dataset* WestSuperstore memiliki 1.598 *record* dan merupakan *sequence dataset*, sehingga dapat dianalisis menggunakan algoritme GSP. Proses pembangkitan *sequence pattern* membutuhkan waktu selama 666.47 detik, di mana menggunakan nilai *minimum support* sebesar 2. Hasil pembangkitan *sequence pattern* tersebut ditunjukkan pada Gambar 6.

```
389 <{'Home/Office Personal File Carts'}> count: 2 support: 0.003898635477582846 confidence: 1.0
390 <{'Wilson Jones Turn Tabs Binder Tool for Ring Binders'}> count: 2 support: 0.003898635477582846 confidence: 1.0
391 <{'Avaya 4621SM VoIP phone'}> count: 2 support: 0.003898635477582846 confidence: 1.0
392 <{'Master Caster Door Stop, Brown'}> count: 2 support: 0.003898635477582846 confidence: 1.0
393
394 No of length :2 Frequent sequences: 1
395 <{'Carina Double Wide Media Storage Towers in Natural & Black', 'GBC Wire Binding Combs'}> count: 2
396 support: 0.003898635477582846 confidence: 0.5
397
398 No of length :3 Frequent sequences: 0
```

Gambar 6. Hasil pembangkitan *sequencepattern*

Jumlah *sequence pattern* yang berhasil dibangkitkan adalah 391 aturan, di mana terdiri dari 390 aturan  $L_1$  dan 1 aturan  $L_2$ . Berdasarkan hasil tersebut, hanya ada 1 *subsequence* saja pada  $L_2$  dan merupakan *frequencepattern*, di mana konsumen memiliki kecenderungan untuk membeli produk *Carina Double Wide Media Storage Towers in Natural & Black* bersamaan dengan membeli produk *GBC Wire Binding Combs*. Hasil ini didukung dengan nilai *support* sebesar 0.00389 dan nilai *confidence* sebesar 0.5. Pihak supermarket dapat menggunakan aturan tersebut untuk strategi pemasaran, misalnya dengan memberikan potongan harga, membuat *bundling product* maupun sebagai acuan dalam penempatan produk pada halaman toko *online*.

#### 4. Kesimpulan

Berdasarkan hasil pembahasan dan pengujian yang telah dilakukan, dapat disimpulkan bahwa penelitian ini berhasil mengimplementasikan algoritme GSP dalam bahasa pemrograman Python. Hasil pengujian *Blackbox* menunjukkan bahwa hasil pembangkitan *sequence pattern* dari kode program Python sudah sesuai dengan perhitungan yang dilakukan dalam algoritme GSP.

Pada analisis terhadap *sequence dataset* West Superstore didapatkan hasil bahwa semakin banyak jumlah transaksi penjualan yang dianalisis, maka waktu komputasi pembangkitan *sequence pattern* juga semakin lama. Berdasarkan hasil pengujian, *dataset sequence* yang terdiri dari 23 *record* membutuhkan 0.01 detik untuk membangkitkan *sequence pattern*, sedangkan pada pengujian kedua di mana melibatkan 1.598 *record* data membutuhkan 666.47 detik untuk pembangkitan *sequence pattern*.

Algoritme GSP berhasil membangkitkan 2 jenis aturan, yaitu *frequentitem* dan *sequencepattern*. Pada percobaan yang sudah dilakukan pada *dataset West Superstore* hanya menghasilkan keluaran berupa *frequentitem* saja, yaitu dengan nilai *minimumsupport* sebesar 2. Aturan yang dihasilkan tersebut masih bisa untuk dimanfaatkan untuk penentuan strategi penjualan,

misalnya dengan memberikan potongan harga, penyusunan *bundling product* maupun sebagai acuan dalam *supermarket layouting*.

#### 5. Saran

Berikut ini beberapa saran penelitian lanjutan:

- Kualitaskomputasi pada algoritme GSP dapat diperbaiki melalui pengkodean yang lebih baik, sehingga menghasilkan waktu komputasi yang lebih efisien.
- Penelitian ini diharapkan menjadi acuan untuk pengembangan berikutnya, misalnya dengan memperhitungkan jumlah item yang dibeli untuk proses pembangkitan *rule*.

