

ANALISIS KETERKAITAN DATA TRANSAKSI PENJUALAN BUKU MENGUNAKAN ALGORITMA APRIORI DAN ALGORITMA CENTROID LINKAGE HIERARCHICAL METHOD (CLHM)

Nurani¹, Hamdan Gani²

¹nurani_nanni@yahoo.com, ²hamdan.gani.inbox@gmail.com²

^{1,2}AMIK Rizky Makassar

Abstrak

Pengambilan keputusan merupakan faktor yang menentukan didalam sebuah perusahaan. Kenyataannya pemanfaatan *big data* dalam sebuah pengambilan keputusan masih kurang efektif, salah satu contohnya adalah pemanfaatan *big data* perilaku belanja konsumen untuk menjadi sebuah pengetahuan yang mendukung pengambilan keputusan. Salah satu implementasi *big data* ini adalah pengambilan keputusan dalam penempatan barang pada rak toko buku. Berdasarkan permasalahan tersebut penelitian ini bertujuan mencari keterkaitan antara satu buku dengan buku yang lain didalam suatu set data dengan menggunakan teknik data mining, Penelitian ini memanfaatkan dua teknik data mining yaitu implementasi algoritma *Apriori* yang berfungsi untuk mendapatkan pola-pola item yang saling berkaitan kemudian algoritma CLHM (*Centroid Linkage Hierarchical Method*) untuk klusterisasi data. Penelitian ini menggunakan sampel data 15 kategori item dari data belanja konsumen pada toko buku tahun 2014 selama 3 bulan. Hasil akhir penelitian adalah sebuah pengetahuan baru yang dapat dijadikan rekomendasi pengambilan keputusan dalam sebuah toko buku.

Kata kunci: *Apriori, Big Data, CLHM, Data Mining, Pendukung Pengambil Keputusan*

1. Pendahuluan

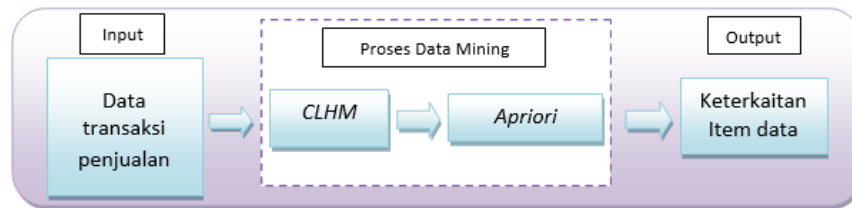
Pengolahan informasi menggunakan data mining *association rule mining* bertujuan untuk menemukan hubungan diantara data atau bagaimana suatu kelompok data mempengaruhi suatu keberadaan data yang lain [1]. Metode ini dapat membantu mengenali pola-pola tertentu di dalam kumpulan data yang besar. Aturan keterkaitan (*association rules*) adalah suatu prosedur untuk mencari hubungan antar item dalam suatu data set yang ditentukan [2]. *Association rule mining* dipergunakan untuk mencari kaidah asosiasi antara suatu kombinasi item. Mendeteksi kumpulan-kumpulan atribut yang muncul bersamaan (*co-occur*) dalam frekuensi yang sering, dan membentuk sejumlah kaidah dari kumpulan-kumpulan tersebut. *Association rules* seringkali disebut dengan "*Market Basket Analysis*", yang digunakan untuk menemukan relasi atau korelasi diantara himpunan item-item. *Market Basket Analysis* adalah Analisis dari kebiasaan membeli *customer* dengan mencari asosiasi dan korelasi antara item-item berbeda yang diletakkan *customer* dalam keranjang belanjanya.

Salah satu potensi penerapan *association rules* adalah pada pengaturan penempatan jenis-jenis buku pada rak di toko buku. Saat ini penempatan barang pada etalase barang di toko buku tidak sesuai dengan perilaku belanja konsumen. Hal inilah yang menjadi masalah yang sering dihadapi oleh perusahaan. Perusahaan mengalami masalah dalam manajemen produk serta kesulitan dalam hal pengaturan penempatan letak produk, sehingga perputaran barang tidak terjadi secara maksimal. Aturan keterkaitan (*association rules*) merupakan suatu prosedur untuk mencari hubungan antar item dalam suatu set data yang ditentukan [2] dan [3]. Dengan menggunakan teknik pengolahan informasi Data Mining yaitu *association rule mining* [4] yang digunakan untuk menemukan hubungan antara data atau bagaimana suatu kelompok data yang mempengaruhi suatu keberadaan data yang lain [1]. Metode ini dapat membantu mengenali pola-pola tertentu di dalam kumpulan data yang besar. Umumnya data yang digunakan untuk proses *association rules* atau aturan keterkaitan sangat besar dan terdapat variasi data yang sangat tinggi. Hal ini dapat mengurangi kualitas dari kaidah yang dihasilkan. Oleh karena itu, pada penelitian ini digunakan teknik clustering terlebih dahulu menggunakan algoritma CLHM (*Clustering Centroid Linkage Hierarchical Method*) [5], agar data yang digunakan untuk aturan asosiasi sudah terkluster dengan baik [6] dan [7].

