

# Association Rules

## *(Kaidah Asosiasi)*

Ardytha Luthfiarta

# Association Rules

- Association Rule Mining disebut juga Frequent Itemset Mining
- Adalah proses mendeteksi kumpulan atribut-atribut yang muncul bersamaan (co-occur) dalam frekuensi yang sering (itemset), dan membentuk sejumlah kaidah(rules) dari kumpulan-kumpulan tersebut.
- Karena aplikasi association rules yang sangat luas untuk menganalisa keranjang belanja di pasar swalayan, association rules sering juga disebut sebagai **market basket analysis**.
- Tujuan dari algoritma association rules adalah untuk menghasilkan algoritma yang efisien untuk analisa pola frekuensi tinggi (**frequent pattern mining**)
- Contoh : 90% orang yang berbelanja di suatu supermarket yang membeli roti juga membeli selai, dan 60% dari semua orang yang berbelanja membeli keduanya.
- Contoh 2 : "70% dari orang-orang yang membeli mie, juice dan saus akan membeli juga roti tawar".

# Definisi Association Rules (dari bbrp pakar)

- *Association rule mining* adalah analisa dari kebiasaan belanja konsumen dengan mencari asosiasi dan korelasi antara item-item berbeda yang diletakkan konsumen dalam keranjang belanjanya (Yang, 2003)
- Dengan kemajuan teknologi, data penjualan dapat disimpan dalam jumlah besar yang disebut dengan "*basket data*."
- *Aturan asosiasi yang didefinisikan* pada basket data tersebut, dapat digunakan untuk menganalisa data dalam rangka :
  - keperluan desain katalog promosi,
  - proses pembuatan keputusan bisnis,
    - Kampanye pemasaran dengan diskon atau potongan harga
    - Pengaturan tata letak display barang, (barang A didekatkan dengan barang B)
  - segmentasi konsumen dan
  - target pemasaran.

# Contoh aplikasi kaidah asosiasi

## ○ Marketing and Sales Promotion

### ○ Misal :

- Ketergantungan {bagels, ... } → {Potato Chips}
- Potato Chips sebagai consequent → dapat digunakan untuk menentukan apa yang dilakukan untuk meningkatkan penjualan
- Bagels in the antecedent → dapat digunakan untuk melihat produk mana yang akan terkena dampak jika toko tersebut tidak lagi menjual bagels.
- Bagels in antecedent and Potato chips in consequent → Dapat digunakan untuk melihat produk apa yang harus dijual dengan bagels untuk mempromosikan penjualan potato chips.

# Contoh aplikasi kaidah asosiasi

## ○ Supermarket Shelf Management

- Tujuan untuk mengenali item2 yang dibeli bersama-sama(dalam sekali transaksi) oleh beberapa pelanggan.
- Pendekatan : memproses data point of sale dengan pemindai barcode untuk dicari ketergantungan antar item.
- Implementasi real pada promosi di supermarket atau swalayan, akan jamak kita jumpai pembelian 6 pack keju cheedar yang dibundling dengan 1 pack roti tawar.
- Atau kita jumpai, penataan pampers yang berdekatan dengan tissue,

## ○ Inventory Management

- Tujuan : seorang pelanggan perusahaan perbaikan peralatan mengharapkan keaslian dari perbaikan produk konsumen dan menjaga pelayanan dengan menggunakan suku cadang yang baik untuk mengurangi jumlah kunjungan ke rumah pelanggan.
- Pendekatan : memproses data peralatan dan suku cadang yang dibutuhkan pada perbaikan sebelumnya di tempat pelanggan yang berbeda dan menemukan pola kejadian yang berulang.

# Association Rules Pattern

- Bentuk dari aturan asosiasi umumnya dinyatakan dalam bentuk :

$X \Rightarrow Y$  , dimana X dan Y adalah itemset  
Contoh : {roti, mentega}  $\rightarrow$  {susu} (support = 40%, confidence = 50%)

- Aturan tersebut berarti “50%” dari transaksi di database yang memuat item roti dan mentega juga memuat item susu. Sedangkan 40% dari seluruh transaksi yang ada di database memuat ketiga item itu.
- Dapat juga diartikan : seorang konsumen yang membeli roti dan mentega punya kemungkinan 50% untuk juga membeli susu. Aturan ini cukup signifikan karena mewakili 40% dari catatan transaksi selama ini.

