**PROSES MENEMUKAN POLA DATA BISNIS RITAIL DENGAN PENGEMBANGAN ALGORITMA APRIORI MENGGUNAKAN PENDEKATAN *K-WAY JOIN***

Joko Aryanto, S.Kom., M.Kom.

Informatika, Fakultas Teknologi Informasi dan Elektro Universitas Teknologi Yogyakarta

Jl.Siliwangi(Ring Road Utara), Jombor, Sleman, Yogyakarta Joko.Aryanto@uty.ac.id

**ABSTRAK**

Persaingan sesama pelaku bisnis retail membuat para pemilik usaha ritel harus bersaing satu sama lain untuk memperoleh pelanggan guna meningkatkan pendapatan, dengan demikian pengelola dituntut untuk melakukan improvisasi dengan langkah yang efektif dan efisien. Algoritma Apriori merupakan metode dasar di dalam *data mining*. Prinsip yang digunakan algoritma Apriori adalah jika sebuah *itemset* sering muncul, maka seluruh *subset* dari *itemset* tersebut juga harus sering muncul. Hal ini mengakibatkan pengecekan yang berulang-ulang dan akan membutuhkan waktu yang tidak sebentar. Sebuah metode diusulkan untuk dapat memecahkan masalah ini, yaitu dengan melakukan pengembangan atau modifikasi pada algoritma Apriori tersebut, khususnya dalam proses penghitungan nilai *support* dan *confidance*. Metode yang digunakan adalah memanipulasi penggunaan bahasa query memanfaatkan teknik *k-Way* untuk optimasi susunan yang sesuai.

Hasil yang didapat pada penelitian ini adalah waktu eksekusi relatif lebih cepat, dan hasil aturan asosiasi yang sama dengan yang dihasilkan oleh metode Apriori tanpa adanya pengembangan atau modifikasi.

Kata kunci : *Data Mining, Improvisasi Apriori, Algoritma Association rule*, *Pengembangan Apriori, k-way*

***ABSTRACT***

Competition among retail businesses makes retail business owners have to compete with each other to get customers to increase revenue, thus managers are required to improvise with effective and efficient steps. The Priori algorithm is the basic method in data mining. The principle used by the Apriori algorithm is that if an itemset appears frequently, then all subset of itemset must also appear frequently. This results in repeated checks and will require a short time. A method is proposed to be able to solve this problem, namely by developing or modifying the Apriori algorithm, especially in the process of calculating the value of support and confidence. The method used is to manipulate the use of query languages using the k-Way technique for optimization of the appropriate arrangement. The results obtained in this study are relatively faster execution times, and the results of the association rules are the same as those produced by the Priori method without any development or modification.

Keywords: Data Mining, Priori Improvisation, Association rule Algorithm, Priori Development, k-way

# PENDAHULUAN

Metode *Market Basket Analysis* adalah proses analisa pada level data transaksi yang dapat meningkatkan produktifitas bisnis. Gambarannya berupa kecendrungan kelompok item (terdiri dari 2 item atau lebih) yang dibeli oleh seorang konsumen dalam satu transaksi penjualan. Metode ini mampu menemukan sebuah trend penjualan, pengelompokan produk umum atau mudah ditebak dan bahkan dapat menemukan trend yang tidak pernah terpikirkan sebelumnya. *Market Basket Analysis* didasarkan pada analisis dari kebiasaan konsumen dengan mencari asosiasi dan korelasi antar item berbeda yang dilakukan pada satu transaksi yang sama oleh konsumen[1].

## Data Mining

Data mining merupakan proses komputasi untuk menemukan pola dalam data set yang besar melibatkan interseksi metode dari kecerdasan buatan, pembelajaran mesin, statistik, dan sistem database. Tujuan keseluruhan dari proses data mining adalah untuk mengekstrak informasi dari kumpulan data dan mengubahnya menjadi struktur yang dapat dimengerti untuk digunakan lebih lanjut. Selain dari langkah analisis mentah, melibatkan database dan manajemen data aspek, data yang pra-pengolahan, model dan pertimbangan inferensi, metrik interestingness, pertimbangan kompleksitas, pasca-pengolahan struktur ditemukan, visualisasi, dan memperbarui secara online [17].

