

APLIKASI DATA MINING MARKET BASKET ANALYSIS PADA TABEL DATA ABSENSI ELEKTRONIK UNTUK MENDETEKSI KECURANGAN ABSENSI (CHECK-LOCK) KARYAWAN DI PERUSAHAAN

Gregorius Satia Budhi; Felicia Soedjianto

Jurusan Teknik Informatika, Fakultas Teknologi Industri, Universitas Kristen Petra
Jl. Siwalankerto 121-131 Surabaya 60236
Email: greg@petra.ac.id; felis@petra.ac.id

ABSTRAK: Absensi pegawai selama ini selalu menjadi permasalahan yang pelik bagi bagian HRD di perusahaan – perusahaan yang ada. Walaupun telah ada peralatan absensi otomatis, alat ini masih memiliki kelemahan yaitu, tidak dapat mendeteksi kecurangan pegawai untuk menitipkan kartu absensinya pada karyawan lain untuk diabsenkan. Peneliti berkeinginan untuk mengatasi permasalahan absensi tersebut dengan memanfaatkan metode data mining, khususnya metode market basket analysis, untuk mendeteksi kecurangan ini. Perangkat lunak yang dibuat ini akan mentransformasikan data absensi pegawai menggunakan metode MaxDiff Histogram menjadi format compact transaction yang selanjutnya akan diproses menggunakan Algoritma Pincer Search menjadi frequent itemset. Pada akhirnya dari data frequent itemset ini didapat association rule pegawai untuk disajikan kepada pengguna, yaitu bagian HRD perusahaan. Dari hasil pengujian dapat diketahui bahwa metode Data Mining Market Basket Analysis dapat dimanfaatkan untuk menggali pattern kebiasaan absensi (check-lock) pegawai sebuah perusahaan. Dari sini kemungkinan terjadinya kecurangan saat melakukan absensi masuk / pulang dapat dideteksi.

Kata kunci: absensi/check-lock, data mining, MaxDiff histogram, market basket analysis, pincer search

ABSTRACT: Taking attendance from employees always becomes a problem for Human Resource Department (HRD) in many companies lately. Although there is an automatic check-lock machine, it still has a weakness. This machine can't detect some frauds like the employee swipes double identity card, his card and the others card. Reseachers want to solve this problem by using data mining method, especially market basket analysis. This software will transform the attendance data to compact transaction format by using MaxDiff Histogram method. And it will be processed into frequent itemset with Pincer Search Algoritm. At the final process the employee's association rule will got from frequent itemset. This output will be served to user that is the HRD of a firm. Testing result shows that Data Mining Market Basket Analysis can be used to get pattern of employee's check-lock from a company. And this pattern can help user to detect fraud that is done by employee.

Keywords: taking attendance/check-lock, data mining, MaxDiff histogram, market basket analysis, pincer search

PENDAHULUAN

Absensi pegawai selama ini selalu menjadi permasalahan yang pelik bagi bagian HRD di perusahaan – perusahaan yang ada, khususnya di Indonesia yang terkenal dengan budaya jam karet. Banyak cara yang telah diambil untuk mengatasi masalah ini, misal dengan menyediakan alat absensi otomatis (check-lock) dalam bentuk mekanik maupun elektronik. Hal ini memang dapat lebih mengurangi kemungkinan kecurangan dalam absensi. Walaupun demikian peralatan absensi otomatis ini memiliki kelemahan, yaitu, tidak dapat mendeteksi kecurangan pegawai yang menitipkan kartu absensinya pada pegawai lain sehingga pegawai dapat masuk/pulang kerja tidak pada jam yang telah ditentukan.

Peneliti berkeinginan untuk mengatasi permasalahan absensi tersebut dengan memanfaatkan metode *Data Mining*, khususnya metode *Market Basket Analysis*, untuk menganalisa database absensi elektronik secara otomatis, guna mendeteksi dan menemukan informasi tentang pegawai-pegawai yang seringkali melakukan absensi dalam waktu yang sangat berdekatan. Asumsi dari pemikiran ini adalah bila pada data absensi ditemukan pola / *pattern* dengan frekuensi cukup tinggi tentang dua atau lebih pegawai yang absen dalam waktu yang hampir bersamaan dalam kurun waktu tertentu, besar kemungkinan dalam realitasnya, pegawai yang melakukan absen hanya satu orang sedangkan lainnya "titip absen" kepada pegawai tersebut.

Informasi yang didapat dari Analisa Otomatis terhadap data absensi pegawai pada kurun waktu tertentu ini akan disajikan pada pihak HRD dalam bentuk daftar pegawai-pegawai yang sering absen dalam waktu yang hampir bersamaan, lengkap dengan prosentase frekuensinya. Informasi ini diharapkan dapat menjadi masukan bagi bagian HRD untuk memutuskan langkah-langkah lebih lanjut dalam penanganan pegawai di perusahaan.

TINJAUAN PUSTAKA

Knowledge Discovery dan Data Mining

Knowledge Discovery secara sederhana dapat dikatakan sebagai proses meng-ekstrak atau menggali (*mining*) pengetahuan/informasi yang berharga (*interesting knowledge*) dari sejumlah besar data baik yang disimpan di dalam *database*, *data warehouse* maupun media penyimpanan informasi lainnya. Sementara itu *data mining* merupakan salah satu tahap yang terdapat di dalam *Knowledge Discovery* [4, 8].

Discretization Histogram Analysis

Histogram menggunakan metode *binning* guna memperkirakan distribusi data dan juga untuk mereduksi jumlah data. Pada *histogram*, data dibagi – bagi menjadi subset – subset data yang dinamakan *bucket*. Sebuah *bucket* bisa berupa sebuah *value* tertentu atau bisa juga merepresentasikan *continuous ranges* dari sekelompok data [3].

MaxDiff Histogram adalah salah satu metode *Discretization Histogram*. Pada metode ini batasan tiap *bucket* ditentukan dengan cara memperhitungkan selisih dari sepasang data yang bersebelahan. Isi tiap *bucket* ditentukan dengan rumus $\beta - 1$. Selama selisih dua pasang data masih lebih kecil atau sama dengan $\beta - 1$, maka sekelompok data tersebut ada pada *bucket* yang sama, bila tidak, saatnya membuat *bucket* baru [4].

Pada penelitian ini, sebuah *bucket* yang dihasilkan oleh *MaxDiff Histogram* menjadi dasar pembuatan sebuah transaksi pada tabel *compact transaction*.

