

PENERAPAN *MARKET BASKET ANALYSIS* PADA SWALAYAN UNTUK PENGENALAN POLA TRANSAKSI PENJUALAN DENGAN MENGGUNAKAN METODE *FUZZY C-COVERING*

David Surya Ganda¹

Dian Eka R, S.Si., M.Kom, Dewi Yanti L, S.Kom, M.Kom²

¹⁾ Mahasiswa Jurusan Matematika Program Studi Ilmu Komputer

²⁾ Staf Pengajar Jurusan Matematika Program Studi Ilmu Komputer

Universitas Brawijaya Malang

Jalan Mayjen Haryono 169, Malang 65145, Indonesia

Abstract

Supermarket has a big amount of data which can be used to know the people's habit of buying something. By knowing this habit may lead us to know the linkage pattern of product which bought. This kind of process is called association rules. The resulting pattern can be used to set the layout, promotion, or sales forecast. To determine association rules pattern we can use Fuzzy c-Covering method. Fuzzy c-Covering is a method which can be used for classificate the elements from universal set to fuzzy sets partitions. The result of trial using Fuzzy c-Covering method using 98 transactions is lift score 16.33.

Keyword : *association rules, fuzzy c-covering, data mining*

PENDAHULUAN

Pertumbuhan yang pesat dari akumulasi data telah menciptakan kondisi "rich of data but poor of information" karena data yang terkumpul tidak dapat digunakan untuk aplikasi yang bermanfaat, data tersebut hanya dibiarkan saja menjadi tumpukan-tumpukan data yang tidak bermanfaat. Data-data yang sudah sangat banyak tersebut dapat digunakan menjadi informasi yang berguna dengan *data mining*. *Data mining*, yang didefinisikan sebagai proses untuk mengenali suatu pola dari data, menjadi sarana pemecahan bagi masalah penumpukan data tersebut (Witten, 2000).

Istilah *market basket* digunakan untuk menggambarkan kelompok *item* (terdiri dua atau lebih *item*) yang cenderung dibeli oleh satu konsumen sewaktu berbelanja di swalayan dalam satu transaksi pembelian.

Fuzzy c-Covering merupakan salah satu metode yang dipakai untuk mengklasifikasikan elemen-elemen dari suatu himpunan universal menjadi partisi-partisi berupa *fuzzy sets*. Dalam mencari hubungan antar *item*, metode *Fuzzy c-Covering* ini berdasarkan pada persepsi

bahwa semakin banyak *item* yang dibeli dalam suatu transaksi, maka hubungan antar *item* dalam transaksi itu semakin lemah (Klir, 2001).

Manfaat dari penelitian ini diharapkan dapat menghasilkan sebuah sistem yang dapat menghasilkan informasi tentang hubungan keterkaitan antar *item* berbeda yang seringkali dibeli bersamaan sehingga dapat digunakan untuk mengenali pola pembelian konsumen dan membantu manajer dalam mengambil keputusan pada suatu swalayan. Misalnya pengambilan keputusan tata letak barang atau dalam penentuan pemberian hadiah untuk meningkatkan penjualan.

TINJAUAN PUSTAKA

Data Mining

Menurut Edelstein, *data mining* adalah suatu proses analisa data untuk menemukan pola dan hubungan di dalam data yang kemungkinan dapat digunakan untuk melakukan suatu prediksi yang valid (Edelstein, 1999).

Association rules

Association rules mining adalah suatu prosedur untuk mencari hubungan antar *item* dalam suatu set data yang telah ditentukan (Kamber,2001).

Umumnya ada dua ukuran kepercayaan (*interestingness measure*) yang digunakan dalam menentukan suatu *association rules*, yaitu *support* dan *confidence*. *Support* adalah suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu *item* dari keseluruhan transaksi. Sedangkan *confidence* adalah suatu ukuran yang menunjukkan hubungan antar 2 atau lebih *item* secara *conditional*. Kedua ukuran ini nantinya berguna dalam menentukan *interesting association rules*, yaitu untuk dibandingkan dengan batasan yang telah ditentukan. Batasan tersebut terdiri dari *minsup* dan *minconf*. Secara sederhana perhitungan untuk mendapatkan *support* dapat dilihat pada persamaan (2.1) dan *confidence* dapat dilihat pada persamaan (2.2).