Data mining merupakan bagian dari *Knowledge Discovery In Databases* (KDD). Sedangkan data mining memiliki tiga metode, yakni *clustering, classification* dan *Association Rule*. KDD adalah proses multi langkah yang mengubah data mentah menjadi pengetahuan yang berguna (Bagga and Singh, 2011). Tahapan-tahapan yang terjadi pada proses *data mining* atau *knowledge discovery* menurut Kenneth Collier (1998) dibagi menjadi 5 tahapan yaitu :



## Seleksi Data

Tujuan dari fase ini adalah ekstraksi dari gudang data yang besar menjadi data yang relevan dengan analisis *data mining*. Proses ekstraksi data

membantu untuk merampingkan dan mempercepat proses.

### *Data Preprocessing*

Fase ini berkaitan dengan pembersihan data dan persiapan tugas yang diperlukan untuk memastikan hasil yang benar. Menghilangkan *missing value* dalam data, memastikan bahwa nilai-nilai kode memiliki arti seragam dan memastikan bahwa tidak ada nilai data palsu adalah tindakan khas yang terjadi selama fase ini.

1. **Transformasi *Data***

Tahap ini mengubah data ke dalam bentuk atau format yang sesuai untuk kebutuhan *data mining*. Proses normalisasi biasanya diperlukan dalam tahap data transformas.

### *Data mining*

Tujuan dari tahap *data mining* adalah untuk menganalisis *database* sesuai algoritma yang digunakan sehingga menemukan pola atau aturan yang bermakna serta menghasilkan model prediksi. *Data mining* adalah elemen inti dari siklus KDD.

## Interpretasi dan Evaluasi

Sementara algoritma *data mining* memiliki potensi untuk menghasilkan jumlah yang tidak terbatas dari pola tersembunyi dalam data, banyak hasil dari proses tersebut mungkin tidak bermakna atau berguna. Tahap akhir ini bertujuan untuk memilih model-model yang valid dan berguna untuk membuat keputusan bisnis masa depan.

## Aturan Asosiasi

*Association rule mining* merupakan salah satu teknik yang digunakan dalam data mining yang bertujuan untuk menemukan aturan- aturan asosiasi yang terdapat dalam sekumpulan data. Teknik *association rule mining* akan menemukan pola hubungan antar data yang saling berasosiasi. Teknik ini juga dikenal dengan nama *market basket analysis* yang mendefinisikan itemset dalam suatu transaksi secara bersamaan, Contohnya dalam proses transaksi penjualan barang. Prosesnya diawali dengan mencari sejumlah *frequent itemset* dengan dilanjutkan pembentukan aturan-aturan asosiasi yang diperoleh [18].

Aturan yang menyatakan asosiasi antara beberapa atribut sering disebut *affinity analysis* atau *market basket analysis*. Metodologi dasar analisis asosiasi terbagi menjadi dua tahap:

1. Analisa pola frekuensi tinggi

Tahap ini mencari kombinasi item yang memenuhi syarat minimum dari nilai support dalam database.

Nilai support sebuah item diperoleh dengan rumus berikut:

𝐽𝑢𝑚𝑙𝑎ℎ 𝑇𝑟𝑎𝑛𝑠𝑎𝑘𝑠𝑖 𝑀𝑒𝑛𝑔𝑎𝑛𝑑𝑢𝑛𝑔 𝐴

𝑆𝑢𝑝𝑝𝑜𝑟𝑡 (𝐴) =

𝑇𝑜𝑡𝑎𝑙 𝑇𝑟𝑎𝑛𝑠𝑎𝑘𝑠𝑖

Sedangkan nilai support yang mengandung 2 itemset dapat dituliskan dalam rumus sebagai berikut :

𝑆𝑢𝑝𝑝𝑜𝑟𝑡 (𝐴ᴖ𝐵) = 𝐽𝑚𝑙 𝑇𝑟𝑎𝑛𝑠𝑎𝑘𝑠𝑖 𝑀𝑒𝑛𝑔𝑎𝑛𝑑𝑢𝑛𝑔 𝐴 𝑑𝑎𝑛 𝐵

𝑇𝑜𝑡𝑎𝑙 𝑇𝑟𝑎𝑛𝑠𝑎𝑘𝑠𝑖

merupakan item ke- j dari l1, dapat ditunjukkan sebagai berikut :

Di mana l1 dan l2 adalah k-item dari bagian kumpulan set Lk, l1[k] ≠ l2[k] adalah untuk memastikan dapat menghasilkan duplikat itemsets k-. Itemset yang dihasilkan oleh l1 dan l2 koneksi sebagai berikut:

1. Langkah pemangkasan : Untuk memilih itemset yang sering muncul Lk+1 berasal dari kandidat Ck+1 , karena kandidat set Ck + 1 merupakan superset dari
2. Pembentukan aturan asosiasi