Market Basket Analysis

Market Basket Analysis merupakan sebuah analisis terhadap kebiasaan *customer* berbelanja pada supermarket dengan cara menemukan asosiasi dan korelasi di antara berbagai macam *item* yang dimasukkan *customer* di dalam *shopping basket* mereka. Secara lebih spesifik *Market Basket Analysis* bertujuan untuk mengetahui *item* apa saja yang sering dibeli bersamaan oleh *customer*. *Item* di sini diartikan sebagai berbagai macam produk atau barang pada supermarket [5, 6].

Association Rule Analysis

Association Rule Analysis adalah suatu prosedur dalam *Market Basket Analysis* untuk mencari hubungan antar *item* dalam suatu *data set* dan menampilkannya dalam bentuk *Association Rules* [4].

Contoh proses dari *Association Rules Analysis* dapat dilihat seperti berikut ini [4]:

Tabel 1. Contoh Compact Transaction dari buku Data Mining: Concepts and Techniques [4]

TID	List of item_IDs
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

Setelah proses pencarian *frequent Itemset* dan pembuatan *association rule* dilakukan, akan menghasilkan data *association rule* seperti pada Tabel 2:

Tabel 2. Association Rule yang dihasilkan dari Compact Transaction pada tabel 1 [4]

No.	Rule	No.	Rule
1.	$I1 \Rightarrow I2$	13.	$I1 \wedge I2 \Rightarrow I5$
2.	$I2 \Rightarrow I1$	14.	$I1 \wedge I5 \Rightarrow I2$
3.	$I1 \Rightarrow I3$	15.	$I2 \wedge I5 \Rightarrow I1$
4.	$I3 \Rightarrow I1$	16.	$I1 \Rightarrow I2 \wedge I5$
5.	$I1 \Rightarrow I5$	17.	$I2 \Rightarrow I1 \wedge I5$
6.	$I5 \Rightarrow I1$	18.	$I5 \Rightarrow I1 \wedge I2$
7.	$I2 \Rightarrow I3$	19.	$I1 \wedge I2 \Rightarrow I5$
8.	$I3 \Rightarrow I2$	20.	$I1 \wedge I5 \Rightarrow I2$
9.	$I2 \Rightarrow I4$	21.	$I2 \wedge I5 \Rightarrow I1$
10.	$I4 \Rightarrow I2$	22.	$I1 \Rightarrow I2 \wedge I5$
11.	$I2 \Rightarrow I5$	23.	$I2 \Rightarrow I1 \wedge I5$
12.	$I5 \Rightarrow I2$	24.	$I5 \Rightarrow I1 \wedge I2$

Dalam *association rule*, terdapat suatu *interestingness measure* (ukuran bahwa suatu informasi/ *knowledge* dianggap menarik) yang didapatkan dari perhitungan tertentu, yaitu [1, 2]: *Support*, yaitu suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu *item/itemset* dari keseluruhan transaksi.

$$\text{support}(A \Rightarrow B[s, c]) = p(A \cup B) = \text{support}(\{A, B\}) \quad (1)$$

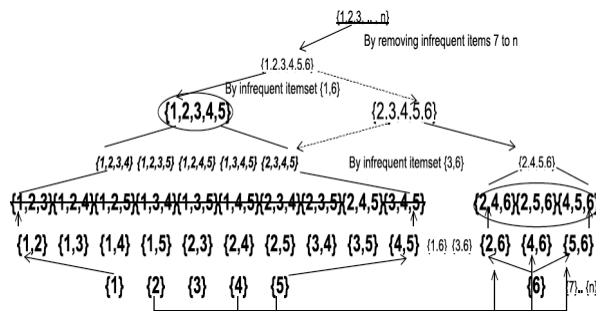
Confidence, yaitu suatu ukuran yang menunjukkan hubungan kondisional antar 2 item

$$\text{confidence}(A \Rightarrow B[s, c]) = p(B|A) = \frac{p(A \cup B)}{p(A)} = \frac{\text{support}(\{A, B\})}{\text{support}(\{A\})} \quad (2)$$

Kedua ukuran ini berguna untuk menentukan *interestingness* sebuah *association rule*. Dua batasan ini, yaitu *minimum support* dan *minimum confidence*, biasa digunakan dalam proses *Market Basket Analysis*.

Pincer Search

Pincer-Search dikembangkan oleh Dao-I Lin dan Zvi M. Kedem pada tahun 1997 [7]. Algoritma ini dinamakan *Two-Way Search* karena memakai 2 cara pendekatan, yaitu *top down* serta *bottom up*. Dalam prosesnya arah pencarian utama dari *Pincer search* adalah *bottom up*, sama seperti algoritma *Apriori*, kemudian secara *concurrent* dilakukan pencarian secara *top down*, yaitu *Maximum Frequent Candidat Set (MFCS)* yang menghasilkan output berupa *Maximum Frequent Set*. Untuk lebih jelasnya, proses *Pincer Search* dapat dilihat pada Gambar 1.



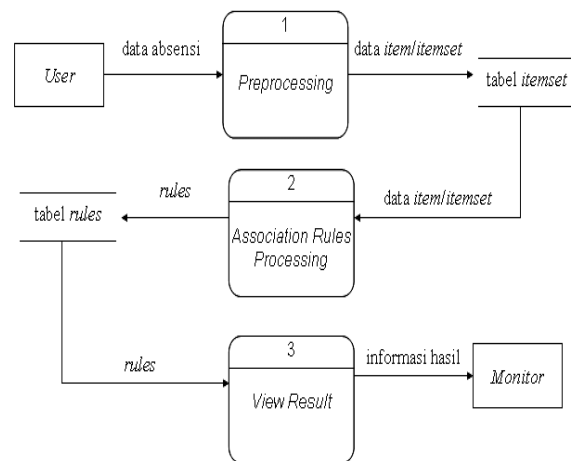
Gambar 1. Hasil generate data dari Pincer Search [7]

Keterangan gambar:

- * Angka bercetak tebal mewakili kandidat yang frequent
- * Angka bergaris bawah mewakili kandidat MFCS pertama yang infrequent
- * Angka bercetak miring adalah kandidat yang diprune dengan menggunakan MFCS
- * Tulisan dengan coretan ditengahnya mewakili kandidat yang frequent tapi dihilangkan dengan MFS, dan tidak dikembalikan oleh recovery procedure
- * Itemset dengan garis yang melingkarinya mewakili Maximum Frequent Itemset.

