

## ANALISIS KLUSTER UNTUK DATA *BINARY* DAN PENERAPANNYA PADA PEMILIHAN TIPE MOBIL

Zulaekhah Pratiwi<sup>1</sup>

### **ABSTRACT**

*The purpose of this study is to describe the cluster analysis for binary data using the single linkage method, to determine the best type of car and to determine the consumption of foods containing protein in several European countries using the single linkage method. There are several stages in determining the cluster analysis, to determine the matrix of data, transforming data in to binary data, determining the similarity matrix, the formation of clusters, determine the number of clusters, interpret the clusters formed. association coefficients for use similarity measure four variable of binary data. The results obtained on the selection of the best type of car that is the type of car Toyota Corrola with economic characteristics, given good service, good design, sporty, safe and comfortable although the price is a bit expensive. While the consumption of protein in some countries lead to the conclusion that most countries that consume protein is Country Czech protein consume die eggs, milk, fish, cereal, nuts and fruits.*

**Key words:** *Analysis cluster, Binary data, coefficients of similarity, matches, mismatches*

### **PENDAHULUAN**

Analisis multivariat merupakan salah satu jenis analisis statistik yang digunakan untuk menganalisis data lebih dari dua variabel secara bersamaan. Variabel-variabel itu saling berkorelasi satu sama lain. Teknik analisis multivariat secara umum diklasifikasi menjadi dua, yaitu analisis dependensi dan analisis interdependensi. Analisis dependensi berfungsi untuk menerangkan atau memprediksi variabel tergantung dengan menggunakan dua atau lebih variabel bebas. Yang termasuk dalam klasifikasi ini adalah analisis regresi linear berganda, analisis

---

<sup>1</sup> Mahasiswa Program Studi Matematika Jurusan Pendidikan Matematika Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Negeri Yogyakarta

diskriminan, analisis varian multivariat dan analisis korelasi kanonik. (Hair, Joseph F., et.al., 1998)

Analisis interdependensi berfungsi untuk memberikan makna terhadap seperangkat variabel atau membuat kelompok-kelompok secara bersama-sama. Yang termasuk dalam klasifikasi ini adalah analisis faktor, analisis kluster dan multidimensional scaling. Metode interdependensi diklasifikasi berdasarkan pada jenis masukan variabel dengan skala pengukuran bersifat matriks atau nonmatriks. Jika masukan data berskala matriks maka dapat digunakan teknik analisis faktor, analisis kluster dan multidimensional scaling. Jika masukan data berskala nonmatriks maka hanya digunakan teknik analisis multidimensional scaling.

Didalam analisis kluster untuk data binary dapat diterapkan dalam pemilihan tipe mobil, banyaknya mobil yang beredar di pasaran semakin menjadikan konsumen lebih selektif dalam memilih mobil yang tepat dan sesuai untuk digunakan. Mobil yang beredar memiliki keunggulan dan kelemahannya sendiri. Misalkan mobil Toyota Corolla harganya ekonomis serta layanan dalam service baik, namun nilai jual mobil ini tidak stabil. Ukuran kesesuaian atau kualitas suatu mobil dapat ditentukan dengan menggunakan beberapa variabel, antara lain ekonomis, layanan, nilai jualnya stabil, harga mobil, desain, mobil sporty, keamanan dan kenyamanan. Jika suatu mobil telah ditentukan kesesuaiannya, barulah mobil itu bisa dibandingkan dengan mobil lainnya untuk ditentukan kelebihan baikannya.

Selain pemilihan tipe mobil analisis kluster bisa diterapkan dalam pemilihan konsumsi protein makanan dalam 25 Negara Eropa. Data tersebut merupakan data kategori dengan skala ordinal. Ada sembilan variabel (karateristik) yang digunakan untuk membandingkan sebuah objek dengan objek lainnya yaitu tentang makanan yang mengandung protein, disini akan diteliti tentang seberapa banyak jumlah makanan yang mengandung protein yang dikonsumsi di setiap Negara di Eropa yaitu daging sapi, daging ayam, telur, susu, Ikan, cereal, makanan bertepung, kacang-kacangan, dan buah-buahan. Memberikan sumbangan terhadap perkembangan ilmu pengetahuan dan menambah wawasan dalam bidang praktis.