$$\begin{aligned}
 support(u) &= \frac{\sum_{t=1}^n \frac{1}{C_{|T_t|}^k s(u, T_t)}}{n} \\
 &= \frac{\sum_{t=1}^n \frac{1}{k!(|T_t|-k)!} s(u, T_t)}{n} \quad (2.1)
 \end{aligned}$$

Dimana *u* adalah kombinasi *k-item* yang dicari *support*-nya. Jika *I* adalah *universal set of items*, maka $u \subseteq I$. $|u| = k$, yaitu jumlah *item* dalam *u*. T_t adalah transaksi ke-*t* ($T_t \subseteq I$). $|T_t|$ adalah jumlah *item* dalam T_t . $C_{|T_t|}^k$ adalah kombinasi *k-item* terhadap $|T_t|$. *N* adalah jumlah *record* dalam *QT*. $s(u, T_t) \in \{0,1\}$ adalah suatu *function*, jika $u \subseteq T_t$, maka $s(u, T_t) = 1$, selain itu $s(u, T_t) = 0$ (Budhi,2005).

$$conf(A \Rightarrow B) = \frac{P(A \cap B)}{P(A)} \quad (2.2)$$

Minimum support adalah parameter yang digunakan sebagai batasan *item/itemset* yang harus dipenuhi dalam suatu kelompok data untuk dapat dijadikan aturan, sedangkan *minimum confidence* adalah parameter yang mendefinisikan minimum level dari *confidence* (kepercayaan) yang harus dipenuhi.(Budhi,2005)

Sebagai contoh, dari suatu himpunan data transaksi, seseorang mungkin menemukan suatu hubungan berikut, yaitu jika konsumen yang membeli roti tawar, biasanya juga akan membeli keju dalam suatu transaksi yang sama ditunjukkan sebagai berikut :

Roti tawar \rightarrow keju [*support* 2%, *confidence* 60%]

Artinya adalah 60% dari transaksi di *database* yang memuat *item* roti tawar juga memuat *item* keju. Sedangkan jumlah transaksi yang memuat kedua *item* itu adalah 2% dari seluruh transaksi yang ada di *database*.

Market Basket Analysis

Market Basket Analysis adalah salah satu cara yang digunakan untuk menganalisis data penjualan dari suatu perusahaan. Proses ini menganalisis kebiasaan belanja konsumen dengan menemukan asosiasi antar *item* yang berbeda yang dibeli konsumen. Hasil yang telah didapatkan ini nantinya dapat dimanfaatkan oleh perusahaan retail seperti toko atau swalayan untuk mengembangkan strategi pemasaran dengan melihat *item-item* yang sering dibeli secara bersamaan oleh konsumen (Ibig,2008).

Fuzzy c-covering

Fuzzy c-covering merupakan salah satu metode yang dipakai untuk mengklasifikasikan elemen-elemen dari suatu himpunan universal menjadi partisi-partisi berupa *fuzzy sets*. *Fuzzy c-covering* sendiri merupakan generalisasi dari metode *fuzzy c-partition* yang telah dikenal sebelumnya. *Fuzzy c-partition* dapat



didefinisikan sebagai berikut (Klir, Yuan, 2001):

Misalkan $I = \{i_1, i_2, \dots, i_n\}$ adalah domain dari data. Fuzzy c-partition dari I adalah fuzzy subset atau fuzzy classes dari T, ditunjukkan oleh $T = \{t_1, t_2, \dots, t_c\}$, yang memenuhi persamaan (2.3) berikut ini.

$$\sum_{m=1}^c \mu_{tm}(i_k) = 1, \text{ untuk semua } k \in N_n \quad (2.3)$$

dimana : c adalah positive integer ($0 < c \leq n$), $\mu_{tm}(i_k)$ adalah derajat keanggotaan item i_k pada transaksi ke-m, $\mu_{tm}(i_k) \in [0,1]$, dan $0 < \sum_{k=1}^n \mu_{tm}(i_k) < n$, untuk semua $m \in N_c$

Berikut ini contoh untuk memperjelas rumus di atas.