Sehingga berdasarkan latar belakang masalah yang dijelaskan, maka dapat dirumuskan tujuan penelitian ini adalah untuk mencari keterkaitan atau hubungan antara item-item yang berbeda dalam suatu set data dengan menggunakan teknik data mining, dalam hal ini keterkaitan antara satu buku dengan buku lain pada sebuah toko buku, tujuan akhir penelitian yaitu solusi-solusi pendukung pengambilan keputusan dan kebijakan di dalam sebuah toko buku.

2. Metode

Secara umum penelitian yang dilakukan adalah merancang suatu aplikasi menggunakan teknik data mining algoritma *Apriori* untuk mendapatkan pengetahuan berupa pola asosiasi pembelian item oleh konsumen yang bisa digunakan sebagai strategi penjualan. Pada penelitian ini proses pencarian kaidah asosiasi menggunakan algoritma kombinasi antara *CLHM* dan *Apriori*. Proses dari pencarian keterkaitan item tersebut dapat dilihat pada gambar 1.



Gambar 1. Rancangan Penelitian

Proses analisis data dimulai dari implementasi algoritma *CLHM* untuk mengkluster 15 kategori item buku dengan pemilihan secara acak K , dimana K disini merupakan banyaknya kluster yang ingin dibentuk (5 kluster). Kemudian ditetapkan nilai-nilai K secara random, untuk sementara nilai tersebut menjadi pusat dari kluster atau biasa disebut dengan *centroid*, yang memiliki *mean* atau "*means*". Hitung jarak setiap data yang ada terhadap masing-masing *centroid* menggunakan rumus *Euclidian* hingga ditemukan jarak yang paling dekat dari setiap data dengan *centroid*. Klasifikasikan setiap data berdasarkan kedekatannya dengan *centroid*. Lakukan langkah tersebut hingga nilai *centroid* tidak berubah (stabil) [8]. Hasil cluster dari algoritma *CLHM* ini selanjutnya dimasukkan kedalam algoritma *Apriori* dimana algoritma ini bekerja untuk melakukan pencarian frequent itemset dengan menggunakan teknik *association rule* [9] dan [10].

2.1. Lokasi dan Waktu Penelitian

Penelitian ini dilaksanakan disalah satu toko buku gramedia yang ada di kota Makassar dan dilakukan selama 6 (Enam) bulan dimulai bulan Oktober 2013 sampai bulan Maret 2014.

2.2. Sumber Data

Data yang digunakan pada penelitian ini adalah *real dataset* yaitu data transaksi penjualan yang diambil dari laporan harian penjualan toko buku gramedia selama tiga bulan (Oktober-December) pada tahun 2014.

2.3. Proses Data Mining

Input pertama adalah data transaksi penjualan ini berisi tentang informasi penjualan item buku. Data transaksi ini nantinya dimasukkan ke dalam suatu database. Selanjutnya data tersebut diolah menggunakan teknik data mining. *Proses data mining* selanjutnya adalah kombinasi antara algoritma *CLHM* dan *Apriori* dimana akan dilakukan proses klustering terlebih dahulu menggunakan algoritma clustering *CLHM*, kemudian menggunakan algoritma asosiasi *Apriori* untuk mengetahui keterkaitan antar item buku. Input atau variabel dari algoritma *CLHM* yaitu (Jenis Kategori Item, Nama Buku dan Data Transaksi Penjualan Buku), kemudian hasil dari kluster *CLHM* menjadi input dari algoritma *Apriori* yaitu Nama Buku, Data Transaksi (Penjualan Buku). *Output* dari proses data mining atau keluaran dari sistem adalah keterkaitan antara satu kategori item buku dengan kategori item buku yang lain pada sebuah toko buku, dimana output sistem ini tujuannya adalah memberikan informasi tentang solusi-solusi pendukung pengambilan keputusan dan kebijakan di dalam sebuah toko buku.