Pengelompokan metode hirarki dibedakan menjadi empat berdasarkan konsep jarak antar kluster yaitu metode pautan tunggal (*single linkage*), metode pautan lengkap (*complete linkage*), metode pautan rata-rata (*average linkage*), metode *Centroid method* dan metode varians (*sum of squares methods*). Pada skripsi ini metode yang akan digunakan adalah metode *single linkage*. Salah satu alasan menggunakan *single linkage* ini karena teknik ini menggunakan jarak terdekat antar pasangan data (objek) yang terdapat pada kedua kluster tersebut.

Tujuan yang akan dicapai dalam penelitian ini adalah menjelaskan analisis kluster untuk data *binary* dengan menggunakan metode *single linkage*, menjelaskan metode pembentukan kluster dan ukuran kemiripan yang sesuai digunakan berdasarkan jenis datanya, dan menjelaskan penerapan analisis kluster untuk data *binary* pada pemilihan tipe mobil dan konsumsi protein di suatu negara.

Salah satu manfaat dari penelitian ini adalah memberikan pengetahuan tentang aplikasi ilmu statistik khususnya analisis *cluster*

### **Pengertian Analisis Kluster**

Analisis kluster merupakan suatu teknik analisis multivariat yang bertujuan untuk mengelompokkan unit-unit pengamatan ke dalam beberapa kelompok (kluster) dimana setiap unit pengamatan dalam satu kelompok akan mempunyai ciri yang relatif sama sedangkan antar kelompok memiliki sifat yang berbeda. Pengelompokan dilakukan berdasarkan kemiripan (*similarities*) atau jarak (*distance, dissimilarities*). Input yang diperlukan adalah ukuran kemiripan atau data yang kemiripannya dapat dihitung. (Simamora, 2005)

Langkah-langkah analisis kluster dilakukan dengan beberapa tahap, yaitu :

1. Menentukan matriks data.
2. Standarisasi matriks data.
3. Menentukan matriks jarak atau matriks similaritas.
4. Melakukan prosedur / metode pembentukan kluster.
5. Menentukan banyaknya kluster.

6. Menginterpretasi kluster-kluster yang dibentuk.

### Ukuran Kemiripan Untuk Data Binary

Data multivariat dapat juga diperoleh dengan membentuk sistem kode biner (binary) pada masing-masing variabel (digunakan sistem skala kualitatif). Suatu jenis ukuran kemiripan yang dirumuskan untuk jenis data ini adalah koefisien asosiasi. koefisien asosiasi adalah koefisien yang menunjukkan adanya hubungan atau keterkaitan antar satu variabel. Untuk mengukur kemiripan antara objek-objek selalu dibandingkan pasangan observasi atau objek  $(x_i, x_j)$  dimana  $x_i^T = (x_{i1}, \dots, x_{ip})$ ,  $x_j^T = (x_{j1}, \dots, x_{jp})$ , dan  $x_{ik}, x_{jk} \in \{0,1\}$ . (Hardley, and Simar, 2003: 303)

berarti ada empat kasus

$$x_{ik} = x_{jk} = 1; \quad x_{ik} = 0, x_{jk} = 1; \quad x_{ik} = 1, x_{jk} = 0 \text{ dan } x_{ik} = x_{jk} = 0$$

Koefisien asosiasi dirumuskan berdasarkan kecocokan (*matches*) dan ketidakcocokan (*mismatches*) yang biasanya disusun dalam tabel kontingensi 2x2. Untuk membahas ukuran kemiripan untuk data binary selanjutnya akan dibahas terlebih dahulu tentang pemilihan metode (Algoritma) pembentukan kluster.

### Analisis Kluster untuk Data Binary

Jika objek-objek tidak dapat ditampilkan dengan pengukuran  $p$  dimensi, pasangan-pasangan objek sering dibandingkan berdasarkan karakteristik tertentu. Objek yang similar mempunyai karakteristik yang sama lebih banyak dari pada objek yang disimilar. Suatu karakteristik tersebut dapat digambarkan secara matematis dengan memperkenalkan suatu variabel dan nilai binary, dengan asumsi bahwa nilai 1 jika karakteristik tersebut ada dan nilai 0 jika karakteristik tersebut tidak ada. Ukuran tipe *matching* yang dikenal dengan koefisien asosiasi ini merupakan ukuran yang tepat jika datanya berupa skala nominal. Tipe ukuran similaritas ini mengambil nilai dalam range 0 sampai 1. (Johnson, R.,A., and Wichern, D. W., 1992: 578)