Setelah semua pola frekuensi tinggi itemukan, langkah berikutnya dilakukan pencarian aturan assosiatif yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan assosiatif A >B. Nilai confidence dari aturan A >B diperoleh dari rumus berikut:

𝑆𝑢𝑝𝑝𝑜𝑟𝑡 𝑃(𝐵/𝐴) = 𝐽𝑢𝑚𝑙𝑎ℎ 𝑇𝑟𝑎𝑛𝑠𝑎𝑘𝑠𝑖 𝑀𝑒𝑛𝑔𝑎𝑛𝑑𝑢𝑛𝑔 𝐴 𝑑𝑎𝑛 𝐵

𝑇𝑜𝑡𝑎𝑙 𝑇𝑟𝑎𝑛𝑠𝑎𝑘𝑠𝑖 𝑚𝑒𝑛𝑔𝑎𝑛𝑑𝑢𝑛𝑔 𝐴

Analisis asosiasi didefenisikan suatu proses untuk menemukan semua aturan asosiasi yang memenuhi syarat minimum untuk *support (minimum support)* dan syarat minimum untuk *confidence* (*minimum confidence*).

## Algoritma Apriori

Dari penelitian yang dilakukan D. M. Tank, dijelaskan bahwa algoritma Apriori adalah algoritma klasik untuk memperoleh aturan asosiasi dari data yang sering muncul. Ide dasar dari algoritma ini adalah dengan menggunakan pendekatan berulang lapisan demi lapis untuk menemukan yang frekuensi kemunculan yang sering muncul. Proses pertama algoritma akan mendapatkan k-itemset, dan kemudian menggunakan itemsets k- untuk mengeksplorasi (k + 1) itemset. Pertama, memperkenalkan pengetahuan apriori itemset yang sering muncul, setiap bagian dari frekuensi itemset juga itemset yang sering muncul. Algoritma Apriori menggunakan pengetahuan sebelumnya dari itemset yang sering muncul, pertama dengan menemukan koleksi keseringan 1-itemset, dinotasikan

1. Kemudian gunakan 2 itemset L1 untuk mendapatkan L2, dan kemudian L, dan seterusnya, sampai tidak dapat ditemukan frekuensi k-itemset. Algoritma Apriori terutama terdiri dari tiga langkah berikut :

**1.** Langkah menghubungkan : menghubungkan k- frekuensi itemset untuk menghasilkan kandisdat (k

+ 1), dinotasikan dengan Ck+1. Kondisi selanjutnya dari langkah menghubungkan adalah bahwa dua k- itemset memiliki item pertama (k-1) yang sama dan item k-selanjutnya yang berbeda. Ditunjukkan l1[j]

itemset L yang sering muncul. Menurut sifat Apriori: setiap himpunan bagian dari frekuensi set juga harus sering muncul, yaitu setiap (k-1) item himpunan bagian dari k-item juga harus sering muncul. Dengan properti ini kita dapat mengetahui apakah item k- subset dari Ck + 1 berada di Lk + 1, jika tidak, maka hapus calon (k + 1) - itemset dihapus dari C.

1. Langkah menghitung : scanning data dalam database, menumpuk jumlah calon k + 1 muncul dalam database. Jika jumlah kemunculan dari calon kurang dari ambang batas minimum *support* yang diberikan, calon itemset akan dihapus[9].

Algoritma yang digunakan dalam pencarian pola antara lain Apriori, FP-Growth, dan CT- Pro. Penggunaan Algoritma Apriori dikenal dengan metode asosiasi, dengan ide dasar menghitung pola kemunculan *item* yang muncul dalam data transaksi dengan beberapa iterasi sehingga akan diperoleh pola dan aturan. Kendala mendasar dalam proses pencarian hubungan antar data yaitu dibutuhkannya waktu yang tidak sebentar. [5].

## Struktur Kombinasi

Struktur dari *itemset* disini adalah mengikuti suatu bentuk dari kombinasi. Pengertian kombinasi adalah menggabungkan beberapa objek dari suatu grup tanpa memperhatikan urutan. Di dalam kombinasi, urutan objek tidak diperhatikan sebagai contoh dimana {1,2,3} adalah sama dengan {2,3,1} dan {3,1,2}. Kombinasi dapat dibagi menjadi dua yaitu kombinasi dengan pengulangan dan kombinasi tanpa pengulangan. Kombinasi tanpa pengulangan ketika urutan tidak diperhatikan akan tetapi setiap objek yang ada hanya bisa dipilih sekali maka jumlah kombinasi yang ada adalah:

𝒏! 𝒏

𝒓! (𝒏 − 𝒓)! = (𝒓)

Dimana **n** adalah jumlah objek yang bisa dipilih dan **r** adalah jumlah yang harus dipilih. Sebagai contoh, terdapat 5 pensil warna dengan warna yang berbeda

yaitu merah, kuning, hijau, biru dan ungu. Pensil warna tersebut hanya boleh dipilih dua warna. Banyak cara untuk mengkombinasikan pensil warna yang ada dengan menggunakan rumus di atas adalah :

## 5!/(5-2)!(2)! = 10 kombinasi.

Kombinasi dengan pengulangan jika urutan tidak diperhatikan dan objek bisa dipilih lebih dari sekali, maka jumlah kombinasi yang ada adalah ditunjukkan pada rumus berikut :

penggabungan terakhir, tidak dapat dihitung dari rumus di atas sebagai nilai s untuk itemset dengan panjang k tidak diketahui, karena itemset dari panjang k yang dihasilkan tidak sering muncul. Tetapi relasi yang diperoleh dari join terakhir akan memiliki catatan tupel sebagai jumlah dukungan untuk setiap itemset di set Ck. Menggunakan notasi join ini dapat direpresentasikan sebagai join (Ck \* sk-1, R, S (Ck)). Oleh karena itu untuk pendekatan ini dapat dinotasikan sebagai: [21]

(𝒏+𝒓−𝟏)! = ((𝒏 + 𝒓 − 𝟏)

) = (

𝒓!(𝒏−𝟏)! 𝒓

(𝒏 + 𝒓 − 𝟏))

𝒏 − 𝟏

# METODE PENELITIAN

Di mana *n* adalah jumlah objek yang bisa dipilih dan *r* adalah jumlah yang harus dipilih. Sebagai contoh adalah terdapat 10 jenis kue donat berbeda pada suatu toko donat. Kombinasi yang dihasilkan jika ingin untuk membeli tiga buah donat adalah :

(10+3-1)!/3!(10-1)! = 220 kombinasi.

## Konsep K-Way

Pendekatan menggunakan metode *K-Way* merupakan salah satu metode untuk mengurangi beban kerja terutama dalam menghitung nilai *support* dan *confidance* dalam sekumpulan data yang sangat banyak. Tujuan optimasi dan analisis menggunakan pendekatan ini adalah untuk mengetahui dampak dari optimasi pada dataset yang memiliki karakteristik seperti banyaknya transaksi rata-rata, dukungan, kepercayaan, dan lainnya untuk mendapatkan heuristik yang dipengaruhi oleh input dataset. Untuk penghitungan *support*, setiap k yang sesuai kriteria akan di gabungkan dalam Ck



Total jumlah item yang dihasilkan dari penggabungan Ck dengan T sama dengan jumlah data dalam Ck \* rata- rata support dari item pertama dalam Ck. Dengan menggunakan notasi *“join”* , nilai join dapat di wakili sebagai gabungan (Ck, R, Ck\*S1). Dengan demikian, gabungan dari T dengan C akan menghasilkan sebuah tabel, yang berisi jumlah jumlah support untuk item pertama itemset di Ck. Dalam hal notasi join ini bisa diwakili sebagai join in (Ck\*sm-1, R, Ck\*sm). Nilai

Dalam pengembangannya, untuk dapat diperoleh hasil yang diinginkan, sebagai acuan digunakan bagan kerja sebagai pemandu langkah- langkah yang akan dilakukan adalah sebagai berikut,

**PENGEMBANGAN**

Database

Menemukan Item Kandidat L1

Di atas support

**1**

**PEMANGKASAN**

Penghapusan Data Ck Itemset di Bawah *Support* untuk mendapatkan data

**PENGEMBANGAN**

**2**

1. Memperkecil ruang

Menemukan Kandidat

Lk itemset Dan menghitung nilai *Support* & *Confidence*

lingkup data dasar.

2. Mengurangi atribut data

untuk

kebutuhan memori.

memperkecil

ruang

1. Proses Penentuan data

Lk dilakukan secara terpisah.

1. Struktur DMQL tidak adanya relasi antar tabel

Langkah pengembangan dijelaskan sebagai berikut,

1. Langkah pertama memperkecil ruang lingkup data dasar yang memiliki jumlah besar menjadi lebih sedikit sehingga proses pencarian data menjadi lebih mudah.
2. Data dengan ruang lingkup lebih kecil, karena sudah pasti bahwa data tersebut yang akan dilakukan pemrosesan, dimungkinkan untuk mengurangi atributnya sehingga dapat mengurangi permasalah filter data.
3. Penentuan kandidat item set dilakukan secara terpisah setelah data diletakkan dalam tabel sementara.
4. Proses DMQL(*Data Mining Query Language*) dalam satu buah tabel tanpa adanya relasi antar tabel akan mempengaruhi kinerja waktu proses.