# Association Rule Mining

- Jika terdapat sebuah himpunan transaksi  $T$ , maka tujuan dari association rules mining adalah untuk menemukan semua aturan yang mempunyai support  $\geq \textit{minsup}$  dan confidence  $\geq \textit{minconf}$

# Tabel Representasi Biner

Tabel pembelian

Id_Trans	Id_Cust	Tanggal	Item	Jumlah
111	201	5/1/2007	Pena	2
111	201	5/1/2007	Tinta	1
111	201	5/1/2007	Susu	3
111	201	5/1/2007	Jus	6
112	105	6/3/2007	Pena	1
112	105	6/3/2007	Tinta	1
112	105	6/3/2007	Susu	1
113	106	5/10/2007	Pena	1
113	106	5/10/2007	Susu	1
114	201	6/1/2007	Pena	2
114	201	6/1/2007	Tinta	2
114	201	6/1/2007	Jus	4
114	201	6/1/2007	Air	1



Tabel representasi biner

Id_Trans	Pena	Tinta	Susu	Jus	Air
111	1	1	1	1	0
112	1	1	1	0	0
113	1	0	1	0	0
114	1	1	0	1	1

Jika diamati ada redundancy pada table pembelian.

Pembuatan table 'denormalized' untuk mempermudah data mining dilakukan pada tahap data preparation pada CRISP-DM.



# Support - Minimum Support - Frequent itemset

Catatan : Itemset dapat berisi hanya satu item.

Id_Trans	Pena	Tinta	Susu	Jus	Air
111	1	1	1	1	0
112	1	1	1	0	0
113	1	0	1	0	0
114	1	1	0	1	1

- Pada table diatas, dapat kita hitung :
- Support count ( $\sigma$ ), merupakan jumlah transaksi yang berisi suatu itemset tertentu atau merupakan frekuensi kejadian dari suatu itemset.
- **Support** dari suatu item adalah perbandingan dari transaksi dalam basisdata yang berisi semua item dalam itemset.
- Dalam contoh diatas, itemset {pena,tinta} memiliki support 75% dalam table pembelian. Itemset {susu,jus} supportnya hanya 25%
- Frequent itemset menunjukkan itemset yang memiliki frekuensi kemunculan lebih dari nilai minimum yang telah ditentukan ( $\phi$ )
- Jika missal, ditentukan **minimum support** adalah **70%**, maka **frequent-itemset** pada contoh diatas adalah {pena}, {tinta}, {susu}, {pena,tinta}, dan {pena,susu}

# Support and Confidence

- Support (s) dan Confidence (c) merupakan metrik yang digunakan pada Association Rule.
- Support menunjukkan persentasi jumlah transaksi yang berisi X dan Y.
- Sedangkan Confidence menunjukkan persentasi banyaknya Y pada transaksi yang mengandung X.
- Bentuk persamaan matematisnya dapat dituliskan seperti ini:  
$$\text{support}(X \Rightarrow Y) = P(X \cup Y)$$
$$\text{confidence}(X \Rightarrow Y) = P(Y | X)$$
- Contoh : {Milk, Diaper} => {Beer}

$$\text{support}(\{\text{Milk}, \text{Diaper}\} \Rightarrow \{\text{Beer}\}) = \frac{2}{5} = 0.4 = 40\%$$

$$\text{confidence}(\{\text{Milk}, \text{Diaper}\} \Rightarrow \{\text{Beer}\}) = \frac{2}{3} = 0.667 = 66.7\%$$

# Support and Confidence

- Confidence menyatakan seberapa sering item-item dalam Y muncul dalam transaksi yang berisi X

- $S(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N}$ , Dimana S = Support, N = total transaksi

- $C(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)}$ , Dimana c = confidence

