

Hierarchical Market Basket Analysis berbasis Algoritma Apriori

David Boy Tonara¹

Abstrak— Market Basket Analysis adalah salah satu metode pada data mining yang menitik beratkan pada kecenderungan sifat yang sama pada sebuah kondisi. Pada implementasinya Market Basket Analysis sering digunakan untuk menganalisa kecenderungan pembelian seseorang pada saat yang sama. Salah satu kekurangan penggunaan Algoritma Apriori pada Market Basket Analysis adalah ketidakmampuannya untuk memperhatikan hirarki dalam membentuk model pengetahuan yang dihasilkan. Hal ini dikarenakan Algoritma Apriori hanya bekerja pada jajaran level yang sama. Artikel ini akan membahas mengenai modifikasi yang dapat dilakukan pada Algoritma Apriori sehingga dapat digunakan untuk menganalisa kecenderungan dengan memperhatikan hirarki. Modifikasi Algoritma Apriori ini diimplementasikan pada data dummy toko roti. Modifikasi algoritma Apriori menjadi memperhatikan hirarki atau pengkategorian menghasilkan rule-rule yang lebih inovatif dan tidak dapat dihasilkan dari algoritma Apriori biasa. Modifikasi algoritma Apriori ini tetap membutuhkan pengujian pada kondisi data dunia nyata, walaupun di atas kertas sudah terlihat kelebihan dibandingkan algoritma Apriori standart.

Kata Kunci: Market Basket Analysis, Apriori, Data Mining

Abstract— Market Basket Analysis is one method in Data Mining which focused on identifying behavior tendency in the same condition. In the real world, Market Basket Analysis often used to analyse customer buying behavior. One of the algorithm which often used in Market Basket Analysis is Apriori, but Apriori can't accomodate hierarchy in their process. Standart Apriori only works on the same level. This article will discuss about modification in Apriori to be used to do Hierarchical Market Basket Analysis. This article will use bakery dummy data. This modification in Apriori algorithm can produce a more inovative association rule and can't be produced by standart apriori. Eventough this modification in Apriori algorithm need to be implemented in real world data.

Keywords: Market Basket Analysis, Apriori, Data Mining.

I. PENDAHULUAN

Market Basket Analysis adalah metode Data Mining yang paling banyak digunakan di dunia nyata. Kemampuan Market Basket Analysis untuk dapat menemukan kecenderungan perilaku pembeli membuat

metode ini menarik digunakan oleh para pelaku bisnis. Pengetahuan yang dihasilkan oleh metode ini memiliki banyak alternatif solusi yang dapat dipilih sesuai dengan kondisi yang dimiliki oleh pemilik bisnis.

Apriori adalah salah satu algoritma yang kerap digunakan pada metode Market Basket Analysis. Apriori dapat menemukan pola kecenderungan perilaku pembeli dengan cepat dan akurat.

Salah satu pola kecenderungan temuan Apriori yang dinyatakan sebagai temuan terbesar pada Market Basket Analysis adalah “bir maka popok”. Pola kecenderungan pembeli yang tidak terpikir oleh tenaga pemasaran manapun, bahwa orang yang membeli bir memiliki kecenderungan untuk juga membeli popok. Tentunya temuan pengetahuan ini hanya berlaku untuk lingkup yang kecil, tidak serta merta dapat diaplikasikan pada skala yang lebih luas atau ruang lingkup yang berbeda. Namun temuan pengetahuan yang akhirnya dapat divalidasi ini harus diakui adalah merupakan temuan yang menarik.

Namun tentu Apriori tidak dapat langsung digunakan pada data penjualan di supermarket contohnya. Pada awal implementasi Apriori membutuhkan kesepakatan mengenai level item yang digunakan. Apakah Apriori akan digunakan pada level item terendah (Pasta Gigi merk tertentu beraroma Mint), atau level yang lebih tinggi (Pasta Gigi merk tertentu atau Pasta Gigi beraroma Mint), atau level selanjutnya yang lebih tinggi (Pasta Gigi), atau bahkan level yang lebih tinggi lagi (Peralatan Mandi). Tentunya pengkategorian ini memakan waktu yang lama dan sangat berpengaruh pada keakuratan hasil Algoritma Apriori. Pengkategorian ini juga menyebabkan hasil pengetahuan yang didapatkan tidak dapat berlaku lintas kategori misal: Orang yang membeli Pasta Gigi merk tertentu beraroma Mint cenderung juga membeli Peralatan Masak.

