

SEQUENTIAL PATTERN MINING DENGAN SPADE UNTUK PREDIKSI PEMBELIAN SPARE PART DAN AKSESORIS KOMPUTER PADA KEDATANGAN KEMBALI KONSUMEN

Riqky Juliastio dan Gunawan

Teknologi Informasi

Sekolah Tinggi Teknik Surabaya

rickynetonline@yahoo.com dan gunawan@stts.edu

ABSTRAK

Trend penjualan suatu produk atau barang pada Toko Seagate Computer yang dimaksud saat ini terus meningkat sesuai dengan kebutuhan masyarakat. Jumlah pengunjung memiliki pengaruh yang besar terhadap transaksi penjualan. Semakin banyak pengunjung, semakin besar kemungkinan transaksi yang dapat diramalkan. Salah satu pola yang dapat dicari adalah sequential pattern, dimana dicari pola urutan transaksi konsumen. Algoritma SPADE (Sequential PAttern Discovery using Equivalence classes) adalah sebuah algoritma baru untuk penemuan cepat pola sekuensial dalam sebuah database besar. Algoritma ini mencari frequent sequence dari data transaksi penjualan menggunakan vertical database dan proses join sequence. Hasil dari algoritma SPADE ini adalah frequent sequence yang kemudian digunakan untuk membentuk rule. Sehingga Rule yang telah dibentuk dapat digunakan sebagai prediksi kedatangan kembali konsumen melakukan transaksi. Rule ini diuji kekuatannya menggunakan nilai lift rasio.

Berdasarkan hasil percobaan, algoritma SPADE akurat dan dapat diterapkan pada prediksi kedatangan konsumen. Dengan menggunakan Metode SPADE dapat diketahui prediksi kedatangan konsumen yang akan datang diketahui dengan hasil akurasi 75%. Kekuatan rule yang dihasilkan terpengaruh pada input nilai minimum support dan jumlah data transaksi, sedangkan nilai minimum confidence tidak berpengaruh terhadap kekuatan rule.

Kata kunci: Data Mining, Sequential Pattern Mining, Algoritma SPADE, Data Transaksi

ABSTRACT

The trend of sale of products or items at Seagate Computer today continues to increase in accordance with people's needs. The great number of visitors influences the sales transaction. The more visitors are there, the more likely that the transaction can be predicted. One of the patterns that can be searched is the sequential pattern, which searches the pattern sequence of the consumer transactions. Algorithm SPADE (Sequential Pattern Discovery using Equivalence classes) is a new algorithm for rapid discovery of sequential patterns in a big database. This algorithm searches the frequent sequences of sales transaction data using vertical database and join sequence process. The results of the SPADE algorithm is frequent sequences which are then used to form rules. The rules that have been established can be used as a prediction of the coming consumer transactions. The rule strength is tested using lift ratio.

Based on the experimental results, the algorithm SPADE is proved to be accurate and can be applied to the next consumer coming pattern. By using a SPADE method, the next consumer's coming pattern is known with an accuracy of 75%. The power of the resulting rule is affected by the minimum value of the input support and the number of transaction data, while the minimum confidence value does not affect the strength of the rule.

Keywords: Data Mining, Sequential Pattern Mining, SPADE Algorithm, Transactions Data.

I. PENDAHULUAN

Trend penjualan suatu produk atau barang pada perusahaan-perusahaan yang dimaksud saat ini terus meningkat sesuai dengan kebutuhan masyarakat. Penjualan yang terus meningkat akan mempengaruhi pendapatan sebuah perusahaan, sehingga diperlukan strategi penjualan yang matang. Sistem penjualan pada umumnya, melihat besarnya omset atau pendapatan dari penjualan suatu barang. Pendapatan tersebut dapat dijadikan tolak ukur keberhasilan suatu perusahaan dengan melihat sebuah transaksi perharinya.

Jumlah pengunjung memiliki pengaruh yang besar terhadap transaksi penjualan. Semakin banyak pengunjung, semakin besar kemungkinan transaksi yang dapat diramalkan. Jumlah pengunjung setiap harinya berbeda-beda dan memiliki persentase yang tidak sama dalam melakukan transaksi penjualan. Salah satu cara dalam meningkatkan pendapatan penjualan adalah dengan memprediksi penjualan berdasarkan rata-rata jumlah pengunjung sehingga perencanaan strategi penjualan dapat tepat sasaran.

