

# **PENGELOMPOKAN DATA KATEGORIK DENGAN ALGORITMA K-MODUS**

**SKRIPSI**

MPM 31/56

Asm.

P



**JUNAEDI ASMORO**

**JURUSAN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS AIRLANGGA  
SURABAYA  
2006**



## LEMBAR PENGESAHAN

**Judul** : **Pengelompokan Data Kategorik Dengan Algoritma K-Modus**  
**Penyusun** : **Junaedi Asmoro**  
**NIM** : **080112405**  
**Tanggal Ujian** : **17 Februari 2006**

**Disetujui Oleh :**

**Pembimbing I,**



**Ir. Dyah Herawatie, M.Si.**  
**NIP. 132 061 804**

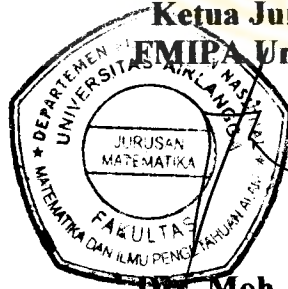
**Pembimbing II,**



**Drs. Eto Wuryanto, DEA.**  
**NIP. 131 933 015**

**Mengetahui :**

**Ketua Jurusan Matematika  
EMIPA Universitas Airlangga**



**Drs. Moh. Imam Utoyo, M.Si.**  
**NIP. 131 801 397**

## **PEDOMAN PENGGUNAAN SKRIPSI**

Skripsi ini tidak dipublikasikan, namun tersedia di perpustakaan dalam lingkungan Universitas Airlangga. Diperkenankan untuk dipakai sebagai referensi kepustakaan, tetapi pengutipan seijin penulis dan harus menyebutkan sumbernya sesuai kebiasaan ilmiah.

**Dokumen skripsi ini merupakan hak milik Universitas Airlangga.**



*SATU-SATUNYA HAL  
YANG HARUS KITA TAKUTI  
ADALAH KETAKUTAN ITU SENDIRI*



*BERANTKANLAH DIRIMU  
SEBAB BERJUTA-JUTA ORANG YANG BERHASIL  
BUKAN KARENA PINTAR,  
TAPI KARENA DIA BERANI*

## UCAPAN TERIMA KASIH

Dengan rasa syukur alhamdulillah, penulis menyampaikan ucapan terima kasih kepada :

1. Drs. H. A. Latief Burhan, M.S selaku Dekan FMIPA Unair.
2. Ir. Dyah Herawatie, M.Si dan Drs. Eto Wuryanto, DEA selaku pembimbing I dan pembimbing II yang dengan kesabaran dan ketulusannya telah membimbing penulis dalam menyelesaikan skripsi ini.
3. Toha Saifudin, S.Si, M.Si dan Drs. Ardi Kurniawan, M.Si selaku dosen penguji dan dosen wali yang telah dengan ikhlas membimbing dan mengarahkan penulis dalam menyelesaikan kuliah di FMIPA Unair.
4. Bapak dan ibu dosen jurusan Matematika atas ilmu, pendidikan dan nasehat yang diberikan kepada penulis.
5. *Mas Edi, Mas Milan, dan Mas Udin* makasih atas lancarnya semua urusan pendidikan dan administrasiku.
6. Bapakku *Harsojo* dan Ibuku *Sa'diyah* yang tidak henti-hentinya memberikan doa, restu dan segalanya demi kesuksesan dan kelancaran penulis selamanya.
7. Kakak-kakakku *Ana* dan *Nina* serta iparku *Moh. Tri* dan *Aris* terimakasih atas suport dan dukungannya. Buat keponakan-keponakanku *andri, nanda & hanafi* nurut sama ortu, jangan lupa belajar.

8. "Nduk" makasih buat masa-masa yang indah selama ini, Motivasi serta semangatmu benar-benar aku butuhkan dalam setiap langkahku.
9. Sahabatku PHB (Pemuda Harapan Bangsa) : *hamsah* (u are one of all best programmer), *sugenk* (makasih buat belajar barengnya), *dee\_deen*, *apet*, *paman* : keep fighting, bangsa & negara nggak sabar nunggu kehadiran kita.
10. teman-teman TDC : *ipunk*, *grespho*, *cuplies*, *yanto*, *oyot*, *kurniawan* (cepatan maju skripsi), *haryoko*, *fauzi*, *dodol*, *bogel*, *miko* mungkinkah kita masuk timnas Indonesia?? .Teman-teman angkatan 00, 01 dan 02 thanx's buat kebersamaannya. Makasih buat teman canda dan tawa *sibaway*, *nyamun*, *bambang*, *rizal*, *colied*
11. Tim Bola Volley Unair, LIVOMA 2006 menanti kita.
12. Semua pihak yang telah berperan dan membantu dalam penyelesaian skripsi ini.

## PRAKATA



Puji syukur yang sedalam – sedalamnya penulis panjatkan kehadiran Allah SWT atas berkat, rahmat serta hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Pengelompokan Data Kategorik Dengan Algoritma K-Modus”.

Sholawat serta salam semoga tetap tercurahkan atas Nabi Besar Muhammad SAW, pemimpin serta panutan kami, yang senantiasa sebagai inspirator serta pendorong segala gerak langkah kami untuk menyelesaikan skripsi ini.

Pada kesempatan ini, penulis juga menyampaikan ucapan terimakasih yang sebesar – besarnya kepada Ibu Ir. Dyah Herawatie, M.Si dan Bapak Drs. Eto Wuryanto, DEA sebagai pembimbing yang telah meluangkan waktunya untuk memberikan bimbingan dan nasehat serta rekan – rekan yang telah membantu dalam penyelesaian skripsi ini.

Akhir kata penulis mengharap, semoga skripsi ini dapat memberikan sesuatu yang bermanfaat bagi pembaca pada umumnya dan jurusan Matematika pada khususnya.

Surabaya, Februari 2006

Penyusun



Junaedi Asmoro, 2005. *Pengelompokan Data Kategorik Dengan Algoritma K-Modus*. Skripsi ini dibawah bimbingan Ir. Dyah Herawatie, M.Si dan Drs. Eto Wuryanto, DEA. Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Airlangga Surabaya.

---

## ABSTRAK

Skripsi ini bertujuan untuk melakukan pengelompokan data berjenis kategori dengan menggunakan metode non-Hirarki berdasarkan algoritma K-Modus dalam analisis kelompok. Algoritma K-Modus merupakan pengembangan dari algoritma K-Mean dengan mengganti mean dengan modus sebagai pusat kelompoknya.

Pengelompokan data dengan algoritma K-Modus dilakukan secara iterasi dengan empat langkah dasar. Langkah yang pertama yaitu menentukan pusat kelompok (modus) awal. Langkah kedua, menghitung jarak antar obyek pengamatan dengan pusat kelompok. Langkah ketiga yaitu menempatkan obyek-obyek pengamatan pada kelompok yang mempunyai jarak terdekat (minimal) dengan pusat kelompoknya. Langkah yang keempat menentukan modus yang baru dari masing-masing kelompok, proses selanjutnya kembali ke langkah kedua. Kemudian proses ini diulang sampai tidak ada obyek yang berpindah ke kelompok lain.

Program S-plus yang telah dibuat diterapkan pada dua data sekunder. Pada pengelompokan data I baik menggunakan metode I dan II didapatkan hasil pengelompokan yang sama jika nilai  $k = 2$ , sedangkan pada pengelompokan data II didapatkan hasil pengelompokan yang berbeda jika nilai  $k = 3$ . Perbedaan nilai awal dalam menetapkan pusat kelompok dapat memberikan hasil yang berbeda dalam proses pengelompokan

Kata kunci : K-Mean, K-Modus, Data kategorik, Analisis kelompok



Junaedi Asmoro, 2006. Clustering of Categorical Data by using K-Modes Algorithm. The *skripsi* was under guidance of Ir. Dyah Herawatie, M.Si. dan Drs. Eto Wuryanto, DEA. Mathematics Department. Faculty of Mathematics and Natural Science. Airlangga University.

---

## ABSTRACT

The purpose of this *skripsi* is to cluster the categorical data by using non-hierarchical method based on K-Modes algorithm in cluster analysis. K-Modes algorithm is the improvement of K-Means algorithm by changing means with modes as the center of cluster.

Clustering of data by using K-Modes algorithm can be applied by iteration with four basic steps. The first step is to determine the first center of cluster (modes). Determination of the first center of cluster can be done by using two kind of methods. The second step is to calculate the distance between observation object with the center of cluster. The third step is to allocate observation objects in the cluster which has nearest distance (minimum) with the center of cluster. The fourth step is to determine the new modes from each cluster, the next process is back to the second step. These process is repeated until there is no object that move to another cluster.

The S-Plus program which have been made applied at two secondary data. In the clustering of data I by using method I or II was obtained the same result of clustering if the value of  $k$  is 2, whereas in clustering of data II was obtained the different result of clustering if the value of  $k$  is 3. The different of first value in determining the center of clustering can give different result in process of clustering.