{Pena, Tinta}  $\rightarrow$  Jus

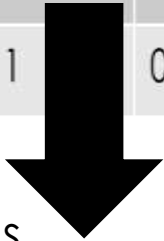
Contoh :  $S(X \rightarrow Y) = \sigma(\{Pena, Tinta, Jus\}) / 4 = 2/4 = 0.5$

$C(X \rightarrow Y) = \sigma(\{Pena, Tinta, Jus\}) / \sigma(\{Pena, Tinta\}) = 2/3 = 0.67$

Id_Trans	Pena	Tinta	Susu	Jus	Air
111	1	1	1	1	0
112	1	1	1	0	0
113	1	0	1	0	0
114	1	1	0	1	1

# $\{\text{milk, Diaper}\} \rightarrow \text{beer}$ , tentukan s dan c nya

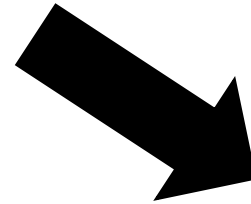
Id_Trans	Pena	Tinta	Susu	Jus	Air
111	1	1	1	1	0
112	1	1	1	0	0
113	1	0	1	0	0
114	1	1	0	1	1



$\{\text{Pena, Tinta}\} \rightarrow \text{Jus}$

Contoh :  $S(X \rightarrow Y) = \sigma(\{\text{Pena, Tinta, Jus}\}) / 4 = 2/4 = 0.5$

$C(X \rightarrow Y) = \sigma(\{\text{Pena, Tinta, Jus}\}) / \sigma(\{\text{Pena, Tinta}\}) = 2/3 = 0.67$



Id trans	Items
1	Bread, milk
2	Bread, diaper, beer, eggs
3	Milk, diaper, beer, coke
4	Bread, milk, diaper, beer
5	Bread, milk, diaper, coker

# Algoritma Association Rules

- Algoritma A priori termasuk dalam association rules.
- Algoritma lainnya yang termasuk kedalam association rules diantaranya :
  - FP-Growth
  - Generalized Rule Induction
  - Hash Based algorithm

# Apriori Algorithm

- Algoritma Apriori pertama kali dikenalkan oleh Agrewal, Imielinski dan Swami.
- Algoritma Apriori merupakan salah satu algoritma yang digunakan di dalam memecahkan persoalan *association rule mining*.
- Yang mengolah suatu database transaksi dengan setiap transaksi adalah suatu himpunan item-item. Kemudian mencari seluruh kaidah apriori yang memenuhi kendala **minimum support** dan **minimum confidence** yang diberikan user.
- Algoritma Apriori dapat digunakan untuk menemukan tren bisnis dengan menganalisa transaksi konsumen.
- Contoh: 30% dari transaksi yang memuat bir juga memuat popok 5% yang artinya 30% merupakan confidence dan 5% merupakan support dari kaidah ini.

# Pseudo-Code Apriori Algorithm

$C_k$ : Candidate itemset dari ukuran  $k$ ;

$L_k$  : Frequent itemset dari ukuran  $k$ .

$L_1 = \{\text{frequent items}\}$ ;