Jika digambarkan bahwa setiap variabel mewakili kehadiran atau ketidakhadiran suatu atribut tertentu, maka data ini disusun dalam tabel kontingensi dua arah. Dalam tabel ini baris dan kolom menghubungkan kehadiran ( $1$ ) dan ketidakhadiran ( $0$ ) dari atribut-atribut untuk individu/objek  $i$  dan  $k$  dan entri-entri sel tersebut merupakan jumlahan dari banyaknya atribut yang ada (atau tidak ada) dari keduanya bersama-sama. (Johnson, R.,A., and Wichern, D. W., 1992: 578).

Untuk variabel binary  $p=5$ , skor untuk dua objek  $i$  dan  $k$  disusun sebagai berikut :

**Tabel 2.2. Data binary objek i dan k untuk variabel  $p=5$**

Objek	Variabel				
	1	2	3	4	5
item i	1	0	0	1	1
Item k	1	1	0	1	0

Dalam kasus ini ada dua *match 1-1*, satu *match 0-0* dan dua *mismatch 1-0* atau *0-1*. Misalkan  $x_{ij}$  adalah skor (1 atau 0) dari variabel binary ke- $j$  pada objek  $i$  dan  $k_{kj}$  adalah skor (1 atau 0) dari variabel binary ke- $j$  pada objek  $k$ ,  $j = 1,2,\dots,p$ , akibatnya :

Dengan mempertimbangkan perlakuan yang berbeda dari *match 1-1* dan *match 0-0* , diberikan beberapa pola untuk mendefinisikan koefisien similaritas untuk memperkenalkan pola ini disebut frekuensi *match* dan *mismatch* objek  $i$  dan  $k$  dalam bentuk tabel kontingensi.

**Tabel 2.3. Kontingensi frekuensi match dan mismatch objek  $i$  dan  $k$**

		Objek $k$		Total
		1	0	
Objek $i$	1	$a$	$b$	$a + b$
	0	$c$	$d$	$c + d$
Total		$a + c$	$b + d$	$p = a + b + c + d$

Definisi 2.1:

Pada tabel di atas  $a$  menunjukkan frekuensi munculnya *match 1-1*,  $b$  adalah frekuensi munculnya *mismatch 1-0*,  $c$  dan  $d$  masing-masing menunjukkan munculnya *mismatch 0-1* *match 0-0*. (Johnson, R.,A., and Wichern, D. W., 1992: 578-579)

## PEMBAHASAN

### Metode Hirarki (*Hierarchical Clustering Methods*)

Pada metode hirarki pengelompokan dimulai dengan dua atau lebih objek yang mempunyai kesamaan paling dekat. Proses dilanjutkan ke objek lain yang mempunyai kedekatan ke dua. Demikian seterusnya sehingga kluster membentuk semacam pohon dimana ada hirarki (tingkatan) yang jelas antar objek, dari yang paling mirip. Secara logika semua objek pada akhirnya hanya akan membentuk sebuah kluster. Proses ini dapat digambarkan dengan dendrogram. Dari tingkat yang paling mendasar ada dua metode hirarki yaitu metode aglomeratif (*agglomerative methods*) dan metode divisif (*divisive methods*). (Johnson, R.,A., and Wichern,D.W., 1992: 584)

Pertama ditentukan jarak terdekat  $D = \{d_{ij}\}$ , dan gabungan objek-objek yang bersesuaian katakan  $U$  dan  $V$  untuk mendapatkan kluster  $(UV)$ . Untuk tahap 3 pada algoritma diatas, jarak antara  $(UV)$  dan kluster  $W$  yang lain dihitung dengan :

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\}$$

Dimana  $d_{UW}$  dan  $d_{VW}$  adalah jarak antara tetangga terdekat kluster-kluster  $U$  dan  $W$ , dan kluster-kluster  $V$  dan  $W$ . Hasil pembentukan kluster pautan tunggal dimana  $d_{UW}$  dan  $d_{VW}$  adalah jarak antara tetangga terdekat kluster-kluster  $U$  dan  $W$ , dan kluster-kluster  $V$  dan  $W$ . Hasil pembentukan kluster pautan tunggal ini dapat digambarkan secara grafik dalam suatu dendrogram atau diagram pohon.

