

PENDEKATAN ALGORITMA APRIORI PADA DATA MINING UNTUK MENEMUKAN POLA BELANJA KONSUMEN

Wahju Tjahjo Saputro¹⁾, Ike Yunia Pasa²⁾

¹⁾Teknologi Informasi, Universitas Muhammadiyah Purworejo

¹⁾wahjusaputro@umpwr.ac.id, ²⁾ikeypasa@umpwr.ac.id

Abstrak

Data mining sering digunakan dalam penelitian terkait pola dan pengetahuan tentang sebuah informasi yang tersimpan dalam *database* skala besar. Saat ini banyak perusahaan atau badan usaha yang memiliki data dalam jumlah jutaan yang tersimpan dalam database. Database skala besar tersebut hanya digunakan untuk menghasilkan informasi yang bersifat tabular untuk kebutuhan rutinitas manajer. Sehingga dapat dikatakan kaya data namun miskin informasi.

Dalam data mining terdapat salah satu metode asosiasi yang mampu menghasilkan pola dan pengetahuan tertentu terhadap data yang memiliki asosiasi antara dua *itemset*, Sehingga memiliki sifat *if-then*. Algoritma yang digunakan untuk menghasilkan *association rule* yaitu apriori.

Hasil penelitian ini menunjukkan bahwa algoritma apriori dapat berjalan dengan baik untuk menghasilkan pola terhadap data transaksi penjualan. Dari hasil pengujian muncul frekuensi transaksi $\{2,3\} \Rightarrow \{28\}$ memiliki *support* 10.5% dan *confidence* 66.6% serta $\{7,8\} \Rightarrow \{22\}$ *support* 10.5% dan *confidence* 66.6%. kedua *rule* tersebut mempunyai frekuensi sering muncul cukup tinggi dengan $\Phi \geq 2$.

Kata kunci: Data Mining, Apriori, *Association Rule*, *Database* Penjualan

Abstract

Data mining is often used in research-related pattern and knowledge of an information is stored in large-scale databases. Today many companies have large amounts of data was stored in the database. The large-scale databases are only used to generate tabular information fo the needs daily of managers. So it can be called rich data but poor information.

Data mining has one association method that can generate certain patterns and knowledge of data have an associate between two itemsets, so it has an if-then property. Algorithm used to produce association rule is called apriori.

The result of research represent that the apriori algorithm can work optimally to generate pattern the sales transaction. This research represents transaction frequency $\{2,3\} \Rightarrow \{28\}$ has support 10.5%, confidence 66.6% and $\{7,8\} \Rightarrow \{22\}$ has support 10.5%, confidence 66.6%. Both of rules have frequencies often represent quite high with $\Phi \geq 2$.

Keywords: data mining, apriori, association rule, sales database

1. PENDAHULUAN

Latar Belakang Masalah

Alasan pokok mengapa data mining menarik perhatian banyak pihak yaitu adanya ketersediaan data dalam jumlah besar dan kebutuhan untuk mengolah data tersebut menjadi informasi yang lebih berguna. Informasi dan pengetahuan yang diperoleh dapat digunakan

dalam berbagai bidang mulai dari manajemen bisnis, kontrol produksi sampai analisa pasar. Untuk memperoleh pengetahuan dalam database skala besar tersebut diperlukan teknik data mining (Han dan Kamber, 2001). Sedangkan menurut Erwin (2009) salah satu cara yang bisa dilakukan untuk mengetahui kondisi pasar dan konsumen adalah mengamati data transaksi penjualan yang tersimpan dalam database

berskala besar. Sebagai contoh salah satu kemampuan yang dapat diperlihatkan oleh data mining yaitu mampu menebak target pasar dengan cara melakukan pengelompokan dari model pembeli dan melakukan klasifikasi setiap konsumen dan kebiasaan konsumen dalam berbelanja dari aspek karakteristik, pendapatan, pekerjaan, usia maupun jenis kelamin (Kusnawi, 2007).

Selain itu pengetahuan mengenai asosiasi antar jenis item yang muncul bersamaan pada setiap transaksi dapat menjadi masukan penting bagi manajer guna melakukan usaha peningkatan penjualan (Yusuf dkk, 2006). Misal dalam melakukan kegiatan promosi penjualan barang dapat pula membeli barang lain dengan harga yang lebih murah. Sehingga menurut Yusuf dkk (2006) pengetahuan antar jenis item yang saling berasosiasi ini dapat memberikan pengetahuan bagi manajer tentang kombinasi item barang yang harus dijual dalam promosi tersebut. Dampak dari promosi ini tentunya juga menambah kenyamanan konsumen dalam melakukan transaksi item barang yang dibeli secara bersamaan.