Artikel ini akan membahas modifikasi yang dibutuhkan oleh algoritma Apriori, sehingga algoritma Apriori dapat mengakomodasi kebutuhan lintas level pengkategorian. Modifikasi ini akan lebih banyak berada pada tahap Preprocessing dan After Process dari Algoritma Apriori. Studi kasus yang digunakan pada artikel ini adalah menggunakan data dummy dari toko roti.

II. LANDASAN TEORI

Pada Bab ini akan membahas mengenai dasar-dasar teori yang digunakan sebagai acuan artikel ini.

A. Market Basket Analysis

Sebuah keranjang belanja dapat memberikan

¹ Jurusan Teknik Informatika, Universitas Ciputra, UC Town, Citraland, Surabaya 60219, Indonesia (telp: 031-745 1699; fax: 031-745 1698; e-mail: dtonara@ciputra.ac.id)

pengetahuan pada kita mengenai pembeli tersebut. Setiap pembeli memiliki produk yang berbeda beda pada keranjang belanja yang dimiliki dengan jumlah yang berbeda beda dan pada saat yang berbeda beda. Market basket analysis menggunakan informasi mengenai pembelian barang yang dilakukan oleh pembeli untuk mendapatkan gambaran mengenai siapa mereka dan mengapa mereka melakukan pembelian tersebut. Gambaran tersebut didapatkan dari hasil pengetahuan yang didapatkan oleh market basket analysis mengenai barang mana yang cenderung dibeli secara bersamaan.

Seringkali, market basket analysis digunakan sebagai titik awal ketika sebuah data transaksi tersedia dan peneliti tidak memahami pola yang akan diambil. Namun patut dipahami bahwa tidak semua pengetahuan yang dihasilkan oleh market basket analysis berguna. Sebuah pengetahuan baru dikatakan berguna ketika memenuhi aspek kebaruan. Contoh hasil pengetahuan penting yang dihasilkan oleh market basket analysis adalah sebagai berikut (Berry & Linoff, 1997):

- Pada hari kamis, terjadi banyak pembelian bir dan popok secara bersamaan.
- Pembeli yang membeli kesepakatan perawatan juga membeli benda elektronik berukuran besar.
- Ketika sebuah toko peralatan baru dibuka, maka barang yang sering terjual adalah saluran kamar mandi.

B. Association Rule

Association Rules adalah sebuah prosedur yang digunakan untuk menentukan hubungan antar item dalam sebuah dataset yang digunakan (Dunham, 2006). Dalam menentukan association rule, dibutuhkan 2 tahap yang harus dilakukan, yaitu: (1) Melakukan pencarian terhadap kombinasi yang paling sering terjadi dari sebuah itemset (frequent itemset); (2) Menghasilkan association rule dari frequent itemset yang telah dibuat sebelumnya.

Dalam menentukan association rule, digunakan dua ukuran kepercayaan, yaitu: (1) Support adalah suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu item dari keseluruhan dataset. Ukuran ini akan menentukan apakah suatu item layak untuk dicari confidence-nya; (2) Confidence adalah suatu ukuran yang menunjukkan hubungan antar 2 item secara conditional.

C. Algoritma Apriori

Algoritma Apriori adalah salah satu algoritma yang melakukan pencarian frequent itemset dengan menggunakan teknik association rule. Algoritma Apriori menggunakan pengetahuan frekuensi atribut yang telah diketahui sebelumnya untuk memproses informasi selanjutnya. Pada algoritma Apriori menentukan kandidat yang mungkin muncul dengan cara memperhatikan minimum support dan minimum confidence. Gambar 1 memuat runtutan logika dari algoritma Apriori.