Key words : K-Mean, K-Modes, Categorical data, Cluster analysis

## DAFTAR ISI

HALAMAN JUDUL.....	i
PRAKATA.....	ii
ABSTRAK .....	iii
ABSTRACT .....	iv
DAFTAR ISI.....	v
DAFTAR TABEL.....	viii
DAFTAR GAMBAR .....	ix
DAFTAR LAMPIRAN.....	x
BAB I. PENDAHULUAN .....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan.....	2
1.4 Manfaat.....	3
BAB II TINJAUAN PUSTAKA.....	4
2.1 Analisis Multivariat.....	4
2.2 Analisis Kelompok.....	4
2.2.1 Ukuran Kemiripan dan Ketakmiripan.....	4
2.2.2 Jarak Chi-square.....	5
2.2.3 Metode Hirarki.....	6
2.2.4 Metode NonHirarki.....	8
2.2.5 Algoritma K-Mean.....	8

2.2.6 Algoritma K-Modus.....	9
2.3 Kategori.....	9
2.3.1 Data Kategorik.....	10
2.4 Frekuensi Kategori dan Frekuensi Relatif Kategori.....	10
2.5 Modus Suatu Himpunan.....	10
2.6 S-Plus.....	11
BAB III METODE PENULISAN .....	12
BAB IV HASIL DAN PEMBAHASAN .....	14
4.1 Jumlah Kelompok ( $k$ ).....	14
4.2 Frekuensi Kategori dan Modus Yang Baru.....	14
4.3 Pusat Kelompok.....	14
4.4 Ukuran Ketakmiripan.....	17
4.5 Algoritma K-Modus.....	17
4.6 Data .....	18
4.7 Penerapan Algoritma ke Program Komputer.....	19
4.8 Pengelompokan Pada Data Secara Manual.....	20
4.8.1 Metode I.....	20
4.8.2 Metode II.....	23
4.9 Pengelompokan Pada Data Dengan program Komputer.....	27
4.9.1 Data I.....	27
4.9.2 Data II.....	30
BAB V KESIMPULAN DAN SARAN .....	34
5.1 Kesimpulan .....	34

5.2 Saran..... 34

DAFTAR PUSTAKA

LAMPIRAN



## DAFTAR TABEL

<b>Nomor</b>	<b>Judul Tabel</b>	<b>Halaman</b>
4.1	Frekuensi kategori dari seluruh variabel .....	14
4.2	Penyajian data I setelah didefinisikan.....	20
4.3	Hasil pengelompokan iterasi pertama pada data I dengan metode I.....	22
4.4	Frekuensi relatif kategori untuk iterasi pertama pada data I....	22
4.5	Hasil pengelompokan akhir data I secara manual menggunakan algoritma K-Modus dengan metode I.....	23
4.6	Frekuensi kategori data I.....	24
4.7	Hasil pengelompokan iterasi pertama pada data I dengan metode II.....	26
4.8	Hasil pengelompokan akhir data I secara manual menggunakan algoritma K-Modus dengan metode II.....	27
4.9	Hasil pengelompokan data I dengan program komputer menggunakan metode I dan metode II.....	29
4.10	Hasil pengelompokan data II dengan menggunakan metode I dan metode II pada algoritma K-Modus.....	31

## DAFTAR GAMBAR

<b>Nomor</b>	<b>Judul Gambar</b>	<b>Halaman</b>
4.1	Dendogram dari data I dengan metode keterkaitan tunggal ( <i>single linkage</i> ).....	28
4.2	Dendogram dari dat II dengan metode keterkaitan tunggal ( <i>single linkage</i> ) .....	30





## DAFTAR LAMPIRAN

<b>Nomor</b>	<b>Judul Lampiran</b>
1.	Daftar program
2.	Data I tentang karakteristik Individu
3.	Data II tentang observasi terhadap jenis mobil
4.	Proses pengelompokan data



# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Untuk mengamati suatu obyek dilakukan dengan lebih dari satu variabel, dengan harapan dapat diperoleh informasi tambahan sehingga dihasilkan kesimpulan yang lebih baik. Analisis terhadap lebih dari satu variabel secara bersama-sama ini disebut Analisis Multivariat. Salah satu metode yang ada dalam Analisis Multivariat adalah Analisis Kelompok (*Cluster Analysis*). Tujuan dari analisis ini adalah untuk mengelompokkan obyek-obyek pengamatan menjadi beberapa kelompok berdasarkan pengukuran variabel-variabel yang ingin diamati.

Di dalam Analisis Kelompok, terdapat beberapa metode pengelompokan, salah satunya adalah metode hirarki. Pada metode ini jumlah kelompok tidak diketahui. Metode yang lain adalah metode non hirarki. Metode non hirarki merupakan suatu pengelompokan yang didasarkan pada jumlah kelompok yang telah ditentukan dengan menempatkan anggota kelompok yang relatif homogen dalam kelompok yang sama. Yang termasuk dalam metode non hirarki antara lain K-Mean, *Partitioning Around Medoids* (PAM), dan K-Median.

K-Mean adalah algoritma yang paling sering digunakan. Algoritma ini bertujuan meminimalkan total jarak antara masing-masing obyek pengamatan dengan pusat kelompoknya (yaitu mean atau rata-rata kelompok). Karena menggunakan jarak dan mean, algoritma K-Mean sesuai digunakan jika datanya berjenis numerik. Yang menjadi masalah dalam hal ini adalah, bagaimana jika

data yang ingin dikelompokkan berjenis kategorik? Data jenis ini sering ditemui pada *data mining*. Salah satu cara pendekatan tradisional yang dapat digunakan adalah mengkonversikan data kategorik ke nilai numerik.

Dalam skripsi ini akan dibahas tentang algoritma pengelompokan untuk data kategorik. Algoritma ini disebut dengan K-Modus, yang merupakan modifikasi dari algoritma K-Mean. Algoritma K-Modus bertujuan sama seperti K-Mean, dengan menggantikan mean dengan modus sebagai pusat kelompok. Kelebihan algoritma K-Modus dari pada algoritma K-Mean adalah algoritma K-Modus bisa digunakan pada data selain data berjenis numerik, misalnya data kategorik.

## 1.2 Rumusan Masalah

1. Bagaimana algoritma K-Modus melakukan pengelompokan pada data dengan variabel berjenis kategori?
2. Bagaimana mengimplementasikan algoritma pengelompokan tersebut dalam program komputer dalam bahasa *S-Plus*?

## 1.3 Tujuan

1. Membahas bagaimana algoritma K-Modus melakukan pengelompokan pada data dengan variabel berjenis kategori.
2. Membahas bagaimana implementasi algoritma K-modus dalam program komputer dalam bahasa *S-Plus*.

#### 1.4 Manfaat

Menambah wawasan baru tentang pengelompokan data dengan menggunakan analisis kelompok terutama pada data dengan variabel berjenis kategori.



## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Analisis Multivariat

Analisis Mutivariat adalah analisis terhadap lebih dari satu variabel. Tujuan dari analisis ini adalah untuk memperoleh informasi tambahan sehingga nantinya didapatkan kesimpulan yang lebih baik. Hampir dari setiap obyek di alam ini dapat diukur dengan lebih dari satu variabel (amatan). (Sharma, 1996)

#### 2.2 Analisis Kelompok

Analisis kelompok merupakan metode yang ada di dalam analisis multivariat yang bertujuan untuk mengelompokkan obyek-obyek pengamatan menjadi beberapa kelompok berdasarkan pengukuran variabel-variabel yang diamati, sedemikian sehingga obyek yang terdapat dalam kelompok yang sama mempunyai karakteristik yang relatif homogen daripada obyek yang berada pada kelompok yang berbeda. (Sharma, 1996)

##### 2.2.1 Ukuran Kemiripan dan Ketakmiripan

Ukuran kemiripan dan ketakmiripan pada dasarnya merupakan kebalikan dalam arti berlawanan arah. Ukuran kemiripan sering disebut koefisien kemiripan yang biasanya bernilai antara 0 dan 1. Misalkan suatu fungsi  $S$  adalah ukuran kemiripan, didefinisikan pada masing-masing obyek pengamatan. Sehingga diperoleh ukuran ketakmiripan yaitu :

$$d = (\text{konstanta} - S)$$

dengan  $d$  = ukuran ketakmiripan

$S$  = ukuran kemiripan

(Chatfield and Collins, 1980)

### 2.2.2 Jarak Chi-square

Jarak *Chi-Square* digunakan untuk mengukur jarak antara obyek pengamatan dengan pusat kelompok yaitu modus. Jarak *Chi-Square* selanjutnya disebut ukuran ketakmiripan. Jika  $X, Y$  adalah dua obyek dengan  $m$  variabel kategorik, ukuran ketakmiripan antara  $X, Y$  didefinisikan oleh total ketidakcocokan yang berkaitan dengan variabel kategorik untuk kedua obyek tersebut.

Secara umum :

$$d(X, Y) = \sum_{j=1}^m \delta(x_j, y_j) \quad (2.1)$$

dimana

$$\delta(x_j, y_j) = \begin{cases} 0, & (x_j = y_j) \\ 1, & (x_j \neq y_j) \end{cases} \quad (2.2)$$

Semakin kecil nilainya, maka kedua obyek semakin mirip. Jika frekuensi dari kategori-kategori dalam data dipertimbangkan, dapat didefinisikan ukuran ketakmiripan sebagai berikut :

$$d_{\chi^2}(X, Y) = \sum_{j=1}^m \frac{(n_{x_j} + n_{y_j})}{n_{x_j} n_{y_j}} \delta(x_j, y_j) \quad (2.3)$$



dengan  $n_{x_j}, n_{y_j}$  masing-masing adalah banyaknya obyek yang mempunyai kategori  $x_j$  dan  $y_j$  untuk atribut yang ke- $j$ .

(Huang, 1997)

### 2.2.3 Metode Hirarki

Metode pengelompokan hirarki digunakan untuk mengelompokkan pengamatan secara terstruktur berdasarkan sifat kemiripannya, dan banyaknya kelompok yang menyusun data belum diketahui. Ada dua cara untuk mendapatkan kelompok dengan metode ini, yaitu agglomeratif dan devisif. Metode pengelompokan hirarki agglomeratif berusaha mengelompokkan data dengan menggabungkan pengamatan atau kelompok secara bertahap sehingga pada akhirnya hanya didapat satu kelompok saja. Sebaliknya, metode devisif dimulai dengan membentuk satu kelompok besar beranggotakan seluruh pengamatan. Kelompok tersebut kemudian dipisah menjadi kelompok yang lebih kecil, sampai didapat  $n$  kelompok yang hanya beranggotakan satu pengamatan saja. (Sharma, 1996)

Pengelompokan dengan algoritma hirarki *agglomerative* mempunyai beberapa macam yang dibedakan menurut jarak antar kelompok yaitu sebagai berikut :

1. Keterkaitan tunggal ( *single linkage, nearest neighbors method* ).

Ukuran jarak yang digunakan adalah :

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

$U$  dan  $V$  obyek yang digabung menjadi kelompok ( $UV$ )

$d_{(U,V)W}$  adalah jarak antara kelompok ( $UV$ ) dan kelompok  $W$

Nilai dari  $d_{UW}$  dan  $d_{VW}$  adalah jarak terkecil antara kelompok  $U$  dan  $W$  dan kelompok  $V$  dan  $W$ .