for ( $k = 1$ ;  $L_k \neq 0$ ;  $k++$ ) do begin

$C_{k+1} = \{\text{kandidat dibangun dari}$

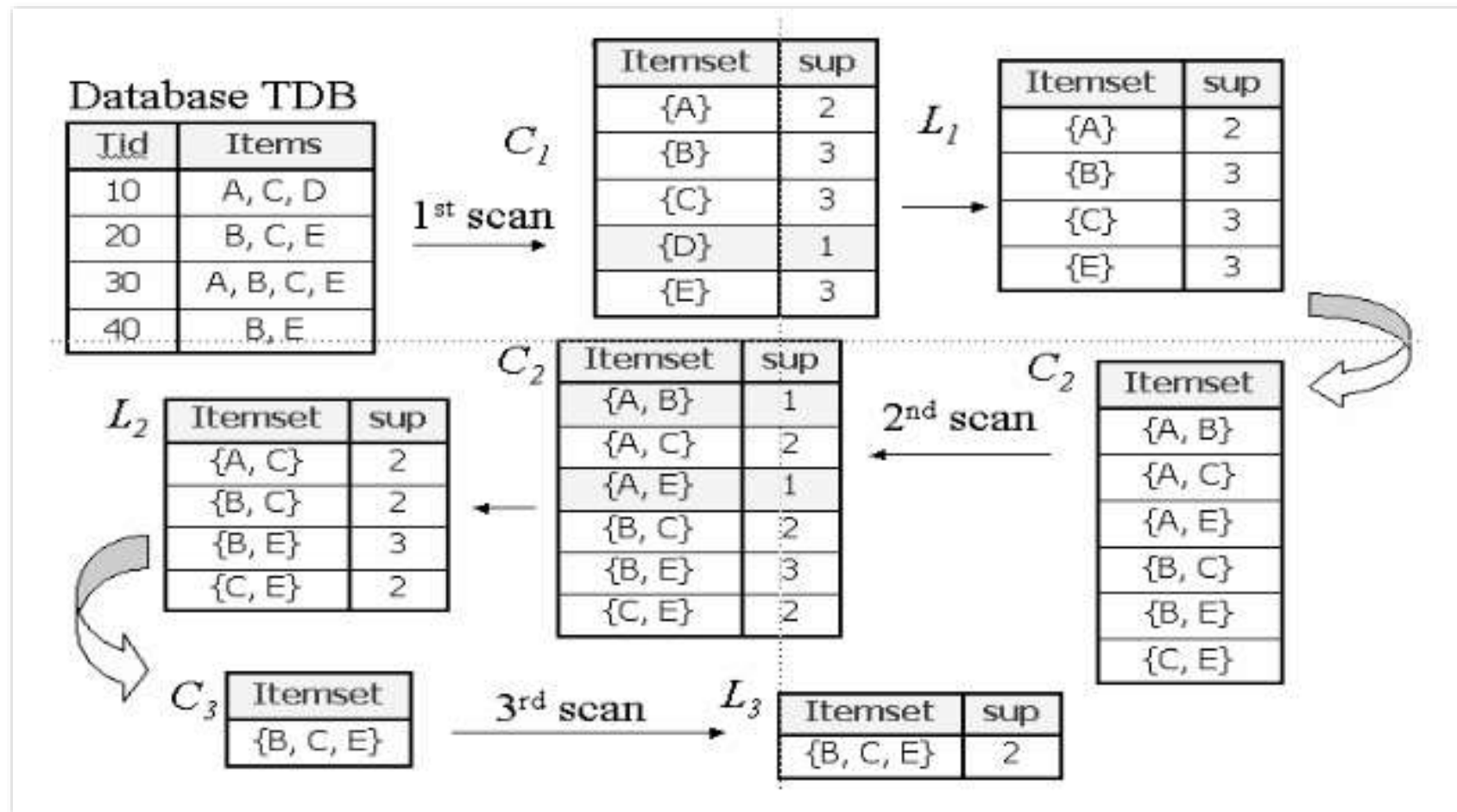
for each transaksi  $t$  yang dimuat dalam database do naikkan hitungan dari seluruh kandidat dalam  $C_{k+1}$  yang dimuat dalam  $t$

$L_{k+1} = \{\text{kandidat dalam } C_{k+1} \text{ dengan } \text{min\_support}\}$

end

return  $\bigcup_{k=1}^{\infty} L_k$ ;

# Ilustrasi Algoritma Apriori





# Contoh Soal:

- Berikut ini Contoh dari 4 Transaksi belanja konsumen, akan dicari hubungan asosiasi antar item dengan minimal support 50%

No	Itemset
1	A.Kopi, C.Gula, D.Bir
2	B.Teh, C.Gula, E.Roti
3	A.Kopi, B.Teh, C.Gula, E.Roti
4	B.Teh, E.Roti

# Contoh Soal ... lanj 2

- Langkah 1:  $L1 = \{\text{large 1-itemset}\}$

Jumlah transaksi = 4.

Min support = 50% artinya 2 dari 4 transaksi.