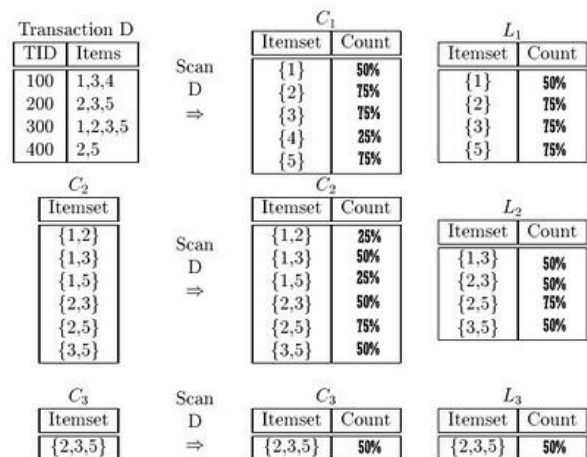
```

1)  $L_1 = \{\text{large 1-itemsets}\};$ 
2) for (  $k = 2; L_{k-1} \neq \emptyset; k++$  ) do begin
3)    $C_k = \text{apriori-gen}(L_{k-1});$  // New candidates
4)   forall transactions  $t \in \mathcal{D}$  do begin
5)      $C_t = \text{subset}(C_k, t);$  // Candidates contained in  $t$ 
6)     forall candidates  $c \in C_t$  do
7)        $c.\text{count}++;$ 
8)   end
9)    $L_k = \{c \in C_k \mid c.\text{count} \geq \text{minsup}\}$ 
10) end
11) Answer =  $\bigcup_k L_k;$ 

```

Gambar 1. Runtutan Logika Algoritma Apriori (Agrawal & Srikant, 1994)

Pada awal mula algoritma menghitung jumlah kemunculan item untuk menentukan large 1-itemsets. Tahap berikutnya terdiri dari 2 fase. Pertama, setiap item dari large 1-itemsets digunakan sebagai kandidat itemset berikutnya. Berikutnya, dilakukan pencarian database sehingga didapatkan nilai support dari setiap kandidat itemset. Untuk setiap kandidat itemset yang memenuhi nilai support yang diharapkan, dijadikan itemset untuk iterasi berikutnya.



Gambar 2. Contoh Penggunaan Algoritma Apriori (Agrawal & Srikant, 1994)

Gambar 2 menunjukkan contoh dari proses pencarian aturan menggunakan algoritma Apriori. Pada contoh di atas, menggunakan nilai support sebesar 50%. Pada proses pertama, algoritma akan melakukan proses eliminasi data terhadap itemset yang memiliki nilai di bawah 50%, pada contoh di gambar, proses eliminasi dilakukan terhadap data dari C1 yang menghasilkan L1. Tahap kedua dilakukan proses eliminasi dari C2 yang merupakan kombinasi dari item L1 dan menghasilkan L2. Dan tahap terakhir dilakukan proses eliminasi dari C3 yang merupakan kombinasi item dari L2 dan menghasilkan L3.

D. Tahap Preprocessing

Tahap Preprocessing adalah tahapan untuk mempersiapkan data sehingga data siap digunakan oleh algoritma yang dipilih untuk proses data mining. Setidaknya terdapat empat jenis tahap preprocessing yang dilakukan, yaitu Data Cleaning, Data Integration and

Transformation, Data Reduction, dan Discretization (Han, Kamber, & Pei, 2011).

Data Cleaning adalah tahapan preprocessing untuk melakukan pembersihan pada data asal. Pembersihan yang dilakukan biasanya dikarenakan oleh kondisi data, meskipun mungkin juga terjadi dikarenakan oleh sebab lain. Beberapa hal yang dilakukan dalam tahapan data cleaning adalah mengatasi missing values. Missing values adalah kondisi di mana data kehilangan satu atau lebih data penting, misal dikarenakan proses input yang tidak sempurna atau mungkin kegagalan proses backup. Penanganan terhadap missing values dapat dilakukan dengan berbagai cara, dua di antaranya adalah penghapusan data dan pengisian dengan data default.

Hal lain yang dilakukan pada tahapan data cleaning adalah pemrosesan noisy data. Noisy data adalah data yang sangat berbeda dibandingkan dengan data lainnya, sehingga dapat menyebabkan tingkat akurasi yang rendah. Noisy data pada statistik sering disebut sebagai outlier dan dapat diatasi dengan melakukan metode binning, clustering, inspeksi oleh manusia, dan metode regresi.

Hal terakhir yang dilakukan pada tahap Data Cleaning adalah mengatasi data yang tidak konsisten. Data yang tidak konsisten biasanya paling banyak berupa data yang terdapat kesalahan ketik sehingga untuk mengatasinya harus dengan campur tangan manusia.

Tahapan data integration and transformation adalah tahapan data yang berfungsi menggabungkan dua atau lebih sumber data sehingga data yang didapatkan memungkinkan untuk diproses secara lebih akurat. Seringkali terjadi dikarenakan penggabungan dilakukan dari dua atau lebih sumber data, maka data perlu diubah ke format yang konsisten di antara dua atau lebih sumber data yang ada.

