



ANALISIS PERBANDINGAN ALGORITMA APRIORI DAN ALGORITMA HASH BASED PADA MARKET BASKET ANALYSIS DI APOTEK UAD

¹Orisky Sitra Arifah Destiyati (08018359), ²Eko Aribowo (0006027001)

^{1,2}Program Studi Teknik Informatika
Universitas Ahmad Dahlan

Prof. Dr. Soepomo, S.H., Janturan, Umbulharjo, Yogyakarta 55164

¹Email: oriskysad@gmail.com

²Email: ekoab@tif.uad.ac.id

ABSTRAK

Dalam aturan asosiasi data mining, terdapat beberapa algoritma diantaranya adalah algoritma apriori, algoritma hash base, algoritma fp growth, dan algoritma fuzy c covering. Masing-masing algoritma memiliki kelebihan dan kekurangan. Penelitian ini bertujuan untuk membandingkan kedua algoritma aturan asosiasi yaitu algoritma apriori dan algoritma hash based dengan studi kasus pemesanan obat di Apotek UAD dimana apotek tersebut secara periodik diharuskan menyuplai obat untuk mahasiswa yang akan melaksanakan KKN. Oleh karena itu, penelitian ini juga dilaksanakan untuk mengetahui kecenderungan mahasiswa KKN UAD dalam memesan obat untuk melaksanakan KKN, sehingga nantinya manager akan menyuplai obat-obat tertentu berdasarkan pemesanan mahasiswa yang akan melaksanakan KKN.

Penelitian dilakukan dengan membangun sebuah sistem yang menghasilkan waktu penambangan data dari masing-masing algoritma. Dari hasil pencatatan waktu tersebut dapat digunakan untuk menganalisis kecepatan masing-masing algoritma dalam menambang data. Teknik pengumpulan data dilakukan dengan wawancara dan observasi langsung ke Apotek UAD. Data diolah dan dikelompokkan berdasarkan merk masing-masing obat.

Hasil penelitian menunjukkan bahwa, pada beberapa uji coba yang telah dilakukan, diketahui algoritma apriori memiliki waktu yang lebih cepat dalam menambang data dibandingkan algoritma hash based. Dan telah diketahui kecenderungan pemesanan obat mahasiswa KKN, dengan nilai support lebih dari 30% dan nilai confidence lebih dari 70%.

Kata kunci : *Algoritma Apriori, Algoritma Hash Based, Market Basket Analysis*

1. PENDAHULUAN

Data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistic dan matematika [1]. Data mining menggunakan *discovery based* yaitu dimana pencocokan pola dan algoritma digunakan untuk menentukan relasi-relasi kunci di dalam data yang akan dieksplorasi. Dalam data mining terdapat beberapa teknik penambangan data, salah satunya adalah *association rule mining* atau dikenal dengan aturan asosiasi. *Association rule mining* adalah teknik *mining* untuk menemukan aturan asosiatif antara suatu kombinasi *item*. Dalam aturan asosiasi sendiri terdapat beberapa algoritma diantaranya adalah algoritma *apriori*, algoritma *hash based*, algoritma *fp growth*, dan algoritma *fuzzy c covering*. Masing-masing algoritma memiliki kelebihan dan kekurangan.

Dalam penelitian ini, dilakukan perbandingan antara dua algoritma aturan asosiasi, yaitu algoritma *apriori* dan algoritma *hash based*. Parameter yang digunakan adalah kecepatan masing-masing algoritma dalam menambang data. Selain itu, akan dilakukan penelitian pola konsumsi obat, dengan objek adalah data transaksi penjualan obat di sebuah apotek. Sebagai pembandingnya akan didasarkan pada pola pemesanan obat oleh mahasiswa UAD yang akan melaksanakan KKN. Dari hasil informasi tersebut, diharapkan dapat membantu pengelola apotek, dalam hal ini adalah *manager* dalam menentukan kebijakan-kebijakan yang nantinya dapat menguntungkan apotek.