DESAIN PERANGKAT LUNAK

Perencanaan Perangkat Lunak secara keseluruhan dapat dilihat pada *Data Flow Diagram Level 0* pada Gambar 2.



Gambar 2. Data Flow Diagram Level 0 dari perangkat lunak

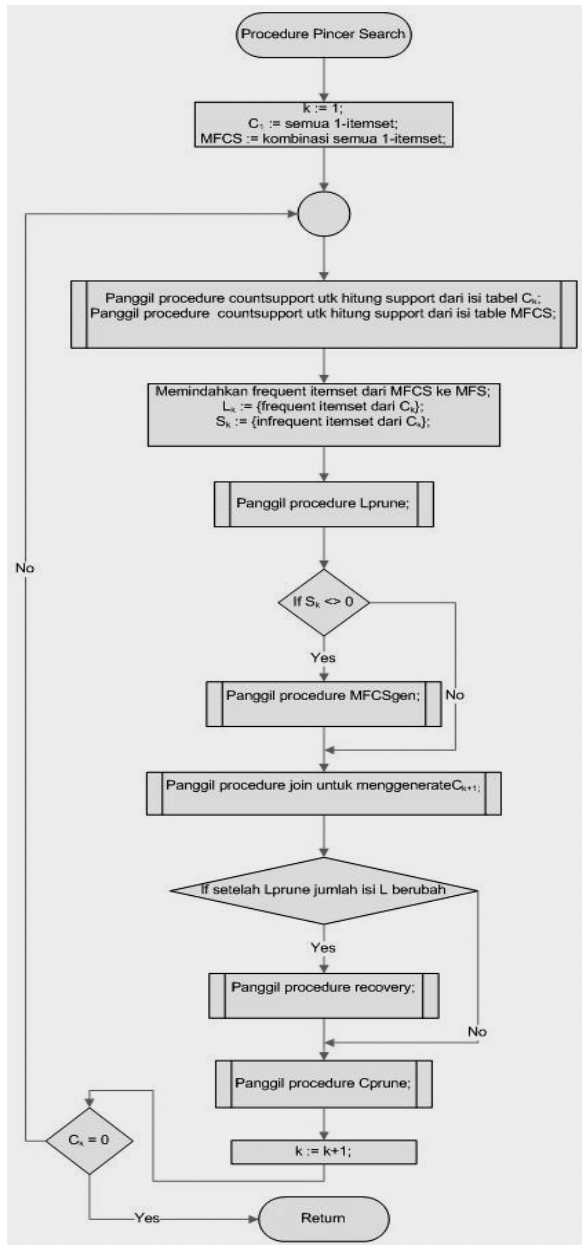
Penjelasan dari desain perangkat lunak yang dibuat adalah sebagai berikut:

Tahap 1: Preprocessing, pada tahap ini data mentah data absensi dapat dipilih oleh user untuk diproses pada bulan tertentu. Setelah itu, dengan didukung oleh data master pegawai, data absensi yang telah dipilih dirubah menjadi data transaksi dengan cara dikelompok-kelompokkan menggunakan metode *Discretization MaxDiff Histogram* berdasarkan waktu absensi pegawai yang berdekatan. Untuk pengelompokan ini user perlu memasukkan nilai β (beta) dalam satuan detik (maks 60 detik). Nilai β (beta) ini harus lebih kecil daripada waktu terpendek 2 orang pegawai melakukan absensi/*check-lock* secara bergantian. Setelah kelompok-kelompok transaksi terbentuk proses *cleaning* dapat dijalankan. Tujuan dari proses *cleaning* untuk menghapus data-data yang tidak valid. Proses terakhir dari tahap *preprocessing* ini adalah merubah data transaksi absensi menjadi *compact transaction data*.

Tahap 2: Association Rules Processing, tahap ini adalah tahapan utama pada perangkat lunak. Ada 2 macam proses pada tahap ini, proses pertama adalah menggali *frequent itemset* dari *compact transaction data* menggunakan algoritma *Pincer Search*. Untuk proses ini user perlu memasukkan nilai *minimum support*. Proses kedua adalah merubah *frequent itemset* yang telah digali menjadi *association rule*. Untuk membatasi jumlah rule yang dihasilkan, user perlu menentukan nilai *minimum confidence*.

Tahap 3: View Result, pada tahap ini *association rule* yang telah dihasilkan dari tahap ke-2 dapat dibuka kembali dan dilihat oleh user.

Desain tahap utama dari perangkat lunak, yaitu *Market Basket Analysis* menggunakan algoritma *Pincer Search*. Flowchart untuk *Pincer Search* pada Gambar 3.



Gambar 3. Flowchart dari Algoritma *Pincer Search*

PENGUJIAN PERANGKAT LUNAK

Setelah selesai dikembangkan, perangkat lunak yang dihasilkan perlu diuji apakah proses didalamnya telah sesuai dengan tujuan penelitian atau tidak. Untuk pengujian pada tahap ini, perangkat lunak diuji menggunakan data ideal yang ada pada Tabel 1.

Untuk membuat data *compact transaction ideal* agar dapat diproses oleh perangkat lunak, pertama-tama perlu dibuat data absensi imajiner. Data absensi

ini dibuat untuk tanggal 1 Nopember 2004 sampai dengan 5 Nopember 2004 dengan meminjam 5 NIP dan nama pegawai asli pada PT. MBS yang telah dikenal baik oleh peneliti dan diketahui secara pasti integritas dan loyalitasnya pada perusahaan, seperti terlihat pada Tabel 2.

Tabel 3. NIP & Nama Pegawai PT. MBS untuk Data Absensi Imajiner

NIP Pegawai	Nama	Item_ID yg diwakili
001308	Meidijanti	I1
001821	Tri Meidiawati	I2
004745	Ir. Tjahjo Wibowo	I3
000422	Henny Anggonowati	I4
000347	Noerhadi	I5

Tabel absensi imajiner yang dibuat, berisi 36 data absensi dari nomor 300001 sampai dengan 300036, dapat dilihat pada Gambar 4 dan Gambar 5.