Langkah 1:  $L1 = \{\text{large 1-itemset}\}$

Itemset	Support
A	50%
B	75%
C	75%
D	25%
E	75%

No	Itemset
1	A.Kopi, C.Gula, D.Bir
2	B.Teh, C.Gula, E.Roti
3	A.Kopi, B.Teh, C.Gula, E.Roti
4	B.Teh, E.Roti



- Langkah 2 : Mencari kandidat itemset untuk L2

- 2.1 Gabungkan itemset pada L1 (algoritma apriori gen) { A B, A C, A D, A E, B C, **B D**, B E, C D, C E, **D E** }

- 2.2 Hapus yang tidak ada dalam itemset      Itemset { B D, DE } dihapus karena tidak ada dalam itemset

# Contoh Soal ... Ianj 3

- Langkah 3 : Hitung Support dari tiap kandidat itemset

No	Itemset
1	A.Kopi, C.Gula, D.Bir
2	B.Teh, C.Gula, E.Roti
3	A.Kopi, B.Teh, C.Gula, E.Roti
4	B.Teh, E.Roti

Itemset	Support
A B	25 %
A C	50 %
A D	25 %
A E	25%
B C	50%
B E	75%
C D	25%
C E	50%

# Contoh Soal ... lanj 4

○ Langkah 4 : L2 {Large 2-itemset}

Itemset	Support
A B	25 %
A C	50 %
A D	25 %
A E	25%
B C	50%
B E	75%
C D	25%
C E	50%

Itemset	Support
A C	50 %
B C	50%
B E	75%
C E	50%

# Contoh Soal ... lanj 5

- Langkah 5 : Ulangi langkah 2-4
  - 5.1 Gabungkan itemset pada L2 dan L2 :,
  - 5.2 Hapus yang tidak ada dalam itemset : { }
- Langkah 6 : Hitung Support dari setiap kandidat itemset L3
- Langkah 7 : L3 {Large 3-Itemset} {BCE}
- Langkah 8 : STOP karena sudah tidak ada lagi

kandidat untuk 4-itemset

No	Itemset
1	A.Kopi, C.Gula, D.Bir
2	B.Teh, C.Gula, E.Roti
3	A.Kopi, B.Teh, C.Gula, E.Roti
4	B.Teh, E.Roti

Itemset	Hasil Gabungan (3 itemset)
A C + B C	A C B
A C + B E	A C B, A C E, A B E B C E
A C + C E	A C E
B C + B E	B C E
B C + C E	B C E
B E + C E	B C E

Itemset	Support
ABC	25%
ACE	25%
ABE	25%
BCE	50%

Langkah  
5

Langkah  
6

No	Itemset
1	A.Kopi, C.Gula, D.Bir
2	B.Teh, C.Gula, E.Roti
3	A.Kopi, B.Teh, C.Gula, E.Roti
4	B.Teh, E.Roti

Goal ... lanjut 6

Confidence =  $\frac{XUY}{X}$

Contoh :

$BC \rightarrow E = \frac{BCE}{BC}$

$A \rightarrow C = \frac{AC}{A}$

○ Hasil Akhir :

L1	L2	L3
A 50%	A C 50%	B C E 50%
B 75%	B C 50%	
C 75%	B E 75%	
D 25%	C E 50%	
E 75%		

○ Untuk mencari aturan asosiasi diperlukan juga min-confidence

○ Misal min-conf : 50%.

○ Aturan yang mungkin terbentuk :

Aturan ( $X \rightarrow Y$ )	Sup( $X \cup Y$ )	Sup(X)	Confidence
B C $\rightarrow$ E	50%	50%	100%
B E $\rightarrow$ C	50%	75%	66.67%
C E $\rightarrow$ B	50%	50%	100%
A $\rightarrow$ C	50%	50%	100 %
C $\rightarrow$ A	50%	75%	66.67%
B $\rightarrow$ C	50%	75%	66.67%
C $\rightarrow$ B	50%	75%	66.67%
B $\rightarrow$ E	75%	75%	100%
E $\rightarrow$ B	75%	75%	100%
C $\rightarrow$ E	50%	75%	66.67%
E $\rightarrow$ C	50%	75%	66.67%