Pengolahan data secara cepat, efisien, dan efektif sangat diperlukan oleh manusia seiring dengan perkembangan zaman. Sedangkan di sisi lain, data mentah yang memerlukan pemrosesan jumlahnya sangat banyak sehingga tidak memungkinkan lagi dilakukan pengolahan data secara manual. Salah satu data mining yang digunakan untuk mengatasi time-series database adalah sequential pattern mining. Sequential pattern mining adalah pencarian frequent pattern dalam time-series database. Salah satu metode sequential pattern mining yaitu algoritma SPADE (*Sequential PAttern Discovery using Equivalence classes* = penemuan pola sekuensial menggunakan kelas yang ekuivalen) adalah sebuah algoritma baru untuk penemuan cepat pola sekuensial dalam sebuah database besar [7].

Toko Seagate Komputer adalah salah satu perusahaan/toko besar yang mempunyai database besar dan bergerak di bidang penjualan spare part dan asesoris komputer yang berlokasi di kabupaten Situbondo, dimana setiap harinya transaksi penjualan mencapai lebih dari ± 70 transaksi. Dapat dibayangkan jumlah pengunjung setiap harinya yang memenuhi Toko Seagate Komputer terbesar ini. Pada awal bulan dan akhir bulan rata-rata penjualan spare part dan asesoris komputer meningkat dikarenakan adanya peningkatan jumlah pengunjung. Pengambilan keputusan oleh manajemen mengenai mempersiapkan jenis spare part dan asesoris komputer serta pengaturan penjadwalan karyawan harus dilakukan jauh hari sebelumnya. Dengan menggunakan prediksi dapat ditentukan jenis spare part dan asesoris komputer apa yang dijual pada waktunya serta penjadwalan yang baik bagi karyawan. Terkait dengan hal ini, penulis

melakukan riset untuk memprediksi transaksi berdasarkan pola yang terjadi dari jumlah pengunjung di Toko Seagate Komputer Situbondo yang diharapkan dapat memberikan masukan baru mengenai masalah penjualan spare part dan asesoris komputer.

Dalam Sequential Pattern Mining, untuk menemukan pola agar mendapat informasi yang berguna, harus dicari frequent sequences atau urutan peristiwa tertentu yang sering muncul. Algoritma yang digunakan sebelumnya untuk mencari frequent sequence ini seperti Generalized Sequential Pattern (GSP) masih membutuhkan pencarian database lengkap untuk mencari frequent sequence yang semakin panjang, sehingga membutuhkan waktu dan proses yang lama [7].

Algoritma Sequential PAttern Discovery using Equivalent classes (SPADE) adalah salah satu algoritma dalam Sequential Pattern Mining yang dapat digunakan untuk mengatasi kekurangan algoritma sebelumnya, dimana pencarian database secara lengkap harus dilakukan berkali - kali. SPADE menggunakan id-list vertikal untuk memudahkan pencarian dalam database. SPADE dapat mencari frequent sequences dengan beberapa kali pencarian database saja. Berdasarkan hasil penelitian Kumar [6], SPADE menghasilkan performa yang jauh lebih baik dalam hal waktu komputasi dibandingkan algoritma-algoritma pencari frequent sequences sebelumnya seperti AprioriAll dan GSP. Sehingga diharapkan algoritma SPADE ini akan cocok diterapkan pada data transaksi penjualan, karena jumlah data transaksi penjualan yang besar. Pada penelitian ini akan mencoba menggunakan algoritma SPADE untuk mencari sequential pattern pada data transaksi penjualan.

II. METODOLOGI PENELITIAN

2.1 Data Mining

Pertumbuhan data pada transaksi yang pesat menyebabkan terciptanya kondisi yang bisa disebut sebagai kaya data tapi miskin informasi. Pada kondisi tersebut data – data hanya menjadi tumpukan data karena tidak dimanfaatkan. Untuk dapat memanfaatkan data–data tersebut dilakukan proses data mining. Data mining merupakan serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual [2].

2.2 Association Rule

Tugas utama dari Association rule adalah untuk mencari hubungan antar item. Himpunan item disebut sebagai itemset. Itemset yang mengandung k items merupakan k-itemset. Kecenderungan kemunculan itemset dalam sejumlah transaksi disebut frequency, support count atau count atau count dari itemset.

Support adalah ukuran yang menunjukkan besar tingkat dominasi suatu item atau itemset keseluruhan transaksi [3]. Support digunakan untuk rule yang memiliki nilai yang kurang dari threshold untuk menentukan support adalah:

$$\text{Support: } S(A \rightarrow B) = (\sigma(A \cup B)) / N \dots\dots\dots (1)$$

Dimana:

$\sigma(A \cup B)$ = Jumlah itemset di semua transaksi

N = Jumlah total transaksi

Association rule mining ini digunakan untuk menemukan semua aturan strong association rule yang memenuhi threshold minimum support (min_sup) dan threshold confidence minimum (min_conf).