Pendekatan data mining mampu memperlihatkan kemampuan dalam menemukan pengetahuan dan pola item yang sering muncul dalam database skala besar salah satunya menggunakan Algoritma Apriori. Dimana Algoritma Apriori mampu memperlihatkan dalam menemukan informasi yang bermanfaat dan menarik pada database skala besar dengan mengidentifikasi hubungan antar item yang disebut juga *association*. Dalam tulisan ini akan digunakan data transaksi penjualan yang akan digali untuk menemukan pengetahuan yang menarik dan pola belanja konsumen serta keterkaitan item barang yang dibeli bersamaan dengan item barang lain.

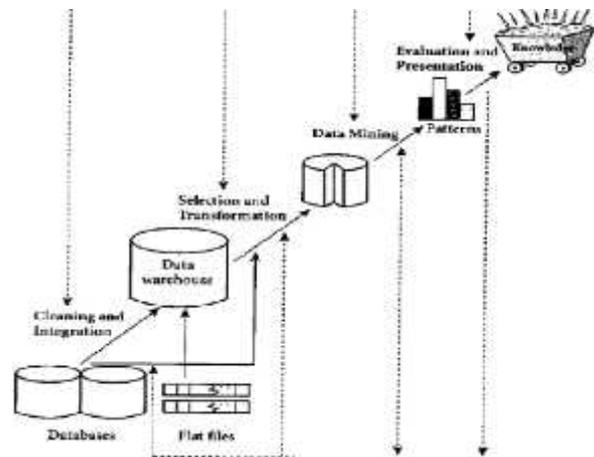
Tujuan dari penelitian ini yaitu mampu memperlihatkan penggunaan Algoritma Apriori pada teknik *association rules* untuk data transaksi penjualan dalam menemukan pengetahuan dan pola keterkaitan antar item barang yang dibeli konsumen.

Data mining harus mampu menunjukkan secara transparan, dimana hasilnya memperlihatkan mampu menyimpan jutaan data. Namun kondisi menyebabkan kaya data tetapi miskin informasi. Data mining mampu melakukan tugas untuk menggali informasi yang bermakna dan memiliki pengetahuan dan pola. Berikut ini

tugas dari data mining (Larose, 2005): pola secara jelas (Larose, 2005). Pertumbuhan data dalam perusahaan saat ini cenderung meningkat berkat teknologi informasi yang

- a. **Deskripsi**, menunjukkan pola dan tren suatu kejadian data yang tersimpan dalam database.
- b. **Estimasi**, mirip dengan fungsi klasifikasi. Dimana yang membedakan adalah variabel pada fungsi estimasi berupa numerik sedangkan klasifikasi variabelnya adalah kategorikal.
- c. **Prediksi**, mirip dengan klasifikasi dan estimasi. Yang membedakan adalah bahwa hasil mining akan memprediksi terjadi di masa mendatang.
- d. **Klasifikasi**, dalam fungsi klasifikasi nilai variabelnya berupa kategorikal. Model data mining memeriksa himpunan data yang besar, setiap data mempunyai informasi variabel target dan variabel *predictor*.
- e. **Clustering**, digunakan untuk mengelompokkan record ke dalam obyek yang mirip. Dalam clustering tidak terdapat variabel target. *Clustering* melakukan segmentasi seluruh himpunan data ke dalam cluster yang relatif homogen.
- f. **Asosiasi**, bertugas menemukan aturan dalam transaksi untuk menentukan hubungan dua atau lebih antar atribut. Asosiasi menghasilkan rules dimana berbentuk IF antaseden THEN konsekuen bersama-sama menentukan ukuran *support* dan *confidence* yang berhubungan dengan aturan.

Proses data mining dalam penelitian ini dilakukan secara bertahap seperti pada Gambar 1.



Gambar 1. Tahap data mining (Han dan Kamber, 2001)

Seperti yang dilakukan oleh Erwin (2009) dengan menganalisis market basket menggunakan Algoritma Apriori dan FP-Growth. Data yang digunakan data transaksi tahun 2009. Hasil pengujian menunjukkan pada data1 dengan Apriori = 10 jam 38 menit dan FP-Growth = 7 jam 20 menit. Untuk data2 pada Apriori = 8 jam 45 menit dan FP-Growth 5 jam 15 menit. Hasil tersebut menunjukkan FP-Growth lebih cepat dalam menemukan *frequent itemset* sehingga waktu yang dibutuhkan relatif singkat dan efisien.

Sedangkan Goswami dkk. (2010) melakukan penelitian algoritma *Frequent Pattern Mining* berbasis Apriori. Dalam penelitian ini ada tiga pendekatan *Frequent Pattern Mining* yang dihasilkan berdasarkan Algoritma Apriori klasik yaitu: *Record Filter*, *Intersection* dan *Proposed Algorithm*. Dengan 2.000 data, pendekatan *Record Filter* memperoleh 132 detik dan Apriori klasik 143 detik, pendekatan *Intersection* 35 detik dan *Proposed Algorithm* 29 detik.