2. KAJIAN PUSTAKA

2.1 Kajian Terdahulu

Penelitian yang dilakukan, mengacu pada penelitian terdahulu yang berjudul “Analisis *Market Basket* dengan Algoritma *Apriori* dan *FP-Growth*” [3]. Penelitian ini bertujuan untuk membandingkan kompleksitas waktu antara kedua algoritma yang digunakan. Dari penelitian ini, kesimpulan yang didapat adalah algoritma *apriori* membutuhkan waktu komputasi yang lama untuk mendapatkan *frequent itemsets* karena berulang kali melakukan pemindaian data. Sedangkan algoritma *fp-growth* hanya membutuhkan dua kali *scanning database* dalam mencari *frequent itemset* sehingga waktu yang dibutuhkan pun menjadi relatif singkat dan efisien.

Dalam penelitian dengan judul “Penggunaan Metode *Fuzzy C-Covering* untuk Analisa *Market Basket* pada Supermarket” [4], dibahas mengenai bagaimana performansi dan rule atau aturan yang dihasilkan algoritma *fuzzy c-covering*. Kesimpulan yang didapat dari kesimpulan ini adalah semakin kecil *minimum support* dan *confidence* yang ditentukan, semakin banyak pula *rule* yang dihasilkan, namun waktu yang diperlukan relatif lebih lama dibandingkan jika *minimum support* lebih besar. Selain itu ditemukan pula bahwa semakin tinggi jumlah kombinasi yang dicari, waktu yang dibutuhkan relatif lebih cepat. Hal ini terjadi karena semakin lama item yang tersaring semakin banyak sehingga kombinasi yang dicari semakin sedikit.

2.2 Landasan Teori

2.2.1. Algoritma Apriori

Algoritma *apriori* menggunakan *knowledge* mengenai *frequent itemset* yang telah diketahui sebelumnya, untuk memproses informasi selanjutnya. Pada algoritma *apriori* untuk menentukan kandidat-kandidat yang mungkin muncul dengan cara memperhatikan *minimum support*.

- Adapun dua proses utama yang dilakukan dalam algoritma *apriori*, yaitu [3]:
- Join* (Penggabungan)
Pada proses ini, setiap *item* dikombinasikan dengan *item* yang lainnya sampai tidak terbentuk kombinasi lagi.
 - Prune* (Pemangkasan)
Pada proses ini, hasil dari *item* yang telah dikombinasikan tadi lalu dipangkas dengan menggunakan *minimum support* yang telah ditentukan *user*.

Algoritma *apriori* dibagi menjadi beberapa tahapan yang disebut *iterasi*. Pada *iterasi* pertama dihasilkan pola frekuensi tinggi dengan panjang 1-*itemset*. Pada *iterasi* pertama pula, nilai *support* dari setiap *item* dihitung dengan melihat *database* yang ada. Setelah didapat nilai *support*, *item* yang memiliki *support* di atas *minimum support* (minsup) dipilih sebagai acuan untuk *iterasi* selanjutnya. *Iterasi* kedua menghasilkan 2-*itemset* yang tiap setnya memiliki dua *item*. Tiap kandidat 2-*itemset* dihitung nilai *support*nya dengan melihat *database* sebelumnya. *Support* disini dimaksudkan untuk jumlah transaksi dalam *database* yang mengandung kedua *item* dalam kandidat 2-*itemset*. Sama seperti 1-*itemset*, setelah didapat nilai *support*nya, maka kandidat 2-*itemset* yang memenuhi minsup dapat ditetapkan sebagai 2-*itemset* yang merupakan pola frekuensi tinggi.

Untuk selanjutnya pada *iterasi* ke-k dapat dibagi menjadi beberapa bagian [5] :

- Pembentukan kandidat *itemset*
Kandidat *k-itemset* dibentuk dari kombinasi (k-1)-*itemset* yang didapat dari *iterasi* sebelumnya.
- Penghitungan *support* dari tiap kandidat *k-itemset*
Support dari tiap kandidat *k-itemset* didapat dengan melakukan *scan database* untuk menghitung jumlah transaksi yang memuat semua *item* di dalam kandidat *k-itemset* tersebut.
- Tetapkan pola frekuensi tinggi
Pola frekuensi tinggi yang memuat *k-item* atau *k-itemset* ditetapkan dari kandidat *k-itemset* yang *support*nya lebih besar dari minsup.
- Bila tidak didapat pola frekuensi tinggi baru, maka seluruh proses dihentikan. Bila tidak, maka k ditambah satu dan kembali ke bagian a.

