

**CENTEROID TERBAIK BERDASARKAN ALGORITMA K-MEANS  
DAN PARTITIONING AROUND MEDOID PADA DATA YANG  
MENGANDUNG OUTLIER**

**SKRIPSI**

oleh :  
**SUYITNO**  
**0510950057-95**

**UNIVERSITAS BRAWIJAYA**



**PROGRAM STUDI STATISTIKA  
JURUSAN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS BRAWIJAYA  
MALANG  
2009**





**CENTEROID TERBAIK BERDASARKAN ALGORITMA K-MEANS  
DAN PARTITIONING AROUND MEDOID PADA DATA YANG  
MENGANDUNG OUTLIER**

**SKRIPSI**

Sebagai salah satu syarat untuk memperoleh gelar  
Sarjana Sains dalam bidang Statistika

oleh :  
**SUYITNO**  
**0510950057-95**



**PROGRAM STUDI STATISTIKA  
JURUSAN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS BRAWIJAYA  
MALANG  
2009**



**LEMBAR PENGESAHAN SKRIPSI**

***CENTROID TERBAIK BERDASARKAN ALGORITMA K-MEANS DAN PARTITIONING AROUND MEDOID PADA DATA YANG MENGANDUNG OUTLIER***

oleh:  
**SUYITNO**  
**0510950057-95**

Setelah dipertahankan di depan Majelis Penguji  
pada tanggal 16 Desember 2009  
dan dinyatakan memenuhi syarat untuk memperoleh gelar  
Sarjana Sains dalam bidang Statistika

**Pembimbing I**

**Pembimbing II**

**Prof. Dr. Ir. Loekito Adi S., MAg**    **Adji Achmad Rinaldo F., SSi., MSc**  
**NIP. 194703271974121001**        **NIP. 198109082005011002**

Mengetahui,  
**Ketua Jurusan Matematika**  
**Fakultas MIPA Universitas Brawijaya**

**Dr. Agus Suryanto., MSc**  
**NIP. 196908071994121001**



**LEMBAR PERNYATAAN**

Saya yang bertanda tangan di bawah ini:

Nama : SUYITNO  
NIM : 0510950057 - 95  
Program Studi : STATISTIKA  
Penulisan Skripsi berjudul :

*Centroid* Terbaik Berdasarkan Algoritma *K-means* dan  
*Partitioning Around Medoid* Pada Data Yang Mengandung  
*Outlier*

Dengan ini menyatakan bahwa:

1. Isi dari Skripsi yang saya buat adalah benar-benar karya sendiri dan tidak menjiplak karya orang lain selain nama-nama yang termaktub di isi dan tertulis di daftar pustaka dalam Skripsi.
2. Apabila di kemudian hari ternyata Skripsi yang saya tulis terbukti hasil jiplakan, maka saya akan bersedia menanggung segala risiko yang akan saya terima.

Demikian pernyataan ini dibuat dengan segala kesadaran.

Malang, 16 Desember 2009  
Yang menyatakan

(SUYITNO)  
NIM. 0510950057

repository.ub.ac

# CENTROID TERBAIK BERDASARKAN ALGORITMA *K-MEANS* DAN *PARTITIONING AROUND MEDOID* PADA DATA YANG MENGANDUNG *OUTLIER*

## ABSTRAK

Analisis kelompok adalah suatu metode untuk mengidentifikasi sekelompok obyek yang mempunyai kemiripan karakteristik tertentu yang dapat dipisahkan dengan kelompok obyek lainnya. Metode analisis kelompok yang sering digunakan adalah metode hierarki dan metode non-hierarki. Secara mendasar, ada dua cara pengalokasian data ke dalam masing-masing kelompok pada saat proses iterasi pengelompokan. Kedua cara tersebut adalah pengalokasian dengan cara tegas (*hard*) dan *fuzzy*. Untuk pengalokasian secara *hard* setiap obyek hanya dapat masuk dalam satu kelompok dan tidak menjadi anggota kelompok lainnya dan yang termasuk dalam pengalokasian *hard* diantaranya adalah algoritma *k-mean* dan *partitioning around medoid* (PAM). Sedangkan pengalokasian secara *fuzzy* di mana masing-masing obyek data diberikan nilai kemungkinan untuk bisa masuk ke setiap kelompok yang ada. Keberadaan pencilan data dapat berpengaruh terhadap hasil analisis. Dalam penelitian ini ingin mencari algoritma yang menghasilkan *centroid* yang terbaik pada data yang mengandung *outlier* dengan menggunakan 3 data sekunder. Berdasarkan hasil analisis diperoleh nilai *Adjusted rand index* (ARI) yang dihasilkan oleh algoritma *k-mean* lebih kecil dibandingkan dengan nilai ARI yang dihasilkan algoritma PAM pada data yang mengandung *outlier*. Sehingga dapat disimpulkan bahwa algoritma *Partitioning Around Medoid* menghasilkan *centroid* yang menghasilkan pengelompokan yang lebih baik dibandingkan dengan algoritma *k-means* pada data yang mengandung *outlier*.

Kata kunci: *k-means*, *partitioning around medoid*, *centroid*, *outlier*

## BEST CENTROID BASED ON K-MEANS ALGORITHM AND PARTITIONING AROUND MEDOID AT DATA THAT CONTAIN OUTLIER

### ABSTRACT

Cluster analysis is a method which is used to identify a group of objects that has certain similarity of characteristics that can be separated with the other cluster object. The cluster analysis that is commonly used is hierarchy method and non hierarchy method. Basically, there are two ways in allocating data in each group at clustering iteration process. Both ways are allocating data by hard way and fuzzy way. In hard way, each object can be only entered into one cluster, and cannot be able to be entered into another group. The methods that are included into hard way or k-means algorithm and partitioning around medoid (PAM). Meanwhile in the fuzzy way, each data object is given probability value to be entered into each cluster. The presence of outlier can be able to influence the analysis's results. The research is conducted to search the algorithm that produce the best centroid at data that contain outlier by using 3 secondary data. Based on analysis's results, it is obtained that the Adjusted rand index (ARI) that produced by k-mean is lower than ARI that is produced by PAM for data that contain outlier. So it can be concluded that Algorithm Partitioning Around Medoid method is better compared with k-means algorithm method.

Keywords: k-means, partitioning around medoid, centroid, outlier.

## KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas rahmat dan hidayah-Nya penulis dapat menyelesaikan skripsi yang berjudul "***Centroid terbaik berdasarkan Algoritma K-means dan Partitioning Around Medoid Pada Data Yang Mengandung Outlier***". Dalam penyusunan skripsi ini, banyak pihak yang telah memberikan bantuan kepada penulis. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan terima kasih kepada :

1. Bapak Prof. Dr. Ir. Loekito Adi S., M.Agr. Selaku dosen pembimbing I dan Bapak Adji Achmad Rinaldo Fernandes, S.Si., M.Sc selaku dosen pembimbing II, yang telah memberikan bimbingan, masukan serta motivasi.
2. Ibu Ir. Heni Kusdarwati, MS, Ibu Ir. Soepraptini, M.Sc dan Ibu Eni Sumarminingsih, S.Si., MM selaku dosen penguji atas saran dan masukan yang telah diberikan.
3. Bapak Dr. Agus Suryanto, M.Sc selaku Ketua Jurusan Matematika Fakultas MIPA Universitas Brawijaya.
4. Bapak, Ibu, Mbak, Mas, Ipik-ipik dan semua keluarga atas dukungan, perhatian dan doa yang diberikan selama ini kepada penulis untuk mencapai prestasi terbaik.
5. Seluruh staf pengajaran dan TU Jurusan Matematika UB.
6. Seluruh personil Statistika 2005 atas perhatian, perjuangan, dukungan, kerjasama dan semangat selama ini.
7. Teman-teman Statistika 2002, 2003, 2004 dan 2006 dan teman-teman kos atas bantuan, dukungan dan perhatiannya.
8. Seluruh pihak yang telah berpartisipasi yang tidak dapat penulis sebutkan seluruhnya.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, penulis menerima saran dan kritik yang membangun demi kesempurnaan skripsi ini. Semoga skripsi ini ini dapat bermanfaat bagi banyak pihak.

Malang, 16 Desember 2009

Penulis



## DAFTAR ISI

	Halaman
<b>HALAMAN JUDUL</b> .....	i
<b>HALAMAN PENGESAHAN</b> .....	ii
<b>HALAMAN PERNYATAAN</b> .....	iii
<b>ABSTRAK</b> .....	iv
<b>ABSTRACT</b> .....	v
<b>KATA PENGANTAR</b> .....	vi
<b>DAFTAR ISI</b> .....	vii
<b>DAFTAR GAMBAR</b> .....	ix
<b>DAFTAR TABEL</b> .....	x
<b>DAFTAR LAMPIRAN</b> .....	xi
<b>BAB I PENDAHULUAN</b>	
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah .....	2
1.4 Tujuan Penelitian .....	3
1.5 Manfaat Penelitian .....	3
<b>BAB II TINJAUAN PUSTAKA</b>	
2.1 Pencilan.....	5
2.1.1 Pengaruh Pencilan.....	5
2.1.2 Pendeteksian Pencilan.....	5
2.2 Korelasi .....	6
2.3 Pembakuan .....	8
2.4 Analisis Kelompok .....	8
2.4.1 Ukuran Kedekatan .....	9
2.4.2 Analisis Kelompok Hirarki dan Non-Hirarki .....	11
2.4.2.1 Algoritma <i>K-means</i> .....	14
2.4.2.2 Algoritma Partitioning Around medoid .....	15
2.5 Metode Perbandingan <i>Adjusted Rand Index</i> (ARI) .....	19
2.6 Validasi Kelompok.....	20
2.7 Analisis Deskriptif.....	22
<b>BAB III METODE PENELITIAN</b>	
3.1 Sumber Data .....	23

3.2 Metode Penelitian.....	23
<b>BAB IV HASIL PEMBAHASAN</b>	
4.1 Deskripsi Data .....	27
4.1.1 Data I.....	27
4.1.2 Data II.....	28
4.1.3 Data III .....	28
4.2 Pendeteksian Pencilan .....	29
4.3 Pendeteksian Korelasi .....	31
4.4 Analisis Kelompok Non-Hirarki .....	33
4.4.1 Algoritma <i>K-means</i> .....	33
4.4.2 Algoritma <i>Partitioning Around Medoid</i> .....	35
4.5 Uji Perbandingan <i>Adjusted rand index</i> (ARI) .....	36
4.6 Validasi Kelompok .....	37
4.7 Interpretasi Karakteristik Pengelompokan .....	38
4.8 Pembahasan .....	40
<b>BAB V KESIMPULAN DAN SARAN</b>	
5.1 Kesimpulan.....	43
5.2 Saran.....	43
<b>DAFTAR PUSTAKA</b> .....	45
<b>LAMPIRAN</b> .....	49



## DAFTAR GAMBAR

	Halaman
Gambar 2.1. Hasi Analisis Kelompok Hirarki .....	12
Gambar 2.2. Hasil Analisis Kelompok Non-hirarki .....	12
Gambar 2.3. Pengelompokan <i>Hard</i> dan <i>Fuzzy</i> .....	13
Gambar 2.4. Pengelompokan berdasarkan Algoritma <i>K-means</i> dan <i>Partitioning Around Medoid</i> .....	13
Gambar 2.5. Kasus Pertama .....	16
Gambar 2.6. Kasus Kedua.....	17
Gambar 2.7. Kasus Ketiga .....	17
Gambar 2.8. Kasus Keempat.....	17
Gambar 3.1. Langkah-langkah Analisis .....	25
Gambar 4.1. Plot Jarak <i>Mahalanobis</i> Data I.....	29
Gambar 4.2. Plot Jarak <i>Mahalanobis</i> Data II.....	30
Gambar 4.3. Plot Jarak <i>Mahalanobis</i> Data III.....	31



## DAFTAR TABEL

	Halaman
Tabel 2.1. Ilustrasi notasi hasil pengelompokan .....	20
Tabel 3.1. Variabel yang terlibat untuk setiap penelitian .....	23
Tabel 4.1. Faktor-faktor penilaian, bobot dan batasan predikat ...	27
Tabel 4.2. Nilai korelasi dan <i>p-value</i> antar peubah pada Data I ..	31
Tabel 4.3. Nilai korelasi dan <i>p-value</i> antar peubah pada Data II .	32
Tabel 4.4. Nilai korelasi dan <i>p-value</i> antar peubah pada Data III	33
Tabel 4.5. <i>Centroid</i> kelompok algoritma <i>k-means</i> pada Data I....	34
Tabel 4.6. <i>Centroid</i> kelompok algoritma <i>k-means</i> pada Data II ..	34
Tabel 4.7. <i>Centroid</i> kelompok algoritma <i>k-means</i> pada Data III .	34
Tabel 4.8. <i>Centroid</i> kelompok algoritma PAM pada Data I .....	35
Tabel 4.9. <i>Centroid</i> kelompok algoritma PAM pada Data II .....	35
Tabel 4.10. <i>Centroid</i> kelompok algoritma PAM pada Data III....	36
Tabel 4.11. Perbandingan Nilai ARI untuk semua Data .....	36
Tabel 4.12. Uji signifikansi peubah antar kelompok Data I.....	37
Tabel 4.13. Uji signifikansi peubah antar kelompok Data II.....	38
Tabel 4.14. Uji signifikansi peubah antar kelompok Data III .....	38
Tabel 4.15. Rata-rata masing-masing kelompok pada Data I.....	39
Tabel 4.16. Rata-rata masing-masing kelompok pada Data II .....	39
Tabel 4.16. Rata-rata masing-masing kelompok pada Data III....	39
Tabel 4.18. Nilai ARI untuk semua Data .....	40