Selain itu Hahsler dkk (2005) melakukan penelitian implikasi model data probabilistik untuk menggali *association rules mining*. Hasil dari penelitian tersebut yaitu menghasilkan konsep probabilistik sederhana pada data transaksi dimana dapat digunakan untuk mensimulasi data transaksi ketika *association* tidak digunakan. Data yang digunakan yaitu database toko kelontong untuk menggali *lift* dan *confidence*. Sedangkan hasil simulasi menunjukkan *dataset* kelontong mempunyai nilai *support* yang lebih tinggi dengan perbandingan rata-rata 0.000203 – 0.000113. Ini mengindikasikan bahwa *dataset* kelontong berisi item yang berasosiasi lebih sering muncul dari yang diharapkan. Dua *itemset* yang digeneralisasi dan menghasilkan rules menunjukkan *confidence* pada *real-word* data rata-rata 0.0086 – 0.0140.

Yang dilakukan oleh Prakash dan Parvathi (2010) melakukan penelitian peningkatan skalabilitas Apriori untuk efisiensi *association rules mining*. Penelitian ini menghasilkan rule baru bernama *informatifve rule set* yang menunjukkan prediksi sequensial sama dengan yang dihasilkan oleh *association rule set* menggunakan prioritas *confidence*. Eksperimen pertama menggunakan data bank berkapasitas 1GB. Total jumlah item 1.000 dan jumlah rata-rata item dalam transaksi ada 15, menggunakan *confidence* 90% dan *support* 0.5.

Eksperimen kedua menggunakan *dataset* dengan 20 item yang berbeda, dimana jumlah rata-rata item per transaksi adalah 6. Total ukuran *dataset* adalah 500Mb dan menggunakan tingkat *confidence* sebesar 90%.

Dalam data mining ada tahap transformasi, dimana proses transformasi adalah melakukan persiapan data agar mudah diproses. Beberapa tipe data yang dapat dihasilkan pada tahap transformasi yaitu:

- **Tipe Nominal**, yaitu tipe data diskrit yang tidak mengenal urutan. Contoh, warna baju: merah, kuning, hijau atau suku bangsa: jawa, sunda, Batak, minangkabau.
- **Tipe Biner**, yaitu tipe data yang hanya memiliki dua nilai. Contoh, bilangan biner: 0 dan 1 atau memiliki rumah sendiri: ya dan tidak.
- **Tipe Ordinal**, yaitu tipe data diskrit yang mengenal urutan. Contoh, ukuran baju: S, M, L, XL atau nilai ujian: A, B, C, D, E.
- **Tipe Continuous**, yaitu tipe data continyu. Contoh, gaji bulanan atau jumlah anak.

Aturan asosiasi (Santoso, 2007) berkaitan dengan “apa bersama apa”. Hal ini bisa berupa studi transaksi di supermarket, misal seseorang membeli susu bayi juga membeli sabun mandi bayi, maka menurut Santoso (2007) aturan asosiasi ini sering disebut *market basket analysis*. Transaksi pembelian diatas dapat ditulis sebagai berikut.

susu bayi \Rightarrow sabun mandi bayi

Ide dasar dari aturan asosiasi adalah memeriksa semua kemungkinan hubungan *if-then* antar item dan memilih yang paling mungkin (*most likely*) sebagai indikator dari hubungan ketergantungan antar item. Biasanya digunakan istilah *antecedent* untuk mewakili bagian “jika” dan *consequent* mewakili bagian “maka”.

Dalam menentukan kandidat dan menemukan frequent itemset pada data sequence, dimisalkan dengan data berikut.

$i = \{ \text{baju anak, baju wanita, asesoris, perabot, sabun mandi, sabun cuci, alat tulis, snack} \}$

Himpunan item yang dibeli oleh pengunjung ke- i disebut transaksi ke- i yang dilambangkan dengan T_i sebagai contoh gambar 2.

$T_1 = \{ \text{asesoris, sabun mandi, sabun cuci, perabot} \}$
$T_2 = \{ \text{baju anak, alat tulis, perabot} \}$
$T_3 = \{ \text{perabot, sabun mandi, snack, baju wanita, asesoris} \}$

Gambar 2. Himpunan item barang (Susanto, 2010)

Misalkan himpunan dari seluruh transaksi pada gambar 2 dapat dilambangkan dengan D sehingga diperoleh seperti gambar 3.

$$D = \{T_1, T_2, T_3, \dots, T_n\}$$

Gambar 3. Representasi data transaksi (Susanto, 2010)

Aturan asosiasi yang ingin dihasilkan kelak akan berbentuk aturan seperti gambar 4.

$$\text{"Jlu Anak B" atau "A } \Rightarrow \text{B"}$$

Gambar 4. Bentuk pola rule yang dihasilkan (Susanto, 2010)

Pada gambar 4, A disebut *antecedent* atau pendahulu sedangkan B disebut *consequent* atau pengikut dari aturan tersebut. Aturan asosiasi yang kelak akan dihasilkan harus memenuhi dua sifat. Pertama, baik A maupun B adalah himpunan bagian murni dari i jadi $A, B \subset i$, kedua, himpunan A dan B adalah dua himpunan yang saling lepas maka $A \cap B = \emptyset$ (Susanto, 2010).