Kemudian akan dibahas proses pembentukan kluster yang termasuk dalam metode hirarki. Pengelompokan metode hirarki dibedakan menjadi lima berdasarkan konsep jarak antar kluster yaitu metode pautan tunggal (*single linkage*), metode pautan lengkap (*complete linkage*), metode pautan rata-rata (*average linkage*), metode *Cendroid method* dan metode varians (*sum of squares methods*). (Hair, Joseph F., et al. 1998: 586-588). Pada skripsi ini metode yang digunakan adalah metode *single linkage*. Salah satu alasan menggunakan metode *single linkage* ini karena teknik ini menggunakan jarak terdekat antar pasangan data (objek) yang terdapat pada kedua kluster tersebut. Selanjutnya jarak dekat lainnya dideteksi, objek ketiga ini bisa saja dikelompokkan bersama dua objek terdahulu atau dijadikan kluster sendiri. Pada setiap tahap, jarak antar dua kluster adalah jarak paling dekat diantara keduanya.

Penerapan analisis kluster pada pemilihan tipe mobil. Data rata-rata nilai 24 tipe mobil dari sampel yang diambil 40 orang. Dengan range 1 (sangat baik) sampai 6 (sangat jelek). Variabel-variabelnya tersebut adalah

$X_1$  : Ekonomis (Dalam penggunaan bensin)

$X_2$  : Layanan ( Fasilitas servis )

$X_3$  : Nilai jualnya stabil (Nilai jualnya tidak naik tidak turun)

$X_4$  : Harga untuk mobil yang murah

$X_5$  : Desain (Bentuk menarik)

$X_6$  : Mobil sporty (Keunggulan mobil)

$X_7$  : Keamanan (Fasilitas Keamanan)

$X_8$  : Kenyamanan (Kenyamanan dalam menggunakan misalkan AC non AC).

Ini adalah data 24 mobil dengan 8 variabel yang ditransformasikan ke data binary:

NO	TIPE	Variabel							
		$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$
1	Audi 100	1	0	0	1	0	0	0	1
2	BMW Seri 5	1	0	0	1	0	0	0	1
3	Citroen AX	0	1	1	0	1	1	1	0
4	Ferrari	1	0	0	1	0	0	1	1
5	Flat Uno	0	1	1	0	1	1	1	0
6	Foid Fiesta	0	1	1	0	0	0	1	1
7	Hyundai	0	1	1	0	1	0	1	0
8	Jaguar	1	0	0	1	0	0	0	1
9	Lada Samark	0	1	1	0	1	1	1	1
10	Mazda 323	0	1	1	0	1	0	1	1
11	Mercedes 200	1	0	0	1	0	0	0	0
12	Mitsubishi Galant	0	0	1	1	0	0	0	0
13	Nissan Sunny	0	1	1	0	1	1	1	0
14	Opei Corsa	0	0	0	0	0	1	0	0
15	Opel Vectro	0	0	0	1	0	0	0	0
16	Peugeot 306	0	1	1	0	0	0	0	0
17	Renault 19	0	1	1	0	0	0	0	0
18	Rover	1	0	0	0	0	0	0	1
19	Toyota Corolla	0	0	1	0	0	0	0	0
20	Volvo	1	0	0	1	0	1	0	0
21	Traban 601	1	1	1	0	1	1	1	1
22	VW Golf	0	0	0	0	0	0	0	0
23	VW Passat	0	0	0	0	1	1	0	0
24	Wartburg 1.3	1	1	1	0	1	1	1	1

### Interpretasi Kluster

Untuk menginterpretasi 4 kluster yang telah diperoleh dan membuat profilnya, digunakan rata-rata (Cendroit) setiap kluster pada setiap variabel

Tabel 3.21. Profil Kluster 1

Negara	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$
Albania	1	0	1	0	0	1	0	1	0
Bulgaria	0	0	0	0	0	1	0	1	1
Hungary	0	1	0	0	0	1	0	1	1
Italy	0	0	0	0	0	1	0	1	1
Poland	0	1	0	1	0	1	1	0	1
Romania	0	0	0	0	0	1	0	1	0
USSR	0	0	0	0	0	1	1	1	0
Yugoslovia	0	0	0	0	0	1	0	1	0
Rata-rata	1.000	0.250	1.000	1.000	0.000	1.000	0.250	0.875	0.500



Pada tabel diatas kluster 1 memiliki rata-rata paling rendah pada variabel 2 dan 5, artinya Negara pada kluster ini kurang mengkonsumsi daging Sapi tetapi banyak mengkonsumsi ikan.