## DAFTAR LAMPIRAN

	Halaman
Lampiran 1. Data I yaitu Rasio keuangan BPR di Daerah Tingkat II Badung Propinsi Bali .....	49
Lampiran 2. Data II yaitu Faktor-faktor yang mempengaruhi penetasan telur burung walet .....	50
Lampiran 3. Data III yaitu Protein Localization Sites.....	52
Lampiran 4. Nilai Jarak <i>Mahalanobis</i> .....	53
Lampiran 5. Matrikas Korelasi Antar Peubah Dengan SPSS.....	56
Lampiran 6. Hasil Output Untuk Algoritma <i>k-means</i> Menggunakan SYSTAT .....	58
Lampiran 7. Hasil Output Untuk Algoritma <i>Partitioning Around Medoid</i> Menggunakan S-PLUS .....	64
Lampiran 8. Uji Perbandingan Peubah Antar Kelompok Dengan SPSS .....	67
Lampiran 9. Analisis Deskriptif Berdasarkan Algoritma <i>k-means</i> Dengan SPSS .....	69





# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Manusia sejak dahulu telah berusaha mengelompokkan fenomena yang ada di alam menjadi berbagai kelompok dan kemudian mengklasifikasikannya menjadi sejumlah kategori. Bagi kehidupan manusia, kegiatan mengelompokkan benda tersebut sangat penting dan bersifat mendasar untuk menghadapi lingkungannya. Sebagai contoh, manusia telah mengelompokkan hewan dan tumbuhan berdasarkan kemiripannya satu sama lain yang kini telah berkembang menjadi taksonomi modern makhluk hidup. Timbul keterbatasan ketika harus mengelompokkan banyak benda, terlebih jika benda-benda tersebut tidak memiliki struktur yang jelas. Oleh karena itu, diperlukan suatu prosedur yang sistematis yang dapat melakukan analisis pengelompokan pada sebuah kumpulan obyek yang dikenal sebagai analisis kelompok.

Sharma (1996), mendefinisikan analisis kelompok sebagai suatu metode untuk mengidentifikasi sekelompok obyek yang mempunyai kemiripan karakteristik tertentu yang dapat dipisahkan dengan kelompok obyek lainnya, pendefinisian kesamaan atau homogenitas kelompok yang ada sangat bergantung kepada tujuan studi atau penelitian. Tujuan utama teknik ini adalah melakukan pengelompokan berdasarkan kriteria tertentu sehingga obyek-obyek tersebut mempunyai ragam di dalam kelompok (*within cluster*) relatif kecil dibandingkan ragam antar kelompok (*between cluster*).

Kelompok adalah sekelompok objek di mana semua titik pada kelompok tersebut lebih dekat dengan pusat (*center*) kelompok tersebut daripada dengan pusat kelompok lain. Umumnya pusat kelompok tersebut adalah *centroid*. Centroid yang telah didefinisikan harus berada pada koordinat yang tepat, agar pengelompokan yang dihasilkan baik (matteua, 2008).

Metode analisis kelompok yang sering digunakan adalah metode hierarki dan metode non-hierarki. Dalam metode hierarki pembagian kelompok dilakukan berdasarkan hierarki yang ada sehingga jumlah kelompok data yang terbentuk sangat bergantung pada karakteristik data, sedangkan pada metode non-hierarki jumlah kelompok ditentukan terlebih dahulu, baru kemudian data dibagi sesuai dengan

jumlah kelompok yang telah ditetapkan (Timm, 2002). Secara mendasar, ada dua cara pengalokasian data ke dalam masing-masing kelompok pada saat proses iterasi (algoritma) pengelompokan. Kedua cara tersebut adalah pengalokasian dengan cara tegas (*hard*), di mana setiap obyek hanya dapat masuk dalam satu kelompok dan tidak menjadi anggota kelompok lainnya, dan dengan cara *fuzzy* di mana masing-masing obyek data diberikan nilai kemungkinan untuk bisa masuk ke setiap kelompok yang ada (Mercer dan College, 2003). Menurut penelitian yang dilakukan oleh (Kirana, 2008), algoritma *fuzzy* lebih baik dari algoritma *k-means* dalam pengelompokan analisis kelompok non-hierarki. Sedangkan pada penelitian ini akan dibahas mengenai *centroid* yang paling baik untuk pengelompokan antara algoritma *k-means* dengan *partitioning around medoid* pada data yang mengandung *outlier*.

*Outlier* merupakan suatu keganjilan dan menandakan suatu titik pengamatan yang sama sekali berbeda dengan pengamatan lain (Marazzi, 1993). Mengabaikan *outlier* dari suatu himpunan data bukanlah tindakan yang tepat, karena *outlier* dapat memberikan pengaruh terhadap hasil analisis secara drastis. Pada analisis kelompok adanya *outlier* dapat menyebabkan satu set obyek terdekat ke suatu *centroid* akan kosong karena *centroid* tidak dapat diperbaharui. Untuk mengetahui hasil kinerja analisis kelompok, baik untuk algoritma *K-means* maupun *partitioning around medoid* digunakan metode perbandingan *Adjusted Rand Index* (Yeung dan Ruzzo, 2001).

## 1.2. Rumusan Masalah

Algoritma apa yang menghasilkan *centroid* terbaik untuk mengatasi data yang mengandung *outlier* pada analisis kelompok non-hierarki berdasarkan algoritma *k-means* dan *partitioning around medoid*.

## 1.3. Batasan Masalah

Batasan masalah pada penelitian ini adalah:

1. Skala data yang digunakan merupakan data interval dan rasio.
2. Ukuran jarak yang digunakan adalah jarak *Manhattan*.
3. Metode perbandingan yang digunakan adalah *Adjusted Rand Index*.