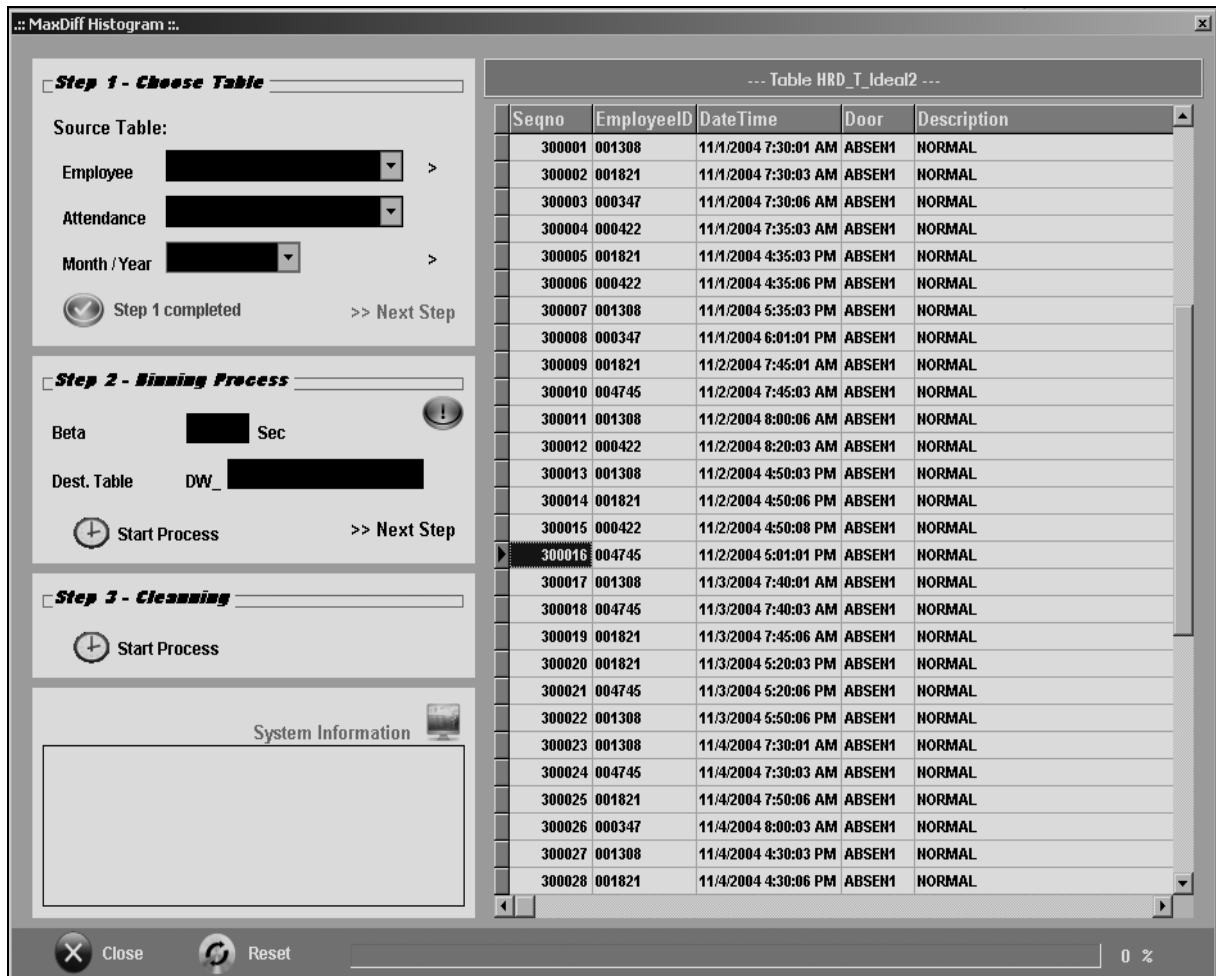
Setelah dilakukan proses *binning* dan *cleaning*, maka didapat data *compact transaction* seperti pada Gambar 6 dan Gambar 7.

Karena batasan bahwa data absensi valid adalah minimal 2 buah per pegawai, yaitu absen masuk dan absen pulang, maka jumlah transaksi yang dihasilkan bukan 9 transaksi, namun 22 transaksi. Hal ini tidak akan jumlah rule yang dihasilkan, karena untuk transaksi - transaksi yang hanya memiliki 1 buah *item* saja tidak akan diproses lebih lanjut menjadi *association rule*. Walaupun demikian nilai *support* dan *confidence* yang dihasilkan perangkat lunak akan berbeda dengan contoh buku, karena pada contoh buku kedua nilai tersebut dihitung terhadap 9 transaksi sementara pada perangkat lunak dihitung terhadap 22 transaksi. Rule yang dihasilkan dari contoh, sebanyak 24 rule, dapat dilihat pada Gambar 8 dan Gambar 9.

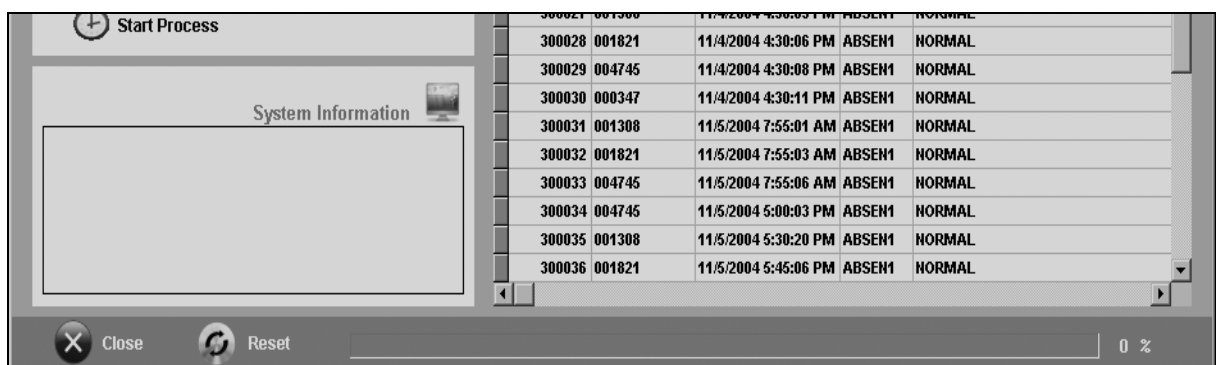
Sementara itu, nilai *support* dan *confidence* tiap - tiap rule dapat pula dilihat dalam bentuk grafik, seperti pada Gambar 10.

Setelah pengujian dengan data ideal dilakukan, dan perangkat lunak memberikan respon sesuai harapan, langkah selanjutnya adalah menguji perangkat lunak ini dengan data *real*, yaitu data absensi real PT. Mulia Batara Semesta pada tanggal 1 sampai dengan 15 Oktober 2004. Tujuan dari pengujian ini untuk melihat kemampuan perangkat lunak dalam memproses data yang besar. *Association rule* yang dihasilkan pengujian dapat dilihat pada Gambar 11.