Confidence adalah ukuran yang menunjukkan hubungan antar 2 item secara kondisional [4]. Nilai confidence keandalan dari rule yang dibuat. Rumus yang digunakan untuk menentukan confidence adalah:

$$\text{Confidence: } C(A \rightarrow B) = (\sigma(A \cup B)) / (\sigma(A)) \dots\dots\dots (2)$$

Dimana:

$\sigma(A \cup B)$ = Jumlah itemset di semua transaksi

$\sigma(A)$ = Jumlah antecedent pada transaksi

Terdapat 2 langkah proses untuk menentukan mining association rule, yaitu:

1. Menemukan frequent itemset. Berdasar definisi, masing-masing dari itemset akan muncul sedikitnya dengan frequency sebesar diberikan dalam minimum support count.
2. Munculkan strong association rule dari frequent itemset. Berdasar definisi, aturan ini harus memenuhi minimum support dan minimum confidence.

Pada prosesnya untuk mencari hubungan antar item tersebut akan dilakukan berulang - ulang untuk menentukan pola hubungan yang berbeda - beda. Proses yang berulang - ulang menyebabkan waktu yang dibutuhkan untuk menemukan hubungan antar item, sehingga dibutuhkan algoritma untuk menyelesaikannya.

2.3 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. 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. Penggalan pola sequential adalah pencarian semua subsequence berulang, yaitu subsequence yang frekuensi kejadiannya lebih besar dari minimum-support [1].

Untuk menyelesaikan permasalahan sekuensial ini dapat dilakukan dengan beberapa metode antara lain Generalizes Sequential Pattern atau disebut GSP, FreeSpan, dan PrefixSpan [5]. Sebagai contoh proses sequential pattern mining, terdapat tabel transaksi penjualan yang berisikan customer, tanggal dan item. Dari tabel transaksi tersebut kemudian dibentuk sequence transaksi berdasarkan customer dan diurutkan berdasar tanggal sehingga membentuk beberapa sequence [8].

2.4 Lift Rasio

Lift rasio adalah nilai minimum maka rule antecedent berpengaruh negatif pada rule consequent. Jika nilai lift rasio sama dengan 1 maka rule tersebut sering muncul bersamaan tetapi independen. Rule yang independen merupakan rule dimana untuk mendapatkan consequent tidak tergantung pada antecedent. Pada Lift rasio digunakan untuk mengukur seberapa kuat rule yang dibentuk dari algoritma sequential pattern mining. Nilai lift rasio berkisar antara 0 sampai dengan tak terhingga. Nilai minimum

dari lift rasio tidak ditentukan seperti halnya support atau confidence. Jika nilai lift rasio kurang dari 1 dalam hal lift rasio, rule yang direkomendasikan adalah jika lift rasio lebih dari 1 karena antecedent memiliki pengaruh positif pada consequent. Berikut rumus untuk menentukan lift rasio [3]:

$$\text{Expected Confidence: } EC(A \rightarrow B) = (\sigma(B)) / m \dots\dots\dots (3)$$

Dimana:

Lift = Confidence/(Expected Confidence)

$\sigma(B)$ = Jumlah consequent dalam transaksi

m = Jumlah transaksi

2.5 Algoritma SPADE (Sequential Pattern Discovery using Equivalent classes)

Sequential Pattern Discovery using Equivalent classes adalah salah satu algoritma sequential pattern mining yang menggunakan format data vertikal pada database sequence. Dalam format data vertikal, database sequence menjadi berbentuk kumpulan urutan yang formatnya [itemset :(sequence_ID, eventID)]. Dengan kata lain, untuk setiap itemset akan disimpan sequence identifier dan event identifier yang berkoresponden. Event identifier berguna sebagai timestamp atau penanda waktu dari itemset tersebut. Sepasang (sequence_ID, eventID) untuk setiap itemset membentuk ID_list dari itemset tersebut.

Langkah-langkah algoritma SPADE dalam mencari frequent sequence kemudian menentukan rule dari frequent sequence tersebut adalah sebagai berikut [7]:

1. Menghitung frequent 1-sequence

Untuk mencari frequent 1-sequence dari sequence database yang harus dilakukan adalah dilakukan scan untuk setiap itemset dalam sequence database. Untuk masing-masing itemset, simpan id-listnya (pasangan sid dan eid). Kemudian scan id-list dari masing-masing id-list tersebut, setiap ditemui sid yang sebelumnya belum ada maka nilai supportnya ditambah. Sequence yang dimasukkan dalam frequent 1-sequence adalah yang supportnya lebih dari min_sup.