Tabel 3.22. Profil Kluster 2

Negara	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>
Austria	0	1	1	1	0	0	0	0	1
Belgium	1	1	1	1	0	0	1	0	0
Denmark	1	1	1	1	1	0	1	0	0
E.Germany	0	1	1	0	1	0	1	0	0
Finland	1	0	0	1	1	0	1	0	0
France	1	1	1	1	1	0	1	0	1
Ireland	1	1	1	1	0	0	1	0	0
Netherlands	0	1	1	1	0	0	0	0	0
Norway	0	0	0	1	1	0	1	0	0
Portugal	0	0	0	0	1	0	1	1	1
Spain	0	0	0	0	1	0	1	1	1
Sweden	1	1	1	1	1	0	0	0	0
Switzerland	1	1	0	1	0	0	0	0	1
UK	1	0	1	1	0	0	1	1	0
W.Germany	1	1	1	1	0	0	1	0	0
Rata-rata	0.600	0.666	0.666	0.800	0.533	0.000	0.733	0.200	0.333

Rata-rata terendah pada kluster 2 terdapat pada variabel 6 yaitu 0,0 dan tertinggi pada variabel 4 yaitu 0.800. Artinya Negara ini kurang mengkonsumsi susu dan banyak mengkonsumsi careals.

Tabel 3.23. Profil Kluster 3

Negara	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>
Czech	1	1	0	0	0	1	1	0	0
Rata-rata	1.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000

Pada tabel diatas kluster 3 rata-rata paling rendah yaitu 0,0 terdapat pada variabel 3, 4, 5, 6, 8, 9 artinya negara ini banyak mengkonsumsi Telur, Susu, Ikan, Careals, Kacang-kacangan dan buah-buahan

Pada tabel diatas kluster 4 rata-rata paling rendah yaitu 0,0 terdapat pada variabel 2, 3, 7 artinya negara ini banyak mengkonsumsi daging ayam,telur dan makanan bertepung.

Dari data pengelompokan profil kluster tersebut dapat disimpulkan bahwa negara yang banyak mengkonsumsi protein adalah Negara Czeeh yaitu banyak mengkonsumsi Telur, Susu, Ikan, Cereal, Kacang-kacangan dan buah-buahan.

Penerapan analisis kluster pada konsumsi makanan. Data rata-rata nilai 25 Negara Eropa dari suatu sampel yang diambil dari 40 orang dalam PSP (Persen Sumber Protein)

dengan:

$X_1$  : Daging sapi

$X_2$  : Daging ayam

$X_3$  : Telur

$X_4$  : Susu

$X_5$  : Ikan

$X_6$  : Cereal

$X_7$  : Makanan bertepung

$X_8$  : Kacang-kacangan

$X_9$  : Buah-buahan

Ini adalah data 25 mobil dengan 9 variabel yang ditransformasikan ke data binary:

No	Tipe	Variabel								
		$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$
1	Albania	1	0	1	0	0	1	0	1	0
2	Austria	0	1	1	1	0	0	0	0	1
3	Belgium	1	1	1	1	0	0	1	0	0
4	Bulgaria	0	0	0	0	0	1	0	1	1
5	Czech	1	1	0	0	0	1	1	0	0
6	Denmark	1	1	1	1	1	0	1	0	0
7	E Germany	0	1	1	0	1	0	1	0	0
8	Finland	1	0	0	1	1	0	1	0	0
9	France	1	1	1	1	1	0	1	0	1
10	Greece	1	0	0	1	1	1	0	1	1
11	Hungary	0	1	0	0	0	1	0	1	1
12	Ireland	1	1	1	1	0	0	1	0	0
13	Italy	0	0	0	0	0	1	0	1	1
14	Netherlaads	0	1	1	1	0	0	0	0	0
15	Norway	0	0	0	1	1	0	1	0	0
16	Poland	0	1	0	1	0	1	1	0	1
17	Portugal	0	0	0	0	1	0	1	1	1
18	Romania	0	0	0	0	0	1	0	1	0
19	Spain	0	0	0	0	1	0	1	1	1
20	Swedan	1	1	1	1	1	0	0	0	0
21	Switzerland	1	1	0	1	0	0	0	0	1
22	UK	1	0	1	1	0	0	1	1	0
23	USSR	0	0	0	0	0	1	1	1	0
24	W.German	1	1	1	1	0	0	1	0	0
25	Yogosloslavia	0	0	0	0	0	1	0	1	0