CLHM adalah salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*. Metode ini mempartisi data ke dalam *cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama. Berikut algoritma *CLHM*:

1. Pemilihan secara acak K
2. Inisialisasi k pusat kluster (*centroid*) secara random
3. Tempatkan setiap data atau objek ke kluster terdekat.
Euclidean distance

$$d_{Euclidean}(x, y) = \sqrt{\sum_{j=1}^p (x_j - y_j)^2} \dots\dots\dots (1)$$

4. Hitung kembali pusat kluster dengan keanggotaan kluster yang sekarang. Pusat kluster adalah rata-rata (mean) dari semua data atau objek dalam kluster tertentu.
- a. Pembentukan cluster dan penentuan nilai centroid awal

Tahapan clustering dengan menggunakan *CLHM* dimulai dengan pembentukan cluster. Penulis disini menggunakan contoh hitungan untuk mempermudah untuk memahami kerja algoritma *CLHM*, pembagian *cluster* ini dipilih secara random, langkah pertama yaitu membentuk 5 *cluster* karena penulis menganggap pembentukan *cluster* ini sudah memenuhi pembagian jumlah anggota *cluster*. Proses penghitungan *centroid* awal dimulai dengan pemberian nama awal *cluster* (dari *cluster* pertama sampai dengan *cluster* kelima) secara random seperti pada Tabel 1 berikut.

Tabel 1. Pemberian Nama Awal *Cluster*

Transaksi	C	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	C1	0	2	2	0	2	2	0	0	0	0	0	0	0	0	0
2	C2	0	0	2	0	0	1	0	0	0	0	0	0	1	0	0
3	C3	0	2	2	0	1	2	0	0	1	0	0	0	0	0	0
4	C4	1	0	0	2	1	0	0	0	0	0	0	0	0	0	0
5	C5	0	0	0	0	1	1	0	0	0	0	0	0	2	0	0
6	C1	0	0	2	0	2	2	0	0	0	0	0	0	0	0	0
7	C2	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0
8	C3	0	0	2	0	1	2	0	0	0	0	0	0	0	0	0
9	C4	0	0	0	0	3	1	0	0	0	0	0	0	2	0	0
10	C5	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0
11	C1	0	2	2	0	1	0	0	0	0	0	0	0	0	0	0
12	C2	0	0	0	0	1	1	0	0	0	0	0	0	2	0	0
13	C3	0	2	2	0	2	2	0	0	0	0	0	0	0	0	0
14	C4	0	0	2	0	0	1	0	0	0	0	0	0	1	0	0
15	C5	0	0	2	0	1	2	0	0	0	0	0	0	0	0	0
16	C1	0	0	2	0	2	1	0	0	0	0	0	0	1	0	0
17	C2	0	2	2	0	2	2	0	0	0	0	0	0	0	0	0
18	C3	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0
19	C4	0	2	0	0	1	2	0	0	0	0	0	0	1	0	0
20	C5	0	0	2	0	2	1	0	0	0	0	0	0	1	0	0
21	C1	0	0	1	0	2	2	0	0	1	0	0	0	0	0	0
22	C2	0	2	2	0	2	2	0	0	0	0	0	0	0	0	0
23	C3	0	0	1	0	3	0	1	0	2	0	0	0	0	0	0
24	C4	0	0	2	0	0	1	0	0	0	0	0	0	1	0	0
25	C5	0	2	2	0	1	0	0	0	0	0	0	0	1	0	0
26	C1	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0
27	C2	0	1	2	0	2	0	0	0	0	0	0	0	0	0	0
28	C3	0	2	2	0	2	0	0	0	1	0	0	0	0	0	0
29	C4	0	0	2	0	1	1	0	0	0	0	0	0	1	0	0
30	C5	0	2	2	0	1	2	0	0	0	0	0	0	0	0	0

Setelah pemberian nama *cluster* untuk masing-masing data, selanjutnya akan dilakukan penghitungan untuk mendapatkan nilai *centroid* awal. Penghitungan dilakukan dengan menghitung *mean* (rata-rata) pada masing-masing *cluster* dengan membagi jumlah data yang didapatkan untuk setiap *cluster*-nya. Adapun tujuan dari penghitungan *centroid* awal dengan menggunakan *mean* (rata-rata) agar setiap *cluster* memiliki anggota data pada iterasi pertama yaitu dengan rumus berikut:

$$\bar{x}C_m = \frac{X_1C_m + X_2C_m + X_3C_m + \dots + X_nC_m}{n} \dots\dots\dots (2)$$