2. Menghitung frequent 2-sequence

Dalam mencari frequent 2-sequence, data yang digunakan adalah data dari frequent 1-sequence, sehingga tidak perlu mencari dari sequence database lagi. Untuk setiap masing-masing frequent 1-sequence, gabungkan dengan semua frequent 1-sequence lainnya. Contohnya jika 1-sequence A digabungkan dengan 1-sequence B maka kemungkinan 2 sequence yang terjadi adalah A,B dimana A dan B muncul bersamaan dalam transaksi, $A \rightarrow B$ dimana item B muncul setelah item A, dan $B \rightarrow A$ dimana item B muncul setelah item A. Untuk setiap masing-masing penggabungan frequent 1-sequence ini dilakukan pengecekan apakah dalam id-listnya memiliki sid yang sama, jika sama kemudian dilakukan pengecekan apakah eid dari 1-sequence A sama dengan, kurang dari atau lebih dari eid 1-sequence B. Apabila sama maka id-listnya dimasukkan dalam 2-sequence A,B. Jika eid B lebih besar dari A maka id-listnya dimasukkan dalam 2-sequence $A \rightarrow B$ dan jika eid A lebih besar dari B maka id-listnya dimasukkan dalam 2-sequence $B \rightarrow A$. Kemudian seperti dalam frequent 1-sequence, tambahkan supportnya untuk setiap masing – masing sid yang sebelumnya belum ditemui. Dari 2-sequence itu kemudian dilakukan pengecekan apakah supportnya lebih dari min_sup. Jika memenuhi syarat maka dimasukkan dalam frequent 2-sequence.

3. Menentukan frequent k-sequence.

Setelah mencari frequent 2-sequence, untuk mencari frequent sequence–frequent sequence berikutnya dilakukan proses yang sama, yaitu mencari frequent k-sequence. Untuk mencari frequent k-sequence ini dilakukan join pada frequent (k-1) sequence yang memiliki prefix yang sama. Contohnya untuk mencari 3-sequence, gabungkan frequent sequence dari 2-sequence yang memiliki prefix yang sama, untuk mencari 4-sequence, gabungkan frequent sequence dari 3-sequence yang memiliki prefix yang sama, dan seterusnya. Untuk mencari prefix frequent (k-1) sequence, hilangkan item terakhir dari sequence tersebut. Contoh jika terdapat 4-sequence $A \rightarrow B \rightarrow C \rightarrow D$, maka prefixnya adalah $A \rightarrow B \rightarrow C$. Untuk setiap penggabungan ini ada 3 kemungkinan hasil:

- a. Jika A,B digabungkan dengan A,C, maka kemungkinan hasilnya hanya A, B, C.
- b. Jika A,B digabungkan dengan $A \rightarrow C$, maka kemungkinan hasilnya hanya $A, B \rightarrow C$.
- c. Jika $A \rightarrow B$ digabungkan dengan $A \rightarrow C$, maka ada 3 kemungkinan hasil: $A \rightarrow B, C$, dan $A \rightarrow B \rightarrow C$ dan $A \rightarrow C \rightarrow B$.

Dari setiap kemungkinan ini, periksa supportnya apakah memenuhi min_sup , jika ya maka sequence itu termasuk dalam frequent k-sequence. Pencarian frequent sequence dihentikan apabila tidak ada frequent (k-1) sequence yang bisa dijoin atau sudah tidak ditemukan frequent k-sequence lagi.

4. Pembentukan Rule

Setelah ditemukan semua frequent sequence, ditentukan rule dari sequence – sequence tersebut. 1-sequence tidak digunakan untuk membentuk rule karena hanya terdiri dari 1 item. Untuk 2-sequence yang menjadi antecedent adalah item pertama dan consequentnya adalah item keduanya. Contoh untuk sequence $A \rightarrow B$ maka rule yang dibentuk adalah $A \Rightarrow B$. Sedangkan untuk sequence yang panjangnya lebih dari 2 atau k- sequence, yang dijadikan consequent adalah item terakhir, sedangkan antecedentnya adalah semua item sebelum item terakhir. Contohnya pada 4-sequence $A \rightarrow B \rightarrow C \rightarrow D$, maka rule yang dihasilkan adalah $A \rightarrow B \rightarrow C \Rightarrow D$. Untuk masing–masing rule dihitung nilai confidencenya menggunakan rumus 2.2. Jika rule tersebut memenuhi batas min_conf , maka rule itu diterima. Kemudian dari rule yang diterima tersebut, hitung nilai lift rasionya menggunakan rumus 2.4. Nilai lift rasio semakin besar semakin baik, dengan batas 1. Apabila rule memiliki nilai lift lebih dari atau sama dengan 1, maka dalam rule itu antecedent memiliki pengaruh positif terhadap consequent. Sehingga rule dinyatakan baik, sebaliknya apabila nilai lift kurang dari 1 maka rule dianggap kurang baik.