2.2.2. Algoritma Hash-Based

Algoritma *Hash Based* menggunakan teknik *hashing* untuk menyaring keluar *itemset* yang tidak penting untuk pembangkitan *itemset* selanjutnya. Ketika *support count* untuk kandidat *k-itemset* dihitung dengan menelusuri basis data, algoritma *hash based* mengumpulkan informasi mengenai (k+1)-*itemset* dengan cara seluruh kemungkinan (k+1)-*itemset* dihash ke dalam *hash table* dengan menggunakan fungsi *hash* (menggunakan sebuah bilangan prima untuk operasi modulo) [5].

Algoritma *Hash Based* terbagi menjadi tiga bagian utama yang masing-masing bagian melakukan proses yang berbeda. Bagian pertama akan menghasilkan kandidat 1-*itemset* yang disebut C_1 dan *large* 1-*itemset* yang disebut L_1 dari basis data. Untuk kandidat 1-*itemset*, seluruh transaksi ditelusuri untuk menghitung *support count* dari *itemset* ini. Pada tahap ini *hash tree* untuk C_1

dibangun dengan tujuan mengefisienkan penghitungan *support count*. Pada bagian ini juga algoritma akan membangun *hash table* (dengan fungsi *hash*) untuk 2 *itemset* yang akan berguna mengurangi banyaknya kandidat 2-*itemset* C_2 .

Pada bagian kedua, kumpulan kandidat *itemset* C_k dibangkitkan berdasarkan *hash table* yang telah dibuat pada *iterasi* sebelumnya. Lalu ditentukan *large itemset* L_k dan mengurangi ukuran basis data untuk pembangkitan *itemset* selanjutnya. Bagian algoritma ini terbagi menjadi dua fase. Fase pertama untuk membangkitkan kandidat *k-itemset* berdasarkan *hash table*. Fase kedua akan menghitung *support* pada kandidat *itemset* dan mengurangi ukuran dari setiap transaksi.

Bagian ketiga sama seperti bagian kedua tetapi tidak menggunakan *hash table* sehingga mirip dengan algoritma *apriori*. Bagian kedua dilakukan selama nilai *hash* buket lebih besar dari *minimum support*. Setelah batasan ini terlewati, algoritma *hash based* diganti dengan algoritma *apriori* [2].

Berikut penjelasan dari rumus *hash table* (*hash bucket*) [5].

$H(X,Y) = [(order\ of\ X) * penambahan\ ctr\ hash\ table + (order\ of\ y)] \text{ mod prima}$

Keterangan :

- *Order of X* adalah perwakilan nilai X
- Penambahan *ctr hash table* adalah nilai modulus bilangan prima, apabila terjadi *collision* nilai tersebut ditambah 1 (+ 1) sampai tidak terjadi *collision*.
- *Order of Y* adalah perwakilan nilai Y
- Prima adalah bilangan prima yang terdekat dan yang lebih besar dari jumlah kombinasi 2-*itemset* (C_2).

3. METODE PENELITIAN

Subyek penelitian dalam penelitian ini adalah dua algoritma data mining yaitu algoritma *apriori* dan algoritma *hash based* juga data transaksi obat di apotek uad. Adapun alat yang digunakan dalam penelitian adalah Borland Delphi 7 dan untuk database menggunakan My SQL. Data transaksi obat tersebut diperoleh langsung dari apoteker dengan melakukan wawancara dan observasi data obat. Selain itu dilakukan studi kepustakaan dengan menggunakan buku-buku dan referensi dari penelitian-penelitian yang telah dilakukan sebelumnya terhadap algoritma-algoritma dalam aturan asosiasi untuk membandingkan data dan fakta-fakta yang ada guna memperkuat landasan teori.

4. HASIL DAN PEMBAHASAN

Data utama yang digunakan pada penelitian ini adalah transaksi data obat yang terdiri dari 129 transaksi. Pada Tabel 1 dapat dilihat sampel data transaksi obat di apotek UAD. Sebelumnya data telah diinisiasi sehingga tidak memakan banyak ruang.