Dimana :

n : jumlah data

$\bar{x}C_m$: rata-rata nilai x pada *centroid* m

Adapun penghitungan nilai *centroid* awal pada masing-masing *cluster* sebagai berikut, dimana *item* 3 diambil sebagai contoh:

1. *Centroid* awal C1 (*item*3) = 1.5
2. *Centroid* awal C2 (*item*3) = 1.33
3. *Centroid* awal C3 (*item*3) = 1.83
4. *Centroid* awal C4 (*item*3) = 1.00
5. *Centroid* awal C5 (*item*3) = 1.33

Selanjutnya perhitungan nilai *centroid* awal setiap *cluster* untuk *item* 13, 6, 1, 2, 5, 4, 7, dan 9, dilakukan sama seperti menghitung nilai *centroid* awal setiap *cluster* untuk *item* 3 di atas. Adapun hasil dari *centroid* awal dari masing-masing *cluster* dapat dilihat pada tabel 2.



Tabel 2. Nilai *Centroid* Awal

Item	C1	C2	C3	C4	C5
1	0	0.17	0	0.17	0.17
2	0.67	0.83	1.17	0.33	0.67
3	1.5	1.33	1.83	1	1.33
4	0	0.33	0	0.33	0.33
5	1.5	1.17	1.5	1	1
6	1.17	1	1	1	1
7	0.17	0	0.17	0	0
8					
9	0.33	0	0.67	0	0
10					
11					
12					
13	0.33	0.5	0	1	0.67
14					
15					

b. Perhitungan jarak terdekat menggunakan *Euclidean Distance*

Setelah penghitungan nilai *centroid* awal pada masing-masing *cluster*, tahap selanjutnya adalah melakukan penghitungan untuk menentukan jarak setiap data dengan *centroid* awal yang telah dibentuk dengan menggunakan rumus *euclidean distance*. Hasil dari penghitungan jarak dengan rumus *euclidean distance* ini akan berpengaruh pada penempatan setiap data ke *cluster* tertentu. Adapun rumus dari *Euclidean distance*:

$$d_{Euclidean}(x, y) = \sqrt{\sum_{j=1}^p (x_j - y_j)^2} \dots\dots\dots (3)$$

Dimana :

- d : jarak
- x : x1, x2, x3, ..., xp : berupa jumlah item yang dibeli
- y : y1, y2, y3, ..., yp : nilai *centroid*
- j : mempresentasikan nilai atribut
- p : dimensi data

Penghitungan jarak data awal terhadap nilai *centroid* masing-masing *cluster* menggunakan rumus:

1. Jarak antara transaksi pertama dengan *centroid* pertama (C1) **d11 = 1.79**
2. Jarak antara transaksi pertama dengan *centroid* kedua (C2) **d21 = 1.97**
3. Jarak antara transaksi pertama dengan *centroid* ketiga (C3) **d31 = 1.56**
4. Jarak antara transaksi pertama dengan *centroid* keempat (C4) **d41 = 2.63**
5. Jarak antara transaksi pertama dengan *centroid* kelima (C5) **d51 = 2.19**

Penghitungan jarak data awal dengan *centroid* masing-masing *cluster* selanjutnya dilihat pada tabel 3:

Tabel 3. Nilai *Euclidean distance* Setiap *Cluster*

Transaksi	d1	d2	d3	d4	d5						
1	1.79	1.97	1.56	2.63	2.19	16	1.3	1.5	1.8	1.5	1.5
2	1.89	1.7	2.26	1.5	1.46	17	1.8	2	1.6	2.6	2.2
3	1.88	2.06	1.45	2.63	2.19	18	2.1	1.8	2	2.1	1.8
4	3.09	2.69	3.38	2.57	2.67	19	2.4	2.1	2.6	2.2	2.2
5	2.43	2.21	3.07	1.5	2.03	20	1.3	1.5	1.8	1.5	1.5
6	1.37	1.79	1.77	2.06	1.87	21	1.5	2	1.9	2.1	2
7	3.4	2.92	3.67	2.75	2.85	22	1.8	2	1.6	2.6	2.2
8	1.37	1.6	1.77	1.8	1.58	23	2.8	3.3	2.8	3.4	3.3
9	2.81	2.86	3.38	2.5	2.85	24	1.9	1.7	2.3	1.5	1.5
10	3.4	2.92	3.67	2.75	2.85	25	2.1	1.8	1.9	2.2	1.9
11	1.97	1.8	1.56	2.43	1.95	26	2.8	2.7	3.1	2.3	2.5
12	2.43	2.21	3.07	1.5	2.03	27	1.5	1.6	1.3	2.1	1.8
13	1.79	1.97	1.56	2.63	2.19	28	2.1	2.2	1.5	2.8	2.4
14	1.89	1.7	2.26	1.5	1.46	29	1.3	1.3	1.8	1.1	1.1
15	1.37	1.6	1.77	1.8	1.58	30	1.8	1.8	1.6	2.4	2