III. UJI COBA

Pengujian dilakukan di Toko Seagate Computer Situbondo yang menyediakan spare part dan aksesoris komputer. Pada penelitian ini data transaksi penjualan spare part dan aksesoris komputer di toko Seagate Computer Situbondo ini menggunakan minimum support 2 untuk mengetahui kedatangan kembali konsumen yang akan datang dengan melakukan pembelian sesuai jenis kategori barang dengan data training mulai pada tanggal 5 April 2011 sampai dengan tanggal 5 Mei 2011 (2 bulan) dengan rincian data transaksi penjualan spare part dan aksesoris komputer di toko Seagate Computer

Situbondo yang akan ditraining 452 customer dan 1212 transaksi. Sedangkan data transaksi yang akan digunakan untuk testing pada data transaksi penjualan spare part dan aksesoris komputer di toko Seagate Computer Situbondo yaitu tanggal 1 Juni 2011 sampai dengan tanggal 23 Februari 2013) sehingga dapat menghasilkan dan membuktikan prediksi kedatangan kembali konsumen dengan melakukan transaksi berdasarkan jenis kategori barang.

Tabel 1. Data Transaksi untuk Training

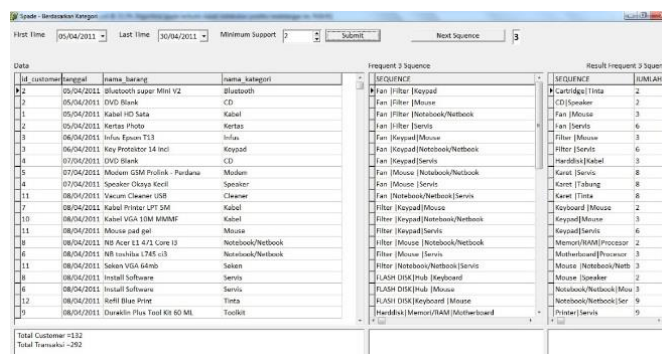
Id Customer (SID)	Tanggal (EID)	Kategori Barang
1	2011-04-05	Kabel
2	2011-04-05	Bluetooth
2	2011-04-05	CD
2	2011-04-05	Kertas
3	2011-04-06	Infus
3	2011-04-06	Keypad

3.1 Menghitung Frequent 1-Sequence Data Transaksi Penjualan

Pada data transaksi penjualan spare part dan aksesoris komputer toko Seagate Computer Situbondo adalah berformat biasa, sehingga harus dirubah terlebih dahulu menjadi format vertical id-list database sesuai dengan penerapan algoritma SPADE (*Sequential Pattern Discovery using Equivalence classes*). Pada vertical id-list database jenis kategori barang menjadi acuan dari id customer (SID) dan tanggal (EID) pada data transaksi penjualan spare part dan aksesoris komputer toko Seagate Computer Situbondo.

Tabel 2. Vertical Id-List Data Transaksi Penjualan

id_customer	tanggal1	nama_kategori	sequence
443	24-05-2011	Access Point	Access Point
274	26-05-2011	Access Point	Access Point
98	20-04-2011	Battery	Battery
397	20-05-2011	Battery	Battery
473	27-05-2011	Battery	Battery
492	28-05-2011	Battery	Battery



id_customer	tanggal	nama_barang	nama_kategori
1	05/04/2011	Bluetooth paper Mini V2	Bluetooth
2	05/04/2011	DVD Blank	CD
1	05/04/2011	Kabel HP Sata	Kabel
2	05/04/2011	Kertas Photo	Kertas
3	06/04/2011	Infus Epson T13	Infus
3	06/04/2011	Key Prohibitor 14 InC	Keypad
4	07/04/2011	DVD Blank	CD
5	07/04/2011	Modem GSM ProLink - Panda	Modem
4	07/04/2011	Speaker Okya Kecil	Speaker
11	08/04/2011	Vacuum Cleaner USB	Cleaner
7	08/04/2011	Kabel Printer LPT SM	Kabel
10	08/04/2011	Kabel VGA 10M MAME	Kabel
11	08/04/2011	Mouse pad gel	Mouse
10	08/04/2011	HP Acer E2 472 Core I3	Netbook/Netbook
6	08/04/2011	NB Toshiba 1745 i13	Netbook/Netbook
11	08/04/2011	Sekam VGA 64mb	Sekam
6	08/04/2011	Install Software	Service
6	08/04/2011	Install Software	Service
12	08/04/2011	Reali Blue Print	Tinta
9	08/04/2011	Charwin Plus Tost K1 80 ML	Tooklit