2. Keterkaitan lengkap ( *complete linkage, farthest neighbor method* )

Ukuran jarak yang digunakan adalah :

$$d_{(U,V)W} = \max \{d_{UW}, d_{VW}\}$$

$U$  dan  $V$  obyek yang digabung menjadi kelompok ( $UV$ )

$d_{(U,V)W}$  adalah jarak antara kelompok ( $UV$ ) dan kelompok  $W$

Nilai dari  $d_{UW}$  dan  $d_{VW}$  adalah jarak terjauh antara kelompok  $U$  dan  $W$  dan kelompok  $V$  dan  $W$ .

3. Keterkaitan rata-rata ( *average linkage method* ).

Ukuran jarak yang digunakan adalah :

$$d_{(U,V)W} = \frac{\sum_i \sum_k d_{ik}}{N_{(UV)}N_w}$$

$U$  dan  $V$  obyek yang digabung menjadi kelompok ( $UV$ )

$d_{(U,V)W}$  adalah jarak antara kelompok ( $UV$ ) dan kelompok  $W$

$d_{ik}$  adalah jarak antara obyek  $i$  dalam kelompok ( $UV$ ) dan obyek  $k$  dalam kelompok  $W$ .

$N_{(UV)}$  dan  $N_w$  adalah banyaknya anggota kelompok dalam kelompok ( $UV$ ) dan dalam kelompok  $W$ .

4. Metode centroid ( *Centroid method* )

Pengelompokan dengan metode centroid pengukuran jarak dilakukan dengan menghitung kembali nilai rata – rata obyek pengamatan.

(Johnson, 1982)

Untuk penentuan jumlah kelompok dalam penelitian ini, digunakan metode *single linkage*. Hasil dari metode ini dapat disajikan dalam bentuk diagram pohon atau dendogram. Dari dendogram ini dapat dilakukan pengelompokan obyek-obyek pengamatan.

#### 2.2.4 Metode Non Hirarki

Metode non hirarki adalah metode pengelompokan yang didasarkan pada jumlah kelompok yang telah ditentukan. Dengan menempatkan anggota kelompok yang relatif homogen dalam kelompok yang sama, salah satu metode pada analisis kelompok non hirarki yang sering digunakan adalah algoritma K-Mean.

(Sharma, 1996)

#### 2.2.5 Algoritma K-Mean

Algoritma K-Mean merupakan salah satu metode pengelompokan non hirarki yang digunakan untuk mengelompokkan obyek-obyek pengamatan ke dalam  $k$  kelompok dengan jumlah kelompok ditentukan terlebih dahulu. Prinsip dari algoritma K-Mean adalah sebagai berikut:

1. Memilih sejumlah  $k$  mean awal untuk  $k$  kelompok.
2. Menghitung ketakmiripan (jarak) antar suatu obyek dengan mean dari suatu kelompok, dan mengalokasikan masing-masing obyek ke kelompok yang mempunyai mean terdekat. Setelah semua obyek teralokasikan, mean dari masing-masing kelompok yang terbentuk dihitung kembali.

3. Mengalokasikan kembali setiap obyek pengamatan ke kelompok lain, jika terdapat obyek yang mempunyai jarak yang lebih dekat ke mean pada kelompok lain.
- 4 Langkah 2. dan 3. diulang sampai tidak ada obyek berpindah kelompok. **(Sharma, 1996)**

### 2.2.6 Algoritma K-Modus

Dalam algoritma K-modus obyek-obyek pengamatan dibagi menjadi k kelompok. Prinsip dari algoritma k-modus adalah sebagai berikut :

1. Memilih sejumlah k modus awal untuk k kelompok.
2. Menghitung ketakmiripan (jarak) antar suatu obyek dengan modus dari suatu kelompok, dan mengalokasikan masing-masing obyek ke kelompok yang mempunyai modus terdekat.
3. Setelah semua obyek teralokasikan, modus dari masing-masing kelompok yang terbentuk dihitung kembali. Mengalokasikan kembali setiap obyek pengamatan ke kelompok lain, jika terdapat obyek yang mempunyai jarak yang lebih dekat dengan ke modus pada kelompok lain dibandingkan dengan kelompoknya sendiri.
4. Langkah 3. diulang sampai tidak ada obyek berpindah kelompok.

**(Huang, 1997)**

## 2.3 Kategori

Kategori bisa diartikan sebagai bagian atau jenis, bisa juga diartikan sebagai cara untuk mengungkapkan makna dengan berbagai potensi yang ada.

Kategori biasa digunakan dalam ilmu sosial atau biomedis. Sering juga ditemui dalam ilmu perilaku, ekologi, pendidikan dan marketing.

(Sulchan Yasyin, 1997)

### 2.3.1 Data Kategorik

Data kategorik adalah data yang variabelnya berjenis kategori. Dalam Agresti (1990) dijelaskan bahwa variabel kategori adalah variabel yang skala pengukurannya terdiri dari sekumpulan kategori. Variable ini biasa digunakan dalam ilmu sosial atau biomedis. Sering juga ditemui dalam ilmu perilaku, ekologi, pendidikan dan marketing.

## 2.4 Frekuensi Kategori dan Frekuensi Relatif Kategori

Misal  $n_{c_{k,j}}$  adalah jumlah obyek yang berkategori  $c_{k,j}$  dalam variabel  $A_j$ ,

$$\text{maka } f_r = (A_j = c_{k,j} | X) = \frac{n_{c_{k,j}}}{n} \quad (2.4)$$

merupakan frekuensi relatif dari kategori  $c_{k,j}$  dalam X.

Dimana :  $f_r$  = Frekuensi relatif dari kategori  $c_{k,j}$

$n_{c_{k,j}}$  = Jumlah obyek yang berkategori  $c_{k,j}$  dalam variabel  $A_j$

= Frekuensi kategori

$n$  = Jumlah obyek pengamatan

## 2.5 Modus Suatu Himpunan

Misal  $X$  adalah suatu himpunan  $n$  obyek yang memiliki variabel kategori

$A_1, A_2, \dots, A_m$

### Definisi 2.1 Modus

Modus dari  $X$  adalah vektor  $Q = [q_1, q_2, \dots, q_m] \in \Omega$  sedemikian hingga

$$d(Q, X) = \sum_{i=1}^m d(X_i, Q) \text{ minimal.}$$

Dimana  $X = \{X_1, X_2, \dots, X_n\}$  dan  $d$  dapat ditentukan dari persamaan (2.1)

atau (2.3)

(Huang, 1997)

### Teorema 2.1

Fungsi  $d(Q, X)$  adalah minimum jika  $f_r(A_j = q_j | X) \geq f_r(A_j = c_{k,j} | X)$

untuk  $q_j \neq c_{k,j}$  dan  $j = 1, \dots, m$

(Huang, 1997)

### 2.6 S-Plus

S-plus adalah suatu paket program yang memungkinkan membuat program sendiri walaupun di dalamnya sudah tersedia banyak program internal yang siap digunakan. Kelebihan dari paket program ini adalah baik program internal maupun program yang pernah dibuat dapat digunakan sebagai sub program dari program yang akan dibuat.

(Anonim, 1993a & 1993b)



### BAB III

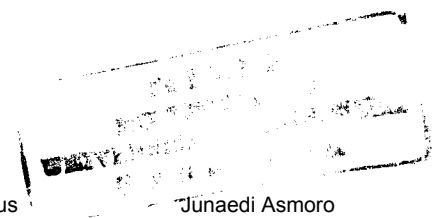
#### METODE PENULISAN

Dalam penulisan ini akan digunakan metode penulisan sebagai berikut :

1. Melakukan penelusuran pustaka yang berkaitan dengan analisis kelompok untuk data kategorik dan algoritma pendukungnya, yaitu K-Mean, K-Median dan K-Modus.
2. Menyusun algoritma untuk mengelompokkan data dengan menggunakan K-Modus. Algoritma ini dimodifikasi dari algoritma K-Mean. Garis besar algoritma K-Modus adalah sebagai berikut :
  - a. Memilih sejumlah  $k$  modus awal untuk  $k$  kelompok. Dalam langkah ini digunakan salah satu metode dari dua metode di bawah ini, yaitu :
    - i. Metode I ( $k$  obyek yang pertama dari suatu set data).
    - ii. Metode II (penugasan kategori) dengan langkah sebagai berikut :
      - Menghitung frekuensi semua kategori untuk seluruh variabel yang ada dan disimpan dalam suatu *array* kategori dalam urutan frekuensi yang menurun.

$$\begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,j} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,j} \\ c_{3,1} & c_{3,2} & \cdots & c_{3,j} \\ \vdots & \vdots & & \vdots \\ c_{i,1} & c_{i,2} & \cdots & c_{i,j} \end{bmatrix}$$

$c_{i,j}$  menyatakan kategori ke- $i$  dari variabel ke- $j$ , dan frekuensi ( $c_{i,j}$ )  $\geq$  frekuensi ( $c_{i+1,j}$ ).