Peneliti membagi menjadi beberapa tahapan untuk dapat mengurangi waktu proses yaitu :

## Proses pemilihan kandidat item.

Proses ini dapat mengurangi waktu dikarenakan akan langsung mengambil data yang memiliki nilai *support* sama dengan atau lebih besar dari nilai yang sudah kita tentukan.

Proses di atas akan mendapatkan hasil item, *support* atau jumlah kontribusi, dan nilai *confidence* dari item- item yang menjadi kandidat. Dalam proses selanjutnya tidak akan dilakukan seleksi ulang untuk dilakukan pemangkasan data. Berbeda dengan Apriori tradisional, dalam pencarian data kandidat item akan ditampilkan semua item tanpa memperhatikan item mana yang memiliki nilai support dan *confidance* dengan batasan yang kita tentukan. S*yntax* pencarian kandidat untuk algortima tradisional adalah seperti yang terlihat di bawah ini :

//Seleksi Item select \* from

(SELECT j.ITEM,count(J.trans) as juml, (COUNT(J.trans)/@ttl)\*100 as pross FROM D\_Item j,TRANs TJ where Tj.trans=j.trans and tj.TGL >= @tgl1 and tj.TGL<= @tgl2 GROUP BY ITEM) as DataTabel where pross>=@suport

Pada syntax di atas terdapat dua proses yaitu menampilkan semua kandidat yang muncul, kemudian dilanjutkan dengan seleksi data yang masuk kriteria batasan *support* yang sudah ditentukan sebelumnya.

1. **Dalam proses menentukan kandidat itemset** Proses dalam menemukan pasangan kandidat itemset yang dengan menentukan jumlah *support* dan *confidance* merupakan proses yang membutuhkan waktu tidak sedikit. Dalam proses algoritma tradisional dilakukan proses untuk menghasilkan data Ck yaitu menemukan semua pasangan kandidat item set yang selanjutnya dilakukan proses untuk menemukan data Lk dimana

data dalam Lk merupakan data yang sudah dilakukan pemangkasan data dengan nilai *support* dan *confidance* dibawah nilai yang telah kita tentukan sebelumnya.

Berbeda dengan proses pengembangan algoritma yang peneliti lakukan yaitu secara langsung menemukan data dalam Lk tanpa mencari terlebih dahulu data Ck . Jadi proses yang dilakukan tanpa melewati proses scanning Ck terlebih dahulu. *syntax* query di atas digunakan dalam pengembangannya untuk mencari kandidat itemset yang dengan menemukan Lk bukan untuk menemukan Ck terlebih dahulu. Sehingga tidak akan melewati proses pemangkasan data Ck untuk mendapatkan data Lk. Dengan demikian proses di

sini akan mengurangi waktu dalam menemukan semua kandidat itemset, dan menghilangkan waktu proses pemangkasan. Berikut adalah bentuk pseudocode

L1:={Kandidat itemset} K:=2

**While** (Lk-1 <>0) do

**Begin**

**Forall** transaksi Lk-1 **do**

Lk := select count(\*) from kandidat pasangan yang

berasal dari Lk-1 K :=K+1;

**End**

Answer :=Uk Lk

**Insert into** L k-itemset

**Select \* from (select** p.item1,p.item2,…,p.itemk-1, p.n\_sup,p.n\_conf **From** L(k-1)-itemset p, L(k-1)-itemset q **where** p.item1=q.item1,…, p.itemk-2=q.itemk-2 , p.itemk-1 < q.itemk-1) as Tabel

**where** n\_sup>=sup and n\_conf>=conf

# HASIL DAN PEMBAHASAN

Data yang digunakan berasal dari basisdata sistem Informasi Manajemen Retail SFA Toserba. Data yang digunakan adalah data transaksi dari bulan Oktober 2014 hingga November 2016 dengan jumlah *record* untuk tabel transaksi mencapai lebih dari 1328282 *record.* Terdapat tiga partisi penting yang mempengaruhi proses penelitian ini yaitu,

|  |  |  |
| --- | --- | --- |
| **Atribut** | **Type Data** | **Keterangan** |
| **Tgl\_jual** | Datetime | Kelompok Transaksi |
| **Kode\_jual** | Varchar | Tid |
| **Kode\_brg** | Varchar | Titem |