Contoh 2.1 : Dimisalkan ada suatu fuzzy classes yang diketahui jumlah elemen yang terlibat ($n = 6$) dan jumlah partisinya ($c = 3$) :

$$\begin{aligned} \mu_{t1} &= \{1/i_1, 0.8/i_2, 0.2/i_3\}, \\ \mu_{t2} &= \{0.2/i_2, 0.8/i_3, 0.4/i_4\}, \\ \mu_{t3} &= \{0.6/i_4, 1/i_5, 1/i_6\}. \end{aligned}$$

Fuzzy class tersebut dapat dikatakan sebagai fuzzy 3-partition dari suatu domain $I = \{i_1, i_2, i_3, i_4, i_5, i_6\}$ karena memenuhi persamaan (2.3).

Di dalam Fuzzy c-Covering, persamaan (2.3) dari teori fuzzy c-partition tersebut digeneralisasi. Persamaan hasil generalisasi tersebut dapat dilihat pada persamaan (2.4) berikut (Intan, Mukaidono, 2003):

$$\sum_{m=1}^c \mu_{tm}(i_k) \geq 1, \text{ untuk semua } k \in N_n \quad (2.4)$$

Untuk memperjelasnya, diberikan contoh sebagai berikut:

Contoh 2.2 : Dimisalkan ada suatu fuzzy classes sebagai berikut:

$$\begin{aligned} \mu_{t1} &= \{1/i_1, 0.8/i_2, 0.4/i_3\}, \\ \mu_{t2} &= \{1/i_2, 0.8/i_3, 0.6/i_4, 0.2/i_6\}, \\ \mu_{t3} &= \{1/i_4, 1/i_5, 1/i_6\}. \end{aligned}$$

Dari contoh tersebut dapat diketahui $n = 6$ dan $c = 3$. Fuzzy class tersebut dapat dikatakan sebagai fuzzy 3-covering dari suatu domain $I = \{i_1, i_2, i_3, i_4, i_5, i_6\}$ karena memenuhi persamaan (2.3) dan (2.4) (Budhi,2005).

Degree of Similarity in Fuzzy c-Covering

Contoh 2.1 dan 2.2 merepresentasikan fuzzy classes (T) sebagai fuzzy set terhadap crisp element dari data pada domain I. Sebaliknya, setiap crisp element dari data pada I juga dapat direpresentasikan sebagai fuzzy set dari fuzzy classes (T) dengan menggunakan persamaan (2.5) (Intan, Mukaidono, 2003):

$$\mu_{im}(t_j) = \frac{\mu_{tj}(i_m)}{\sum_{i \in I} \mu_{tj}(i)} \quad (2.5)$$

Dimana i_m adalah item-m dan t_j adalah transaksi ke-j.

Fuzzy Conditional Probability Relation

Umumnya, dalam mengukur derajat kesamaan antara dua fuzzy set digunakan fuzzy conditional probability relation. Rumusnya dapat dilihat pada persamaan (2.6). (Intan, Mukaidono, 2003)

$$\begin{aligned} R(X, Y) &= \text{conf}(Y \rightarrow X) = P(X | Y) = \frac{|X \cap Y|}{Y} \\ &= \frac{\sum_{i \in I} \{\mu_x(i), \mu_y(i)\}}{\sum_{i \in I} \mu_y(i)} \quad (2.6) \end{aligned}$$

dimana: μ_x dan μ_y adalah fungsi anggota dari domain I untuk masing-masing X dan Y dari himpunan universal U.

Dalam prakteknya, fuzzy conditional probability relation dapat digunakan sebagai dasar untuk merepresentasikan degree of similarity relationship antara dua fuzzy set dalam universe U. Dalam definisi fuzzy conditional probability relation, nilai probabilitas bisa diperkirakan berdasarkan hubungan antar fuzzy set dengan

menggunakan pandangan subjektif dari teori probabilitas (Budhi,2005).

Algoritma Fuzzy c-Covering

Dimulai dari menentukan *minimum support* dan *minimum confidence* sebagai *threshold* bagi kombinasi *item* dari seluruh transaksi.

Langkah kedua adalah mencari *support* dari setiap *item* yang memungkinkan yang ada di dalam transaksi tersebut dengan persamaan (2.1).

Langkah ketiga yaitu melakukan penyaringan terhadap kombinasi *item* yang ada di dalam transaksi tersebut yang tidak memenuhi $support \geq minimum\ support$.

Langkah keempat, adalah mencari kombinasi 2-*item* yang memungkinkan dari tiap 1-*item* yang memenuhi *minimum support* dan *minimum confidence* yang telah ditentukan.