Tabel 1 : Tabel Transaksi Obat Apotek UAD

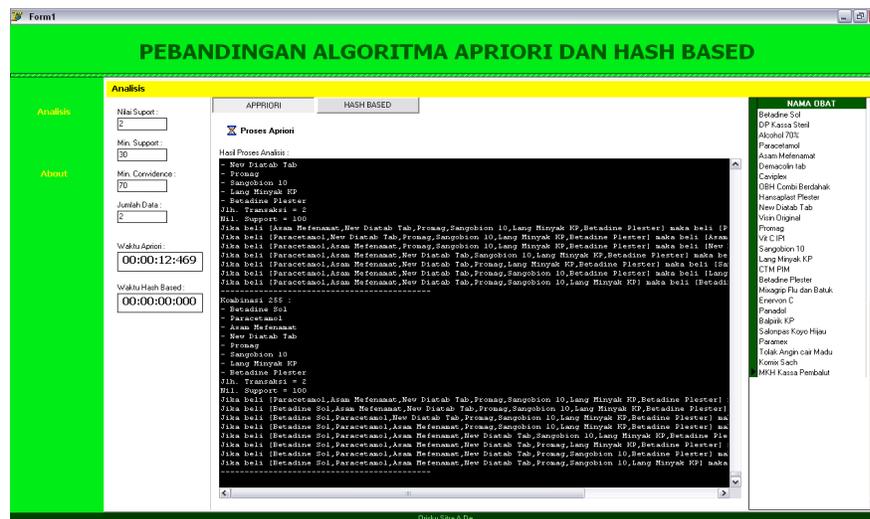
Transaks i	Item
1	BS,KS,ALK,PCM,AM,DMC,CAV,OBH,HP,NDT,VO,PMG,VIT C,SG,MKP,CTM,BP,MFB
2	AM,PCM,MKP,EVC,PND,PMG,BS,HP,BKP,SKH,PRM,NDT,T AM,KMK,SGB,KP,BP
3	MS,BS,HP,DCG,VIC,RHT,TAM,MYL,AS,BM,SGB,MKP,PND,

	ECS,BP
4	MKP,FTG,MFB,PND,BS,ASK,BP,KS,HP,MS,PMG,AM,NDT,S GB,OBH
5	MKP,MYL,ULT,GGB,BS,ASM,CTS,RTD,EPX,PCM,OBH,MK, NETS,AM,CAV,SYN,RHT,BP,DP

Setelah dilakukan *running* program, maka dihasilkan output sebagai berikut.

4.1. Hasil Algoritma *Apriori*

Dalam form ini terdapat parameter nilai *support* (ϵ), nilai *minimum support*, nilai *minimum confidence*, dan jumlah data. Parameter tersebut juga akan digunakan pada analisis algoritma *hash based*. Data yang ditampilkan pada form ini merupakan keseluruhan data yang memiliki nilai *support* dan *confidence* sebesar 100%, mulai dari kombinasi pertama hingga kombinasi terakhir yang dapat dilakukan. Hal tersebut mengindikasikan bahwa data obat tersebut adalah obat yang sering dipesan oleh mahasiswa dalam pelaksanaan KKN. Hasil penambangan data menggunakan algoritma *apriori* dapat dilihat pada Gambar 1.



Gambar 1: Form Analisis Algoritma *Apriori*

4.2. Hasil Algoritma *Hash Based*

Sama seperti pada algoritma *apriori*, Dalam form ini terdapat parameter nilai *support* (ϵ), nilai *minimum support*, nilai *minimum confidence*, dan jumlah data. Parameter tersebut juga akan digunakan pada analisis algoritma *hash based*. Data yang ditampilkan pada form ini merupakan keseluruhan data yang memiliki nilai *support* dan *confidence* sebesar 100%, mulai dari kombinasi pertama hingga kombinasi terakhir yang dapat dilakukan. Hal tersebut mengindikasikan bahwa data obat tersebut adalah obat yang sering dipesan oleh mahasiswa dalam pelaksanaan KKN. Hasil dari penambangan data menggunakan algoritma *hash based* dapat dilihat pada Gambar 2 di bawah ini.