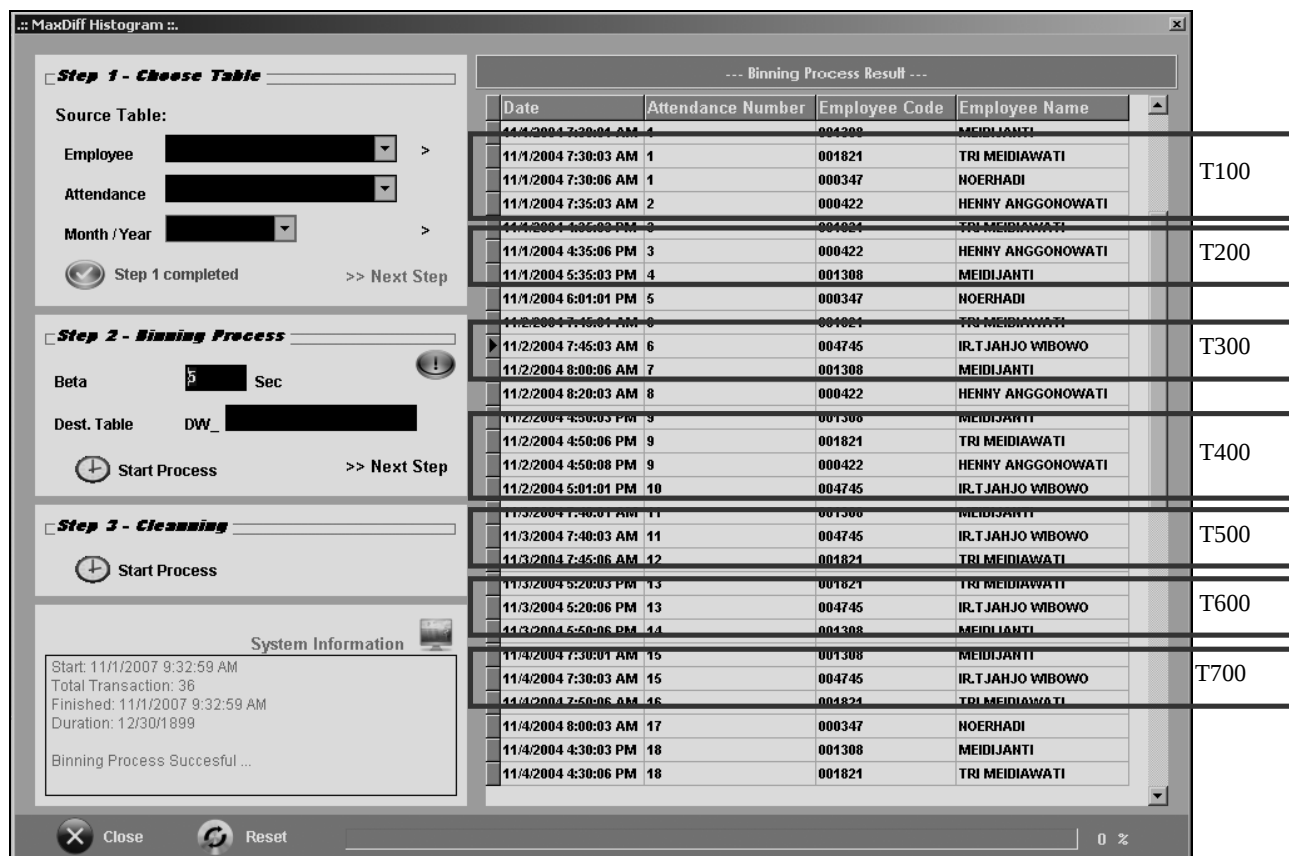
Pengujian selanjutnya yang dilakukan adalah melihat lama waktu proses masing-masing sub-proses dari perangkat lunak. Hasil pengujian dapat dilihat pada Tabel 4.