SEQUENCE	Jumlah
Filter Filter Keypad	2
Filter Filter Mouse	2
Filter Filter Netbook/Netbook	3
Filter Filter Service	6
Filter Keypad Mouse	3
Filter Keypad Netbook/Netbook	6
Filter Keypad Service	3
Filter Mouse Netbook/Netbook	8
Filter Mouse Service	8
Filter Netbook/Netbook Service	8
Filter Keypad Mouse	2
Filter Keypad Netbook/Netbook	3
Filter Keypad Service	6
Filter Mouse Netbook/Netbook	2
Filter Mouse Service	3
Filter Netbook/Netbook Service	3
FLASH DISK Hub Keyboard	2
FLASH DISK Hub Mouse	3
FLASH DISK Keyboard Mouse	9
Harddisk Memori RAM Motherboard	9
Printer Service	9

Gambar 1. Hasil Program frequent 1-sequence

3.2 Menghitung frequent 2-sequence data transaksi penjualan

Dari table vertical-to-horizontal data transaksi penjualan spare part dan aksesories komputer pada Toko Seagate Computer Situbondo, frequent 2-sequence dapat dibentuk dengan langkah-langkah sebagai berikut :

1. Dibentuk support array untuk frequent 2-sequence pada data transaksi penjualan spare part dan aksesories komputer pada Toko Seagate Computer Situbondo dengan panjang $n \times n$ untuk setiap sequence dengan bentuk dengan salah satu array kategori barang misalnya CD -> Flashdisk dan $[n \times (n-1)]/2$ untuk setiap sequence dengan bentuk CD Flashdisk. Array diindex berdasarkan frequent 1-sequence, n adalah jumlah frequent 1-sequence.
2. Pembentukan frequent 2-sequence dilakukan pada masing-masing baris id customer (SID) dari tabel vertical-to-horizontal database recovery yaitu dengan join antar pasangan item dan tanggal (EID) yang ada dalam satu baris dengan id customer (SID) tersebut. Misal kategori barang yang dijoin adalah CD dan Flashdisk, maka hasil join ada tiga kemungkinan yaitu CD -> Flashdisk, CD Flashdisk, dan Flashdisk -> CD tergantung dari tanggal (EID) jenis kategori barang CD dan Flashdisk. Kemudian dilakukan update pada support array yang telah disiapkan pada langkah pertama.

Dari langkah kedua, dilakukan pengecekan pada nilai support (dalam hal ini penulis menggunakan nilai support 2). Untuk supportnya lebih besar dari minimum support maka sequence tersebut merupakan frequent 2-sequence.

Tabel 3. Frequent 2-Sequence

id_customer	tanggal1	tanggal2	sequence
184	04-05-2011	18-05-2011	Cartridge -> Printer
278	09-05-2011	30-05-2011	Cartridge -> Printer
66	04-05-2011	28-05-2011	FLASH DISK -> Mouse
274	09-05-2011	26-05-2011	FLASH DISK -> Mouse

tanggal	id_customer	nama_kategori	SEQUENCE	SEQUENCE	Result Frequent 2 Sequence
05/04/2011	2	Blueooth	Cartridge&FLASH DISK	Cartridge&Tinta	Cartridge&Tinta
05/04/2011	2	CD	Cartridge&Memori/DAM	Cartridge&Tinta	CD&Speaker
05/04/2011	1	Kabel	Cartridge&Servis	CD&Speaker	Fan & Mouse
05/04/2011	2	Kertas	Cartridge&Tinta	CD&Speaker	Fan & Servis
04/04/2011	3	Infes	Cartridge&Tinta	Fan & Mouse	Filter & Mouse
05/04/2011	3	Keyboard	CD&Speaker	Fan & Mouse	Filter & Servis
07/04/2011	4	CD	CD&Speaker	Fan & Mouse	Hardisk&Kabel
07/04/2011	5	Modem	Fan & Filter	Fan & Servis	Karet & Servis
07/04/2011	4	Speaker	Fan & Filter	Fan & Servis	Karet & Tabung
08/04/2011	11	Cleaner	Fan & Filter	Fan & Servis	Karet & Tinta
08/04/2011	7	Kabel	Fan & Filter	Fan & Servis	Keyboard & Mouse
08/04/2011	10	Kabel	Fan & Filter	Fan & Servis	Keyboard&Mouse
08/04/2011	11	Mouse	Fan & Filter	Fan & Servis	Keyboard&Servis
08/04/2011	8	Netbook/Netbook	Fan & Filter	Filter & Mouse	Memori/InRAM/Processor
08/04/2011	6	Netbook/Netbook	Fan & Filter	Filter & Mouse	Motherboard&Processor
08/04/2011	11	Sekens	Fan & Filter	Filter & Mouse	Mouse & Netbook/Netbook
08/04/2011	8	Servis	Fan & Filter	Filter & Servis	Mouse & Speaker
08/04/2011	6	Servis	Fan & Keyboard	Filter & Servis	Netbook/Netbook&Servis
08/04/2011	12	Tinta	Fan & Keyboard	Filter & Servis	Printer&Servis
08/04/2011	9	Teblak	Fan & Keyboard	Filter & Servis	Printer&Tabung