Gambar 2: Form Analisis Algoritma Hash Based

4.3. Analisis Hasil

Hasil dari penelitian ini adalah dengan menjalankan sistem yang telah dibangun untuk melakukan penghitungan waktu dari kedua algoritma yang digunakan. Dengan menggunakan nilai minimum yang telah ditentukan sebesar 8 dan 9, hasil dari *running* program dapat dilihat pada Tabel 2 dan Tabel 3 di bawah ini.

Tabel 2 : Hasil Uji Coba I dengan nilai $\phi = 8$

Jumlah Data	Waktu Apriori	Waktu Hash Based
2	01 : 19 : 547	03 : 31 : 390
5	00 : 01 : 844	00 : 16 : 156
8	00 : 00 : 297	00 : 30 : 406
48	00 : 02 : 183	01 : 38 : 764
53	00 : 02 : 946	02 : 56 : 218
58	00 : 04 : 527	05 : 20 : 813
96	00 : 05 : 642	06 : 51 : 219
112	00 : 06 : 108	07 : 38 : 813
129	00 : 06 : 265	08 : 54 : 421

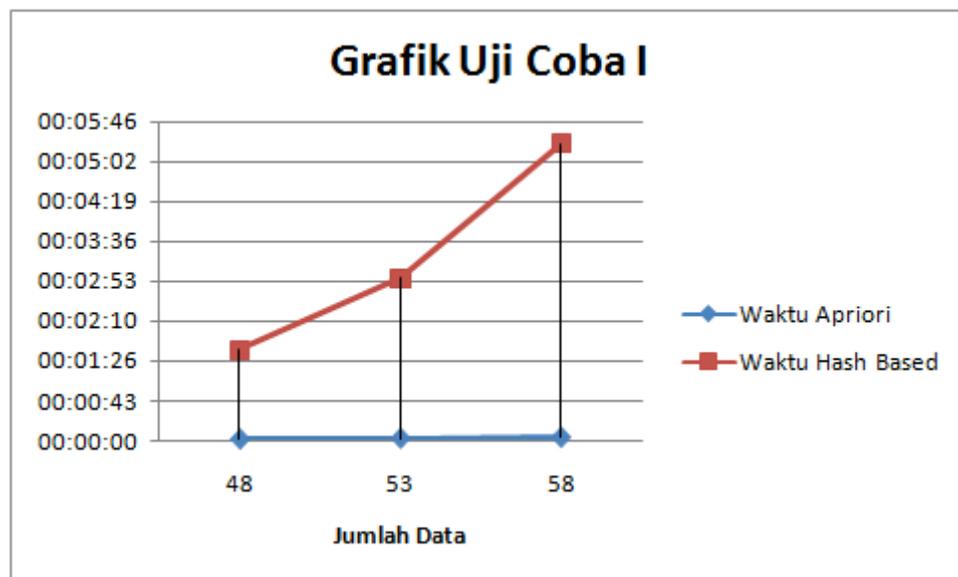
Tabel 3 : Hasil Uji Coba II dengan nilai $\phi = 9$

Jumlah Data	Waktu Apriori	Waktu Hash Based
2	01 : 58 : 219	05 : 00 : 235
5	00 : 01 : 906	00 : 22 : 656
8	00 : 00 : 328	00 : 37 : 937
48	00 : 06 : 824	08 : 19 : 590