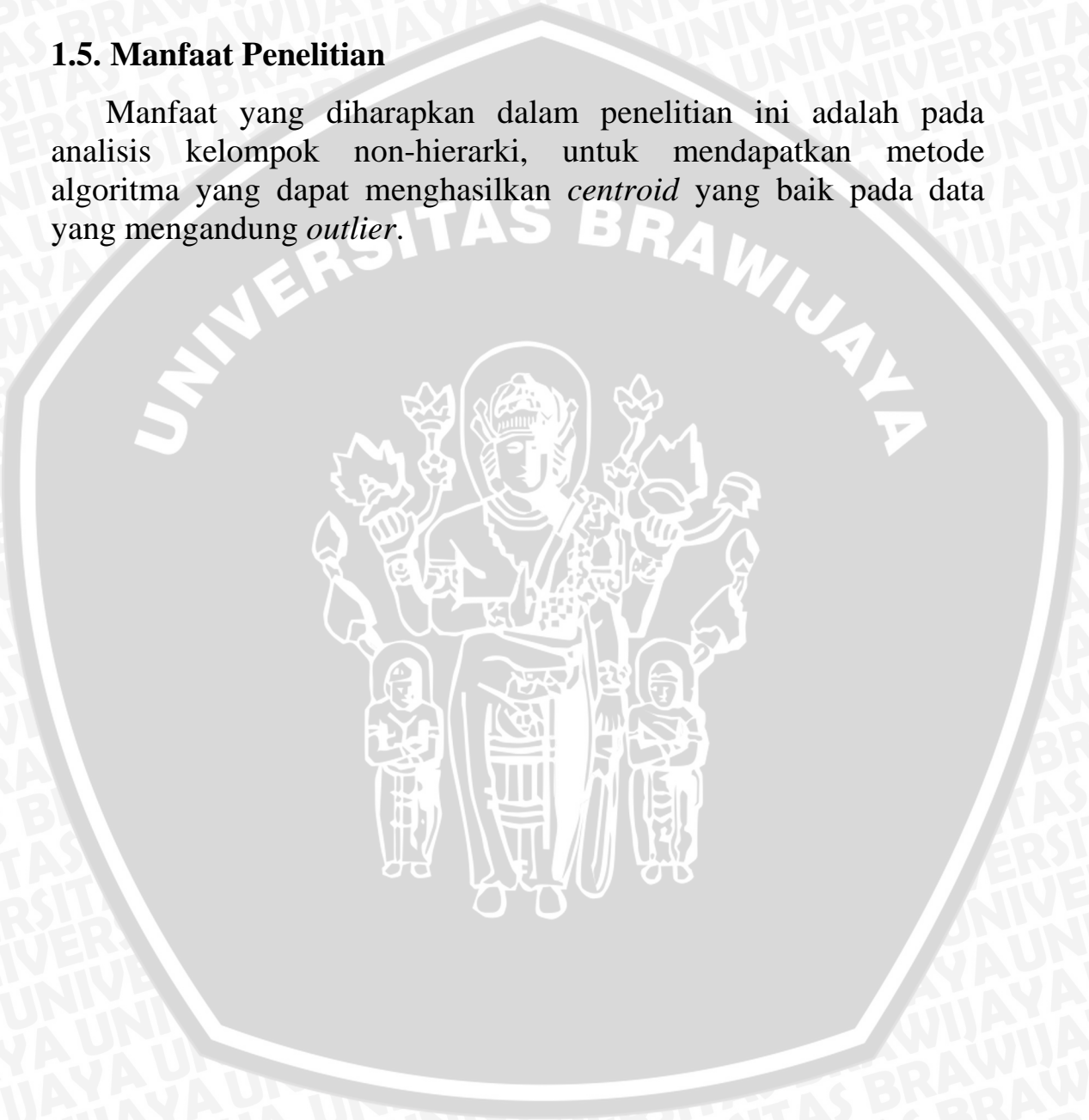


#### 1.4. Tujuan Penelitian

Mengetahui Hasil analisis kelompok non-hierarki yang menghasilkan *centroid* terbaik antara algoritma *k-means* dengan algoritma *partitioning around medoid* berdasarkan *Adjusted rand Index*.

#### 1.5. Manfaat Penelitian

Manfaat yang diharapkan dalam penelitian ini adalah pada analisis kelompok non-hierarki, untuk mendapatkan metode algoritma yang dapat menghasilkan *centroid* yang baik pada data yang mengandung *outlier*.





## BAB II TINJAUAN PUSTAKA

### 2.1. *Outlier*

*Outlier* adalah pengamatan dengan kombinasi unit dari karakteristik yang dapat diindikasikan berbeda jelas dari pengamatan yang lain (Hair *et al*, 1998). Keberadaan *outlier* dapat sebagai indikasi ciri populasi yang tidak normal. Keberadaan *outlier* membuat suatu data memiliki distribusi yang berbeda. Distribusi pertama disebut sebagai distribusi dasar (*basic distribution*) yang menghasilkan pengamatan baik. Distribusi kedua disebut sebagai distribusi kontaminan (*contaminating distribution*) yang menghasilkan pengamatan tidak baik. Jumlah maksimum *outlier* dalam data yang diperbolehkan adalah 50 persen (Rousseeuw dan Zomeron, 1990).

#### 2.1.1. Pengaruh *Outlier*

*Outlier* pada suatu himpunan data tidak dapat diabaikan begitu saja, karena keberadaan *outlier* dapat menimbulkan beberapa masalah, antara lain (Anonymous, 2009):

1. *Outlier* dapat mengubah kesimpulan atau mengaburkan kesimpulan yang dibuat berdasarkan penelitian.
2. *Outlier* dapat mempengaruhi hasil analisis secara drastis. Artinya, jika sebuah titik yang merupakan *outlier* tidak dipakai dalam analisis maka hasil analisis bisa berbeda. Dengan kata lain, keberadaan *outlier* dalam analisis membuat hasil analisis tidak lagi sah.
3. *Outlier* dapat mengindikasikan ketidaknormalan distribusi data.

Oleh karena hasil suatu analisis dapat berubah karena adanya *outlier*, maka diperlukan suatu metode yang lebih kekar terhadap adanya *outlier*.

#### 2.1.2. Pendeteksian *Outlier*

*Outlier* yang disebabkan oleh variabel *independen* dinamakan *leverage*. *Leverage* sangat sulit diketahui sejak awal karena:

1. Visualisasi seperti *scatter diagram* tidak mampu menggambarkan secara utuh dalam satu gambar.
2. Beberapa *outlier* dalam data membentuk efek *masking*.

Pada data univariat pendeteksian *outlier* digunakan untuk mendeteksi ada tidaknya *outlier* pada setiap peubah secara terpisah. Pendeteksian *outlier* tersebut dapat menggunakan Box Plot. Tetapi, dalam proses perhitungan analisis peubah ganda yang melibatkan lebih dari dua peubah, peneliti membutuhkan alat untuk mengukur posisi multidimensi tiap pengamatan secara relatif terhadap beberapa titik. Metode yang paling banyak digunakan untuk mendeteksi *outlier* pada peubah ganda adalah jarak Mahalanobis (Hair *et al*, 1998).

Jarak Mahalanobis dapat dirumuskan sebagai berikut:

$$MD(x_i) = d(\mathbf{x}_i, \bar{\mathbf{x}}, \mathbf{S}) = \sqrt{(\mathbf{x}_i - \bar{\mathbf{x}})' (\mathbf{S})^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})} \quad (2.1)$$

di mana :  $MD(x_i)$  = jarak Mahalanobis (Mahalanobis Distance)

$\bar{\mathbf{x}}$  = vektor rata-rata sampel

$\mathbf{S}$  = matriks kovarian sampel

Menurut Rousseuw dan Zomeren (1990) pengamatan dipandang sebagai *outlier* apabila  $MD(x_i) > \sqrt{\chi^2_{p,1-\alpha/2}}$  atau bisa juga dinyatakan sebagai  $(MD(x_i))^2 > \chi^2_{p,1-\alpha/2}$

## 2.2. Korelasi

Korelasi adalah tingkat keeratan hubungan dua peubah atau lebih yang digambarkan oleh besarnya koefisien korelasi. Koefisien korelasi adalah koefisien yang menggambarkan tingkat keeratan hubungan antara dua peubah atau lebih. Besaran dari koefisien korelasi tidak menggambarkan hubungan sebab akibat antar dua peubah atau lebih, tetapi semata-mata menggambarkan keterkaitan hubungan linier antar peubah (Mattjik dkk, 2002).