Langkah kelima adalah menyimpan *rule* yang dihasilkan dari 2 kombinasi *item* yang memiliki *support* dan *confidence* lebih besar dari nilai *minimum* yang ditentukan. Jika tidak ada lagi kombinasi yang memungkinkan yang memenuhi *minimum support* yang telah ditentukan maka proses akan dilanjutkan dan dicari kombinasi 3-*item*.

Langkah keenam yaitu mencari *candidate rules* dengan cara menghitung *confidence* dari setiap kombinasi 3-*item* yang memenuhi *minimum support* kombinasi 3 *item* yang telah didapat pada langkah 5 dengan persamaan (2.6).

Lift Rasio

Lift Rasio digunakan untuk mengukur seberapa kuat *rule* yang dibentuk dari algoritma *Fuzzy c-Covering*. Nilai *lift* rasio berkisar antara 0 sampai tak terhingga. Nilai minimum dari *lift* rasio tidak ditentukan seperti halnya *support* atau *confidence*. Jika nilai *lift* rasio lebih dari 1,dalam hal ini adalah nilai minimum, maka *rule antecedent* berpengaruh negatif pada *rule consequent*. Jika nilai *lift* rasio sama dengan 1 maka *rule* tersebut sering muncul bersamaan tetapi independen. *Rule*

yang independen adalah *rule* dimana untuk mendapatkan *consequent* tidak tergantung pada *attendant*. Pada *lift* rasio, *rule* yang direkomendasikan adalah jika *lift* rasio lebih dari 1 karena *attendant* memiliki pengaruh positif pada *consequent*. Berikut rumus untuk menentukan *lift* rasio (Foby,2011):

$$Expected\ confidence, EC(A \rightarrow B) = \frac{\sigma(B)}{m}$$

$$Lift = \frac{Confidence}{Expected\ Confidence}$$

Dimana :

$\sigma(B)$ = Jumlah *consequent* dalam transaksi

m = Jumlah transaksi

METODE PENELITIAN

Sistem ini bertujuan untuk menemukan asosiasi antar *item* dari data transaksi penjualan pada swalayan. Dari asosiasi yang dihasilkan dapat diketahui keterkaitan antar *item* yang terdapat dalam transaksi penjualan. Data yang digunakan untuk penelitian ini adalah data transaksi penjualan yang diambil dari Indomaret selama bulan Desember 2011. Tahapan pengembangan sistem dimulai dari penentuan nilai *minimum support* dan *minimum confidence* yang digunakan sebagai batasan dalam pencarian *rule*.

PEMBAHASAN

Perangkat keras yang digunakan dalam penelitian ini adalah:

- Intel Core i3 2.26GHz
- Memory 2GB
- Monitor 14"
- Harddisk kapasitas 320GB
- Keyboard
- Mouse

Database pada program ini diimplementasikan pada *database* Microsoft Access. Pada program ini terdapat lima tabel yang dibutuhkan, tabel tersebut adalah tabel untuk menyimpan transaksi yaitu *ttransaksi* dan *titems*, serta tabel untuk menyimpan hasil yaitu *tresult*,

tsupport, dan tsaverule. Dalam implementasi program terdapat beberapa langkah yaitu menentukan *list item*, menghitung *support* 1 *item*, menghitung *support* 2 dan 3 kombinasi *item*, menghitung *confidence* dan *lift ratio*.

Untuk mengukur kekuatan *rule* dilakukan dua skenario pengujian yaitu :

1. Pengujian dengan nilai *support* sama dan *confidence* yang berbeda. Pada proses uji coba ini digunakan parameter *minimum support* 0.5% dengan *minimum confidence* 40% kemudian *minimum confidence* dinaikkan menjadi 50% dan 60%. Jumlah transaksi yang digunakan adalah 98 transaksi.
2. Pengujian yang kedua dilakukan dengan nilai *confidence* yang sama dan *support* yang berbeda. Pada proses uji coba ini digunakan parameter *minimum confidence* 60% dengan *minimum support* 0.5%, 0.6%, dan 0.7%. Jumlah transaksi yang digunakan adalah 98 transaksi.