Selain tahap pengambilan dan penggabungan data dari sumber data sekunder, transformasi data juga dilakukan dengan empat metode yaitu normalization, smoothing, aggregation dan generalization. Normalisasi adalah metode transformasi data yang mengubah data numerical menjadi dalam rentang tertentu yang bertujuan supaya tingkat akurasi menjadi lebih baik. Smoothing adalah metode tahapan transformasi data yang menjadi alternatif solusi untuk menghilangkan noise selain dilakukan pembersihan data pada tahapan data cleaning. Agregasi adalah metode transformasi data yang melakukan perubahan data dari satuan lebih kecil ke satuan yang lebih besar, misal terjadi untuk mengubah data penjualan satuan menjadi data penjualan harian, mingguan, bulanan, atau tahunan. Generalisasi adalah metode transformasi data untuk mengelompokkan data numerical menjadi data nominal atau ordinal. Data nominal atau ordinal adalah data categorical dari data numerical, misal muda, setengah baya, tua atau kurus, sedang, gemuk. Keseluruhan dari tahapan integrasi data dan transformasi data pada umumnya adalah supaya akurasi dari proses data mining dapat meningkat dengan signifikan.

Tahapan berikutnya pada Preprocessing adalah Data Reduction yang bertujuan untuk mengurangi load

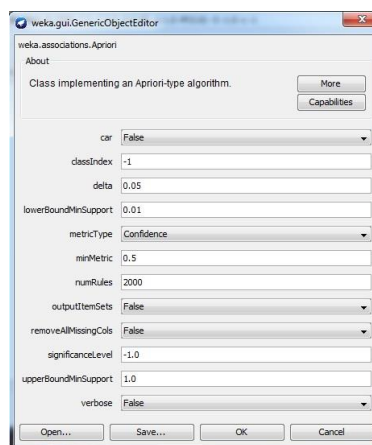
algoritma sehingga pengetahuan bisa didapatkan dengan lebih cepat dengan tetap memperhatikan akurasi. Reduksi data biasa dilakukan dengan agregasi data ke satuan yang lebih besar (seperti telah dijelaskan pada tahap transformasi data), penghapusan atribut yang tidak signifikan, dan kompresi data.

E. WEKA (Waikato Environment for Knowledge Analysis)

WEKA adalah sebuah software pembelajaran yang memiliki koleksi algoritma untuk melakukan proses dalam data mining (Frank, et al., 2005). Software WEKA pertama kali dikembangkan oleh Universitas Waikato di Selandia Baru. WEKA merupakan alat untuk melakukan pre-processing data, klasifikasi, regresi, association rule dan visualisasi. Interface utama dari aplikasi WEKA adalah explorer.

Explorer memiliki beberapa panel yang digunakan untuk generate data dengan fungsi yang berbeda. Preprocess digunakan untuk memasukkan data ke dalam WEKA dan melakukan filter terhadap data tersebut. Filter yang dilakukan oleh WEKA terhadap data yang dimasukkan berdasarkan attribute dan instance dari data tersebut. Dalam WEKA juga terdapat classify panel yang berfungsi untuk menerapkan algoritma klasifikasi dan regresi terhadap dataset yang dihasilkan. Cluster panel untuk memberikan akses terhadap fungsi clustering dalam WEKA. Association panel digunakan untuk mencari association rules dari kumpulan data dengan menggunakan beberapa algoritma yang didukung oleh WEKA. Select attribute panel menyediakan algoritma untuk melakukan identifikasi dengan tujuan mendapatkan prediksi paling besar terhadap sebuah data. Visualize panel digunakan untuk menampilkan scatter plot matrix dari kumpulan data, dimana scatter plot matrix tersebut dapat dipilih, diperbesar dan dianalisa lebih lanjut dengan berbagai operasi yang disediakan oleh WEKA (Witten, Frank, & Hall, 2011).

Pada Gambar 3 ditampilkan tampilan antar muka yang dimiliki WEKA untuk pemrosesan dengan algoritma Apriori.