- Menugaskan kategori dengan frekuensi tertinggi berturut-turut (pola diagonal) sebagai K-Modus awal.
  - Memulai dengan  $Q_1$ . Memilih obyek yang paling mirip dengan  $Q_1$ , dan  $Q_1$  disubstitusi dengan obyek tersebut sebagai modus awal yang pertama. Selanjutnya memilih obyek yang paling mirip dengan  $Q_2$ , kemudian  $Q_2$  disubstitusi dengan obyek tersebut sebagai modus awal yang kedua. Langkah ini dilanjutkan sampai  $Q_k$  tersubstitusi.
- b. Menghitung ketakmiripan (jarak) antar obyek dengan modus dari suatu kelompok, dan mengalokasikan masing-masing obyek ke kelompok yang mempunyai modus terdekat. Kemudian menentukan kembali modus dari masing-masing kelompok yang sudah terbentuk, dan menghitung jaraknya dengan semua obyek yang ada.
  - c. Mengalokasikan kembali setiap obyek pengamatan ke kelompok lain, jika terdapat obyek yang mempunyai jarak yang lebih dekat ke modus pada kelompok lain dibandingkan dengan kelompok sendiri.
  - d. Langkah b. dan c. diulang sampai tidak ada obyek yang berpindah kelompok.
3. Menerapkan algoritma yang telah dibuat ke dalam program komputer yang menggunakan software **S-Plus**.
  4. Menerapkan program yang telah dibuat dengan menggunakan data sekunder.

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Jumlah Kelompok ( $k$ )

Untuk menentukan jumlah kelompok, digunakan salah satu metode yang ada dalam metode hirarki, pada pembahasan ini jumlah kelompok ditentukan berdasarkan dendogram yang diperoleh dari proses pengelompokan dengan metode *single linkage*.

#### 4.2 Frekuensi Kategori Dan Modus Yang Baru

Proses perhitungan frekuensi kategori dalam suatu variabel sangat penting. Proses ini dibutuhkan untuk penentuan modus yang baru dari masing-masing kelompok. Untuk menentukan frekuensi kategori dicari dengan menghitung jumlah suatu kategori dalam variabel tertentu. Kemudian menentukan frekuensi relatif dengan menggunakan persamaan (2.4). Setelah mendapatkan frekuensi relatif dari tiap-tiap kategori, lalu menentukan modus yang baru sesuai dengan sub bab (2.2) dan teorema 2.1. Kategori dalam suatu vektor obyek pengamatan dapat dijadikan modus baru, jika frekuensi relatifnya paling besar daripada frekuensi relatif kategori yang lain dalam suatu variabel.

#### 4.3 Pusat Kelompok

Beberapa cara dalam menentukan nilai awal pusat kelompok :

1. Metode I ( $k$  obyek pertama)

Memilih  $k$  obyek yang pertama sebagai pusat kelompok awal, dengan  $k$  adalah jumlah kelompok yang sudah ditentukan. Misalnya dari seluruh data mulai dari satu sampai  $n$  obyek, pusat kelompok 1 yaitu obyek pertama. Pusat kelompok 2 yaitu obyek yang kedua, begitu seterusnya sampai diperoleh  $k$  pusat kelompok.

## 2. Metode II (penugasan kategori).

Memilih pusat kelompok awal dari penugasan kategori dengan langkah sebagai berikut :

- Menghitung frekuensi kategori dari seluruh variabel yang ada kemudian disusun menjadi matrik kategori dalam urutan frekuensi yang menurun.

Tabel 4.1 frekuensi kategori dari seluruh variabel

Kategori \ variabel	1	2	...	j
1	$C_{1,1}$	$C_{1,2}$	...	$C_{1,j}$
2	$C_{2,1}$	$C_{2,2}$	...	$C_{2,j}$
3	$C_{3,1}$	$C_{3,2}$	...	$C_{3,j}$
⋮	⋮	⋮	⋮	⋮
i	$C_{i,1}$	$C_{i,2}$	...	$C_{i,j}$

Dari tabel 4.1 dapat dibentuk matrik kategori sebagai berikut :

$$\begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,j} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,j} \\ c_{3,1} & c_{3,2} & \cdots & c_{3,j} \\ \vdots & \vdots & & \vdots \\ c_{i,1} & c_{i,2} & \cdots & c_{i,j} \end{bmatrix}$$

$c_{i,j}$  menyatakan kategori ke- $i$  dari variabel ke- $j$ , dan frekuensi  $(c_{i,j}) \geq$  frekuensi  $(c_{i+1,j})$ .

- Menugaskan kategori dengan frekuensi tertinggi berturut-turut sebagai K-Modus awal. Untuk penugasan kategori dilakukan dengan cara pengambilan kategori dari matrik kategori secara diagonal. Dimulai dari  $Q_1$ , kategori  $Q_1$  adalah  $[c_{1,1}, c_{2,2}, c_{3,3}, \dots, c_{i,j}]$ . Kemudian kategori  $Q_2$  adalah  $[c_{2,1}, c_{3,2}, c_{4,3}, \dots, c_{i,j}]$ . Masing-masing variabel mempunyai kategori yang tidak sama banyaknya. Jika variabel yang kategorinya terpilih untuk dijadikan modus awal tidak ada, maka untuk kategori yang terpilih digantikan dengan kategori yang pertama dari variabel tersebut. Begitu seterusnya sampai mendapatkan  $Q_k$  yang diinginkan. Untuk lebih jelasnya akan diberikan suatu contoh sebagai berikut : Misal diberikan suatu data kategorik yang mempunyai 4 variabel yang masing-masing mempunyai 4, 2, 5 dan 3 kategori.

$$\begin{bmatrix} c_{1,1} & c_{1,2} & c_{1,3} & c_{1,4} \\ c_{2,1} & c_{2,2} & c_{2,3} & c_{2,4} \\ c_{3,1} & & c_{3,3} & c_{3,4} \\ c_{4,1} & & c_{4,3} & \\ & & c_{5,3} & \end{bmatrix}$$

Jika ingin dibentuk 3 kelompok, maka modus awalnya adalah sebagai berikut :

$$\begin{aligned} Q_1 &= [c_{1,1}, c_{2,2}, c_{3,3}, c_{1,4}], \\ Q_2 &= [c_{2,1}, c_{1,2}, c_{4,3}, c_{2,4}], \\ Q_3 &= [c_{3,1}, c_{2,2}, c_{5,3}, c_{3,4}]. \end{aligned}$$

- Memilih obyek yang mirip dari kategori yang sudah ditugaskan. Misal  $Q_1$ , dipilih obyek yang paling mirip dengan  $Q_1$ , dan  $Q_1$  disubstitusi dengan obyek tersebut sebagai modus awal yang pertama. Selanjutnya dipilih

obyek yang paling mirip dengan  $Q_2$ , kemudian  $Q_2$  disubstitusi dengan obyek tersebut sebagai modus awal yang kedua, begitu seterusnya sampai  $Q_k$  tersubstitusi.

### 3. Metode III (Random).

Untuk penentuan pusat kelompok awal dalam penelitian ini digunakan hanya dua metode, yaitu metode I dan metode II.

## 4.4 Ukuran Ketakmiripan

Untuk ukuran ketakmiripan dalam penelitian ini digunakan jarak Chi-square (sub bab 2.1.2). Untuk selanjutnya ukuran ketakmiripan dalam pembahasan disebut jarak.

## 4.5 Algoritma K-Modus

Algoritma K-Modus merupakan algoritma pengelompokan *Non-hirarki* yang mendasarkan gagasan pengelompokannya pada ukuran ketakmiripan (*dissimilarity*). Anggota-anggota setiap kelompok ditentukan dengan menempatkan obyek-obyek pengamatan ke dalam sebuah kelompok yang paling sesuai, dalam hal ini jarak yang paling kecil atau minimal.

Algoritma K-Modus secara terperinci adalah sebagai berikut :

1. Menginput data set dan nilai  $k$ . Penentuan jumlah kelompok ( $k$ ) sesuai dengan sub bab 4.1.
2. Memilih sejumlah  $k$  modus awal untuk  $k$  kelompok sesuai dengan sub bab (4.3).



3. Menghitung jarak antara obyek dengan modus kelompok seperti yang didefinisikan pada persamaan (2.1) dan (2.3).
4. Menempatkan obyek pengamatan ke kelompok terdekat, sebagai anggota kelompok.
5. Menentukan modus yang baru dari masing-masing kelompok sesuai dengan sub bab (4.2).
6. Pengelompokan selesai jika modus tidak berubah atau tidak ada obyek - obyek pengamatan yang pindah kelompok. Jika ada obyek pengamatan yang berpindah kelompok, maka kembali ke langkah 3.

#### 4.6 DATA

Program yang telah dibuat akan diuji dengan menggunakan data sekunder dan pada pembahasan ini akan dipilih data sebagai berikut :

- Data I adalah data yang diambil dari **Johnson (1975)** tentang karakteristik suatu individu. Data ini terdiri dari 5 individu yang mempunyai karakteristik yang berbeda dengan memperhatikan ukuran tinggi badan, berat badan, warna mata, warna rambut, tangan yang aktif, dan jenis kelamin ( lampiran 2 ).
- Data II adalah data yang diambil dari data **Jobson (1992)**. Data ini terdiri dari 20 observasi terhadap jenis mobil yang berbeda dengan memperhatikan banyaknya bahan bakar yang digunakan, ukuran silinder, berat mobil, ukuran mesin serta negara pemproduksi (lampiran 3).

#### 4.7 Penerapan algoritma ke program komputer

Program komputer dibuat dengan menggunakan paket program S-Plus. Algoritma K-Modus yang telah tersusun dijabarkan dalam 6 program yang saling terkait. Program 2a sampai dengan 2g merupakan sub program dari program utama (lampiran 1).

- Program utama untuk mengelompokkan obyek-obyek pengamatan algoritma K-Modus berdasarkan metode I dan metode II dengan terlebih dahulu melalui proses program 2a sampai 2g dengan
- Program 2a untuk menghitung frekuensi kategori kemudian disusun menjadi suatu matrik kategori dalam urutan frekuensi yang menurun.
- Program 2b untuk menentukan pusat awal kelompok menggunakan metode I yaitu memilih  $k$  obyek yang pertama dari suatu set data .
- Program 2c untuk menentukan pusat awal kelompok menggunakan metode II yaitu menugaskan kategori dengan frekuensi tertinggi berturut-turut.
- Program 2d untuk memperoleh jarak antar obyek-obyek pengamatan dengan pusat kelompoknya.
- Program 2e untuk menghitung banyaknya kelompok
- Program 2f untuk menempatkan obyek pengamatan pada kelompok yang sesuai.
- Program 2g untuk menentukan pusat kelompok yang baru dari masing-masing kelompok.
- Program 2h untuk memperoleh matrik jarak antar obyek-obyek pengamatan.