Setelah melakukan penghitungan jarak data dengan masing-masing *cluster*, tahap selanjutnya adalah mengelompokkan jarak terkecil disetiap masing-masing *cluster*. Seperti pada Tabel untuk data yang pertama atau struk pertama jarak yang terkecil yaitu 1.56 dan terletak pada (C3). Maka data pertama menjadi anggota dari *cluster* ketiga. Berdasarkan tabel diatas yang merupakan jumlah anggota yang didapatkan oleh masing-masing *cluster* sebagai berikut:



1. Pada *cluster* pertama (C1) memiliki jumlah anggota 6 data transaksi yaitu, 6, 8, 15, 16, 20, 21.
 2. Pada *cluster* kedua (C2) memiliki jumlah anggota 2 data transaksi yaitu, 19 dan 25.
 3. Pada *cluster* ketiga (C3) memiliki jumlah anggota 10 data transaksi yaitu, 1, 3, 11, 13, 17, 22, 23, 27, 28, dan 30.
 4. Pada *cluster* keempat (C4) memiliki jumlah anggota 8 data transaksi yaitu, 2, 4, 5, 7, 9, 10, 12, 26.
 5. Pada *cluster* kelima (C5) memiliki jumlah anggota 4 data transaksi yaitu 14, 18, 24 dan 29.
- Adapun hasil akhir *clustering* dengan menggunakan *CLHM* dapat dilihat pada tabel 4 di bawah.

Tabel 4. Hasil *Cluster* dan Data Transaksi yang telah diklaster

C1	C2	C3	C4	C5	Transaksi	Item
6	19	1	2	14	1	{2,3,5,6}
8	25	3	4	18	3	{2,3,5,6,9}
15		11	5	24	11	{2,3,5}
16		13	7	29	13	{2,3,5,6}
20		17	9		17	{2,3,5,6}
21		22	10		22	{2,3,5,6}
		23	12		23	{3,5,7,9}
		27	26		27	{2,3}
		28			28	{2,3,5,9}
		30			30	{2,3,5,6}

Dari hasil *cluster* pada tabel 4, diperoleh kelompok data yang nantinya digunakan untuk penerapan kaidah asosiasi menggunakan algoritma *Apriori*. Kelompok yang diambil untuk menjadi data awal yang digunakan pada algoritma *Apriori* adalah kelompok yang memiliki anggota terbanyak yaitu klaster C3. Maka data input untuk algoritma *Apriori* dari hasil klaster dapat dilihat pada tabel 4.

Algoritma *Apriori* adalah algoritma keterkaitan data. *Apriori* digunakan untuk menemukan kaidah asosiatif atau pola antara suatu kombinasi item. *Interestingness measure* yang dapat digunakan dalam data mining adalah :

1. *Support*, adalah suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu item atau itemset dari keseluruhan transaksi.
2. *Confidence*, adalah suatu ukuran yang menunjukkan hubungan antar dua item secara conditional (berdasarkan suatu kondisi tertentu). Metodologi dasar analisis asosiasi terbagi menjadi 2 tahap, yakni melakukan analisa pola frekuensi tinggi (*frequent pattern*) dan berikutnya adalah proses pembentukan aturan asosiasi.

Dalam penelitian ini, penulis mengambil 15 kategori *item* sebagai bahan penelitian. Sehingga pada tahap ini akan mengambil data transaksi yang mengandung kategori *item* yang sudah ditentukan oleh penulis dan transaksi yang tidak mengandung kategori item yang diteliti akan dibuang. Adapun 15 kategori *item* yang akan diteliti antaranya :

Tabel 5. Kategori Item

Kode <i>item</i>	Nama <i>item</i>
1	Kesehatan
2	Pendidikan Dasar
3	Pelajaran
4	Sains
5	Teknologi
6	Cerita
7	Kejiwaan
8	Agama
9	Resep
10	Fiksi
11	Hidup_Sehat
12	Pemerintahan
13	Ensiklopedi
14	Hiburan
15	Majalah

Pada pembahasan ini diambil contoh sederhana dengan menggunakan data transaksi sebanyak 10 transaksi. Diberikan batasan awal nilai minimum *support* 0.2 atau sama dengan 20 % dan *confidence* 0.75 atau sama dengan 75%. Nilai minimum *support* menunjukkan suatu ukuran yang

menunjukkan seberapa besar tingkat dominasi itemset dalam keseluruhan transaksi. Sedangkan nilai *confidence* menunjukkan suatu ukuran keterkaitan antar item-itemset.

a. *Data transaksi*

Tabel 6. Data Transaksi Yang Telah Ditransformasi dan Frekuensi Kemunculan Item.