Ukuran korelasi antar dua peubah  $x$  dan  $y$  di duga dengan koefisien korelasi seperti persamaan di bawah ini (Walpole, 1995)

$$r_{x,y} = \frac{n \sum_{i=1}^n x_i y_i - \left( \sum_{i=1}^n x_i \right) \left( \sum_{i=1}^n y_i \right)}{\sqrt{\left( n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2 \right) \left( n \sum_{i=1}^n y_i^2 - \left( \sum_{i=1}^n y_i \right)^2 \right)}} \quad (2.2)$$

Di mana  $x$  dan  $y$  merupakan peubah. Nilai korelasi berkisar antara -1 sampai dengan 1. Nilai -1 menandakan terdapat hubungan

negatif (berkebalikan) yang sempurna, nilai 0 menandakan tidak terdapat hubungan positif yang sempurna, sedangkan nilai 1 menandakan terdapat hubungan positif yang sempurna. Nilai koefisien korelasi 0 menunjukkan tidak adanya hubungan linier antara dua peubah, namun bukan menunjukkan bahwa antara dua peubah tersebut tidak terdapat hubungan (Atom, 2008).

Untuk mengetahui adanya korelasi antar dua peubah maka hipotesis yang diuji adalah:

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

Statistik uji yang digunakan adalah:

$$t_{hit} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

di mana  $\rho$  merupakan nilai korelasi yang diduga dari  $r$ . Kriteria keputusan adalah jika nilai  $|t_{hit}| \leq t_{n-2}^{\sigma/2}$  maka terima  $H_0$  dengan taraf nyata  $\alpha$  dan dapat disimpulkan bahwa kedua peubah saling bebas, sebaliknya  $H_0$  ditolak jika  $|t_{hit}| > t_{n-2}^{\sigma/2}$  (Afifi dan Clark, 1990).

Selain menggunakan statistik uji salah satu indikator pengambilan keputusan dalam pengujian hipotesis adalah dengan menggunakan *p-value* yang merupakan taraf keberartian terkecil sehingga nilai statistik uji yang diperoleh masih berarti dan dihitung berdasarkan peluang eksak berlandaskan statistik uji yang digunakan (Walpole dan Myers, 1995). Keuntungan *p-value* adalah tidak hanya menentukan hasil dalam suatu pengambilan keputusan tentang  $H_0$ , tetapi juga memberikan informasi tentang kuatnya keputusan yang diambil dan dapat diketahui besarnya resiko salah secara eksak dalam pengambilan keputusan. Sehingga dalam berbagai pengujian *p-value* sering digunakan sebagai indikator dalam pengambilan keputusan. Besarnya nilai  $\alpha$  yang digunakan sebagai pembanding bagi *p-value* tergantung pada peneliti mengenai seberapa resiko salah yang masih dapat ditolerir menurut jenis penelitian yang dilakukan. Semakin kecil *p-value* maka semakin kecil kesalahan peluang untuk membuat kesalahan dengan menolak  $H_0$  (Hines dan Montgomery, 1990).

### 2.3 Pembakuan

Format pembakuan yang umum adalah mengubah dari tiap nilai variabel ke nilai baku (biasa disebut Z score) dengan pengurangan rata-rata dan membagi dengan simpangan baku untuk masing-masing variabel. Pembakuan adalah bentuk normal fungsi jarak, yang menggunakan ukuran jarak *euclid* untuk transformasi bentuk normal dari bentuk data mentah. Proses pembakuan adalah dengan mengubah masing-masing data mentah ke nilai baku dengan rata-rata 0 dan ragam 1, sehingga menghapus perbedaan skala beberapa atribut atau variabel yang digunakan dalam analisis. Dengan menggunakan nilai baku menghapus pengaruh dalam kaitanya dengan perbedaan skala, tidak hanya untuk variabel lain tetapi juga untuk variabel yang sama. Nilai baku sangat penting digunakan jika skala variabel yang digunakan berbeda karena akan mempengaruhi ragam dari hasil analisis kelompok. Untuk menghitung nilai baku setiap obyek adalah sebagai berikut:

$$Z_{ij} = \frac{(x_{ij} - \bar{x}_j)}{S_j} \quad (2.3)$$

di mana :  $x_{ij}$  = obyek ke-i pada variabel ke-j

$\bar{x}_j$  = rata-rata dari variabel ke-j

$S_j$  = simpangan baku dari variabel ke-j

(Hair *et al*, 1998)

### 2.4. Analisis Kelompok

Analisis kelompok merupakan teknik mereduksi informasi. Informasi dari sejumlah obyek akan direduksi menjadi sejumlah kelompok, di mana jumlah kelompok lebih kecil dari jumlah obyek. Obyek-obyek yang sama dikelompokkan dalam suatu kelompok sehingga mempunyai tingkat kesamaan yang tinggi dibandingkan dengan obyek dari kelompok lain.

Sharma (1996), mendefinisikan analisis kelompok adalah suatu metode untuk mengidentifikasi sekelompok obyek yang mempunyai kemiripan karakteristik tertentu yang dapat dipisahkan dengan kelompok obyek lainnya, pendefinisian kesamaan atau homogenitas kelompok yang ada sangat bergantung kepada tujuan studi atau

penelitian. Tujuan utama teknik ini adalah melakukan pengelompokan berdasarkan kriteria tertentu sehingga obyek-obyek tersebut mempunyai ragam di dalam kelompok (*within cluster*) relatif kecil dibandingkan ragam antar kelompok (*between cluster*).

Analisis kelompok mengelompokkan obyek menjadi sejumlah kelompok. Hasil analisis harus mencerminkan struktur dasar pada data tersebut. Sebagai contoh, analisis kelompok dapat digunakan untuk menemukan gen dan protein yang memiliki fungsi yang mirip.

Data yang berasal dari hasil pengukuran atau pengamatan suatu obyek biasanya direpresentasikan sebagai titik atau vektor pada ruang multidimensi. Setiap dimensi mewakili atribut yang berbeda (mislnya variabel atau sifat) yang menggambarkan obyek tersebut.

Secara umum proses dimulai dengan merumuskan masalah pengelompokan dengan mendefinisikan variabel-variabel yang dipergunakan untuk dasar pembentukan kelompok. Kemudian pengambilan  $p$  pengukuran peubah pada  $n$  obyek pengamatan. Data tersebut dijadikan matriks data mentah berukuran  $n \times p$ . Matrik tersebut ditransformasikan ke dalam bentuk matriks *similaritas* atau *distance* berukuran  $n \times n$  yang dihitung berdasarkan pasangan-pasangan obyek  $p$  peubah. Konsep dasar pengukuran analisis kelompok adalah konsep pengukuran jarak (*distance*) dan kesamaan (*similarity*). *Distance* adalah ukuran tentang jarak pisah antar obyek sedangkan *similarity* adalah ukuran kedekatan. Konsep ini penting karena pengelompokan pada analisis kelompok didasarkan pada kedekatan. Pengukuran jarak digunakan untuk data-data yang bersifat *metriks*, sedangkan pengukuran kesesuaian (*matching type measure*) digunakan untuk data-data yang bersifat kualitatif (Laboratorium Data Mining Jurusan Teknik Industri Fakultas Teknologi Industri Universitas Islam Indonesia, 2009).