#### Referensi

- Agrawal, R. S. (1996). Mining Sequential Patterns: Generalization and Performance Improvements. In *5th International Conference Extending Database Technology (EDBT)* (pp. 3–17). Berlin, Heidelberg: Springer.  
<https://doi.org/https://doi.org/10.1007/BFb0014140>
- Efrain Turban, Jay E. Aronson, T.-P. L. (2007). *Decision Support Systems and Intelligent Systems (Sistem Pendukung Keputusan dan Sistem Cerdas) (Jilid 1) (Edisi 7)* (7th ed.). Yogyakarta: Andi.
- Eko Prasetyo. (2012). *Data Mining-Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: Andi.
- Gregorius Satia Budhi, Andreas Handojo, C. O. W. (2009). Algoritma Generalized Sequential Pattern untuk Menggali Data Sekuensial Sirkulasi Buku pada Perpustakaan UK Petra. In *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*.
- Gunawan Gunawan Gunawan, Alex Xandra Albert Sim, M Hawari Simanullang, M Firkhan Siregar, F. H. (2015). Pengembangan Aplikasi Market Basket Analysis Menggunakan Algoritma Generalized Sequential Pattern pada Supermarket. In *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)* (pp. 1–6).
- Haskett. (2000). An Introduction to Data Mining, Part 1: Understanding The Critical Data Relationship in The Corporate Data Warehouse. *Enterprise System Journal*, 15, 32–34.
- Jiawei Han, Jian Pei, dan Y. Y. (2000). Mining Frequent Patterns without Candidate

- Generation. In *SIGMOD '00 Proceedings of the 2000 ACM SIGMOD international conference on Management of data*. Dallas, Texas.
- Jiawei Han, Micheline Kamber, J. P. (2012). *Data Mining: Concepts and Techniques*. Waltham, USA: Morgan Kaufmann. Retrieved from <https://www.sciencedirect.com/book/9780123814791/data-mining-concepts-and-techniques>
- Larose, D. T. (2005). *Data Mining Methods and Models*. New Jersey: John Wiley & Sons, Inc. <https://doi.org/10.1002/0471756482>
- Manpreet Kaur, S. K. (2016). Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining. *Procedia Computer Science*, 85, 78–85. <https://doi.org/https://doi.org/10.1016/j.procs.2016.05.180>
- Nilam Ramadhani, Abd Wahab Syahroni, Arin Supikar, W. Z. (2020). Penerapan Market Basket Analysis Menggunakan Metode Multilevel Association Rules dan Algoritma ML\_T2L1 Pada Data Order PT. Unirama. *Jurnal Nasional Informatika Dan Teknologi Jaringan*, 4(2), 71–84. <https://doi.org/https://doi.org/10.30743/infotekjar.v4i2.2405>
- Rizky Bayu Novrianto, M. (2020). Penerapan Metode Market Basket Analysis pada Minimarket Toko Baru. *Indonesian Journal of Data and Science*, 1(1), 01–05. <https://doi.org/https://doi.org/10.33096/ijodas.v1i1.2>
- Srikant, R. A. R. (1995). Mining Sequential Patterns. In *Proceedings of the Eleventh International Conference on Data Engineering*. Taipei, Taiwan, Taiwan: IEEE. <https://doi.org/10.1109/ICDE.1995.380415>
- Srinivasa Kumar, R. Renganathan, C. VijayaBanu, dan I. R. (2018). Consumer Buying Pattern Analysis using Apriori Association Rule. *International Journal of Pure and Applied Mathematics*, 119(7), 2349–2350. Retrieved from <https://acadpubl.eu/jsi/2018-119-7/articles/7c/54.pdf>
- Tresna Yudha Prawira, Sunardi Sunardi, A. F. (2020). Market Basket Analysis To Identify Stock Handling Patterns & Item Arrangement Patterns Using Apriori Algorithms. *Jurnal Ilmu Komputer Dan Informatika*, 6(1), 33–41. <https://doi.org/https://doi.org/10.23917/khif.v6i1.8628>
- Zaki, M. (1997). *Fast Mining of Sequential Patterns in Very Large Databases*. New York.