Transaksi	Item	Item	Frekuensi
1	{2,3,5,6}	3	10
2	{2,3,5,6,9}	5	9
3	{2,3,5}	2	9
4	{2,3,5,6}	6	6
5	{2,3,5,6}	9	3
6	{2,3,5,6}	7	1
7	{3,5,7,9}		
8	{2,3}		
9	{2,3,5,9}		
10	{2,3,5,6}		

Dapat dijelaskan bahwa pada transaksi pertama, konsumen membeli item 2, 3, 5 dan 6. pada transaksi kedua, konsumen membeli item 2,3,5,6 dan 9, seterusnya sampai transaksi 10. Tahap selanjutnya adalah penelusuran database untuk mengetahui jumlah frekuensi kemunculan tiap item berdasarkan frekuensi kemunculan paling besar. Frekuensi kemunculan artinya jumlah kemunculan tiap item dalam keseluruhan transaksi. Hasilnya dapat dilihat pada tabel 6 yaitu untuk *item* 3, frekuensi kemunculannya dalam 10 transaksi adalah sebanyak 10 kali.

a. Pemangkasan data menggunakan *Support Count*

Cara menghitung minimum *support* dapat dilihat dibawah ini :

$$\text{Support (A)} = \frac{\text{Jumlah transaksi mengandung A}}{\text{Total transaksi}} \times 100\%$$

$$\begin{aligned} \text{Contoh menghitung support (3)} &= \frac{\text{Jumlah transaksi mengandung item 3}}{\text{Total transaksi}} \times 100\% \\ &= \frac{10}{30} \times 100\% = 0.3 \end{aligned}$$

Pada rumus diatas menjelaskan bahwa nilai *support* diperoleh dengan cara mencari jumlah transaksi yang mengandung nilai A (satu *item*) dibagi dengan jumlah keseluruhan transaksi. Sedangkan nilai *support* dari 2 *item* diperoleh dari rumus berikut :

$$\text{Support (A,B)} = P(A \cap B) = \frac{\text{Jumlah transaksi mengandung A dan B}}{\text{Total transaksi}} \times 100\%$$

Contoh menghitung *support* (3, 6)

$$\begin{aligned} \text{Support (3,6)} &= P(3 \cap 6) = \frac{\text{Jumlah transaksi mengandung 3 dan 6}}{\text{Total transaksi}} \times 100\% \\ &= \frac{3}{30} \times 100\% = 10\% \end{aligned}$$

Tabel 7. Nilai *Support* dan *Frequent List*

item	nilai support	Transaksi	Item
3	0.3	1	{2,3,5,6}
5	0.3	3	{2,3,5,6}
2	0.3	11	{2,3,5}
6	0.2	13	{2,3,5,6}
9	0.1	17	{2,3,5,6}
7	0	22	{2,3,5,6}
		23	{3,5}
		27	{2,3}
		28	{2,3,5}
		30	{2,3,5,6}

Dari tabel diatas dapat diperoleh Itemset yang memiliki frekuensi di atas minimum *support count* $\geq 0,2$ yaitu 3, 5, 2 dan 6 yang kemudian diberi nama *Frequent List* seperti terlihat pada gambar 11, keempat item ini akan berpengaruh saat pembuatan Algoritma Apriori. Sementara item 9 dan 7 dibuang karena tidak memenuhi minimum *support*. Setelah pemangkasan data menggunakan *support count*, diperoleh data baru yang terdiri dari item yang memiliki nilai *support count* diatas nilai minimum

support count, dan membuang item yang tidak memenuhi seperti pada gambar 8 frequent list. Tabel *frequent list* diatas mendata kemunculan item yang *frequent* dalam setiap transaksi, diurut berdasarkan yang frekuensinya paling tinggi.