Gambar 3. Tampilan Antar Muka WEKA untuk Algoritma Apriori

Terdapat beberapa konfigurasi yang dapat dilakukan, yaitu lowerBoundMinSupport untuk menentukan nilai support yang dibutuhkan. Metrictype untuk menentukan jenis metric yang digunakan, sebagai contoh adalah confidence, karena confidence digunakan oleh algoritma Apriori sebagai nilai acuan menentukan aturan. Minmetric untuk menentukan nilai minimal dari jenis metric yang digunakan, minimal nilai confidence yang ditetapkan oleh user. NumRules untuk menentukan jumlah rules yang dibutuhkan oleh user, sebagai default aplikasi WEKA menetapkan sepuluh rules. Semantara option lain adalah car untuk menentukan atribut yang akan digunakan, pilihan yang diberikan adalah TRUE atau FALSE, class index untuk menetapkan index dari class atribut, default yang diberikan adalah -1, delta berfungsi untuk mengurangi nilai support sampai nilai minimal support didapatkan atau sampai generate rules selesai dilakukan, outputItemSets memiliki pilihan TRUE dan FALSE, jika diaktifkan memungkinkan itemset untuk menjadi keluaran, removeAllMissingClos untuk menghapus kolom dengan nilai kosong, significanceLevel digunakan untuk memasukkan nilai signifikan khusus untuk metric confidence, upperBoundMinSupport untuk menentukan nilai paling tinggi untuk nilai support. Data siap untuk diproses dan siap untuk mencari rules dari data tersebut dan WEKA melakukannya secara otomatis.

III. Hierarchical Apriori

Pada Bab ini akan membahas mengenai modifikasi Apriori yang dilakukan beserta dengan data yang digunakan.

A. Data

Data yang digunakan adalah data dummy dari sebuah toko roti. Studi kasus toko roti dipilih dikarenakan jumlah item yang relatif lebih sedikit dibandingkan dengan bentuk bisnis lain dan sifat kategorikal yang mendukung penggunaan algoritma Apriori yang mengakomodasi hirarki. Hirarki yang digunakan pada contoh kasus ini adalah hirarki dengan 1 level kategorikal. Berikut pada tabel 1 adalah daftar item yang digunakan.

TABEL I
DAFTAR ITEM

Kode Barang	Kategori	Nama Barang
0001	Roti Manis	Roti Coklat
0002	Roti Manis	Roti Oreo
0003	Roti Manis	Roti Selai Strawberry
0004	Roti Manis	Roti Selai Srikaya
0005	Roti Manis	Roti Selai Nanas
0006	Roti Manis	Roti Sate Gula Pasir
0007	Roti Manis	Cupcake
0008	Roti Asin	Roti Keju
0009	Roti Asin	Pizza Vegetarian
0010	Roti Asin	Roti Susu Keju

0011	Roti Daging	Hotdog
0012	Roti Daging	Pizza
0013	Roti Daging	Roti Abon Ikan
0014	Roti Daging	Roti Abon Ayam
0015	Roti Daging	Roti Abon Daging
0016	Roti Daging	Roti Isi Daging
0017	Roti Daging	Roti Isi Ayam
0018	Roti Tawar	Roti Tawar Kulit
0019	Roti Tawar	Roti Tawar Pandan
0020	Roti Tawar	Roti Tawar Gandum

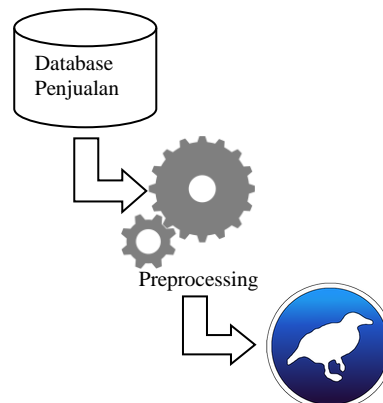
Data item tersebut digenerate secara random untuk mendapatkan 100.000 nota penjualan dengan jumlah roti per pembelian yang juga random. Adapun berikut adalah contoh penjualan yang berhasil tergenerate.

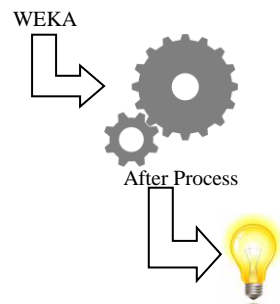
TABEL III
CONTOH DAFTAR PENJUALAN

Kode Penjualan	Kode Barang
0001	0003
0001	0012
0002	0020
0003	0001
0003	0009
0003	0013
0004	0005
0004	0008
0005	0015
0005	0020
0006	0003
0006	0008
0006	0012
0007	0008
0007	0013
0008	0007
0008	0020
0009	0019
0010	0001
0010	0013

B. Arsitektur Sistem

Perbedaan modifikasi algoritma apriori untuk dapat mengakomodasi hirarki terdapat pada tahap preprocessing dan penambahan satu tahap setelah proses yang diberi nama tahap after process. Gambaran Arsitektur Sistem dijabarkan pada gambar 4.