Salah satu ukuran kinerja bagi aturan asosiasi $A \Rightarrow B$ adalah besaran *support* (dukungan) yang dilambangkan dengan $S(A \Rightarrow B)$. *Support* adalah probabilitas (nilai kemungkinan/peluang) pelanggan yang membeli beberapa produk secara bersamaan dari seluruh transaksi. Ukuran *support* ditunjukkan dengan prosentase dari keranjang belanja dimana sisi kiri A sebagai *antecedent* dan B sebagai *consequent* keduanya ditemukan secara bersamaan (Yusuf dkk, 2006). Dapat diartikan *support* merupakan prosentase dukungan pembeli yang belanja barang A juga akan belanja barang B dari total transaksi, sehingga didefinisikan seperti gambar 5.

$$S(A \Rightarrow B) = \frac{P(A \cap B)}{P(D)} = \frac{\text{jumlah transaksi yang mengandung item } A \cap B}{\text{jumlah total transaksi } D}$$

Gambar 5. Menentukan support (Susanto, 2010)

Confidence (nilai kepastian) adalah suatu ukuran yang menunjukkan kuatnya hubungan antar item dalam aturan asosiasi, dapat pula diartikan *confidence* berguna untuk mengukur tingkat

kebenaran dari kesimpulan yang diambil dari aturan yang dihasilkan (Susanto, 2010). Misal pada aturan $A \Rightarrow B$, jika nilai *confidence* rendah maka kemungkinan munculnya B yang memuat A juga semakin rendah. Hal ini didefinisikan seperti gambar 6.

$$C(A \Rightarrow B) = \frac{P(A \cap B)}{P(A)} = \frac{\text{jumlah transaksi yang mengandung item } A \cap B}{\text{jumlah transaksi yang mengandung item } A}$$

Gambar 6. Menentukan confidence (Susanto, 2010)

Suatu *itemset* adalah himpunan yang beranggotakan sebagian atau seluruh item yang menjadi anggota i (Susanto, 2010). Sebagai contoh {baju anak} atau {asesoris, sabun mandi} demikian pula dengan {baju anak, asesoris, sabun mandi}. Suatu *itemset* yang beranggotakan k buah item dapat ditulis atau disebut k -*itemset* (Agrawal dan Srikant, 1994). Misal:

1. Himpunan {baju anak} merupakan *1-itemset* karena hanya beranggotakan satu buah item saja.
2. Himpunan {asesoris, sabun mandi} merupakan *2-itemset* karena beranggotakan dua buah item.
3. Himpunan {baju anak, asesoris, sabun mandi} merupakan *3-itemset* karena beranggotakan tiga buah item.

Besaran *frequent itemset* yaitu mengukur berapa kali sebuah *itemset* muncul sebagai bagian atau keseluruhan transaksi yang menjadi anggota daftar transaksi D (Susanto, 2010). Sedangkan menurut Yusuf dkk (2006) *frequent itemset* adalah *itemset* yang terjadi pada jumlah tertentu mempunyai *itemset* frekuensi $\geq \Phi$. *Itemset* sering yang memiliki k buah anggota disebut *frequent k-itemset*. Notasi *set of frequent itemset k-itemsets* ditulis F_k (Susanto, 2010).

2. METODE

Tahap pertama dimulai dari *Data Preprocessing*, dimana menurut Larose (2005) bagian ini membutuhkan waktu hampir 60% dari proses data mining mulai dari aspek waktu sampai aspek aktifitas selama proses data mining secara keseluruhan. Di dalam *data preprocessing* terdiri dari beberapa tahapan:

- **Selection**, yaitu proses memilih dan memisahkan data berdasarkan beberapa kriteria. Misal berdasarkan kota atau jenis pembeli.
- **Preprocessing**, yaitu mempersiapkan data dengan cara membersihkan data berisi nilai

yang tidak dibutuhkan, melakukan penyeragaman format data.

- **Transformasi**, merupakan proses pada data yang telah dipilih sehingga data tersebut sesuai untuk proses *data mining*. Proses ini merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basisdata.
- **Data mining**, tahap ini difokuskan untuk mendapatkan pola dari data.
- **Interpretasi dan Evaluasi**, dalam proses ini pola-pola yang telah diidentifikasi oleh sistem diterjemahkan ke dalam bentuk *knowledge* yang mudah dimengerti oleh user.