b. Menghitung nilai *Confidence*

Dari keseluruhan *Frequent itemsets* tersebut, tidak semua dihitung. Karena dalam menghasilkan *Association Rule*, minimal *Frequent itemsets* yang dihitung terdapat 2 *item* dimana jika membeli *item* A maka akan membeli *item* B. Sehingga yang layak dihitung *confidence*-nya adalah 8 *subsets*, diantaranya : {2,3,5,6}, {2,3,5}, {3,5}. Setelah mendapatkan *frequent itemsets* dihitung, selanjutnya adalah membuat *rule* dengan menghitung *confidence*-nya. Hanya pola yang nilai *confidence*-nya ≥ 0.75 yang diambil. Karena perhitungannya sangat banyak, maka penulis mengambil salah satu contoh dari *frequent itemsets* Untuk {2, 3, 5, 6} untuk dicari kombinasinya dan dihitung nilai *confidence*-nya. Adapun rumus dari *confidence* adalah :

$$Confidence(A \rightarrow B) = \frac{Support(A, B)}{Support(A)} \times 100\% \quad \dots\dots\dots (4)$$

Dari proses pencarian kombinasi untuk *frequent itemsets* {2, 3, 5, 6} didapat 36 pola. Dari perhitungan *confidence* terhadap pola itu maka *Association Rule* yang memenuhi syarat *confidence* ≥ 0.75 adalah seperti pada tabel 8.

Tabel 8. Daftar Item-item yang memenuhi aturan keterkaitan
Untuk {2,3,5,6}

Item-item	Nilai confidence	Keterangan
2->3	1 atau 100%	Jika konsumen membeli <i>item</i> 2 (Pendidikan Dasar) maka ia membeli <i>item</i> 3 (Pelajaran)
3->2	0.9 atau 90 %	Jika konsumen membeli <i>item</i> 3 (Pelajaran) maka ia membeli <i>item</i> 2 (Pendidikan Dasar)
2->5	0.8 atau 80 %	Jika konsumen membeli <i>item</i> 2 (Pendidikan Dasar) maka ia membeli <i>item</i> 5 (Teknologi)
5->2	0.8 atau 80 %	Jika konsumen membeli <i>item</i> 5 (Teknologi) maka ia membeli <i>item</i> 2 (Pendidikan Dasar)
6->2	1 atau 100 %	Jika konsumen membeli <i>item</i> 6 (Cerita) maka ia membeli <i>item</i> 2 (Pendidikan Dasar)
3->5	0.9 atau 90 %	Jika konsumen membeli <i>item</i> 3 (Pelajaran) maka ia membeli <i>item</i> 5 (Teknologi)
5->3	1 atau 100 %	Jika konsumen membeli <i>item</i> 5 (Teknologi) maka ia membeli <i>item</i> 3 (Pelajaran)
6->3	1 atau 100 %	Jika konsumen membeli <i>item</i> 6 (Cerita) maka ia membeli <i>item</i> 3 (Pelajaran)
6->5	1 atau 100 %	Jika konsumen membeli <i>item</i> 6 (Cerita) maka ia membeli <i>item</i> 5 (Teknologi)
2,3->5	1 atau 100 %	Jika konsumen membeli <i>item</i> 2 (Pendidikan Dasar) dan <i>item</i> 3 (Pelajaran) maka ia membeli <i>item</i> 5 (Teknologi)
2,5->3	0.8 atau 80 %	Jika konsumen membeli <i>item</i> 2 (Pendidikan Dasar) dan <i>item</i> 5 (Teknologi) maka ia membeli <i>item</i> 3 (Pelajaran)
2,5->6	0.75 atau 75 %	Jika konsumen membeli <i>item</i> 2 (Pendidikan Dasar) dan <i>item</i> 5 (Teknologi) maka ia membeli <i>item</i> 6 (Cerita)
3,5->2	1 atau 100 %	Jika konsumen membeli <i>item</i> 3 (Pelajaran) dan <i>item</i> 5 (Teknologi) maka ia membeli <i>item</i> 2 (Pendidikan Dasar)

3. Hasil dan Pembahasan

Penelitian ini menunjukkan bahwa dari data yang di-*cluster*, dengan menggunakan algoritma *CLHM* yang kemudian dicari keterkaitannya menggunakan algoritma Apriori menghasilkan nilai *confidence* yang bervariasi dan mayoritas menghasilkan nilai *confidence* yang tinggi. Ini menunjukkan bahwa proses clustering akan menghasilkan kaidah-kaidah asosiasi dengan kualitas yang lebih baik. Sesuai dengan tujuan penelitian ini yaitu menghasilkan kaidah asosiasi multi obyektif yang baik. Dari sampel data transaksi toko buku yang digunakan dan digabungkan dengan proses clustering didapatkan bahwa kaidah-kaidah asosiasi yang dihasilkan lebih baik.