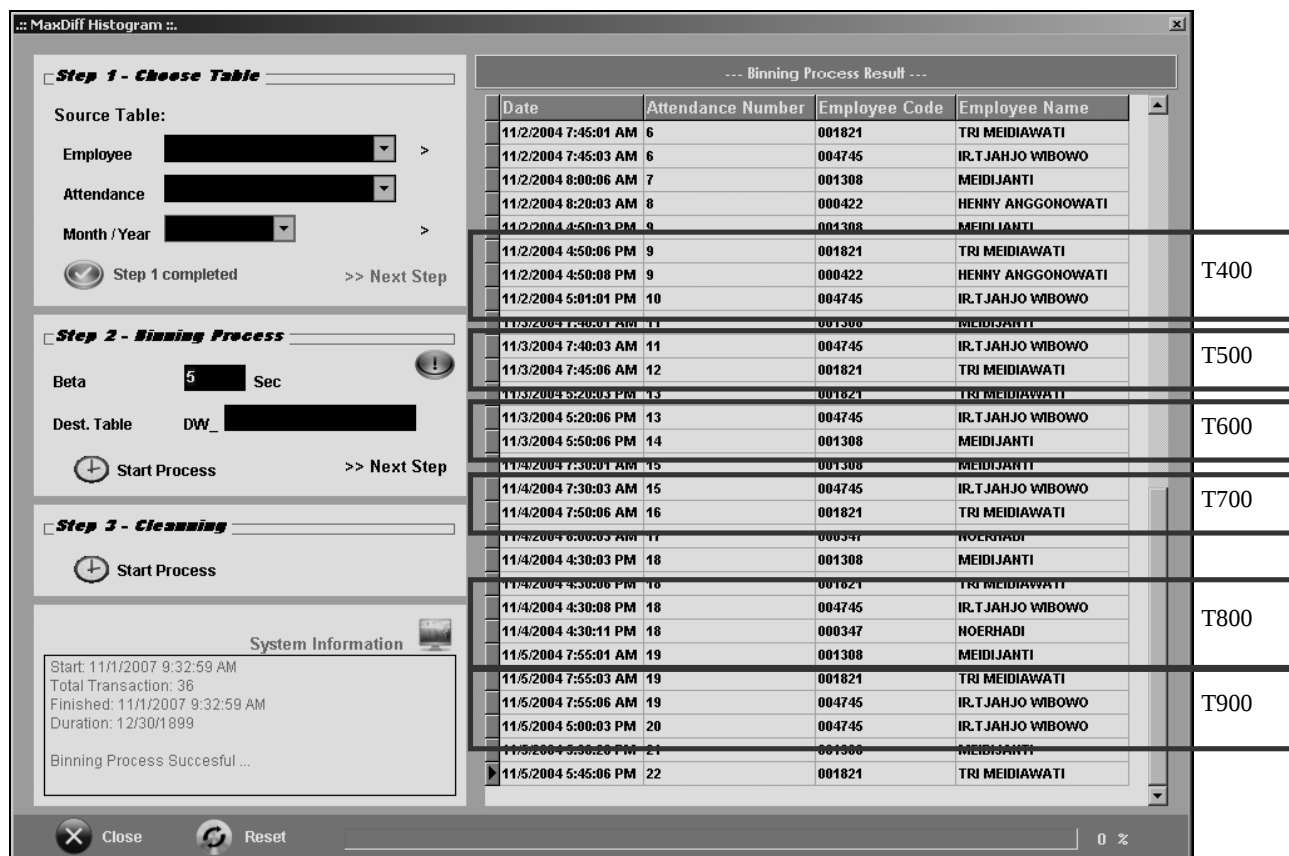
Gambar 4. Isi Tabel Absensi Imajiner



Gambar 5. Isi Tabel Absensi Imajiner (lanjutan)



Gambar 6: Compact Transaction yang Dihasilkan dari Proses Binning dan Cleaning



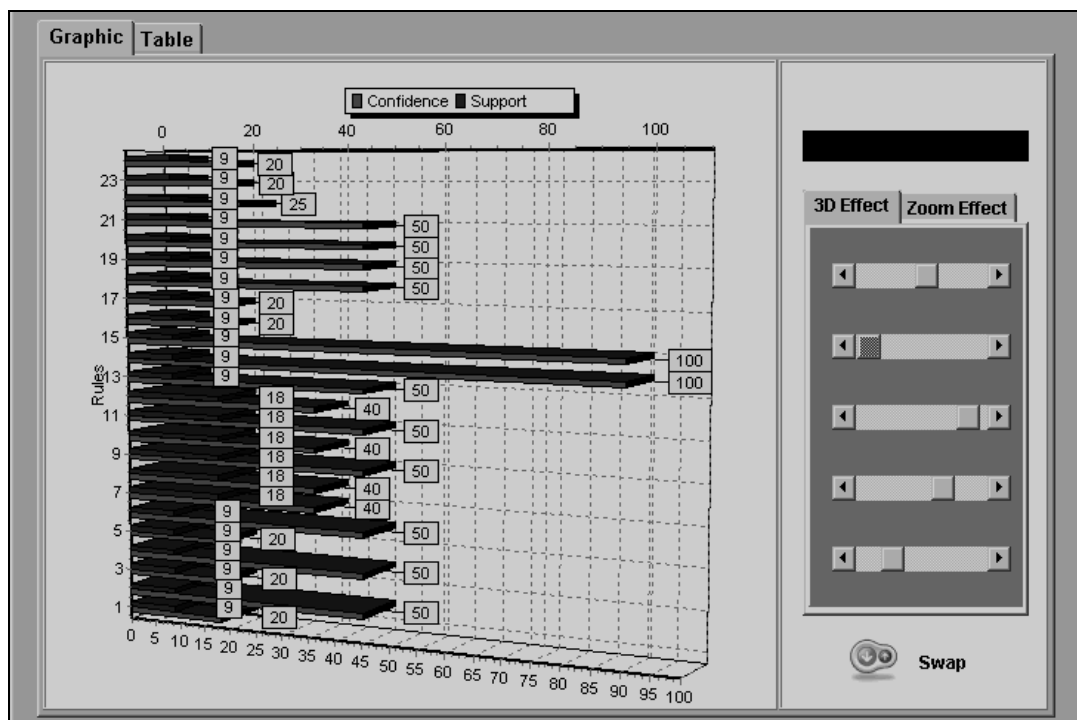
Gambar 7. Compact Transaction yang Dihasilkan dari Proses Binning dan Cleaning (lanjutan)

Graphic		Table			
id	comb	rules	supp_perc	conf_perc	
1	2	MEIDIJANTI -> NOERHADI	9	20	
2	2	NOERHADI -> MEIDIJANTI	9	50	
3	2	TRI MEIDIAWATI -> NOERHADI	9	20	
4	2	NOERHADI -> TRI MEIDIAWATI	9	50	
5	2	TRI MEIDIAWATI -> HENNY ANGGONOWATI	9	20	
6	2	HENNY ANGGONOWATI -> TRI MEIDIAWATI	9	50	
7	2	TRI MEIDIAWATI -> MEIDIJANTI	18	40	
8	2	MEIDIJANTI -> TRI MEIDIAWATI	18	40	
9	2	IR.TJAHJO WIBOWO -> MEIDIJANTI	18	50	
10	2	MEIDIJANTI -> IR.TJAHJO WIBOWO	18	40	
11	2	IR.TJAHJO WIBOWO -> TRI MEIDIAWATI	18	50	
12	2	TRI MEIDIAWATI -> IR.TJAHJO WIBOWO	18	40	
13	3	MEIDIJANTI & TRI MEIDIAWATI -> NOERHADI	9	50	
14	3	NOERHADI & TRI MEIDIAWATI -> MEIDIJANTI	9	100	
15	3	NOERHADI & MEIDIJANTI -> TRI MEIDIAWATI	9	100	
16	3	TRI MEIDIAWATI -> NOERHADI & MEIDIJANTI	9	20	
17	3	MEIDIJANTI -> NOERHADI & TRI MEIDIAWATI	9	20	
18	3	NOERHADI -> MEIDIJANTI & TRI MEIDIAWATI	9	50	
19	3	TRI MEIDIAWATI & IR.TJAHJO WIBOWO -> MEIDIJANTI	9	50	
20	3	MEIDIJANTI & IR.TJAHJO WIBOWO -> TRI MEIDIAWATI	9	50	
21	3	MEIDIJANTI & TRI MEIDIAWATI -> IR.TJAHJO WIBOWO	9	50	