#### 4.8 Pengelompokan pada data secara manual

Data pengelompokan diambil dari **Johnson (1975)** tentang karakteristik individu. Untuk mempermudah proses pengelompokan pada data kategorik, maka kategori-kategori dalam variabel pada data tersebut didefinisikan terlebih dahulu.

Pendefinisian variabel dalam bentuk *binary* sebagai berikut :

$$V_1 = \begin{cases} 1 & \text{height} \geq 72 \text{ in} \\ 0 & \text{height} < 72 \text{ in} \end{cases} \quad V_4 = \begin{cases} 1 & \text{blond hair} \\ 0 & \text{brown hair} \end{cases}$$

$$V_2 = \begin{cases} 1 & \text{weight} \geq 150 \text{ lb} \\ 0 & \text{weight} < 150 \text{ lb} \end{cases} \quad V_5 = \begin{cases} 1 & \text{right handed} \\ 0 & \text{left handed} \end{cases}$$

$$V_3 = \begin{cases} 1 & \text{brown eyes} \\ 0 & \text{yang lain} \end{cases} \quad V_6 = \begin{cases} 1 & \text{female} \\ 0 & \text{male} \end{cases}$$

Berdasarkan pendefinisian variabel dan data pada lampiran 2, didapatkan hasil data pada tabel 4.2 sebagai berikut :

Tabel 4.2 penyajian data I setelah didefinisikan

obyek	Variabel					
	V <sub>1</sub>	V <sub>2</sub>	V <sub>3</sub>	V <sub>4</sub>	V <sub>5</sub>	V <sub>6</sub>
individu 1	0	0	0	1	1	1
individu 2	1	1	1	0	1	0
individu 3	0	1	0	1	1	0
individu 4	0	0	1	0	1	1
individu 5	1	1	1	0	0	0

Langkah-langkah algoritma K-Modus secara manual dilakukan dengan menggunakan dua metode, berdasarkan dari penentuan pusat awal kelompoknya.

##### 4.8.1 Metode I

1. Pilih sejumlah K-Modus awal untuk  $k$  kelompok.

Pilih  $k$  obyek yang pertama dari suatu set data. Misal  $k = 2$

sehingga  $Q_1$  adalah obyek ke 1 yaitu  $[0 \ 0 \ 0 \ 1 \ 1 \ 1]$

$Q_2$  adalah obyek ke 2 yaitu  $[1 \ 1 \ 1 \ 0 \ 1 \ 0]$

dengan  $Q_1$  dan  $Q_2$  masing-masing adalah modulus 1 dan modulus 2.

- Menghitung jarak antar suatu obyek dengan modulus dari suatu kelompok. Dalam hal ini dapat dipilih salah satu jenis jarak pada persamaan (2.1) atau (2.3).

$$\text{Modus awal : } \begin{aligned} Q_1 &= [0 \ 0 \ 0 \ 1 \ 1 \ 1] \\ Q_2 &= [1 \ 1 \ 1 \ 0 \ 1 \ 0] \end{aligned}$$

$$* \quad d_{x^2}(S_1, Q_1) = \left\{ \left( \frac{3+3}{9} \right) \times 0 + \left( \frac{2+2}{4} \right) \times 0 + \left( \frac{2+2}{4} \right) \times 0 + \left( \frac{2+2}{4} \right) \times 0 + \left( \frac{4+4}{16} \right) \times 0 + \left( \frac{2+2}{4} \right) \times 0 \right\} = 0$$

$$d_{x^2}(S_1, Q_2) = \left\{ \left( \frac{2+3}{6} \right) \times 1 + \left( \frac{3+2}{6} \right) \times 1 + \left( \frac{3+2}{6} \right) \times 1 + \left( \frac{3+2}{6} \right) \times 1 + \left( \frac{4+4}{16} \right) \times 0 + \left( \frac{3+2}{6} \right) \times 1 \right\} = 4,17$$

Dengan cara yang sama, diperoleh :

$$* \quad d_{x^2}(S_2, Q_1) = 4,17$$

$$d_{x^2}(S_2, Q_2) = 0$$

$$* \quad d_{x^2}(S_3, Q_1) = 1,67$$

$$d_{x^2}(S_3, Q_2) = 2,5$$

$$* \quad d_{x^2}(S_4, Q_1) = 1,67$$

$$d_{x^2}(S_4, Q_2) = 2,5$$

$$* \quad d_{x^2}(S_5, Q_1) = 5,42$$

$$d_{x^2}(S_5, Q_2) = 1,25$$

dengan  $S_i$  : obyek pengamatan ke- $i$ ,

$$i = 1, 2, \dots, n$$

Menempatkan masing-masing obyek pengamatan ke kelompok yang mempunyai jarak terdekat . Obyek 1 menjadi anggota kelompok 1, obyek 2 menjadi anggota kelompok 2, obyek 3 menjadi anggota kelompok 1, obyek 4 menjadi anggota kelompok 1 dan obyek 5 menjadi anggota kelompok 2. Untuk lebih jelas lihat pada tabel 4.3.

Tabel 4.3 Hasil pengelompokan iterasi pertama pada data I dengan metode I

kelompok	Obyek yang bergabung	V1	V2	V3	V4	V5	V6
1	1	0	0	0	1	1	1
	3	0	1	0	1	1	0
	4	0	0	1	0	1	1
2	2	1	1	1	0	1	0
	5	1	1	1	0	0	0

Kemudian menentukan modus yang baru dari masing-masing kelompok, yaitu dengan cara menghitung frekuensi kategori yang sama dari seluruh variabel yang ada sesuai dengan sub bab 4.2. Frekuensi kategori dari masing-masing kelompok dapat dilihat pada tabel 4.4.

Tabel 4.4 frekuensi relatif kategori untuk iterasi pertama pada data I

kelompok	variabel	Frekuensi relatif kategori	
		0	1
I	V1	1	0
	V2	0,67	0,33
	V3	0,67	0,33
	V4	0,33	0,67
	V5	0	1
	V6	0,33	0,67
II	V1	0	1
	V2	0	1
	V3	0	1
	V4	1	0
	V5	0,5	0,5
	V6	1	0

Modus baru dari tiap kelompok, dibentuk dengan mengambil kategori dari tiap variabel yang mempunyai frekuensi relatif terbesar. Dari tabel 4.4 didapatkan modus yang baru untuk masing-masing kelompok yaitu : kelompok 1  $Q_1 = [0 \ 0 \ 0 \ 1 \ 1 \ 1]$ ,

$$\text{kelompok 2 } \begin{aligned} Q_{2,1} &= [1 \ 1 \ 1 \ 0 \ 0 \ 0] \\ Q_{2,2} &= [1 \ 1 \ 1 \ 0 \ 1 \ 0] \end{aligned}$$

Karena pada kelompok 2, peluang untuk menjadi modus ada 2, dapat kita pilih salah satu. Misal  $Q_2 = [1 \ 1 \ 1 \ 0 \ 0 \ 0]$ .

Setelah mendapatkan modus yang baru, proses dilanjutkan dengan mengulang langkah 2 dan seterusnya. Iterasi akan selesai jika sudah tidak ada obyek pengamatan yang berpindah kelompok. Dengan cara tersebut, didapatkan hasil pengelompokan akhir pada tabel 4.5.

Tabel 4.5 Hasil pengelompokan akhir data I secara manual menggunakan algoritma K-Modus dengan metode I

Kelompok	Obyek yang bergabung
1	1 3 4
2	2 5

#### 4.8.2 Metode II

1. Pilih sejumlah K-Modus awal untuk  $k$  kelompok, dengan langkah sebagai berikut :
  - i. Hitung frekuensi semua kategori untuk seluruh variabel yang ada, kemudian disimpan dalam suatu matrik kategori dalam urutan



frekuensi menurun. Hasil frekuensi kategori dapat dilihat pada tabel 4.6

Tabel 4.6 frekuensi kategori data I

Variabel	Frekuensi kategori	
	0	1
V1	3	2
V2	2	3
V3	2	3
V4	3	2
V5	1	4
V6	3	2

Dari tabel 4.6 didapatkan matrik kategori :

$$\begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

ii. Menugaskan kategori dengan frekuensi tertinggi berturut-turut sebagai K-Modus awal. Misal  $k$  ditentukan sebesar 2, maka

didapat:  $Q_1 = [0 \ 0 \ 1 \ 1 \ 1 \ 1]$

$$Q_2 = [1 \ 1 \ 0 \ 0 \ 0 \ 0]$$

iii. Pilih obyek yang paling mirip dengan  $Q_1$  dan  $Q_2$ . Dengan cara yang sama pada perhitungan manual data I, diperoleh:

$$\begin{aligned}
 * \quad d_{\chi^2}(S_1, Q_1) &= 0,83 \\
 d_{\chi^2}(S_1, Q_2) &= 4,58 \\
 * \quad d_{\chi^2}(S_2, Q_1) &= 3,33 \\
 d_{\chi^2}(S_2, Q_2) &= 2,08 \\
 * \quad d_{\chi^2}(S_3, Q_1) &= 2,5 \\
 d_{\chi^2}(S_3, Q_2) &= 2,92 \\
 * \quad d_{\chi^2}(S_4, Q_1) &= 0,83 \\
 d_{\chi^2}(S_4, Q_2) &= 4,58 \\
 * \quad d_{\chi^2}(S_5, Q_1) &= 4,58 \\
 d_{\chi^2}(S_5, Q_2) &= 0,83
 \end{aligned}$$