## Pemasangan Kandidat Item Set k-way Join

Dalam pengembangan model yang dilakukan dengan memperkecil ruang lingkup pencarian dalam tabel ”JUAL” tersebut. Langkahnya akan mengambil data transaksi yang terjadi pada batasan tanggal yang dimaksud dengan tujuan agar jumlah total data akan lebih sedikit dibandingkan dengan total data secara keseluruhan. Langkah- langkah yang dilakukan dari syntax tersebut adalah sebagai berikut :

1. Membuat tabel penampungan untuk menampung data pasangan itemset dalam hal ini adalah tabel @L1 .
2. Dengan menggunakan perintah perulangan “*while”* dilakukan memasangkan kandidat item dari tabel C1.
3. Dalam proses memasangkan item, sekaligus dilakukan penghitungan nilai *support* dan *confidance* dari pasangan tersebut.
4. Jika nilai *support* dan *confidance* sesuai dengan batasan yang telah ditentukan, maka akan dimasukkan ke dalam tabel L1.

## Hitung Support dan Confidance

Dalam tahapan ini merupakan proses yang membutuhkan waktu paling lama, karena akan dilakukan *scanning* data secara keseluruhan dari item set yang terbentuk. Implementasi rumus untuk menghitung nilai support dan confidance dilakukan dengan pseudocede,

@tgl1=batas awal tanggal transaksi @tgl2=batas akhir tanggal transaksi

@ttl= select count(kode\_tran) as jml FROM @C1 WHERE tanggal>= @tgl1 AND tanggal<=@tgl2 AND jumlah item<=3;

@transak= select count(kode\_tran) as jml FROM @C1 WHERE tanggal>= @tgl1 AND tanggal<=@tgl2 A ND jumlah item<=3 AND (item1,item2)

|  |  |  |
| --- | --- | --- |
| **Item** | **Jumlah Trans** | **Dalam %** |
| **100038** | 1310 | 11.02 |
| **100201** | 1940 | 16.32 |
| **100336** | 1220 | 10.26 |
| **100338** | 2030 | 17.08 |
| **753797** | 1280 | 10.76 |

Hasil pemasangan kandidat item yang terbentuk aturan asosiasi L1 sebagai berikut :

* 1. Jika membeli 100201 maka 100338 jumlah transaksi sebanyak 140 transaksi dan nilai confidance sebesar 72.16 %.
	2. Jika membeli 100338 maka 100201 jumlah transaksi sebanyak 140 transaksi dan nilai confidance sebesar 68.96 %.

|  |  |  |  |
| --- | --- | --- | --- |
| **Item 1** | **Item 2** | **Jumlah****Trans** | **Nil Confidance****(%)** |
| **100201** | 100338 | 140 | 72.16 |
| **100338** | 100201 | 140 | 68.96 |

## 2. Uji coba kelompok data 2

@support=(@transak/@ttl)\*100

Setelah diperoleh nilai jumlah transaksi yang mengandung A dan B maka dapat di hitung nilai s*upport*nya,

//memasukkan data ke tabel @L1 yang sesuai kriteria Jika (item1<>item2) and (item2<>item1) and

(confidance>=batas)

Insert @L1 @itm1,@itm2,@jml,(@support),(@conf))

## Perbandingan Proses Pengujian Data

Proses selengkapnya dilakukan pengujian metode yang dilakukan dengan pembagian 3 kelompok data berdasarkan rentang waktu transaksi.

## 1. Uji coba kelompok data 1

Menghasilkan data kandidat itemset berjumlah 5 item merupakan data transaksi dalam rentang waktu tanggal 1-10 bulan November 2016, jumlah transaksi yang terjadi sebanyak 11885 dengan data C1=5 item, L1=2 item, L2=0 item menghasilkan waktu tempuh untuk *syntax* tradisional dalam hitungan detik adalah 69 detik sedangkan untuk pengembangan *syntax* menghasilkan waktu 65 detik dengan kandidat item,

Merupakan data transaksi dalam rentang waktu

tanggal 1-20 bulan November 2016, jumlah transaksi yang terjadi sebanyak 23631 dengan data C1=4 item, L1=2 item, L2=0 item menghasilkan waktu tempuh untuk *syntax* tradisional dalam hitungan detik adalah 85 detik sedangkan untuk pengembangan *syntax* menghasilkan waktu 77 detik.

|  |  |  |
| --- | --- | --- |
| **Item** | **Jumlah Trans** | **Dalam %** |
| **100038** | 2650 | 11.21 |
| **100201** | 3760 | 15.91 |
| **100338** | 3900 | 16.50 |
| **753797** | 2440 | 10.32 |