Gambar 8. Hasil Association Rule Hasil Pengujian

19	3	TRI MEIDIAWATI & IR.TJAHJO WIBOWO -> MEIDIJANTI	9	50
20	3	MEIDIJANTI & IR.TJAHJO WIBOWO -> TRI MEIDIAWATI	9	50
21	3	MEIDIJANTI & TRI MEIDIAWATI -> IR.TJAHJO WIBOWO	9	50
22	3	IR.TJAHJO WIBOWO -> MEIDIJANTI & TRI MEIDIAWATI	9	25
23	3	TRI MEIDIAWATI -> MEIDIJANTI & IR.TJAHJO WIBOWO	9	20
24	3	MEIDIJANTI -> TRI MEIDIAWATI & IR.TJAHJO WIBOWO	9	20

Gambar 9. Hasil Association Rule Hasil Pengujian (lanjutan)



Gambar 10. Antar Muka untuk Melihat Nilai Support dan Confidence Tiap Rule dalam Grafik



Gambar 11. Association Rule Hasil Pengujian dengan Data Real Tanggal 1 sampai dengan 15 Oktober 2004 dengan Setting Minimum Confidence = 60%

Tabel 4. Hasil pengujian kecepatan proses dari perangkat lunak

Data Imaginer (1 Minggu)			
Proses		Lama Proses	Total Transaksi & Rule
Binning		1 detik	38
Cleaning		2 detik	38
Converting		1 detik	22
Gen. Frequent Itemset	Min. Support = 2	1 detik	8
Gen. Association Rules	Min. Conf. = 0%	1 detik	24
Data Real (1 Minggu)			
Binning		6 detik	988
Cleaning		13 detik	983
Converting		3 detik	566
Gen. Frequent Itemset	Min. Support = 6	4 jam 14 menit 59 detik	566
Gen. Association Rules	Min. Conf. = 50%	4 detik	2
Data Real (2 Minggu)			
Binning		8 detik	1964
Cleaning		17 detik	1957
Converting		5 detik	1102
Gen. Frequent Itemset	Min. Support = 10	6 jam 18 menit 46 detik	1102
Gen. Association Rules	Min. Conf. = 50%	19 detik	224

Kunci dari berhasil tidaknya perangkat lunak yang dibuat guna mendeteksi kecurangan absensi adalah dengan mengetahui waktu normal 2 orang pegawai melakukan proses absensi secara bergantian. Waktu tersebut nantinya digunakan untuk mengisi nilai β (beta) pada proses pengelompokan/binning. Untuk mengetahui hal itu peneliti melakukan pengamatan pada data absensi yang telah diurutkan berdasarkan waktunya. Dari hasil pengamatan didapat hal-hal sebagai berikut:

- Secara umum 2 buah data absensi yang berdekatan memiliki selisih waktu lebih dari 10 detik.
- Bila ada 2 buah data absensi berdekatan yang memiliki selisih waktu kurang dari 5 detik, nama dari pegawai yang absen pada 2 data tersebut sama.
- Beberapa data absensi yang memiliki selisih waktu absen kurang dari 10 detik dan nama pegawai kedua data tersebut tidak sama. Kedua pegawai itu, biasanya berada pada lokasi kerja yang sama atau berdekatan.

Dari ketiga fakta yang didapat dapat diasumsikan nilai β (Beta) yang disarankan untuk PT. MBS adalah 10 detik. Untuk lebih meyakinkan hal ini, peneliti menguji sendiri proses absensi pada PT. MBS. Ada 3

skenario yang dicoba yaitu:

- 1 orang melakukan 2 kali absensi dengan kartu pegawai yang sama. Langkah yang dilakukan: *Slashing* pertama langsung disusul dengan *slashing* kedua tanpa menunggu respon perangkat absensi. Selisih waktu rata-rata yang didapat di bawah 5 detik.
- 1 orang melakukan 2 kali absensi dengan 2 kartu pegawai yang berbeda. Langkah yang dilakukan: *Slashing* dengan kartu pertama, menunggu respon perangkat absensi sambil menukar kartu, setelah respon didapat melakukan *slashing* kedua. Selisih waktu rata-rata yang didapat di atas 5 detik dan di bawah 10 detik.
- 2 orang melakukan absensi secara bergantian. Langkah yang dilakukan: Orang pertama melakukan *slashing* dengan kartu pegawainya, menunggu respon dari perangkat absensi, setelah respon didapat segera meninggalkan perangkat absensi, orang kedua maju dan segera melakukan *slashing* dengan kartu pegawainya. Selisih waktu rata-rata yang didapat di atas 10 detik.

Hasil pengujian pada data *real* untuk $\beta = 10$ dapat dilihat pada Gambar 12 dan Gambar 13. *Minimum support* yang digunakan pada pengujian ini adalah 1.

Graphic		Table		
id	comb	rules	supp_perc	conf_perc
31	2	ACHMAD ASHORI -> SOEHARYO	6	100
32	2	SOEHARYO -> ACHMAD ASHORI	6	100
33	2	GUNADI -> IMAN SUCI ROSES	6	100
34	2	IMAN SUCI ROSES -> GUNADI	6	100
35	2	ACHMAD ZAKARIA -> SRI WAHYOE BODROWATI	6	100
36	2	SRI WAHYOE BODROWATI -> ACHMAD ZAKARIA	6	100
37	2	ANDRI ARIJADI -> MERY KUSUMAWARDHANI	6	100
38	2	MERY KUSUMAWARDHANI -> ANDRI ARIJADI	6	100
39	2	KRISTANTIONO -> HADI PADMONO	6	100
40	2	HADI PADMONO -> KRISTANTIONO	6	100
41	2	BAMBANG AGUNG BINTORO -> LUKMAN HAKIM	6	100
42	2	LUKMAN HAKIM -> BAMBANG AGUNG BINTORO	6	100
43	2	TEGUH SADEWO -> SUGIYONO	6	100
44	2	SUGIYONO -> TEGUH SADEWO	6	100
45	2	GATOT SUMARDIYANTO -> BUDI SANTOSO	6	100
46	2	BUDI SANTOSO -> GATOT SUMARDIYANTO	6	50
47	2	BAMBANG SUDJOKO -> BUDI SANTOSO	6	100
48	2	BUDI SANTOSO -> BAMBANG SUDJOKO	6	50
49	2	BAMBANG SUDJOKO -> GATOT SUMARDIYANTO	6	100
50	2	GATOT SUMARDIYANTO -> BAMBANG SUDJOKO	6	100

Gambar 12. Association Rule 2 Kombinasi dari Pengujian dengan Data *real* untuk $\beta = 10$