Setelah disubstitusi, yang paling mirip dengan  $Q_1$  adalah S4 dan yang paling mirip dengan  $Q_2$  adalah S5. Sehingga S4 merupakan pusat kelompok awal dari kelompok 1 ( $k_1$ ) dan S5 merupakan pusat kelompok awal dari kelompok 2 ( $k_2$ ).

$$\begin{aligned}
 Q_1 &= [0 \ 0 \ 1 \ 0 \ 1 \ 1] \\
 Q_2 &= [1 \ 1 \ 1 \ 0 \ 0 \ 0]
 \end{aligned}$$

2. Menghitung jarak antar suatu obyek dengan modus dari suatu kelompok, dan mengalokasikan masing-masing obyek ke kelompok yang mempunyai modus terdekat.

$$\begin{aligned}
 * \quad d_{\chi^2}(S_1, Q_1) &= 1,67 & * \quad d_{\chi^2}(S_4, Q_1) &= 0 \\
 d_{\chi^2}(S_1, Q_2) &= 5,42 & d_{\chi^2}(S_4, Q_2) &= 3,75 \\
 * \quad d_{\chi^2}(S_2, Q_1) &= 2,5 & * \quad d_{\chi^2}(S_5, Q_1) &= 3,75 \\
 d_{\chi^2}(S_2, Q_2) &= 1,25 & d_{\chi^2}(S_5, Q_2) &= 0 \\
 * \quad d_{\chi^2}(S_3, Q_1) &= 3,33 \\
 d_{\chi^2}(S_3, Q_2) &= 3,75
 \end{aligned}$$

Mengalokasikan masing-masing obyek ke kelompok yang mempunyai modus terdekat. Hasil dari pengelompokan dapat dilihat pada tabel 4.7.

Tabel 4.7 hasil pengelompokan iterasi pertama pada data I dengan metode II

kelompok	Obyek yang bergabung	V1	V2	V3	V4	V5	V6
1	1	0	0	0	1	1	1
	3	0	1	0	1	1	0
	4	0	0	1	0	1	1
2	2	1	1	1	0	1	0
	5	1	1	1	0	0	0

Kemudian menghitung frekuensi kategori dari masing-masing kelompok sesuai dengan tabel 4.4 untuk menentukan modus yang baru. Modus yang baru untuk masing-masing kelompok adalah  $Q_1 = [0 \ 0 \ 0 \ 1 \ 1 \ 1]$ ,  $Q_2 = [1 \ 1 \ 1 \ 0 \ 0 \ 0]$

Dengan cara yang sama pada pengelompokan data menggunakan metode I untuk menghitung jarak dan mengalokasikan obyek-obyek pengamatan didapatkan hasil pengelompokan akhir pada tabel 4.8. Dikatakan sebagai kelompok akhir, dikarenakan sudah tidak ada lagi obyek yang berpindah kelompok.

Tabel 4.8 Hasil pengelompokan akhir data I secara manual menggunakan algoritma K-Modus dengan metode II

Kelompok	Obyek yang bergabung
1	1 3 4
2	2 5

Hasil pengelompokan pada data I menggunakan algoritma K-Modus metode I mempunyai persamaan pada hasil pengelompokan dengan algoritma K-Modus metode II.

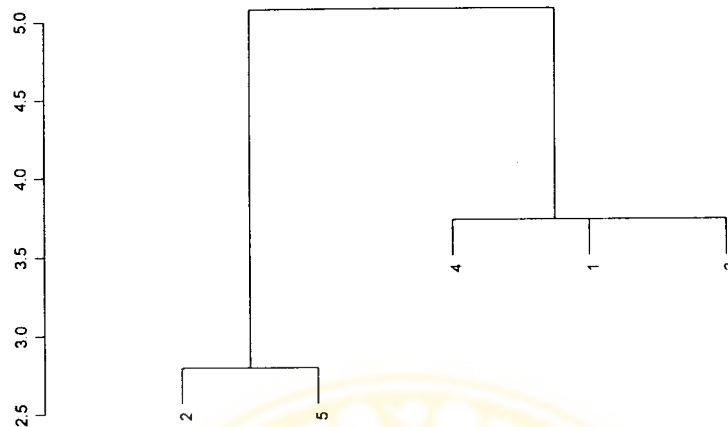
#### 4.9 Pengelompokan pada data dengan program komputer

Untuk mengetahui hasil pengelompokan algoritma K-Modus menggunakan program komputer, maka dilakukan pengelompokan pada Data I dan data II.

##### 4.9.1 Data I

Sebagai gambaran awal untuk mengetahui bentuk penggabungan obyek pengamatan dapat dilihat pada gambar 4.1 yaitu dendogram yang diperoleh dengan metode *single linkage* (keterkaitan tunggal) dengan terlebih dahulu menghitung jarak antar obyek-obyek pengamatan. Dari dendogram tersebut diketahui jumlah kelompok yang baik adalah dua kelompok.

Selanjutnya dilakukan proses pengelompokan algoritma K-Modus dengan menggunakan program komputer (S-Plus). Pusat awal dari masing-masing kelompok dengan menggunakan metode I dan II.



Gambar 4.1 Dendrogram dari data I dengan metode keterkaitan tunggal (*single linkage*).

### 1. Metode I

Untuk data I proses pengelompokan dimulai dengan menginput jumlah kelompok yang diinginkan yaitu 2 kelompok. Kemudian menentukan modulus awal yang ingin digunakan yaitu metode I (program 2b). Setelah didapat 2 modulus awal yaitu 2 obyek yang pertama dari suatu set data, proses selanjutnya yaitu menghitung jarak (program 2d) antar suatu obyek pengamatan dengan modulus dari suatu kelompok, dan mengalokasikan/menempatkan masing-masing obyek pengamatan ke kelompok yang mempunyai modulus terdekat (program 2f). Pengelompokan data dengan menggunakan metode I selesai.

Hasil pengelompokan diperoleh pada tabel 4.9 dengan menetapkan jumlah kelompok sebanyak 2, maka kelompok I beranggotakan obyek 1, obyek 3 dan obyek 4 sedangkan kelompok II beranggotakan obyek 2 dan obyek 5.

## 2. Metode II

Selanjutnya pengelompokan dengan menggunakan metode II. Algoritma dimulai dengan menghitung frekuensi semua kategori untuk seluruh atribut yang ada dan disimpan dalam suatu matrik kategori dalam urutan frekuensi menurun (program 2a). Kemudian menentukan modus awalnya, proses penentuan modus awal sesuai dengan algoritma K-Modus (program 2c). Karena jumlah kelompok yang diinput sebanyak 2, maka modus awal yang didapat nantinya juga sebanyak 2. Proses pengelompokan dengan metode II dapat dilihat pada tabel 4.9

Tabel 4.9 Hasil pengelompokan data I dengan program komputer menggunakan metode I dan metode II

	Kelompok	Obyek yang bergabung
Metode I	1	1, 3 dan 4
	2	2 dan 5
Metode II	1	1, 3 dan 4
	2	2 dan 5

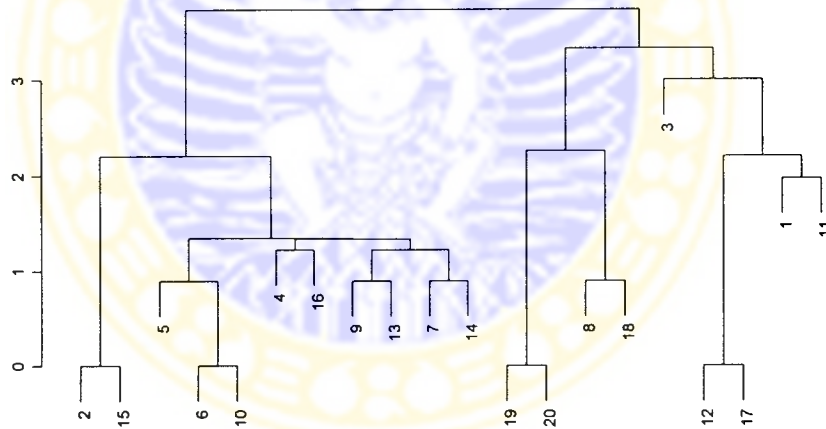
Berdasarkan hasil pengelompokan pada tabel 4.9 diketahui bahwa hasil pengelompokan algoritma K-Modus menggunakan program komputer memberikan hasil yang sama dengan pengelompokan data secara manual. Kelompok pertama memiliki kesamaan dalam hal **Height** (tinggi badan) : kurang dari 72in, **Weight** (berat badan) : kurang dari 150lb, **Eye Colour** (warna mata) : non coklat, **Hair Colour** (warna rambut) : pirang, **Handedness** (tangan aktif) : kanan, **Gender** (jenis kelamin) : pria.



Sedangkan kelompok kedua memiliki kesamaan dalam hal **Height** (tinggi badan) : lebih dari sama dengan 72in, **Weight** (berat badan) : lebih dari sama dengan 150lb, **Eye Colour** (warna mata) : coklat, **Hair Colour** (warna rambut) : non pirang, **Handedness** (tangan aktif) : kiri, **Gender** (jenis kelamin) : wanita.

#### 4.9.2 Data II

Untuk mengetahui hasil pengelompokan pada Data II dengan metode I dan metode II algoritma K-Modus maka dilakukan pengelompokan dengan menggunakan program komputer. Dalam pembahasan ini diambil 20 obyek.



Gambar 4.2 Dendrogram dari data II dengan metode keterkaitan tunggal (*single linkage*).

Berdasarkan dendrogram pada gambar 4.2 didapatkan hasil pengelompokan yang paling baik yaitu 3 kelompok. Hasil dari proses pengelompokan dapat dilihat pada tabel 4.10. Pengelompokan pada data II menggunakan algoritma K-Modus dengan metode I apabila ingin dibentuk

menjadi tiga kelompok didapatkan hasil pengelompokan sebagai berikut : kelompok pertama adalah obyek (1), (8), (11), (12), (17), (19) dan obyek (20) untuk kelompok kedua adalah obyek (2), (4), (7), (14), (15), (16) dan obyek (18) untuk kelompok ketiga adalah obyek (3), (5), (6), (9), (10) dan obyek (13). Sedangkan untuk pengelompokan pada data II dengan menggunakan algoritma K-Modus metode II apabila ingin dibetuk menjadi tiga kelompok maka didapatkan hasil pengelompokan sebagai berikut. Kelompok pertama dengan anggota kelompok obyek (2), (4), dan (15). Kelompok kedua dengan anggota kelompok obyek (3), (8), (12), (17), (18), (19) dan obyek (20). Kelompok ketiga dengan anggota kelompok obyek (1), (5), (6), (7), (9), (10), (11), (13), (14) dan obyek (16).