C. Tahap Preprocessing

Tahap preprocessing adalah tahap yang mutlak dilakukan sebelum proses algoritma Apriori dilakukan. Salah satu bagian dari tahap preprocessing ini adalah mengubah bentuk relational database ke bentuk yang siap diproses oleh aplikasi WEKA yang akan digunakan untuk melakukan data mining. WEKA dapat menerima input berupa file ARFF maupun file CSV. Berikut adalah bentuk format CSV yang dapat diterima oleh WEKA sebagai inputan dari algoritma Apriori.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	0001	0002	0003	0004	0005	0006	0007	0008	0009	0010	0011	0012	0013	0014	0015	0016	0017	0018	0019	0020
2	?	?	t	?	?	?	?	?	?	?	t	?	?	?	?	?	?	?	?	?
3	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	t
4	t	?	?	?	?	?	?	t	?	?	?	t	?	?	?	?	?	?	?	?
5	?	?	?	?	t	?	?	t	?	?	?	?	?	?	?	?	?	?	?	?
6	?	?	?	?	?	?	?	?	?	?	?	?	?	?	t	?	?	?	?	t
7	?	?	t	?	?	?	?	t	?	?	t	?	?	?	?	?	?	?	?	?
8	?	?	?	?	?	?	t	?	?	?	t	?	?	?	?	?	?	?	?	?
9	?	?	?	?	?	?	t	?	?	?	?	?	?	?	?	?	?	?	?	t
10	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	t
11	t	?	?	?	?	?	?	?	?	?	t	?	?	?	?	?	?	?	?	?

Gambar 5. Format CSV yang diterima WEKA

Sesuai pada yang tertampil pada Gambar 5, di mana kolom adalah penanda untuk barang yang dibeli sedangkan baris adalah penjualan per nota tanpa kode penjualan. Nilai setiap cell pada data adalah t,f, dan ?. Notasi t berarti memiliki sifat atau dalam hal ini berarti membeli, notasi f berarti tidak memiliki sifat atau dalam hal ini berarti pasti tidak membeli, sedangkan notasi ? adalah notasi standart yang digunakan selain notasi t dan notasi f. Pada contoh kasus ini tidak menggunakan notasi f karena tidak didapatkan data mengenai kecenderungan tidak membeli dalam satu nota penjualan.

Salah satu modifikasi yang dilakukan untuk mengakomodasi hirarki pada algoritma apriori ini terletak pada tahap preprocessing selain preprocessing standart yang dilakukan untuk mengakomodasi relational database ke bentuk yang siap diproses oleh WEKA. Untuk mengakomodasi hubungan hirarki, maka ditambahkan kolom-kolom tambahan yang memuat hirarki. Sehingga seolah olah jumlah jenis barang pada data menjadi 24 item di mana terdiri dari 20 item barang dan 4 kategori.

Modifikasi data juga dilakukan sehingga seolah-olah setiap penjualan yang membeli roti coklat juga membeli item lain bernama roti manis, setiap penjualan yang membeli roti keju juga membeli item lain bernama roti

asin, setiap penjualan yang membeli pizza juga membeli item lain bernama roti daging, setiap penjualan yang membeli roti tawar kulit juga membeli item lain bernama roti tawar, dan demikian juga diberlakukan pada item-item lain.

D. Tahap Data Mining

Tahap data mining dilakukan dengan menggunakan algoritma apriori berbasisan WEKA. Konfigurasi algoritma Apriori yang digunakan adalah metricType Confidence, minMetric 0.5, numRules 2000, dan MinSupport 1.0. Tahap data mining ini menghasilkan 1219 association rule, beberapa association rule di antaranya adalah:

1. 0002=t 3075 ==> 0021=t 3075 conf:(1)
2. 0020=t 3066 ==> 0024=t 3066 conf:(1)
3. 0005=t 3062 ==> 0021=t 3062 conf:(1)
4. 0009=t 3049 ==> 0022=t 3049 conf:(1)
5. 0008=t 3040 ==> 0022=t 3040 conf:(1)
6. 0006=t 3028 ==> 0021=t 3028 conf:(1)
7. 0016=t 3017 ==> 0023=t 3017 conf:(1)
8. 0012=t 3016 ==> 0023=t 3016 conf:(1)
9. 0015=t 3007 ==> 0023=t 3007 conf:(1)
10. 0018=t 3005 ==> 0024=t 3005 conf:(1)