#### 2.4.1. Ukuran Kedekatan

Pandang variabel yang diukur yaitu  $x_1, x_2, \dots, x_p$  dan pengukuran dilakukan terhadap setiap individu anggota sampel yang besarnya  $n$ , yaitu individu-individu  $Z_1, Z_2, \dots, Z_n$ . Hasil pengukuran individu  $Z_i$  pada variabel  $X_j$  dinyatakan dengan  $X_{ij}$  (untuk setiap individu ke- $i$ , di mana  $i = 1, 2, 3, \dots, n$  dan peubah ke- $j$ , di mana  $j = 1, 2, 3, \dots, p$ )

$$\mathbf{X}_{ij} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

(Johnson dan Wichern, 2002)

Menurut Giudici (2007), bila  $x_1$ ,  $x_2$  dan  $x_3$  adalah baris dari matriks data di mana memiliki fungsi  $d(x_1, x_2)$ , fungsi tersebut dikatakan jarak antar dua obyek data jika memenuhi syarat sebagai berikut (untuk semua nilai  $x_1$ ,  $x_2$  dan  $x_3$ ):

1. Positif :  $d(x_1, x_2) \geq 0$
2. Identitas :  $d(x_1, x_2) = 0 \Leftrightarrow x_1 = x_2$
3. Simetri :  $d(x_1, x_2) = d(x_2, x_1)$
4. Ketidaksamaan segitiga :  $d(x_1, x_2) \leq d(x_1, x_3) + d(x_2, x_3)$

Seperti yang sudah disebutkan di atas bahwa sekumpulan obyek direpresentasikan sebagai matriks berukuran  $n \times p$ . Matriks ini biasa disebut data matriks, tergantung obyek yang direpresentasikannya. Data matriks sering disebut sebagai “data mentah” karena elemennya berupa nilai yang berasal dari pengukuran langsung dari atribut obyek. *Proximity* matriks adalah sebuah matriks berukuran  $n \times n$  yang menggambarkan kedekatan antara sebuah obyek dengan obyek lain dengan  $n$  adalah banyaknya obyek. Umumnya matriks *proximity* berlaku simetris (Toledo, 2005).

Data yang direpresentasikan sebagai data matriks dapat diubah ke dalam bentuk *proximity* matriks dengan melakukan perhitungan *proximity* antara setiap obyek yang ada dengan menggunakan fungsi *similarity* atau *distance*. Namun hal ini tidak berlaku sebaliknya. Data yang sudah berada dalam bentuk *proximity* matriks tidak dapat dikonversi kembali menjadi bentuk data matriks. Pada sebagian kasus, analisis kelompok tetap dapat dilakukan dengan memanfaatkan *proximity* matriks walaupun data matriks semula tidak tersedia (Toledo, 2005).

Sebelum melakukan pengelompokan obyek data, terlebih dahulu menentukan ukuran kedekatan sifat antar obyek data. Beberapa ukuran jarak telah diimplementasikan dalam menghitung jarak antara data dan *centroid* termasuk di antaranya jarak *Manhattan*. Jarak *Manhattan* disebut juga jarak *City Block* karena jarak yang ditempuh



dimisalkan dengan melakukan perjalanan melalui blok-blok kota. Jarak ini berhubungan dengan jarak *Euclid* dimana pada jarak *Euclid* menghitung jarak terpendek antara dua titik, sedangkan pada jarak *Manhattan* menghitung jumlah jarak yang ditempuh. Menghitung jarak ini sama dengan mengukur jarak yang harus ditempuh antara dua tempat dalam sebuah kota, melalui blok-blok kota. Syarat penggunaan jarak Manhattan pada analisis kelompok adalah bahwa peubah yang diamati tidak saling berkorelasi dan mempunyai pengukuran yang sama (de Hoon, 2002). Jarak *Manhattan* sering digunakan karena kemampuannya dalam mendeteksi keadaan khusus seperti keberadaan *outliers* dengan lebih baik (Agusta, 2007).

Jarak antara dua titik  $x_i$  dan  $x_h$  pada jarak *Manhattan* dihitung dengan menggunakan rumus sebagai berikut:

$$d(i, h) = \sum_{j=1}^p |x_{ij} - x_{hj}| \quad (2.4)$$

di mana:  $x_{ij}$  = nilai dari obyek  $i$  variabel ke- $p$

$x_{hj}$  = nilai dari obyek  $h$  variabel ke- $p$

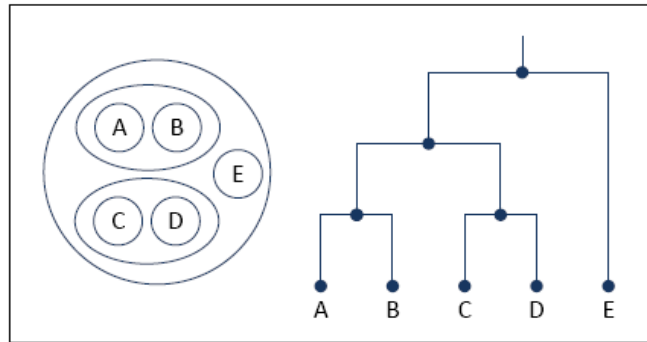
$i = 1, 2, \dots, n$   $h = 1, 2, \dots, n$   $j = 1, 2, \dots, p$

$p$  = banyak variabel

(de Hoon, 2002)

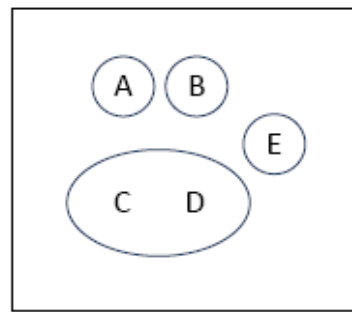
#### 2.4.2. Analisis Kelompok Hierarki dan Non-hierarki

Hasil analisis kelompok dapat bersifat hierarki atau non-hierarki. Pada hasil analisis kelompok yang bersifat hierarki, setiap kelompok dapat memiliki sub kelompok. Sehingga solusinya dapat digambarkan seperti pohon kelompok hierarki yang bertingkat. Sedangkan pada hasil analisis kelompok yang bersifat non-hierarki (*partitional*), hanya ada satu tingkat kelompok. Analisis kelompok hierarki akan menghasilkan kelompok-kelompok yang bersarang (*nested*), sehingga setiap kelompok dapat memiliki sub kelompok. Prinsip utama pada algoritma kelompok hierarki adalah mengatur seluruh obyek dalam sebuah pohon (umumnya pohon biner) berdasarkan suatu fungsi kriteria tertentu. Pohon tersebut biasa disebut dendogram hierarki (bertingkat). Brntuk dari dendogram dapat ditunjukkan pada Gamabar 2.1.