Tabel 4.10 Hasil pengelompokan data II dengan menggunakan metode I dan metode II pada algoritma K-Modus.

Algoritma pengelompokan	Kelompok	Obyek yang bergabung
Metode I	1	1, 8, 11, 12, 17, 19 dan 20
	2	2, 4, 7, 14, 15, 16 dan 18
	3	3, 5, 6, 9, 10 dan 13
Metode II	1	2, 4 dan 15
	2	3, 8, 12, 17, 18, 19 dan 20
	3	1, 5, 6, 7, 9, 10, 11, 13, 14 dan 16

Adapun ciri-ciri atau sifat dari masing-masing kelompok adalah sebagai berikut :

- Pengelompokan dengan metode I.

Kelompok satu memiliki kesamaan dalam hal **Engsize** (ukuran mesin) : antara 38 sampai 50, **Cylind** (banyaknya silinder) : sebanyak 8, **Combrate** (tingkat bahan bakar yang dibutuhkan) : 104 - 110, **Weight** (berat) : antara 3500 sampai 4000, **For** : dibuat di luar Amerika Selatan.

Kelompok kedua kebanyakan memiliki kesamaan dalam hal **Engsize** (ukuran mesin) : antara 15 sampai 18, **Cylind**(banyaknya silinder) : sebanyak 4, **Combrate** (tingkat bahan bakar yang dibutuhkan) : antara 64 sampai 84, **Weight** (berat) : antara 2000 sampai 2750, **For** : dibuat di Amerika Selatan dan diluar Amerika Selatan. Kelompok ketiga kebanyakan memiliki kesamaan dalam hal **Engsize** (ukuran mesin) : antara 20 sampai 24, **Cylind** (banyaknya silinder) : antara 4 sampai 6, **Combrate** (tingkat bahan bakar yang dibutuhkan) : antara 74 – 110, **Weight** (berat) : antara 2500 sampai 3000, **For** : dibuat di luar Amerika Selatan.

- Pengelompokan dengan metode II.

Kelompok satu memiliki kesamaan dalam hal **Engsize** (ukuran mesin) : 15 - 18, **Cylind** (banyaknya silinder) : sebanyak 4, **Combrate** (tingkat bahan bakar yang dibutuhkan) : 64 - 71, **Weight** (berat) : antara 2000 sampai 2750, **For** : dibuat diamerika bagian selatan. Kelompok kedua kebanyakan memiliki kesamaan dalam hal **Engsize** (ukuran mesin) : 28

sampai 30, **Cylind** (banyaknya silinder) : sebanyak 6-8, **Combrate** (tingkat bahan bakar yang dibutuhkan) : 104 sampai 110, **Weight** (berat) : 3500, **For** : dibuat diluar Amerika Selatan. Kelompok ketiga kesamaan dalam hal **Engsize** (ukuran mesin) : antara 20 sampai 24, **Cylind** (banyaknya silinder) : antara 4 , **Combrate** (tingkat bahan bakar yang dibutuhkan) : 74 – 84, **Weight** (berat) : antara 2500 sampai 3000, **For** : luar Amerika Selatan.



## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

Dari hasil penulisan skripsi ini dapat disimpulkan bahwa :

1. Algoritma K-Modus dapat digunakan untuk mengelompokkan data berjenis kategori berdasarkan pada jarak (ukuran ketakmiripan) antar kelompok atau obyek pengamatan.
2. Perbedaan nilai awal dalam menetapkan pusat kelompok awal dapat memberikan hasil yang berbeda dalam proses pengelompokan.
3. Pada data I baik menggunakan metode I atau metode II diperoleh hasil pengelompokan yang sama, jika jumlah kelompok yang diinginkan sebanyak dua kelompok. Kelompok pertama beranggotakan 3 obyek dan kelompok kedua beranggotakan 2 obyek. Sedangkan pada data II, jika jumlah kelompok yang diinginkan sebanyak tiga kelompok, memberikan hasil pengelompokan yang berbeda.

#### **5.2 SARAN**

Adapun saran yang perlu disampaikan pada penulisan ini adalah :

1. Untuk penelitian berikutnya, penelitian ini bisa menjadi bahan masukan untuk mengembangkan metode lain dalam pengelompokan data berjenis kategori.

2. Pada penentuan jumlah kelompok, dapat menggunakan metode yang lain selain metode *Single Linkage*





## DAFTAR PUSTAKA

- Agresti, A., 1990. *Categorical Data Analysis*. John Wiley & Sons, New York
- Anonim, 1993a. *S-Plus Programmer's Manual*. Mathsoft, Seattle.
- Anonim, 1993b. *S-Plus : Guide To Statistical and Mathematical Analysis*. Mathsoft. Seattle.
- Chatfield, C., A.J. Collins, 1980. *Introduction to Multivariate Analysis*. Chapman and Hall, New York.
- Huang, Z., 1997. *A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining*. <http://citeseer.ist.psu.edu/huang97fast.html>. Tanggal akses : 23/3/04.
- O'muirchertaigh, C.A., C. Payne, 1977. *Exploring Data Structures*. John Wiley & sons, New York.
- Sharma, S., 1996. *Applied Multivariate Techniques*. John Wiley & sons Inc, New York.
- Yasyin, S., 1997. *Kamus Lengkap Bahasa Indonesia*, Jakarta.

## Lampiran 1

### Daftar Program

#### 1. Program Utama

**Program untuk mengelompokan obyek-obyek pengamatan dengan algoritma K-Modus berdasarkan metode I dan metode II.**

```
> program.skripsi2
function(x)
{
  x<-as.matrix(x)
  n<-nrow(x)
  cat("\n")
  cat("          S E L A M A T   D A T A N G\n")
  cat("  DI PROGRAM PENGELOMPOKAN DATA KATEGORIK K-MODUS\n")
  cat("-----\n")

  repeat{
    cat("Jumlah kelompok yg anda inginkan? ")
    k<-scan("",n=1)
    if(k<n) break
  }
  repeat{
    cat("Pilih jenis modus awal\n")
    cat("1. Pilih k objek yang pertama dari suatu set data \n")
    cat("2. Penugasan kategori \n")
    cat("Pilihan anda (1 atau 2) ? ")
    pil<-scan("",n=1)
    if(pil==1 || pil==2) break
  }
  if(pil==1) m<-modusawal1(x,k)
  else{
    m<-modusawal2(x,k)
  }
  poskel<-kelompok(x,m,k)
  repeat{
    nkel<-nkelompok(poskel,k)
    if(min(nkel)==0) {
      aa<-0
      break
    }
    else {
      M<-Modusbaru2(x,poskel,nkel,k)
      pos<-kelompok(x,M,k)
      cek<-persamaan(pos,poskel)
      if(cek==1) poskel<-pos
      else {
        aa<-1
        break
      }
    }
  }
  if(aa==1){
    cat("\nHasil pengelompokan data anda :\n")
    for(i in 1:k) {
```

```

        cat("Anggota Kelompok ", i, " : ")
        for(j in 1:(nrow(x)))
            if(pos[j] == i) cat(j, " ")
        cat("\n")
    }
}
else{
    cat("\nPERHATIAN.....\n")
    cat("Terdapat kelompok yang tidak punya anggota\n")
    cat("Gantilah jumlah kelompok yang anda inginkan atau
pilih jenis modus awal yang lain\n")
    cat("\n      T E R I M A      K A S I H\n")
}
}

```

## 2. Sub Program

### a. Program untuk menghitung frekuensi kategori kemudian disusun menjadi suatu matrik kategori dalam urutan frekuensi yang menurun.

```

> frek
function(x)
{
    x<-as.matrix(x)
    maks<-max(x)
    n<-nrow(x)
    m<-ncol(x)
    matrik.frek<-matrix(0,maks,m)
    matrik.frek.relatif<-matrix(0,maks,m)
    for(i in 1:m){
        atribut<-max(x[,i])
        for(j in 1:atribut){
            for(k in 1:n){
                if(x[k,i]==j) matrik.frek[j,i]<-
matrik.frek[j,i]+1
            }
            matrik.frek.relatif[j,i]<-matrik.frek[j,i]/n
        }
    }
    h<-matrik.frek
    a<-nrow(h)
    b<-ncol(h)
    y<-matrix(0,a,b)
    for (i in 1:b){
        for (j in 1:a){
            par<-max(h[,i])
            if(par!=0) {
                k<-0
                repeat{
                    k<-k+1
                    if(h[k,i]==par) break
                }
                y[j,i]<-k
                h[k,i]<-0
            }
        }
    }
}

```

```

    }
    return(matrik.frek,matrik.frek.relatif,y)
}

```

- b. Program untuk menentukan pusat awal kelompok menggunakan metode I yaitu memilih k obyek yang pertama dari suatu set data .**

```

> modusawal1
function(x,k)
{
  x<-as.matrix(x)
  m<-ncol(x)
  modus<-matrix(0,k,m)
  for(i in 1:k) modus[i,]<-x[i,]
  return(modus)
}