**Ganjil = Support 50% Confidence 50%**  
**Genap = Support 30% Confidence 50%**

1	4,5,3
2	6,9,10,4,7
3	6,1,2,3,7
4	6,1,3
5	4,8,3,7
6	4,6,3,7
7	3,5
8	4,7,3
9	4,6,3,1
10	7,1,4

# Suuport=25% , Conf=50%

Ganjil

1	1,5,6
2	4,5,6
3	7,1,5
4	2,6,8
5	2,5,1
6	1,3,8
7	1,5,7
8	2,1,5,6
9	2,5,7
10	1,3,7

Genap

1	1,7,8,5
2	2,5,1
3	3,4,5
4	2,3
5	7,8,9
6	1,5,7,9
7	3,5,1
8	2,4,6
9	1,2,3
10	5,6



# Soal 1. tentukan Rules yang terbentuk jika MinSupport=30% , MinConf=50%

Ganjil

1	1,5,6
2	4,5,6
3	7,1,5,3
4	2,6,8
5	2,5,1,4
6	1,3,8
7	1,5,7
8	2,1,5,6
9	2,5,7
10	1,3,7,2

Genap

1	1,7,8,5
2	2,5,1,6
3	3,4,5
4	2,3,7
5	7,8,9
6	1,3,5,7,9
7	3,5,1
8	2,4,6
9	1,2,3,5
10	5,6,8

# Quiz → Tentukan Nilai Support dan Confidence Dari Rules Berikut ini :

Ganjil

- {Gula,Susu} → {Kopi}
- {Kopi} → {Gula}
- {Kopi} → {Roti,Susu}

Transaksi	Item yang Dibeli
1	Susu, Teh, Gula
2	Teh, Gula, Roti
3	Teh, Gula
4	Susu, Roti
5	Susu, Gula, Roti
6	Teh, Gula
7	Gula, Kopi, Susu
8	Gula, Kopi, Susu
9	Susu, Roti, Kopi
10	Gula, Teh, Kopi

# Soal 2 → Tentukan Nilai Support dan Confidence Dari Rules Berikut ini :

Ganjil

- {Gula,Susu} → {Kopi}
- Kopi → Gula
- {Teh,Gula} → {Roti}

Genap

- {Gula,Kopi} → {Susu}
- {Susu} → {Roti}
- {Kopi,Susu} → Roti

Transaksi	Item yang Dibeli
1	Susu, Teh, Gula
2	Teh, Gula, Roti
3	Teh, Gula
4	Susu, Roti
5	Susu, Gula, Roti
6	Teh, Gula
7	Gula, Kopi, Susu
8	Gula, Kopi, Susu
9	Susu, Roti, Kopi
10	Gula, Teh, Kopi

## Soal 3 →

# Tentukan Cluster yang terbentuk dari

- $M1 = (3, 5.2),$
- $M2 = (4, 3.5),$
- $M3 = (2, 3.5),$
- $M4 = (5.5, 2.7)$
- $M5 = (7, 3.3),$
- $M6 = (3, 4.5),$
- $M7 = (4, 2.5)$
- $C1 = (2,2)$
- $C2 = (4,4)$
- $K=2$
- Stopping Condition, Iterasi maks = 3

# Soal 4 dan 5

- 4 → Clustering merupakan unsupervised classification, Apakah perbedaan dan persamaan dari supervised classification dan unsupervised classification
- 5 → Jelaskan istilah-istilah berikut ini dan sertai dengan contohnya
  - Frequent item set
  - Antecedent dan consequent
  - Euclidian distance
  - Partitioning clustering

# Today's Quote

Please Study Hard for Your Exams.

There are **no secrets** to success. It is the result of **preparation, hard work, learning from failure, and good team discussion**

# References

- Fajar Astutui Hermawati – Data Mining – 2013
- Kusrini – Algoritma Data Mining – 2009
- Ferry Haryono - FCM dan Apriori untuk Segmentasi Produk - MKom Thesis - Udinus - 2011