Hasil pemasangan kandidat item yang terbentuk aturan asosiasi L1 sebagai berikut :

1. Jika membeli 100201 maka 100338 jumlah transaksi sebanyak 270 transaksi dan nilai confidance sebesar 71.80 %.
2. Jika membeli 100338 maka 100201 jumlah transaksi sebanyak 270 transaksi dan nilai confidance sebesar 69.23 %.

|  |  |  |  |
| --- | --- | --- | --- |
| **Item 1** | **Item 2** | **Jumlah****Trans** | **Nil Confidance****(%)** |
| **100201** | 100338 | 270 | 71.80 |
| **100338** | 100201 | 270 | 69.23 |

## Uji coba kelompok data 3

Merupakan data transaksi dalam rentang waktu tanggal 1-30 bulan November 2016, jumlah transaksi yang terjadi sebanyak 34589 dengan data C1=4 item, L1=3 item, L2=0 item menghasilkan waktu tempuh untuk *syntax* tradisional dalam hitungan detik adalah 185 detik sedangkan untuk pengembangan *syntax* menghasilkan waktu 181 detik.

|  |  |  |
| --- | --- | --- |
| **Item** | **Jumlah Trans** | **Dalam %** |
| **100038** | 3810 | 11.01 |
| **100201** | 5140 | 14.86 |
| **100336** | 3530 | 10.20 |
| **100338** | 5680 | 16.42 |

Hasil pemasangan kandidat item yang terbentuk aturan asosiasi L1 sebagai berikut :

* 1. Jika membeli 100201 maka 100338 jumlah transaksi sebanyak 300 transaksi dan nilai confidance sebesar 58.36 %.
	2. Jika membeli 100338 maka 100201 jumlah transaksi sebanyak 300 transaksi dan nilai confidance sebesar 52.81 %.

Perbandingan nilai rata- rata kedua metode ada pada grafik berikut,

Dari ketiga kelompok data yang dilakukan uji algoritma, terjadi nilai penurunan waktu sebesar 4,95

RATA2

105

MODIV

107.6667

110

TRAD

113

115

RATA- RATA WAKTU TEMPUH

WAKTU(S)

%.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **JENIS** | **DATA****1** | **DATA****2** | **DATA****3** | **RATA-****RATA** | **SELISIH****(%)** |
| **TRAD** | 69 | 85 | 185 | 113 | 49,5 |
| **MODIV** | 65 | 77 | 181 | 107,66 |

# PENUTUP

Hasil uji algoritma dalam 2 metode yang dilakukan

|  |  |  |  |
| --- | --- | --- | --- |
| **Item 1** | **Item 2** | **Jumlah****Trans** | **Nil Confidance****(%)** |
| **100201** | 100338 | 300 | 58.36 |
| **100338** | 100201 | 300 | 52.81 |

## Perbandingan Uji Data

Dari ketiga kelompok data yang dilakukan pengujian dilakukan perhitungan nilai tempuh rata- rata diperoleh nilai sebagai berikut :

Dalam bentuk grafik dapat dilihat seperti berikut,

DATA1 DATA2 DATA3

KELOMPOK DATA

TRAD

MODIV

8577

6965

150

100

50

0

181581

200

PERBANDINGAN UJI DATA RATA- RATA

Waktu (S)

diperoleh kesimpulan untuk dapat dijadikan acuan penelitian berikutnya adalah,

1. Penghilangan proses penghapusan itemset yang tidak memenuhi kriteria dapat mengurangi waktu proses *scanning.*
2. Pendekatan menggunakan metode *k-way join* dalam penerapan algoritma apriori, dapat meningkatkan performa dalam iterasi yang terjadi. Pendekatan ini juga dapat mengurangi beban kerja terutama dalam menghitung nilai *support* dan *confidance*. Fokus dari proses kerja *k-way join* adalah DMQL atau *Data Mining Query Language.* Dengan menentukan formula tanda operator dalam penelitian ini dilakukan pengurangan klausa *group by, between,* dan relasi banyak partisi dengan model data yang dipilih menggunakan *vertical data* dapat mengurangi beban dengan hasil rata- rata waktu tempuh untuk apriori tradisional 71,66 detik sedangkan menggunakan pendekatan *k-way join* mencapai 61,11 detik, sehingga terjadi penurunan waktu tempuh selama 10,55 detik.
3. Pendekatan k-way join method mengurangi beban kerja yaitu proses yang terjadi dengan penurunan waktu tempuh selesai dilakukan secara langsung mempengaruhi konsumsi *memory*.