Kemudian hasil akhir dari teknik data mining tersebut adalah sebuah pengetahuan baru tentang pola pembelian konsumen yang selama ini jarang diketahui, penulis berasumsi bahwa hasil tersebut dapat dimanfaatkan untuk membantu membuat strategi bisnis, diantaranya: Menyusun layout yang baik didasarkan pada pola pembelian item yang memiliki nilai *confidence* tinggi atau dengan kata lain meletakkan antar satu item dengan itema yang lain yang nilai *confidencenya/keterkaitannya* tinggi. Selanjutnya toko buku dapat mengatur penempatan item ini agar memudahkan konsumen dalam membeli kedua item tersebut dan sehingga meningkatkan tingkat penjualan kedua item tersebut. Pengetahuan ini juga bisa digunakan dalam membantu manajer sebuah toko buku dalam menentukan keputusan persediaan barang, bisa juga dengan memberikan paket diskon terhadap pola pembelian item yang memiliki nilai *confidence* tinggi.

4. Kesimpulan dan saran

Dengan menerapkan kombinasi algoritma *CLHM* dan *Apriori* pada data transaksi penjualan di toko buku, didapatkan sebuah pengetahuan baru dengan mayoritas nilai-nilai *confidence* yang didapat tinggi (pengetahuan keterkaitan data atau item-item yang sering dibeli secara bersamaan dengan mayoritas nilai *confidence* tinggi). Dari nilai *support* yang dipakai yaitu sebesar 0.2 dan *confidence* sebesar 0.75 didapatkan beberapa pola yang memenuhi syarat *confidence* tinggi yaitu contohnya jika konsumen membeli buku pendidikan dasar dan pelajaran maka ia membeli buku teknologi dengan nilai *confidence* 0.75. Jika konsumen membeli buku pendidikan dasar dan buku pelajaran maka ia membeli buku cerita dengan nilai *confidence* 0.75. ini artinya jika seseorang membeli buku pendidikan dasar dan buku pelajaran maka peluang terbelinya buku teknologi sebesar 75%. Inilah output penelitian yang dapat digunakan sebagai rekomendasi atau pendukung keputusan oleh sebuah took buku. Selanjutnya agar mendapatkan hasil yang lebih baik, saran yang dapat diberikan berkaitan dengan penelitian ini yaitu untuk pengembangan selanjutnya adalah menambah jumlah sampel data yang digunakan untuk uji coba serta menambahkan kriteria item yang digunakan sehingga dapat menghasilkan pola pengetahuan baru yang lebih banyak. Bisa juga dengan menggabungkan algoritma asosiasi apriori dengan algoritma clustering lainnya seperti *K-NN* atau algoritma sejenisnya.

Daftar Pustaka

- [1] Kantardzic, M. 2003. Data Mining Concepts Models, and Algorithms. New Jersey: IEEE.
- [2] Han, J. dan Kamber, M. 2001. Data Mining Concepts and Techniques. Morgan Kaufmann, California.
- [3] Jiawei, H dan Micheline, K. 2000. Data Mining: Concepts and Techniques: Chapter 6. Mining Association Rules in Large Database, Simon Fraser University.
- [4] Ulmer, David. 2002. Mining an Online Auctions Data Warehouse. The Mid-Atlantic Student Workshop on Programming Languages and Systems, Pace University.
- [5] Usmaida, A. 2007. Web Mining Untuk Pencarian Berdasarkan Kata Kunci Dengan Teknik Clustering, *Tugas Akhir Jurusan Teknologi Informasi Politeknik Elektronika Negeri Surabaya*.
- [6] Martiana, E, et al. 2010. Mesin Pencari Dokumen Dengan Pengklasteran Secara Otomatis. *Telkomnika Vol. 8, No. 1*.
- [7] Barakbah, A.R. dan Arai, K. 2004. Identifying Moving To Make Automatic Clustering For Normal Data Set, *In. Proc. IECI Japan Workshop*, Musashi Institute of Technology, Tokyo.
- [8] Witten, et al., 2005. Data Mining Practical Machine Learning Tools and Techniques, Second Edition. Morgan Kaufmann, San Fransisco.
- [9] Moertini, V dan Marsela, Y. 2007. Analisis Keranjang Pasar Dengan Algoritma Hash-Based Pada Data Transaksi Penjualan Apotek. Jurusan Ilmu Komputer, Universitas Katolik Parahyangan, Bandung.
- [10] Yova, R. dan Fahrian, M. 2008. Implementasi Algoritma-Association Rules Sebagai Bagian Dari Pengembangan Data Mining Algorithms Collection. *Konferensi Nasional dan Sistem Informatika*.