Association rule nomor 1 di atas berarti customer yang membeli kode barang 0002 pasti akan membeli kode barang 0021 didukung oleh 3075 transaksi. Kepastian tersebut dinyatakan dari conf(1) yang berarti memiliki kemungkinan 100%. Hasil keluaran dari algoritma Apriori di atas mengacu pada tabel data item yang telah dijelaskan pada tabel I. Namun association rule nomor 1 tersebut nantinya akan dieliminasi pada tahap after process yang akan dijelaskan pada subbab selanjutnya.

E. Tahap After Process

Modifikasi ini juga membutuhkan proses tambahan setelah dilakukan pemrosesan dengan algoritma Apriori. Tahap ini dibutuhkan untuk membuang association rule yang hanya memuat kategorikal dan subsetnya. Misal association rule yang dibuang adalah:

- Roti Oreo → Roti Manis
- Roti Asin → Roti Keju

Dibaca sebagai orang yang membeli roti oreo cenderung juga membeli roti manis dan orang yang membeli roti asin cenderung juga membeli roti keju. Association rule Roti Oreo → Roti Manis dinyatakan pada subbab sebelumnya dalam bentuk kode item yaitu 0002 → 0021. Kedua association rule ini dibuang dari daftar pengetahuan karena hanya mengandung kategorikal dan subsetnya. Berdasarkan daftar item yang ada, roti manis adalah kategori dari item roti oreo dan roti asin adalah kategori dari item roti keju. Sehingga roti manis yang dimaksud pada sisi kanan adalah roti oreo itu sendiri.

Berbeda dengan pengetahuan yang penting seperti:

- Roti Daging → Roti Keju

Association rule di atas dianggap sebagai pengetahuan penting berbeda dengan contoh Roti Oreo → Roti Manis,

dikarenakan roti daging bukan adalah kategori dari item roti keju. Association rule Roti Daging \rightarrow Roti Keju berarti bahwa setiap orang yang membeli roti apapun yang terkategori sebagai roti daging (Hotdog, Pizza, Roti Abon Ikan, Roti Abon Ayam, Roti Abon Daging, Roti Isi Daging, Roti Isi Ayam) memiliki kecenderungan untuk membeli juga Roti Keju.

Proses ini berhasil mengeliminasi 160 rule yang tidak berarti dari total 1219 rule, sehingga menyisakan 1059 rule. Beberapa rule tersisa yang menjadi rule menarik adalah:

- Rule 176 : Roti Keju \rightarrow Roti Manis, conf:(0.93), yang berarti 93% dari pembeli Roti Keju memiliki kecenderungan untuk juga membeli Roti Manis.
- Rule 185 : Roti Keju, Roti Daging \rightarrow Roti Manis, conf (0.92), yang berarti 92% dari pembeli Roti Keju dan Roti Daging secara bersamaan memiliki kecenderungan untuk juga membeli Roti Manis.
- Rule 271 : Roti Abon Ikan \rightarrow Roti Manis, conf (0.92), yang berarti 92% dari pembeli Roti Abon Ikan memiliki kecenderungan untuk juga membeli Roti Manis.
- Rule 277 : Cupcake \rightarrow Roti Daging, conf (0.92), yang berarti 92% dari pembeli Cupcake memiliki kecenderungan untuk juga membeli Roti Daging.
- Rule 286 : Roti Tawar \rightarrow Roti Manis, conf (0.92), yang berarti 92% dari pembeli Roti Tawar memiliki kecenderungan untuk juga membeli Roti Daging.

F. Knowledge Implementation

Tahap ini adalah tahap yang paling menarik dari keseluruhan tahap Data Mining, dikarenakan pada tahap ini tidak ada jaminan mutlak mengenai benar dan salah. Tahap ini mengambil pengetahuan mutlak yang didapatkan dari hasil After Process untuk dipikirkan cara mengambil keuntungan dari pengetahuan itu sehingga keuntungan toko roti meningkat.

Subbab ini akan mengambil contoh kasus jika hanya terdapat 5 association rule menarik yang ditemukan berdasarkan hasil subbab sebelumnya. Setidaknya ada 4 cara yang memungkinkan untuk dilakukan terkait hal ini, yaitu:

- Pemberlakuan paket pembelian antara roti apapun dengan roti yang terkategori sebagai roti manis. Kecenderungan pembeli untuk juga membeli roti manis membuat kategori roti manis adalah kategori yang paling menguntungkan untuk dipaketkan karena banyak menarik minat pembeli.
- Pemberlakuan paket pembelian yang berlaku tanpa roti manis.