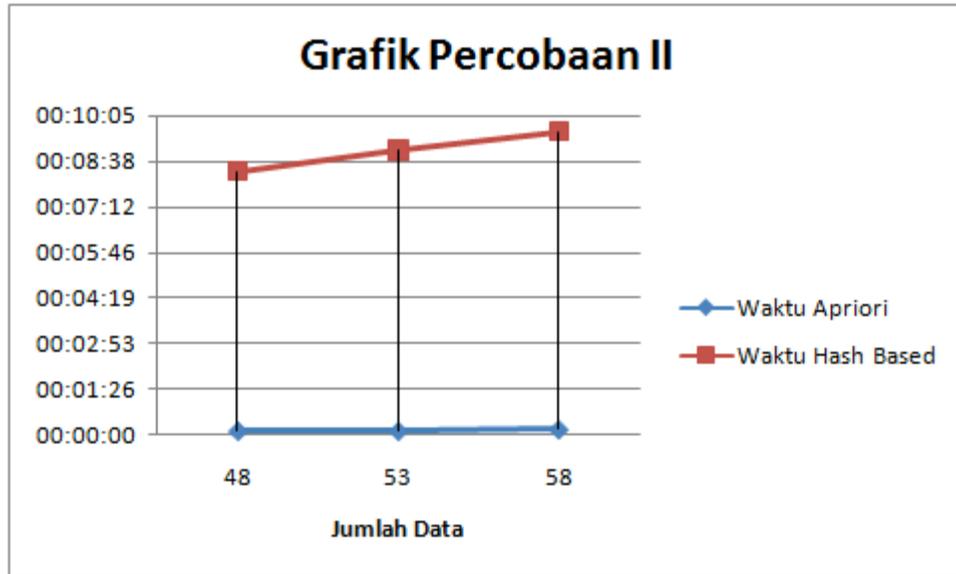
53	00 : 07 : 540	08 : 58 : 942
58	00 : 09 : 754	09 : 34 : 670
96	00 : 10 : 183	10 : 12 : 408
112	00 : 10 : 920	12 : 31 : 917
129	00 : 12 : 057	12 : 58 : 672

Dapat dilihat dari kedua percobaan di atas waktu yang dibutuhkan algoritma *apriori* lebih cepat dibandingkan waktu yang dibutuhkan algoritma *hash based*. Pada algoritma *apriori* waktu yang digunakan untuk penambangan data menunjukkan masih berada di hitungan detik, sedangkan pada algoritma *hash based* waktu menunjukkan berada di hitungan menit.

Selain itu, dari hasil kedua percobaan yang dapat dilihat pada Tabel 2 dan Tabel 3, dibuat juga grafik yang mewakili masing-masing percobaan. Grafik dapat dilihat pada Gambar 3 dan Gambar 4 di bawah ini.



Gambar 3 : Grafik Percobaan I



Gambar 4 : Grafik Percobaan II

Selain perbandingan waktu yang dapat dilihat pada tabel-tabel percobaan dan grafik di atas, dapat dilihat pula dari hasil penambangan data yang dilakukan oleh kedua algoritma yang digunakan, kecenderungan pemesanan obat di Apotek UAD yang dilakukan oleh mahasiswa KKN. Hasil tersebut dapat dilihat pada Gambar 9 dan Gambar 10 di bawah ini.

```

Hasil Proses Analisis :
-----
Kombinasi 4 :
- Betadine Sol
- Lang Minyak KP
Jlh. Transaksi = 9
Nil. Support = 100
Jika beli [Lang Minyak KP] maka beli [Betadine Sol] ==> NKonv = 100
Jika beli [Betadine Sol] maka beli [Lang Minyak KP] ==> NKonv = 100
-----
Kombinasi 5 :
- Betadine Sol
- Betadine Plester
Jlh. Transaksi = 8
Nil. Support = 88,89
Jika beli [Betadine Plester] maka beli [Betadine Sol] ==> NKonv = 88,89
Jika beli [Betadine Sol] maka beli [Betadine Plester] ==> NKonv = 100
-----
Kombinasi 6 :
- Lang Minyak KP
- Betadine Plester
Jlh. Transaksi = 8
Nil. Support = 88,89
Jika beli [Betadine Plester] maka beli [Lang Minyak KP] ==> NKonv = 88,89
Jika beli [Lang Minyak KP] maka beli [Betadine Plester] ==> NKonv = 100
-----
Kombinasi 7 :
- Betadine Sol
- Lang Minyak KP
- Betadine Plester
Jlh. Transaksi = 8
Nil. Support = 88,89
Jika beli [Lang Minyak KP,Betadine Plester] maka beli [Betadine Sol] ==> NKonv = 88,89
Jika beli [Betadine Sol,Betadine Plester] maka beli [Lang Minyak KP] ==> NKonv = 88,89
Jika beli [Betadine Sol,Lang Minyak KP] maka beli [Betadine Plester] ==> NKonv = 100
-----