Sebagai bahan acuan untuk dilakukan penelitian selanjutnya, ada beberapa hal yang harus diperhatikan berdasarkan penelitian ini,

1. Isi atribut “kode\_brg” yang digunakan menggunakan jumlah karakter seminimal mungkin. Dikarenakan akan mempengaruhi proses *searching* menjadi lebih lama. Meskipun hanya akan berpengaruh dalam jumlah record yang sangat besar.
2. Pemilihan tanda operator yang tepat dalam proses DMQL sangat berpengaruh. Perbedaan DBMS yang digunakan menyebabkan perbedaan hasil yang berbeda meskipun menggunakan tanda operator yang sama.

# Daftar Pustaka

* 1. V. Mohan and D. S. Rajpoot, “Matrix-Over- Apriori : An Improvement Over Apriori Using Matrix,” vol. 5, no. 01, pp. 1–6, 2016.
	2. J. H. and M. Kamber, *Data Mining :Concept and Technique, S. Edition*, S.Edition. 500 Sansome Street, Suite 400, San Francisco, CA 94111 This: Diane Cerra, 2006.
	3. D. Edwards, “Data Mining: Concepts, Models, Methods, and Algorithms,” *J. Proteome Res.*, vol. 2, no. 3, pp. 334–334, 2003.
	4. M. B. Nichol, T. K. Knight, T. Dow, G. Wygant, G. Borok, O. Hauch, and R. O’Connor, “Fast Algorithms for Mining Association Rules,” *Ann. Pharmacother.*, vol. 42, no. 1, pp. 62–70, 2008.
	5. C. D. Cleaning, D. D. Transformation, A. T. Analysis, E. D. Mining, and B. D. Selection, “Three phase iterative model of kdd,” vol. 4, no. 2, pp. 695–697, 2011.
	6. S. A. Abaya, “Association Rule Mining based on Apriori Algorithm in Minimizing Candidate Generation,” vol. 3, no. 7, pp. 1–4, 2012.
	7. J. Singh and H. Ram, “Improving Efficiency of Apriori Algorithm Using,” vol. 3, no. 1, pp. 1–4, 2013.
	8. J. Yabing, “Research of an Improved Apriori Algorithm in Data Mining Association Rules,” vol. 2, no. 1, pp. 25–27, 2013.
	9. D. M. Tank, “Improved Apriori Algorithm for Mining Association Rules,” *Int. J. Inf.*

*Technol. Comput. Sci.*, vol. 6, no. 7, pp. 15–

23, 2014.

[10]S. Chaudhari, M. Borkhatariya, A. Churi, and

1. Bhonsle, “Implementation and Analysis of Improved Apriori Algorithm,” pp. 70–78, 2008.

[11]J. Kaur, R. Singh, and R. K. Gurm, “Performance evaluation of Apriori algorithm using association rule mining technique,” vol. 2, no. 5, 2016.

[12]A. Ansari, A. Parab, and S. Kadam, “Apriori - A Big Data Analysis - A Review,” pp. 3517– 3520.

[13]Q. Liu and J. Xin, “An improved Apriori algorithm based on data stream classification,”

*J. Comput. Inf. Syst.*, vol. 10, no. 23, pp. 10259–10266, 2014.

[14]R. S. Dm, V. Saldanha, and S. Sebastian, “Apriori Algorithm and its Applications in The Retail Industry for Analyzing Customer Interests,” vol. 2, no. 3, pp. 46–51, 2015.

[15]B. S. Dhak and M. Sawarkar, “Apriori : a promising data warehouse tool for finding frequent itemset and to define association rules,” vol. 4, no. 1, pp. 60–65, 2016.

[16]N. Gutierrez, “Demystifying Market Basket Analysis,” *http: //www.information- management.com*, 2006. .

[17]ACM SIGKDD, *Data Mining Curriculum*. 2016.

[18]M. Kantardzic, *Data Mining : Concepts, Models, Methods, and Algorithms*, Vol. 2. No. Proteome Research, 2003.

[19]Pete Chapman., *Step-by-step data mining guide*. The CRISP-DM Consortium, 2000.

[20]. Tan, P.N., Steinbach, M., Kumar, V.,

*Introduction to Data Mining*,Addison-

Wesley, Boston.,2006

[21]. L. Notes, C. Science, and C. Growth, “Performance Evaluation and Analysis of K- Way Join Variants for Association Rule Mining,” no. January, 2003.