### Interpretasi Kluster

Untuk menginterpretasi 4 kluster yang telah diperoleh dan membuat profilnya, digunakan rata-rata (Cendroit) setiap kluster pada setiap variable

**Tabel 3.21. Profil Kluster 1**

Negara	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>
Albania	1	0	1	0	0	1	0	1	0
Bulgaria	0	0	0	0	0	1	0	1	1
Hungary	0	1	0	0	0	1	0	1	1
Italy	0	0	0	0	0	1	0	1	1
Poland	0	1	0	1	0	1	1	0	1
Romania	0	0	0	0	0	1	0	1	0
USSR	0	0	0	0	0	1	1	1	0
Yugoslavia	0	0	0	0	0	1	0	1	0
Rata-rata	1.000	0.250	1.000	1.000	0.000	1.000	0.250	0.875	0.500

Pada tabel diatas kluster 1 memiliki rata-rata paling rendah pada variabel 2 dan 5, artinya Negara pada kluster ini kurang mengkonsumsi daging Sapi tetapi banyak mengkonsumsi ikan.

**Tabel 3.22. Profil Kluster 2**

Negara	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>
Austria	0	1	1	1	0	0	0	0	1
Belgium	1	1	1	1	0	0	1	0	0
Denmark	1	1	1	1	1	0	1	0	0
E. Germany	0	1	1	0	1	0	1	0	0
Finland	1	0	0	1	1	0	1	0	0
France	1	1	1	1	1	0	1	0	1
Ireland	1	1	1	1	0	0	1	0	0
Netherlands	0	1	1	1	0	0	0	0	0
Norway	0	0	0	1	1	0	1	0	0
Portugal	0	0	0	0	1	0	1	1	1
Spain	0	0	0	0	1	0	1	1	1
Sweden	1	1	1	1	1	0	0	0	0
Switzerland	1	1	0	1	0	0	0	0	1
UK	1	0	1	1	0	0	1	1	0
W. Germany	1	1	1	1	0	0	1	0	0
Rata-rata	0.600	0.666	0.666	0.800	0.533	0.000	0.733	0.200	0.333

Rata-rata terendah pada kluster 2 terdapat pada variabel 6 yaitu 0,0 dan tertinggi pada variabel 4 yaitu 0.800. Artinya Negara ini kurang mengkonsumsi susu dan banyak mengkonsumsi cereals.

**Tabel 3.23. Profil Kluster 3**

Negara	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>
Czech	1	1	0	0	0	1	1	0	0
Rata-rata	1.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000

Pada tabel diatas kluster 3 rata-rata paling rendah yaitu 0,0 terdapat pada variabel 3, 4, 5, 6, 8, 9 artinya negara ini banyak mengkonsumsi Telur, Susu, Ikan, Careals, Kacang-kacangan dan buah-buahan

Tabel 3.24. Profil Kluster 4

Negara	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>
Greece	1	0	0	1	1	1	0	1	1
Rata-rata	1.000	0.000	0.000	1.000	1.000	1.000	0.000	1.000	1.000

Pada tabel diatas kluster 4 rata-rata paling rendah yaitu 0,0 terdapat pada variabel 2, 3, 7 artinya negara ini banyak mengkonsumsi daging ayam,telur dan makanan bertepung.

Dari data pengelompokan profil kluster tersebut dapat disimpulkan bahwa negara yang banyak mengkonsumsi protein adalah Negara Czech yaitu banyak mengkonsumsi Telur, Susu, Ikan, Cereal, Kacang-kacangan dan buah-buahan.