```

Gambar 9 : Hasil Penambangan Algoritma *Apriori*

```

Hasil Proses Analisis :
Kombinasi 3 :
- Betadine Plester
Jlh. Transaksi = 8
Nilai Support = 88,888888888889%
Nilai Conf = 100,00%
-----
Kombinasi 4 :
- Betadine Sol
- Lang Minyak KP
Jlh. Transaksi = 9
Nilai Support = 88,888888888889%
Nilai Conf = 88,89%
-----
Kombinasi 5 :
- Betadine Sol
- Betadine Plester
Jlh. Transaksi = 8
Nilai Support = 88,888888888889%
Nilai Conf = 100,00%
-----
Kombinasi 6 :
- Lang Minyak KP
- Betadine Plester
Jlh. Transaksi = 8
Nilai Support = 88,888888888889%
Nilai Conf = 100,00%
-----
Kombinasi 7 :
- Betadine Sol
- Lang Minyak KP
- Betadine Plester
Jlh. Transaksi = 8
Nilai Support = 88,888888888889%
Nilai Conf = 100,00%
-----

```

Gambar 10: Hasil Penambangan Algoritma *Hash Based*

Dari tabel-tabel percobaan, grafik, serta hasil penambangan data terlihat bahwa :

- Waktu yang digunakan oleh algoritma *apriori* dalam melakukan penambangan data relatif lebih pendek dari waktu yang digunakan oleh algoritma *hash based*. (Waktu algoritma *apriori* masih dalam hitungan detik untuk menghasilkan keputusan, sedangkan algoritma *hash based* sudah memasuki hitungan menit untuk mengdailkan keputusan)
- Waktu yang dibutuhkan algoritma *hash based* relatif lebih lama dikarenakan dalam melakukan penambangan data, terdapat langkah yaitu ketika mencari alamat *hash* jika terjadi *collision* pada alamat tiap *item*, maka proses dilakukan secara berulang sampai tidak terjadi *collision*.
- Kecenderungan mahasiswa KKN UAD dalam memesan obat di Apotek UAD adalah Betadine sol, Lang minyak kayu putih, dan Betadine plester dengan support sebesar 88% dan confidence sebesar 100%.

5. PENUTUP

5.1. Kesimpulan

Karena pada algoritma *hash based* terjadi *collision*, maka waktu yang dibutuhkan dalam menambang data lebih lama dibandingkan waktu yang dibutuhkan algoritma *apriori*. Selain itu setelah langkah menempatkan *item* pada *hash tabel* sesuai dengan alamatnya, langkah yang dilakukan *hash based* sama dengan langkah yang dilakukan *apriori*. Hal ini membuat algoritma *hash based* tidak lebih efisien dibandingkan algoritma *apriori*.

5.2. Saran

Saran dalam pelaksanaan penelitian ini adalah sebagai berikut :

- Dapat dilakukan penelitian dengan membandingkan masing-masing algoritma dengan algoritma lain (misalnya algoritma *Hash Based* dan Algoritma *FP Growth*)
- Penelitian dapat dilakukan dengan volume data yang lebih besar dan dengan data transaksi lain (misalnya data transaksi pada toserba/supermarket).



6. DAFTAR PUSTAKA

- [1] Kusriani. 2009. *Algoritma Data Mining*. CV Andi Offset. Yogyakarta
- [2] Yulita Marsela, S. Moertini Veronica. 2004. Analisis Keranjang Pasar Dengan Algoritma Hash-Based Pada Dara Transaksi Penjualan Apotek. Skripsi. Jurusan Ilmu Komputer. Universitas Katolik Parahyangan. Bandung.
- [3] Analisis Market Basket dengan Algoritma Apriori dan Algoritma Hash Based <http://eprints.unsri.ac.id/83/1/6-Erwin.pdf>(diunduh pada tanggal 18 April 2012)
- [4] Penggunaan Metode Fuzzy C-Covering untuk Analisa Market Basket pada Supermarket <http://puslit2.petra.ac.id/gudangpaper/files/1091.pdf>(diunduh pada tanggal 20 April 2012)
- [5] Setyawan, Bambang. 2011. Analisa Keranjang Pasar Pada Toko Apotek Menggunakan Algoritma Hash Based. Skripsi. Jurusan Ilmu Komputer. Universitas Pembangunan Nasional Veteran Jakarta. Jakarta.