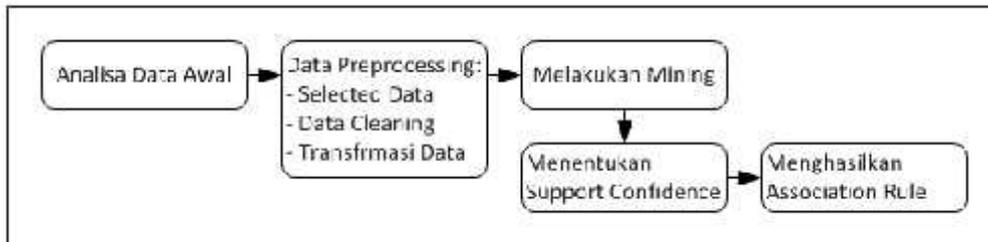
Obyek dari penelitian ini menggunakan data penjualan yang tersimpan pada *database* kasir supermarket. Adapaun data yang diolah sesuai peneliti peroleh yaitu transaksi bulan Agustus

2007 sampai September 2008 sejumlah 1.5 juta *record*.

Metode pengumpulan data primer yang dilakukan dengan cara observasi langsung ke supermarket untuk mengambil data dan mengamati proses transaksi.

Terdapat 4 kasir dengan frekuensi 100 transaksi per hari, memiliki 3000 konsumen sebagai *member* dan 39 kategori barang. Sedangkan pengumpulan data sekunder dilakukan dengan mencari referensi yang relevan sesuai bidang yang diteliti.

Alur penelitian yang dilakukan peneliti untuk menemukan pola aturan asosiasi pada data penjualan dilaksanakan seperti gambar 2.



Gambar 2. Alur association rule database penjualan

3. PEMBAHASAN

1) Algoritma Apriori

Apriori adalah sebuah algoritma yang digunakan untuk menemukan semua *itemset* dengan *support* tidak kurang dari *minimum support*. *Itemset* yang memenuhi *minimum support* disebut *frequent itemset* (Wu dan Kumar, 2009). *Pseudocode* Algoritma Apriori ditunjukkan pada gambar 7 dibawah. Secara garis besar ada dua bagian yang dilakukan oleh Apriori yaitu (Han dan Kamber, 2001):

- a. Menemukan *frequent itemset* menggunakan *candidate generation*. Dalam menemukan *frequent itemset* ada dua tahap yang dilakukan yaitu:
 - Tahap *Join*: menggabungkan *itemset* dalam tabel D.
 - Tahap *Prune*: memangkas *itemset* yang tidak *frequent* dan tidak memenuhi *minimum support*.
- b. Melakukan *generate association rule* dari *frequent itemset*. Dalam melakukan *generate* ada dua hal yang dilakukan yaitu *support* dan *confidence*.

Berikut ini penjelasan Algoritma Apriori dalam menemukan *frequent itemset* transaksi pada database. Menurut Han dan Kamber (2001) Langkah 1 dari Apriori adalah menemukan *frequent 1-itemset* L_1 . Pada langkah 2 dimana baris 2 – 10 terdapat L_{k-1} digunakan untuk menghasilkan kandidat C_k dalam urutan item untuk menemukan L_k .

```

(1)  $L_1 = \text{find frequent 1-itemset}(D);$ 
(2) for  $(k=2; L_{k-1} \neq \emptyset; k++)$  {
(3)    $C_k = \text{apriori\_gen}(L_{k-1}, \text{min\_sup});$ 
(4)   for each transaction  $t \in D$  :
(5)      $C_t = \text{subset}(C_k, t);$ 
(6)     for each candidate  $c \in C_t$  :
(7)        $c.\text{count}++;$ 
(8)   }
(9)    $L_k = \{c \in C_k \mid c.\text{count} \geq \text{min\_sup}\};$ 
(10) }
(11) return  $\bigcup_k L_k;$ 
    
```

Gambar 7. Pseudocode Algoritma Apriori (Han dan Kamber, 2001)

Fungsi *apriori_gen* pada langkah ke 3 menghasilkan kandidat dan kemudian menggunakan properti Apriori untuk menghilangkan *itemset* yang memiliki *subset* dibawah *minimum support*. Pada fungsi

apriori_gen melakukan dua hal yaitu *join* dan *prune*. Dalam komponen *join* L_{k-1} digabung dengan L_{k-1} untuk menghasilkan kandidat yang potensial. Kemudian komponen *prune* menggunakan properti Apriori untuk menghapus kandidat yang memiliki *subset* jarang muncul (*not frequent*).

Pada langkah ke 4, setelah semua kandidat di-*generate*, database di scanning. Langkah ke 5, untuk setiap transaksi fungsi dari *subset* digunakan untuk menemukan semua *subset* dari transaksi yang mengandung kandidat, langkah ke 6 dan 7 menghitung setiap kandidat yang terakumulasi. Terakhir, semua kandidat yang memenuhi *minimum support* dari *frequent itemset* adalah L.