## SIMPULAN DAN SARAN

### Simpulan

1. Langkah-langkah dalam menentukan analisis kluster adalah menentukan matriks data, transformasi data ke data binary, menentukan matriks similaritas, metode pembentukan kluster, menentukan banyaknya kluster, menginterpretasi kluster-kluster yang dibentuk. Untuk menentukan kemiripan antar objek jika datanya berbentuk binary diperlukan ukuran asosiasi (koefisien asosiasi)..
2. Dalam analisis kluster untuk data binary ada 2 penerapan yaitu:
  - a. Dalam pemilihan tipe mobil dengan menggunakan metode *single linkage* menghasilkan kesimpulan bahwa mobil yang paling baik yaitu tipe mobil Toyota Corrola dengan ciri-ciri ekonomis, layanan yang diberikan bagus, harganya agak mahal, desainya bagus, sporti aman dan nyaman.
  - b. Dalam konsumsi protein di beberapa Negara Eropa menghasilkan kesimpulan bahwa Negara yang paling banyak mengkonsumsi protein yaitu Negara Czech

protein yang dikonsumsi adalah Telur, Susu, Ikan, Cereals, Kacang-kacangan dan Buah-buahan.

### Saran

Dalam penulisan skripsi ini hanya terbatas pada pembahasan analisis kluster hirarki dengan menggunakan metode pautan tunggal (*single linkage*). Sedangkan dalam analisis kluster masih terdapat banyak metode yang belum digunakan. Oleh karena itu, sebagai saran untuk pembaca yang tertarik pada topik pembahasan ini dapat membahas lebih lanjut tentang analisis kluster dengan metode yang lain, seperti metode pautan lengkap (*Complete linkage*), pautan rata-rata (*Average linkage*), metode *Cendroid method* dan metode varians (*Sum of Squares methods*).

### DAFTAR PUSTAKA

- Balestre, M., et al., 2008, *Comparison of maize Similarity and Dissimilarity Genetic Coefficient based on Microsatellite Marker*, Journal GMR, 695-705, Lavras, MG, Brasil.
- Borger, F.H., and Barnet, D.C., 1987, *Applying Cluster Analysis in Counseling Psychology Research*, Journal of Counseling, Vol 34. No 4, 456-468, American Psychological Association, Inc.
- Dillon, W.R., and Goldstein, M., 1984, *Multivariate Analysis*, John Wiley & Sons, New York.
- Finch, H., 2005, *Comparison of Distance Measures in Cluster Analysis with Dichotomous Data*, Ball State University.
- Hair, Joseph F., et al., 1998, *Multivariate Data Analysis*, Fifth Edition, Prentice-Hall, Inc., New Jersey.
- Hardley, W., and Simar, L., 2003, *Applied Multivariate Statistical Analysis*, Springer Heidelberg.
- Harry. (2009). Analisis kluster [.http://www.harryonline.info/2009/01/analisis-kluster.html](http://www.harryonline.info/2009/01/analisis-kluster.html). diakses tanggal 12 april 2013

- Johnson, R.,A., and Wichern, *Applied Multivariate Statistical Analysis*, Fifth Edition , Prentice – Hall, Inc., New Jersey.
- Laurenco, F., et. Al., 2004, *Binary Based Similarity Measures For Kategorical Data and Their Application in Self Organising Maps*, Journal , Universidade Nova de Lisboa.
- Muirhead, R. J., 1982, *Aspects of Multivariate Statistical Theory*, John Wiley & Sons Inc.
- Santoso, B., 2007, *Data Mining : Teknik Pemanfaatan Data Untuk Bisnis* , Edisi Pertama , Graha Ilmu Yogyakarta.
- Simamora, B., 2005, *Analysis Multivariate Pemasaran*, Edisi Pertama Penerbit PT Gramedia Utama Pustaka Yogyakarta
- Supama, et. al., 2003, *Kalkulus 1*, Jurusan Matematika dan Ilmu Pengetahuan Alam Universitas Gajah Mada, Yogyakarta.
- Supranto,J.2004. *Analisis Multivariat Arti dan Interpretasi*. Jakarta: Rineka Cipta.
- Yani Soraya. (2011). *Perbandingan Kinerja Metode Single Linkage , Metode Complete Linkage dan Metode K-Means Dala Analisis Kluster*. Skripsi Universitas Negeri Semarang.