Gambar.2.1 Hasil analisis kelompok hierarki

Analisis kelompok non-hierarki biasa juga disebut algoritma pengelompokan *partitional*, algoritma pengelompokan yang non-hierarki menghasilkan kelompok-kelompok yang datar. Obyek-obyek dibagi menjadi sejumlah  $k$  kelompok berdasarkan suatu kriteria optimasi tertentu. Beberapa algoritma memerlukan nilai  $k$  sebagai parameter masukan. Contoh algoritma non-hierarki adalah *k-mean* dan *partitinal around medoid* (Hair et al, 1998)

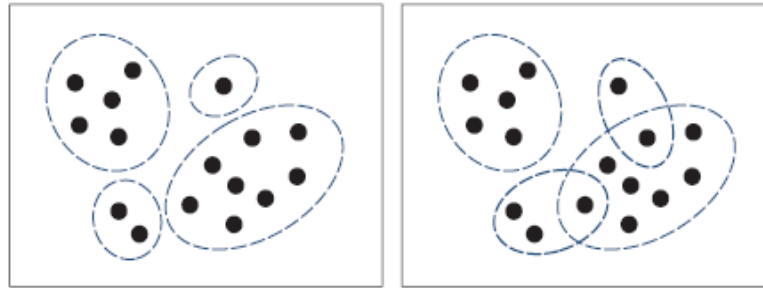


Gambar.2.2 Hasil analisis kelompok non-hierarki

Perbedaan antara Gambar 2.1 dan Gambar 2.2 adalah terletak pada tingkatan atau sub kelompok, di mana pada analisis kelompok hierarki dalam pembentukan kelompok melewati tahapan-tahapan sub kelompok, sedangkan pada Analisis kelompok non-hierarki tidak ada tahapan-tahapan sub kelompok. Selain Analisis kelompok hierarki dan analisis kelompok non-hierarki, ada analisis kelompok lain yang pengelompokannya didasarkan pada nilai kovarian data misalnya algoritma *Gustafson-Kessel* dan analisis kelompok yang pengelompokannya didasarkan pada distribusi peluang data misalnya algoritma *Gath-Geva* (Wolkenhauer, 1998).

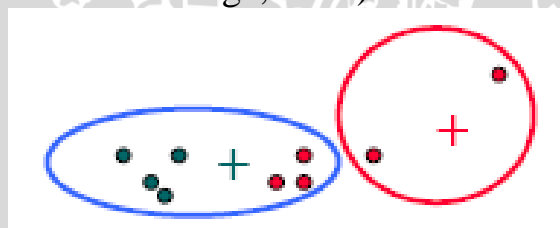
Algoritma *k-means* dan *partitioning around medoid* juga termasuk dalam pengelompokan yang bersifat *hard*, yaitu setiap

obyek hanya dimasukkan tepat ke dalam satu kelompok. Dengan kata lain, setiap obyek hanya menjadi anggota satu kelompok saja.



Gambar.2.3 Pengelompokan *hard* dan *fuzzy*

Dari Gambar 2.3 dapat dijelaskan yaitu untuk gambar sebelah kiri merupakan pengelompokan yang bersifat *hard* sedangkan gambar yang sebelah kanan merupakan pengelompokan yang bersifat *fuzzy* artinya setiap obyek tidak hanya menjadi satu anggota kelompok saja tapi dapat menjadi anggota beberapa kelompok dengan bobot tertentu. Di mana pada penelitian ini akan dibahas mengenai perbandingan algoritma *k-means* dan *algoritma partitioning around medoid*, di mana pada data yang akan dikelompokkan mengandung data *outlier* (Mercer dan College, 2003).



Gambar. 2.4 Pengelompokan berdasarkan algoritma *k-means* dan *partitioning around medoid*

Dari gambar 2.4 di atas dapat dilihat perbedaan *centroid* untuk pengelompokan analisis kelompok non-hierarki antara algoritma *k-means* dan algoritma *partitioning around medoid*, di mana akan dibentuk dua kelompok. Untuk algoritma *k-means* kelompok yang terbentuk adalah elips berwarna biru dan lingkaran berwarna merah dimana untuk *outlier* menjadi kelompok tersendiri. Sedangkan jika menggunakan algoritma *partitioning around medoid* *centroid* yang terbentuk menghasilkan kelompok dengan titik-titik hitam dan titik-titik merah. Tanda *plus* pada Gambar 2.4 menunjukkan *centroid* dari masing-masing kelompok sedangkan titik-titik menunjukkan obyek data yang di kelompokkan (Huan, 2003).

### 2.4.2.1. Algoritma *k-means*

Algoritma *k-means* adalah metode pengelompokan berbasis jarak yang membagi data ke dalam sejumlah kelompok dan hanya bekerja pada atribut numerik. Algoritma *k-means* termasuk *partitioning clustering* yang memisahkan data ke-*k* daerah bagian yang terpisah. Sejumlah *k* kelompok memiliki pusat kelompok (*centroid*). Pusat kelompok yang telah didefinisikan harus berada pada koordinat yang tepat, agar pengelompokan yang didapatkan tidak mengalami perubahan. Sehingga jalan yang terbaik untuk mendapatkan hasil pengelompokan yang maksimal adalah dengan menempatkan setiap koordinat pusat kelompok pada jarak yang cukup jauh dengan koordinat pusat kelompok yang lain (Matteua, 2008). Dalam algoritma *k-means*, setiap data harus termasuk ke kelompok tertentu dan bisa dimungkinkan bagi setiap data yang termasuk kelompok tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke kelompok yang lain. Algoritma *k-means* merupakan metode non-heirarki yang pada awalnya mengambil sebagian dari banyaknya komponen dari populasi untuk dijadikan pusat kelompok awal. Pada langkah ini pusat kelompok dipilih secara acak dari sekumpulan populasi data. Berikutnya *k-means* menguji masing-masing komponen di dalam populasi data dan menandai komponen tersebut ke salah satu pusat kelompok yang telah didefinisikan tergantung dari jarak minimum antar komponen dengan tiap-tiap pusat kelompok. Posisi pusat kelompok akan dihitung kembali sampai semua komponen data digolongkan ke dalam tiap-tiap pusat kelompok dan terakhir akan terbentuk posisi pusat kelompok baru.

Langkah-langkah algoritma *k-means* adalah sebagai berikut

1. Menyusun matriks partisi awal derajat keanggotaan ke dalam kelompok secara acak. Derajat keanggotaan mempunyai nilai 0 dan 1. apabila suatu obyek data merupakan anggota suatu kelompok, maka derajat keanggotaan bernilai satu dan sebaliknya.
2. Menghitung nilai pusat kelompok dengan persamaan:

$$v_{jk} = \frac{\sum_{i=1}^n (x_{ij} \mu_{ik})}{\sum_{i=1}^n \mu_{ik}} \quad (2.5)$$

$$i = 1, 2, \dots, n \quad j = 1, 2, \dots, p \quad k = 1, 2, \dots, c$$

di mana:

$v_{jk}$  = nilai pusat kelompok peubah ke-j kelompok ke-k

$n$  = banyak obyek

$c$  = banyak kelompok

$x_{ij}$  = obyek data ke-i pada peubah ke-j

$\mu_{ik}$  = derajat keanggotaan obyek data ke-i pada kelompok ke-k

Pada iterasi pertama pada algoritma *k-means* ini nilai pusat kelompok dicari secara acak, untuk iterasi selanjutnya digunakan persamaan 2.5.