Tahap berikutnya melakukan *generate frequent itemset*, setelah *frequent itemset* ditemukan semua, kemudian mencari *association rule* yang memenuhi syarat *minimum confidence* dengan menghitung jumlah transaksi yang mengandung $A \cap B$ dibagi jumlah transaksi yang mengandung A. Sedangkan (Larose, 2005) *support* dari *association rule* adalah proporsi dari transaksi dalam database yang mengandung $A \cap B$ dibagi jumlah total transaksi.

2) Menemukan Frequent itemset

Dengan menggunakan transaksi dari database tabel 1 misal ditentukan $\Phi \geq 3$, *itemset* yang terjadi lebih dari 3 disebut *frequent itemset*: dari proses *frequent itemset* tersebut diperoleh sebagai berikut:

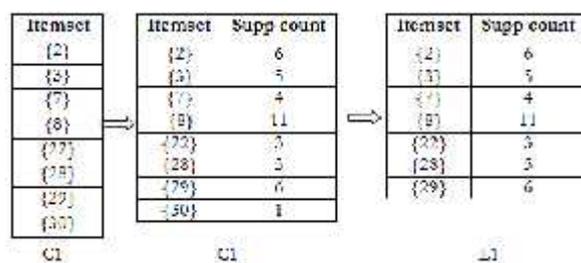
- a. *Frequent itemset* {baju anak} ditunjukkan kode {22} mempunyai $\Phi = 3$ karena himpunan ini menjadi bagian dari tiga transaksi yaitu $D = \{T_5, T_8, T_{12}\}$.
- b. *Frequent itemset* {perabot, asesoris} ditunjukkan kode {2, 3} mempunyai $\Phi = 3$ karena himpunan ini menjadi bagian dari tiga transaksi yaitu $D = \{T_9, T_{18}, T_{19}\}$.
- c. *Frequent itemset* {perabot, alat tulis, perlengkapan bayi} ditunjukkan kode {2, 28, 29} mempunyai $\Phi = 3$ karena himpunan ini menjadi bagian dari tiga transaksi yaitu $D = \{T_9, T_{15}, T_{16}\}$.

Tabel 1. Ada 19 transaksi untuk idmember = 2000339

T	Nofak	Items
1	200708278001	{ {3} }
2	200709069257	{ {8} }
3	200709168183	{ {8} }
4	200709259321	{ {8,30} }
5	200710078355	{ {2,8,22,7} }

6	200710287206	{ {29} }
7	200711049469	{ {8} }
8	200711269034	{ {22,7} }
9	200712108211	{ {2,29,3,28} }
10	200712109192	{ {3,28} }
11	200712139194	{ {7,8} }
12	200712158414	{ {8,7,22} }
13	200712247358	{ {8} }
14	200801017173	{ {29,8} }
15	200801158166	{ {2,29,28} }
16	200803240217	{ {28,29,2} }
17	200805270169	{ {8,29} }
18	200807010350	{ {3,2} }
19	200808030162	{ {8,28,2,3} }

Menggunakan tabel 1 yang diambil dari database, transaksi pertama menentukan besaran *frequent* yang diinginkan misal $\Phi = 2$. Langkah pertama menemukan *frequent 1-itemset* dari database. Misal kejadian transaksi yang diambil untuk idmember = 2000339 dalam satu tahun ada 19 transaksi. Dari gambar 8 setiap item merupakan anggota dari kandidat itemset C1, Selanjutnya menghitung besaran *frequent itemset* yang menghasilkan *frequent 1-itemset* L1. *Itemset* {2} {3} {7} {8} {22} {28} {29} memenuhi $\Phi = 2$. *Itemset* {30} tidak diikutkan dalam perhitungan *2-itemset* karena tidak memenuhi *minimum support*. Hasil *frequent 2-itemset* ditunjukkan gambar 9 dengan menghilangkan *itemset* yang mempunyai nilai *minimum support* $\Phi \leq 2$. Sehingga diperoleh $L2 = \{\{2,3\} \{2,8\} \{2,28\} \{2,29\} \{3,28\} \{7,8\} \{7,22\} \{8,22\} \{8,29\} \{28,29\}\}$ dengan *minimum support* $\Phi \geq 2$.