Tabel hasil uji coba

Sup (%)	Jumlah item	Conf (%)	Jumlah rule	Rata-rata lift rasio
0.5	39	40	58	3,85
0.5	39	50	41	4,18
0.5	39	60	30	4,49
0.6	36	40	41	2,51
0.6	36	50	29	2,60
0.6	36	60	22	2,75
0.7	34	40	34	2,39
0.7	34	50	22	2,44
0.7	34	60	16	2,64

Hasil yang didapat adalah bahwa dengan nilai *confidence* 0,4 didapat nilai *minimum lift* rasio adalah 0,95 pada *rule* ke 9 dan nilai maksimum *lift* rasio adalah 16,3 pada *rule* ke-52 dan 56. *Lift* rasio yang direkomendasikan adalah *rule* yang memiliki nilai *lift* lebih dari 1. Semakin tinggi nilainya maka *rule* tersebut semakin dapat dipercaya. Sebagai contoh terdapat *rule* Minyak Goreng => Air Mineral yaitu jika membeli minyak goreng kemungkinan membeli air mineral dengan nilai *lift* rasio 1,51, maka *rule* tersebut lebih

baik daripada *rule* Minyak Goreng => Air Mineral yang memiliki nilai *lift* rasio 1,23.

Semakin tinggi nilai *minimum support*, semakin kecil nilai *lift* rasio hal ini disebabkan karena nilai *expected confidence* semakin besar. Nilai *lift* rasio berbanding terbalik dengan *expected confidence*. Semakin tinggi nilai *expected confidence*, semakin kecil nilai *lift* rasio. Untuk mendapatkan nilai *expected confidence* sama dengan mendapat nilai *support* yaitu membagi jumlah itemset dalam hal ini *consequent* dengan jumlah transaksi. Jadi dengan batasan *minimum support* semakin besar, maka semakin besar pula nilai *expected confidence*.

KESIMPULAN

Kesimpulan yang diperoleh dalam menyelesaikan skripsi ini adalah :

1. Algoritma *Fuzzy c-Covering* dapat digunakan untuk pengenalan pola transaksi pada sebuah *mini market*. Untuk mencari pola pembelian, dilakukan pemilihan *item* yang memenuhi berdasar *minimum support* dan *minimum confidence* yang ditentukan, kemudian menguji ketepatan *rule* yang didapat dengan menghitung nilai *lift* rasio. Hasil pencarian pola yang menghasilkan *rule* ini dapat digunakan pada Indomaret.
2. Berdasarkan uji coba yang dilakukan, didapat bahwa nilai *lift* rasio dipengaruhi oleh *minimum support*. Pada uji coba tersebut didapat bahwa dengan *minimum support* 0.5%, *minimum confidence* 60%, dan jumlah data 98, memiliki nilai *lift* paling tinggi 16,33 dan paling rendah 1,43. Jika nilai *lift* lebih dari 1 maka *rule* direkomendasikan. Semakin tinggi nilai *lift* rasio maka *rule* tersebut semakin kuat.



DAFTAR PUSTAKA

- Budhi, Gregorius S.; *Aplikasi Data Mining dengan Konsep Fuzzy c-Covering untuk Analisa Market Basket pada pasar swalayan*. Petra. 2005
- Edelstein; *Introduction to Data Mining and Knowledge Discovery*. Two Crows Corp. 1999
- Foby, Tom; *Association Rules (Aka Affinity Analysis or Market Basket Analysis)*. Department of Economics Southern Methodist University Dallas. Texas. 2011
- Intan, Rolly; Mukaidono, Masao; *A Proposal of Fuzzy Thesaurus Generated by Fuzzy Covering*. IEEE. 2003: p. 167-172
- Kamber, Micheline; *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2001
- Kantardzic; *Data Mining: Concepts, Models, Methods, and Algorithms*. IEEE. 2003
- Klir, George J.; Yuan, Bo; *Fuzzy Sets and Fuzzy Logic – Theory and Applications*. India: Prentice Hall, 2001
- Larose, Daniel T.; *Data Mining Methods and Models*. CCSU, 2005
- Prasetyo, Ibig; *Penerapan Fuzzy c-Covering pada Transaksi Belanja Mini Market*. UPN Veteran, 2008
- Riviana, Sitra Lativa; *Analisis Data Penjualan Menggunakan Algoritma Pincer Search*. UB, 2010
- Witten, Ian H.; *Data Mining Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann Publishers, 2000