Total Customer = 132
Total Transaksi = 292
Jumlah Sequence 2 = 236
Jumlah Rule Sequence 2 = 124
Jumlah Rule Sequence 2 = 23

Gambar 2. Hasil Program frequent 2-sequence

3.3 Menghitung frequent 3-sequence data transaksi penjualan

Tahap selanjutnya dari algoritma SPADE adalah penggabungan hasil frequent 1-sequence dan frequent 2-sequence (join) dari data transaksi penjualan spare part komputer di toko Seagate Computer Situbondo dengan ulasan sebagai berikut, dari frequent 1-sequence dan frequent 2-sequence yang ditemukan, dan dikelompokkan

berdasarkan classnya serta enumerate frequent sequence yang dilakukan pada tiap parent class. Ada 2 cara untuk menghasilkan frequent 3-sequence yaitu dengan menggunakan kombinasi array frequent 3-sequence dan menggunakan join hasil frequent 1-sequence dengan hasil frequent 2-sequence.

Tabel 4. Frequent 3-sequence

Id	Tanggal1	Tanggal2	Tanggal3	Sequence
11	20-05-2011	20-05-2011	21-05-2011	Karet & Printer -> Servis
281	20-05-2011	20-05-2011	21-05-2011	Karet & Printer -> Servis
11	20-05-2011	20-05-2011	21-05-2011	Karet & Servis -> Servis
281	20-05-2011	20-05-2011	21-05-2011	Karet & Servis -> Servis
278	09-05-2011	09-05-2011	30-05-2011	Karet & Servis -> Servis

Gambar 3. Hasil Program frequent 3-sequence

3.4 Frequent 4-Sequence (Equivalen Class)

Tahap selanjutnya dari algoritma SPADE adalah equivalen class. Dari frequent 1-sequence, frequent 2-sequence dan frequent 3-sequence data transaksi spare part dan aksesoris komputer di toko Seagate Computer Situbondo yang ditemukan, dan dikelompokkan berdasarkan classnya serta enumerate frequent sequence yang dilakukan pada tiap parent class.

Tabel 5. Frequent 4-sequence

Id	Tanggal1	Tanggal2	Tanggal3	Tanggal4	Sequence
11	20-05-2011	20-05-2011	20-05-2011	21-05-2011	Karet & Printer & Karet -> Servis
281	20-05-2011	20-05-2011	20-05-2011	21-05-2011	Karet & Printer & Karet -> Servis
11	20-05-2011	20-05-2011	20-05-2011	21-05-2011	Karet & Printer & Servis -> Servis
281	20-05-2011	20-05-2011	20-05-2011	21-05-2011	Karet & Printer & Servis -> Servis
11	20-05-2011	20-05-2011	21-05-2011	21-05-2011	Karet & Printer -> Servis -> Servis

IV. HASIL DAN PEMBAHASAN

Berdasarkan hasil id-list yang terdapat pada tabel 5 akan datang kembali dengan barang yang sama sesuai dengan urutan item yaitu id konsumen 11 dan 281 akan diprediksikan yang akan datang akan membeli jenis kategori barang Karet dan Printer dan Karet selanjutnya yang akan datang akan melakukan Servis.

V. AKURASI

Dari tabel 6 membuktikan bahwa algoritma SPADE (*Sequential PAttern Discovery using Equivalence classes*) dapat memprediksi data transaksi penjualan spare part dan aksesories komputer di toko Seagate Computer dengan hasil data training 2 bulan.

Tabel 6. Akurasi

Id_customer	Sequence	Testing				Keterangan
		tanggal1	tanggal2	tanggal3	tanggal4	
1	Karet&Printer&Karet - > Servis					Tidak Terbukti
78	Karet&Servis&Karet - > Servis	11-11-2011				Terbukti
81	Karet&Printer Servis -> Servis	12-08-2011	27-11-2011	14-02-2012		Terbukti
81	Servis -> Tinta -> Karet -> Servis	27-12-2011	24-01-2012	14-02-2012	19-07-2012	Terbukti
59	Servis Karet&Printer&Printer	02-12-2011	14-03-2012			Terbukti

Pada tabel 6 ada 4 customer yang memenuhi support pada hasil frequent 4-sequence transaksi selama 2 bulan terdiri dari prediksi 3 customer terbukti yang akan datang akan melakukan transaksi pembelian kembali dengan jenis kategori barang yang diprediksi dan prediksi 1 customer tidak terbukti melakukan pembelian kembali dengan jenis kategori barang yang diprediksi. Akurasi prediksi pada transaksi selama 2 bulan yaitu:

$$\text{Nilai Akurasi} = \frac{3}{4} \times 100\% = 75\% \dots\dots\dots (4)$$

Dari nilai akurasi membuktikan bahwa algoritma SPADE dapat memprediksi data transaksi penjualan spare part dan aksesories komputer di toko Seagate Computer yaitu dengan akurasi hingga mencapai 75%.