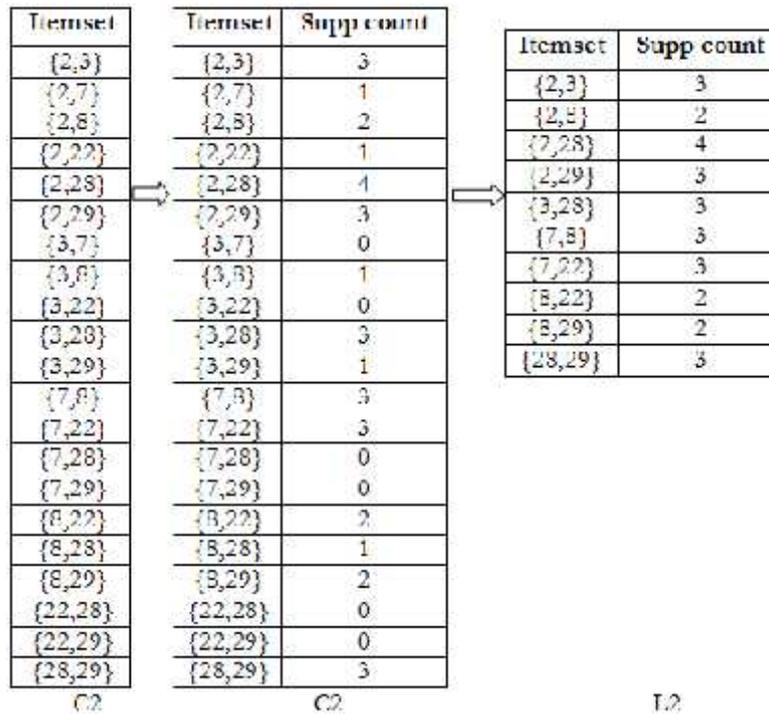


Gambar 8. Hasil frequent 1-itemset

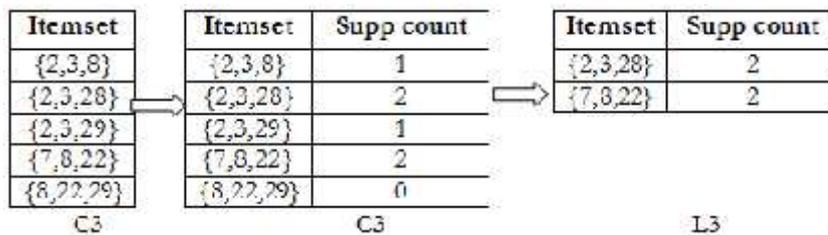
Selanjutnya dari gambar 8 mencari *frequent 2-itemset* menggunakan $L1 \times L1$ (Han dan Kamber, 2001). Hasil *frequent 2-itemset* ditunjukkan gambar 9.

Itemset {2,3} {2,8} {2,28} {2,29} {3,28} {7,8} {7,22} {8,22} {8,29} {28,29} memenuhi $\Phi \geq 2$ dan akan diikutkan dalam proses *3-itemset*. Kemudian langkah yang sama dilakukan terhadap C3 untuk mencari *frequent 3-itemset*. Hasil *3-itemset* ditunjukkan pada gambar 10.

Karena sudah sesuai dengan batasan *minimum support* maka proses berhenti.



Gambar 9. Hasil frequent 2-itemset



Gambar 10. Hasil frequent 3-itemset

3) Menentukan Support dan Confidence

Dari *frequent itemset* gambar 10 dapat diperoleh informasi pola dan pengetahuan belanja konsumen dengan *frequent* yang memenuhi $\Phi = 2$ seperti pada tabel 2.

Tabel 2. Pola belanja konsumen $W = 2$

Itemset	Kategori item	Supp count
{2,3,28}	{Perabot, Asesoris, Alat tulis}	2
{7,8,22}	{Busana Pria, Busana Wanita, Pakaian Anak}	2

Setelah *frequent pattern* ditemukan semua, langkah selanjutnya mencari *association rules* yang memenuhi *minimum support* dan *minimum confidence* (Han dan Kamber, 2001). Sedangkan *support* dari *association rules* adalah proporsi dari transaksi dalam database yang mengandung

A dan B (Larose, 2005). Dari gambar 10 atau tabel 2 pada L_3 diperoleh *frequent 3-itemset* yaitu $\{\{2,3,28\}$ dan $\{7,8,22\}\}$. Misal *minimum support* = 10% dan *minimum confidence* = 20% maka hasil *association rules* seperti pada tabel 3.

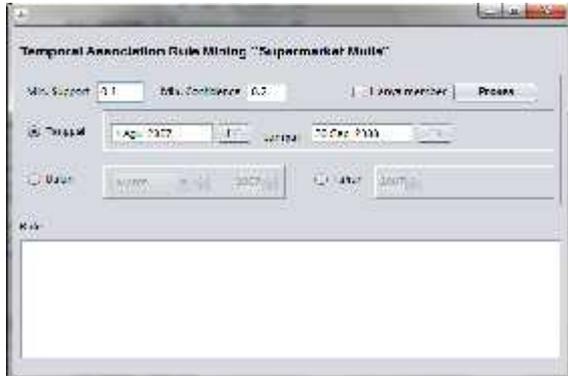
Tabel 3. Hasil association rule

Itemset	Minimum support	Minimum confidence
$\{2,3 \Rightarrow 28\}$	$2/19 = 10.5\%$	$2/3 = 66.6\%$
$\{7,8 \Rightarrow 22\}$	$2/19 = 10.5\%$	$2/3 = 66.6\%$

Dari tabel 3 diketahui bahwa konsumen idmember = 2000339 melakukan pola belanja, jika membeli busana pria dan busana wanita maka juga membeli pakaian anak dengan *support* 10.5% dan *confidence* = 66.6%.