Graphic		Table		
id	comb	rules	supp_perc	conf_perc
86	3	M.RIFAI & ACHMAD ASHORI -> SOEHARYO	6	100
87	3	M.RIFAI & SOEHARYO -> ACHMAD ASHORI	6	100
88	3	ACHMAD ASHORI -> M.RIFAI & SOEHARYO	6	100
89	3	SOEHARYO -> M.RIFAI & ACHMAD ASHORI	6	100
90	3	M.RIFAI -> SOEHARYO & ACHMAD ASHORI	6	100
91	3	GATOT SUMARDIYANTO & BAMBANG SUDJOKO -> BUDI SANTOSO	6	100
92	3	BUDI SANTOSO & BAMBANG SUDJOKO -> GATOT SUMARDIYANTO	6	100
93	3	BUDI SANTOSO & GATOT SUMARDIYANTO -> BAMBANG SUDJOKO	6	100
94	3	BAMBANG SUDJOKO -> BUDI SANTOSO & GATOT SUMARDIYANTO	6	100
95	3	GATOT SUMARDIYANTO -> BUDI SANTOSO & BAMBANG SUDJOKO	6	100
96	3	BUDI SANTOSO -> GATOT SUMARDIYANTO & BAMBANG SUDJOKO	6	50
97	4	SUGIANTO & SUGIYONO & TEGUH SADEWO -> ABDOE SOEKOER.	6	100
98	4	ABDOE SOEKOER. & SUGIYONO & TEGUH SADEWO -> SUGIANTO	6	100
99	4	ABDOE SOEKOER. & SUGIANTO & TEGUH SADEWO -> SUGIYONO	6	100
100	4	ABDOE SOEKOER. & SUGIANTO & SUGIYONO -> TEGUH SADEWO	6	100
101	4	SUGIYONO & TEGUH SADEWO -> ABDOE SOEKOER. & SUGIANTO	6	100
102	4	SUGIANTO & TEGUH SADEWO -> ABDOE SOEKOER. & SUGIYONO	6	100
103	4	SUGIANTO & SUGIYONO -> ABDOE SOEKOER. & TEGUH SADEWO	6	100
104	4	ABDOE SOEKOER. & TEGUH SADEWO -> SUGIANTO & SUGIYONO	6	100
105	4	ABDOE SOEKOER. & SUGIYONO -> SUGIANTO & TEGUH SADEWO	6	100

Gambar 13. Association Rule 3 s/d 4 Kombinasi dari Pengujian dengan Data Real untuk $\beta = 10$

Dari hasil pengujian pada Gambar 12 dan gambar 13 dilakukan penelitian *manual* secara *sampling*, didapat fakta bahwa nama-nama pegawai yang muncul secara bersamaan pada *rule* adalah pegawai yang berlokasi kerja sama, berteman atau bahkan berstatus suami-istri.

KESIMPULAN DAN SARAN

Kesimpulan

Kesimpulan yang dapat ditarik dari penelitian ini adalah sebagai berikut:

1. Dari hasil pengujian perangkat lunak yang mendapatkan hasil sesuai dengan harapan, dapat disimpulkan bahwa metode *Market Basket Analysis* dapat pula dimanfaatkan untuk menggali *pattern* kebiasaan absensi (*check-lock*) pegawai sebuah perusahaan. Dari sini kemungkinan terjadinya kecurangan saat melakukan absensi masuk / pulang dapat dideteksi.
2. Bila nilai β (*beta*) pada saat proses *binning* menggunakan metode *MaxDiff Histogram* diset kurang dari kecepatan rata-rata 2 orang pegawai melakukan absensi secara bergantian (sekitar 10 detik), maka indikasi terjadinya kecurangan absensi dapat terlihat pada *association rule* yang

dihasilkan oleh perangkat lunak. Bila tidak, maka *rule* hanya dapat diartikan bahwa orang-orang yang namanya masuk dalam *rule* tersebut seringkali datang dan pulang bersama - sama.

3. Dari hasil pengujian kecepatan perangkat lunak dapat disimpulkan bahwa penggunaan algoritma *Pincer Search* untuk menggenerasi *frequent itemset* masih belum dapat mengatasi terjadinya 'bottleneck' pada proses tersebut.
4. Mengingat bahwa proses *generate frequent itemset* dilakukan paling cepat 1 minggu sekali, dapat disimpulkan bahwa perangkat lunak telah dapat diaplikasikan pada dunia nyata, terutama pada PT. Mulia Batara Semesta Surabaya.

Saran

1. Perangkat lunak ini dapat dikembangkan fleksibilitasnya terhadap *format* data sumber agar dapat digunakan untuk perusahaan - perusahaan lain.
2. Untuk mempercepat proses generasi *frequent itemset*, dapat dilakukan pengkodean ulang menggunakan kompiler yang lebih optimal dari *Delphi*, misal *Visual C++*.
3. Teknik pengkodean dapat lebih dioptimasi, misal dengan mengganti keterlibatan tabel *database* temporer menggunakan *array* atau *link list*.

DAFTAR PUSTAKA

1. C Agrawal, Rakesh, Tomasz Imielinsky, Arun Swami, *Mining Association Rules between Sets Of Items In Large Databases*, Proceeding Of The 1993 ACM SIGMOD Conference, Washington DC, USA, May 1993
2. Agrawal, Rakesh, Ramakrishnan Srikant, *Fast algorithms for mining association rules*. Proceeding Of The 1994 International Conference Very Large Data Bases, Santiago, Chile, September 1994.
3. Buccafurri, Francesco, Gianluca Lax, Domenico Sacca, Luigi Pontieri and Domenico Rosaci, *Enhancing Histograms by Tree-Like Bucket Indices*, Proceeding of the International Conference on Data Engineering 2002, IEEE Computer Society, 2002.
4. Han, Jiawei, Micheline Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 2001
5. Han, Jiawei, Yongjian Fu, "Discovery Of Multiple-Level Association Rules From Large Databases", *IEEE Transactions On Knowledge And Data Engineering*, Vol. 11, No. 5, September/October 1999
6. Hand, David, Heikki Mannila, Padhraic Smyth, *Principles Of Data Mining*, The MIT Press, 2001.
7. Lin, Dao-I, Zvi M. Kedem, "Pincer-Search: A New Algorithm for Discovering the Maximum Frequent Sets", Proceeding Of The 6th International Conference On Extending Database Technology (EDBT), March 1998.
8. Witten, Ian H., Eibe Frank, *Data Mining Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann Publishers, 2000.