V. PENUTUP

6.1 Kesimpulan

Kesimpulan dari penelitian ini adalah sebagai berikut:

1. Hasil frequent 4-sequence yang diperoleh dalam penelitian ini mendapatkan hasil yang baik dan dapat dijadikan sebuah prediksi pembelian spare part dan aksesoris komputer pada kedatangan kembali konsumen di toko Seagate Computer Situbondo. Uji coba yang pertama dilakukannya dengan menggunakan data transaksi 2 bulan dengan mendapatkan hasil nilai akurasi sebesar 75%.
2. Uji coba kedua dilakukan dengan menggunakan data transaksi 3 bulan dengan mendapatkan hasil nilai akurasi sebesar 75,76%.
3. Dari data transaksi yang digunakan, dicari frequent sequence (pola urutan pembelian barang yang sering muncul) yang dapat dibentuk. Hasil frequent sequence menunjukkan pola belanja konsumen yang sering muncul. Pola belanja konsumen tersebut kemudian dibentuk rule yang diuji menggunakan nilai lift ratio, apabila nilai lift ratio lebih dari 1, maka rule dianggap kuat.
4. Berdasarkan uji coba yang dilakukan menggunakan parameter nilai minimum support, minimum confidence dan jumlah data, kekuatan rule yang dihasilkan terpengaruh pada input nilai minimum support dan jumlah data transaksi, sedangkan nilai minimum confidence tidak berpengaruh terhadap kekuatan rule. Sehingga disimpulkan bahwa algoritma SPADE akurat untuk mencari pola sekuensial dari data transaksi penjualan untuk memprediksi kedatangan kembali konsumen.

6.2 Saran

Melihat kinerja dari sistem prediksi yang dibangun dan implementasi terhadap data transaksi penjualan spare part dan aksesoris komputer di Seagate komputer, saran-saran yang bisa disampaikan adalah sebagai berikut:

1. Sistem prediksi yang dibangun penulis masih dapat dikembangkan lebih lanjut dengan menambahkan jumlah transaksi pembelian yang dilakukan oleh customer untuk mendapatkan hasil akurasi yang lebih akurat.
2. Algoritma SPADE tidak hanya dapat memprediksi kedatangan kembali konsumen tetapi dapat dikembangkan dengan memprediksi hal lain.
3. Aplikasi yang dibangun penulis bekerja relatif lambat sehingga masih dapat dioptimisasi lagi agar proses mining dapat berjalan lebih singkat.

Toko Seagate Computer Situbondo dapat menggunakan sistem prediksi dalam melakukan pengawasan terhadap penjualan barang yang akan datang dengan menambahkan sistem prediksi tersebut dalam sistem penjualan.

VI. DAFTAR PUSTAKA

- [1] Agrawal, R. dan Srikant, R. *Mining sequential patterns*. In 11th Intl. Conf. On Data Engineering, Taiwan. 1995.
- [2] F. Afrati, A. Gionis, and H. Mannila. *Approximating a collection of frequent sets*. Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and data Mining, pages 12-19. 2004.
- [3] Fomby, Tom. *Association Rules (Aka Affinity Analysis or Market Basket Analysis)*. Departemen of Economics Southern Methodist University Dallas. Texas. 2011.

- [4] Han, J. dan Kamber, M. *Data Mining Concepts and Techniques*. Morgan Kaufmann Publishers. 2006.
- [5] Jian Pei et Al., *Prefixspan: Mining sequential patterns efficiently by prefix projection pattern growth*.
- [6] Kumar, M. *Sequential Pattern Mining With Multiple Minimum Support by MS-SPADE*. International Journal of Computer Sciences, Vol.9, Issue 5, No.1. 2012.
- [7] M. J. Zaki. *SPADE: An efficient algorithm for mining frequent sequences*. Machine Learning Journal, 42(1/2):31–60, Jan/Feb 2001. Special issue on Unsupervised Learning (D. Fisher, editor.).
- [8] R. Srikant dan R. Agrawal. *Mining sequential patterns: Generalizations and performance improvements*. 1996.