```

- c. Program untuk menentukan pusat awal kelompok menggunakan metode II yaitu menugaskan kategori dengan frekuensi tertinggi berturut-turut.**

```

> modusawal2
function(x,k)
{
  x<-as.matrix(x)
  n<-nrow(x)
  m<-ncol(x)
  q<-matrix(0,k,m)
  key<-c(rep(1,m))
  for(i in 1:l){
    b<-1
    for(j in 1:m){
      if(x[b,j]==0 || b>n) b<-1
      q[i,j]<-x[b,j]
      b<-b+1
      key[j]<-b
      if(b>n) b<-1
    }
  }
  for(i in 2:k){
    for(j in 1:m){
      if(key[j]>n) key[j]<-1
      if(x[key[j],j]==0) key[j]<-1
      q[i,j]<-x[key[j],j]
      key[j]<-key[j]+1
      if(key[j]>n) key[j]<-1
    }
  }
  return(q)
}

```

**d. Program untuk memperoleh jarak antar obyek-obyek pengamatan dengan pusat kelompoknya.**

```

> Syarat
function(x,y)
{
  x<-as.vector(x)
  y<-as.vector(y)
  m<-length(x)
  s<-c(1:m)
  for(i in 1:m){
    if(x[i]==y[i]) s[i]<-0
    else s[i]<-1
  }
  return(s)
}

> jarak
function(a,b,c)
{
  a<-as.vector(a)
  b<-as.vector(b)
  c<-as.matrix(c)
  m<-length(a)
  panjang.jarak<-0
  s<-syarat(a,b)
  for(i in 1:m){
    d<-(c[a[i],i]+c[b[i],i])/(c[a[i],i]*c[b[i],i])*s[i]
    panjang.jarak<-panjang.jarak+d
  }
  return(panjang.jarak)
}

```

**e. Program untuk menghitung banyaknya kelompok**

```

> nkelompok
function(x,k)
{
  x<-as.vector(x)
  n<-length(x)
  nkel<-c(rep(0,k))
  for(i in 1:k)
  {
    jumlah<-0
    for(j in 1:n)
    {
      if(x[j]==i) jumlah<-jumlah+1
    }
    nkel[i]<-jumlah
  }
  return(nkel)
}

```

**f. Program untuk menempatkan obyek pengamatan pada kelompok yang sesuai.**

```
> Kelompok
function(x,m,k)
{
  x<-as.matrix(x)
  a<-nrow(x)
  t<-frek(x)
  f<-t$matrik.frek
  b<-nrow(m)
  mat.jarak<-matrix(0,a,b)
  for(i in 1:a){
    for(j in 1:b){
      mat.jarak[i,j]<-jarak(x[i,],m[j,],f)
    }
  }
  poskelompok<-c(1:a)
  for(i in 1:a){
    minim<-min(mat.jarak[i,])
    for(j in 1:b){
      if(minim==mat.jarak[i,j])
        poskelompok[i]<-j
    }
  }
  return(poskelompok)
}
```

**g. Program untuk menentukan pusat kelompok yang baru dari masing-masing kelompok.**

```
> Modusbaru
function(x,y,z,k)
{
  x<-as.matrix(x)
  y<-as.vector(y)
  m<-ncol(x)
  n<-nrow(x)
  z<-as.vector(z)
  M<-list()
  MM<-list()
  for(i in 1:k)
  {
    M[[i]]<-matrix(0,z[i],m)
    MM[[i]]<-matrix(0,z[i],m)
  }
  for (j in 1:k)
  {
    b<-1
    for(a in 1:n){
      if(y[a]==j)
      {
        M[[j]][b,]<-x[a,]
        b<-b+1
      }
    }
  }
}
```



```
modus<-matrix(0,k,m)
for(i in 1:k) {
  d<-frek(M[[i]])
  MM[[i]]<-d$y
  modus[i,]<-MM[[i]][1,]
}
return(M,modus)
}
```

**h. Program untuk memperoleh matrik jarak antar obyek-obyek pengamatan.**

```
> matrik.jarak
function(x)
{
  x <- as.matrix(x)
  n <- nrow(x)
  m <- ncol(x)
  t <- frek(x)
  f <- t$matrik.frek
  mat.jarak <- matrix(0, n, n)
  for(i in 1:n) {
    for(j in 1:n)
      mat.jarak[i, j] <- jarak(x[i, ], x[j, ], f)
  }
  return(mat.jarak)
}
```

**Lampiran 2****Data I. Tentang karakteristik individu**

	V1	V2	V3	V4	V5	V6
individu 1	68 in	140 lb	green	blond	right	female
individu 2	73 in	185 lb	brown	brown	right	male
individu 3	67 in	165 lb	blue	blond	right	male
individu 4	64 in	120 lb	brown	brown	right	female
individu 5	76 in	210 lb	brown	brown	left	male

( Sumber : Johnson, 1975 )

**Nama Variabel :**

V1 : Tinggi badan

V2 : Berat badan

V3 : Warna mata

V4 : Warna rambut

V5 : Tangan aktif

V6 : Jenis kelamin

## Lampiran 3

## Data II. Tentang observasi terhadap jenis mobil

	Engsize	Cylind	Combrate	Weight	For
Pontiac Paris	5	3	4	5	0
Honda Civic	1	1	1	1	1
Buick Century	4	2	4	3	0
Subaru GL	1	1	1	2	1
Volvo 740 GLE	2	1	2	3	1
Plymouth Caravel	2	1	2	3	0
Honda Accord	1	1	2	2	1
Chev Camaro	3	2	3	4	0
Plymouth Horizon	2	1	2	2	0
Chrysler Daytona	2	1	2	3	0
Cadillac Fleetw	4	3	4	5	0
Ford Mustang	5	3	4	4	0
Toyota Celica	2	1	2	2	1
Ford Escort	1	1	2	2	0
Toyota Tarcel	1	1	1	1	1
Toyota Camry	2	1	1	2	1
Mercury Capri	5	3	4	4	0
Toyota Cressida	3	2	3	4	1
Nissan 300 zx	3	2	4	4	1
Nissan Maxima	3	2	4	4	1

( Sumber : Jobson, 1992 )

$V_1$  (Engsize) : 15 – 18 = 1; 20 – 24 = 2; 28 – 30 = 3; 38 – 41 = 4; 50 = 5  
( ukuran mesin )

$V_2$  (Cylind) : 4, 6, 8 menjadi 1, 2, 3  
( banyaknya silinder)

$V_3$  (Combrate) : 64 – 71 = 1; 74 – 84 = 2; 93 – 97 = 3; 104 – 110 = 4  
( tingkat bahan bakar yang dibutuhkan )

**V<sub>4</sub> (Weight):** 2000-2250 = 1; 2500-2750 = 2; 3000 = 3; 3500 = 4; 4000 = 5.  
( berat)

**V<sub>5</sub> (For) :** 0 = Buatan luar amerika selatan

1 = Buatan amerika selatan



## Lampiran 4

### Proses pengelompokan data

#### 1. Data I

```
> a<-as.matrix(manual1)
> a
  V1 V2 V3 V4 V5 V6
1  1  1  1  2  2  2
2  2  2  2  1  2  1
3  1  2  1  2  2  1
4  1  1  2  1  2  2
5  2  2  2  1  1  1
> program.skripsi2(a)
```

```
      S E L A M A T   D A T A N G
DI PROGRAM PENGELOMPOKAN DATA KATEGORIK K-MODUS
```

```
-----
Jumlah kelompok yg anda inginkan ? 1: 2
Pilih jenis modus awal
1. Pilih k objek yang pertama dari suatu set data
2. Penugasan kategori
Pilihan anda (1 atau 2) ? 1: 1
```

```
Hasil pengelompokan data anda :
Anggota Kelompok 1 : 1 3 4
Anggota Kelompok 2 : 2 5
> program.skripsi2(a)
```

```
      S E L A M A T   D A T A N G
DI PROGRAM PENGELOMPOKAN DATA KATEGORIK K-MODUS
```

```
-----
Jumlah kelompok yg anda inginkan ? 1: 2
Pilih jenis modus awal
1. Pilih k objek yang pertama dari suatu set data
2. Penugasan kategori
Pilihan anda (1 atau 2) ? 1: 2
```

```
Hasil pengelompokan data anda :
Anggota Kelompok 1 : 1 3 4
Anggota Kelompok 2 : 2 5
```

#### 2. Data II

```
> attach(datamobil)
> b<-as.matrix(datamobil)
> b
  V1 V2 V3 V4 V5
  5  3  4  5  1
  1  1  1  1  2
  4  2  4  3  1
  1  1  1  2  2
  2  1  2  3  2
  2  1  2  3  1
  1  1  2  2  2
```

```

3 2 3 4 1
2 1 2 2 1
2 1 2 3 1
4 3 4 5 1
5 3 4 4 1
2 1 2 2 2
1 1 2 2 1
1 1 1 1 2
2 1 1 2 2
5 3 4 4 1
3 2 3 4 2
3 2 4 4 2
3 2 4 4 2

```

```
> program.skripsi2(b)
```

```

      S E L A M A T   D A T A N G
    DI PROGRAM PENGELOMPOKAN DATA KATEGORIK K-MODUS
-----

```

```

Jumlah kelompok yg anda inginkan ? 1: 3
Pilih jenis modus awal
1. Pilih k objek yang pertama dari suatu set data
2. Penugasan kategori
Pilihan anda (1 atau 2) ? 1: 1

```

```

Hasil pengelompokan data anda :
Anggota Kelompok 1 : 1 8 11 12 17 19 20
Anggota Kelompok 2 : 2 4 7 14 15 16 18
Anggota Kelompok 3 : 3 5 6 9 10 13

```

```
> program.skripsi2(b)
```

```

      S E L A M A T   D A T A N G
    DI PROGRAM PENGELOMPOKAN DATA KATEGORIK K-MODUS
-----

```

```

Jumlah kelompok yg anda inginkan ? 1: 3
Pilih jenis modus awal
1. Pilih k objek yang pertama dari suatu set data
2. Penugasan kategori
Pilihan anda (1 atau 2) ? 1: 2

```

```

Hasil pengelompokan data anda :
Anggota Kelompok 1 : 2 4 15
Anggota Kelompok 2 : 3 8 12 17 18 19 20
Anggota Kelompok 3 : 1 5 6 7 9 10 11 13 14 16

```