4) Aplikasi Association Rule Mining

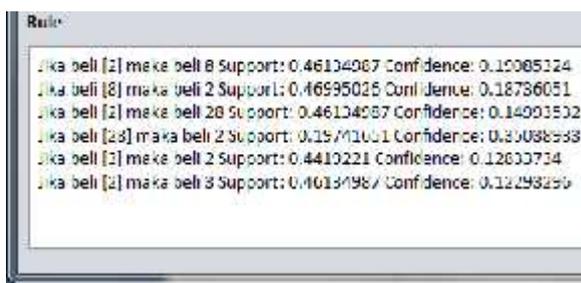
Hasil dari perancangan sistem yang dibuat selama penelitian berupa aplikasi ARM yang mampu menemukan pola *association rule* dari database transaksi penjualan. Bentuk aplikasi ARM ditunjukkan pada gambar 11.



Gambar 11. Tampilan aplikasi ARM

Dari gambar 11, obyek pada *minsup* dan *minconf* digunakan untuk mengisi batasan nilai *support* dan *confidence* berupa desimal. obyek member digunakan untuk menyaring data yang diolah mengikutsertakan konsumen sebagai member atau tidak, sedangkan obyek tanggal digunakan untuk membatasi data yang akan diolah data mining. Obyek bulan dan tahun digunakan untuk membatasi data yang akan diolah berdasarkan periode bulan atau tahun.

Misal diberikan batasan *minsup* = 0.05 (5%) dan *minconf* 0.1 (10%) untuk semua data transaksi selama satu tahun diperoleh *association rule* seperti gambar 12 dengan menghasilkan 6 *rule*.



Gambar 12. Hasil data mining database penjualan

4. KESIMPULAN

Dari pengujian yang dilakukan dan hasil penelitian dapat diambil beberapa kesimpulan sebagai berikut:

- Proses L2 memiliki waktu yang paling lama dalam menghasilkan *frequent itemset* untuk data sejumlah 1.5 juta record.
- Aplikasi ARM berhasil menunjukkan pola *association rule* meskipun dengan

prosentase yang rendah. Hal ini ditunjukkan bila diberikan *minsup* ≥ 0.1 dan *minconf* ≥ 0.2 tidak menghasilkan *rule*.

- Pola *association rule* yang sering muncul terjadi pada item barang {2} \Rightarrow {8}, {2} \Rightarrow {28} dan {3} \Rightarrow {2} untuk *minsup* 5% dan *minconf* 10% untuk semua data.
- Sedangkan pola *association rule* yang sering muncul terjadi pada item barang {7} \Rightarrow {8}, {22} \Rightarrow {8}, {7} \Rightarrow {22} untuk *minsup* 5% dan *minconf* 10% untuk data transaksi yang mengandung member.

DAFTAR PUSTAKA

Erwin, 2009, Analisis Market Basket dengan Algoritma Apriori dan FP-Growth, *Jurnal Generic*, Vol. 4 No. 2, Jur. Teknik Informatika Fak. Ilmu Komputer Universitas Sriwijaya

Goswami D. N., Anshu C. dan Raghuvanshi C. S., 2010, An Algorithm For Frequent Pattern Mining Based On Apriori, *International Journal on Computer Science and Engineering (IJCSSE)*, Vol. 02 No. 04, pp. 942 – 497.

Hahsler M., Hornik K. dan Reutterer T., 2005, Implications of Probabilistic Data Modelling for Mining Association Rules, *Proceeding of The 29th Annual Conference*

Han J dan Kamber M., 2001, *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publiser, San Francisco

Kusnawi, 2007, Pengantar Solusi Data Mining, *Proceeding Seminat Nasional Teknologi (SNT)*, Yogyakarta

Larose Daniel T., 2005, *Discovering Knowledge in Data: An Introduction to Data Mining*, Hoboken, Wiley Interscience, John Wiley and Sons

Prakash S. dan Parvathi R. M. S., 2010, An Enhanced Scaling Apriori for Association Rule Mining Efficiency, *European Journal of Scientific Research*, Vol. 39 No. 2, pp. 257 – 264.

Santoso B, 2007, *Data Mining, Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Teori dan Aplikasi, Graha Ilmu, Yogyakarta

Susanto S., Dedy Suryadi, 2010, *Pengantar Data Mining – Menggali Pengetahuan dari Bongkahan Data*, Andi, Yogyakarta

Yusuf Y. W., Pratikto F. R. dan Gerry T., 2006, Penerapan Data Mining Dalam

Penentuan Aturan Asosiasi Antar Jenis Item, *Proceding SNATI*, pp. E-53 – E-56, 1907-5022.