Kecenderungan pembeli untuk juga membeli roti manis akan membuat load pekerja dan penggunaan bahan untuk roti manis jauh lebih tinggi dibandingkan dengan kategori roti yang lainnya. Pemberlakuan paket diharapkan dapat mendongkrak penjualan dari kategori roti lain selain roti manis.

- Peletakkan rak roti manis di tengah toko. Kecenderungan pembeli untuk juga membeli roti manis akan membuat pembeli cenderung mencari kategori roti manis yang diinginkan. Peletakkan rak roti manis di tengah toko membuat roti manis mudah terlihat dari pembeli yang sedang berada pada rak roti manapun. Sehingga hal ini akan menarik minat pembeli dan meningkatkan profit penjualan toko kue secara signifikan.
- Peletakkan rak roti manis di dekat kasir. Kecenderungan pembeli untuk juga membeli roti manis akan sangat dimudahkan dengan peletakkan rak roti manis di dekat kasir. Letak rak roti manis yang pasti dilewati oleh pengunjung membuat pengunjung cenderung memastikan membeli roti manis sebelum melakukan pembayaran di kasir. Hal ini diharapkan akan meningkatkan minat pembeli menambah jumlah pembeliannya pada saat-saat terakhir.

Tentu alternatif-alternatif yang tersebut di atas harus diuji dengan kondisi di dunia nyata untuk mengetahui potensi keuntungan yang mungkin didapatkan. Hal ini juga untuk mengetahui alternatif mana yang paling tepat diterapkan untuk toko roti. Walaupun pada kenyataannya tidak ada kepastian bahwa alternatif yang baik digunakan pada satu roti cocok juga digunakan pada toko roti lain.

IV. KESIMPULAN DAN SARAN

Dengan sedikit modifikasi pada bagian preprocessing serta menambahkan tahapan after process, memungkinkan algoritma Apriori untuk dapat mengakomodasi data yang memperhatikan hirarki atau pengkategorian.

Modifikasi algoritma Apriori menjadi memperhatikan hirarki atau pengkategorian menghasilkan rule-rule yang lebih inovatif dan tidak dapat dihasilkan dari algoritma Apriori biasa. Rule-rule keluaran dari algoritma Apriori yang telah dimodifikasi sehingga mempertimbangkan hirarki atau pengkategorian membuat alternatif solusi yang dihasilkan lebih beragam dan tetap memuat rule-rule yang dihasilkan dari algoritma Apriori biasa.

Adapun demikian modifikasi algoritma Apriori pada artikel ini belum diuji pada data dunia nyata dikarenakan oleh keterbatasan penulis dalam mendapatkan mitra penelitian. Modifikasi algoritma Apriori ini tetap membutuhkan pengujian pada kondisi data dunia nyata, walaupun di atas kertas sudah terlihat kelebihanannya dibandingkan algoritma Apriori standart.

Selain itu modifikasi algoritma Apriori ini tidak semata-mata tepat digunakan pada semua contoh kasus

yang baik diproses dengan menggunakan algoritma Apriori. Modifikasi ini akan sangat baik jika digunakan pada data dengan jumlah item terbatas dan kategori berjenjang.

DAFTAR PUSTAKA

- [1] Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules in large databases. *Proceedings of the 20th International Conference on Very Large Data Bases*, 487-499.
- [2] Berry, M. J., & Linoff, G. (1997). *Data Mining Techniques: For Marketing, Sales, and Customer Support*. New York: Wiley Computer Publishing.
- [3] Dunham, M. H. (2006). *Data Mining: Introductory and Advanced Topics*. New York: Pearson Education.
- [4] Frank, E., Hall, M. A., Holmes, G., Kirkby, R. B., Pfahringer, B., & Witten, I. H. (2005). Weka: A machine learning workbench for data mining. In O. Maimon, & L. Rokach, *Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers* (pp. 1305-1314). Berlin: Springer.
- [5] Han, J., Kamber, M., & Pei, J. (2011). *Data Mining Concepts and Techniques*. Massachusetts: Morgan Kaufmann.
- [6] Witten, I., Frank, E., & Hall, M. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Massachusetts: Morgan Kaufmann.