3. Menghitung fungsi obyektif dengan persamaan

$$J = \sum_{i=1}^n \sum_{k=1}^c \left( \left[ \sum_{j=1}^p (x_{ij} - v_{jk})^2 \right] \mu_{ik} \right) \quad (2.6)$$

$i = 1, 2, \dots, n \quad j = 1, 2, \dots, p \quad k = 1, 2, \dots, c$

di mana nilai derajat keanggotaan

$$\mu_{ik} = \sum_{j=1}^p |x_{ij} - x_{hj}| \quad (2.7)$$

(Likes *et al*, 2002)

4. Jika nilai fungsi obyektif telah konvergen, pengelompokan dari derajat keanggotaan setiap obyek data dalam suatu kelompok ditentukan berdasarkan jarak terpendek hasil perhitungan jarak. Jika belum konvergen kembali ke langkah kedua Pengalokasian obyek pada kelompok dapat dirumuskan sebagai berikut:

$$a_{ik} = \min \{ D(x_i, v_k) \}$$

di mana:

$a_{ik}$  = keanggotaan obyek ke-i ke kelompok ke-k

$v_k$  = Nilai *centroid* kelompok ke-k

(Timm, 2002)

#### 2.4.2.2. Algoritma *Partitioning Around Medoid*

Pada algoritma *Partitioning around medoid* (PAM) akan berusaha menentukan obyek yang dapat mewakili suatu kelompok yang lebih sering disebut *medoid*. Saat *medoid* telah didapatkan, setiap obyek yang bukan *medoid* akan dikelompokkan dengan *medoid* yang paling dekat jaraknya. Misal jika  $O_j$  adalah obyek *non-medoid* dan  $O_m$  adalah *medoid*, dapat dikatakan bahwa  $O_j$  termasuk

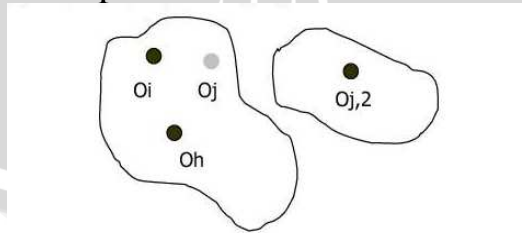
pada kelompok yang diwakili oleh  $O_m$  jika  $d(O_j, O_m) = \min_{O_e} d(O_j, O_e)$ . Untuk mencari sejumlah  $k$  medoid, PAM memulai dengan seleksi *inisial k medoid*. kemudian pada tiap tahap akan melakukan pergantian (*swap*) antara obyek yang dipilih ( $O_m$ ) dengan obyek yang tidak terpilih ( $O_p$ ). Selama proses *swap* tersebut dapat menghasilkan kelompok yang lebih baik (Raymond dan Han, 1994).

Sebelum beralih pada langkah-langkah dari algoritma PAM, diberikan contoh berikut ini. Misal terdapat 2 *medoid*,  $A$  dan  $B$ . Di mana medoid  $A$  akan di gantikan denagan *medoid* baru yaitu  $M$ . Maka, untuk semua obyek  $Y$  yang ada pada kelompok yang diwakili oleh  $A$ , perlu dicari *medoid* terdekat untuk melakukan pergantian. Untuk hal ini, terdapat dua kasus. Kasus pertama,  $Y$  akan berpindah pada kelompok yang diwakili oleh  $B$ , tetapi tidak pada kelompok baru yang diwakili oleh  $M$ . Pada kasus kedua,  $Y$  akan berpindah pada kelompok baru yang diwakili oleh  $M$  dan kelompok yang diwakili oleh  $B$  sama sekali tidak terpengaruh. Selain dari pertimbangan terhadap semua obyek  $Y$  yang terdapat pada kelompok yang diwakili oleh  $A$ , juga perlu mempertimbangkan semua obyek  $Z$  yang ada pada kelompok yang diwakili oleh  $B$ . Pada saat pergantian,  $Z$  dapat tetap berada pada kelompok yang diwakili oleh  $B$  ataupun berpindah ke kelompok baru yang diwakili oleh  $M$ .

Menurut Garg dan Jain (2006) dapat dijelaskan beberapa kasus yang mungkin terjadi saat terjadi pergantian (*swap medoid*), yaitu:

a. Kasus pertama

$O_j$  termasuk pada kelompok yang diwakili oleh  $O_i$ , tetapi meninggalkan kelompok setelah melakukan pergantian (*swap medoid*). Pada kasus ini,  $d(O_j, O_h) \geq d(O_j, O_{j,2})$  jadi biaya (*cost*) dari swapping tersebut adalah  $C_{jih} = d(O_j, O_{j,2}) - d(O_j, O_i)$ . Untuk kasus pertama dapat dilihat pada Gambar 2.5.

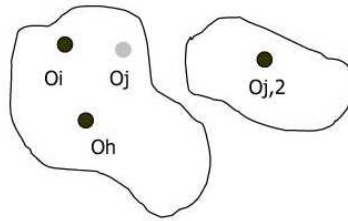


Gambar 2.5 Kasus pertama

b. Kasus kedua

$O_j$  termasuk pada kelompok yang diwakili oleh  $O_i$  dan tetap dalam keadaan tersebut setelah dilakukan pergantian (*swap medoid*). Pada kasus ini,  $d(O_j, O_h) < d(O_j, O_{j,2})$ . Jadi biaya (*cost*) dari swapping

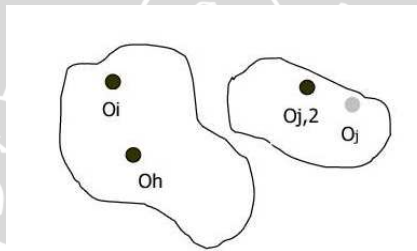
tersebut adalah  $C_{jih} = d(O_j, O_h) - d(O_j, O_i)$ . Untuk kasus kedua dapat dilihat pada Gambar 2.6.



Gambar 2.6 Kasus kedua

c. Kasus ketiga

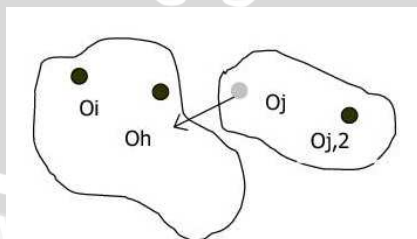
$O_j$  termasuk pada kelompok yang diwakili oleh  $O_{j,2}$  berbeda dengan kelompok yang diwakili oleh  $O_i$  dan tetap dalam kelompok  $O_{j,2}$  setelah dilakukan pergantian (*swap*) *medoid*. Jadi biaya (*cost*) dari *swapping* tersebut adalah  $C_{jih} = 0$ . Untuk kasus ketiga dapat dilihat pada Gambar 2.7.



Gambar 2.7 Kasus ketiga

d. Kasus keempat

$O_j$  termasuk pada kelompok yang diwakili oleh  $O_{j,2}$  tetapi berpindah ke kelompok yang diwakili oleh  $O_h$  setelah dilakukan pergantian (*swap*) *medoid*. Dapat kita lihat bahwa  $d(O_j, O_h) \geq d(O_j, O_{j,2})$ . Jadi biaya (*cost*) dari *swapping* tersebut adalah  $C_{jih} = d(O_j, O_h) - d(O_j, O_{j,2})$ . Untuk kasus keempat dapat dilihat pada Gambar 2.8.



Gambar.2.8 Kasus keempat

Langkah-langkah algoritma *partitioning around medoid* menurut Park *et al* (2005) adalah sebagai berikut:

1. Seleksi *inisial medoid*

- a. Penggunaan jarak sebagai ukuran perbedaan, menghitung jarak antar tiap-tiap pasangan dari semua obyek sebagai berikut:

$$d(i, h) = \sum_{j=1}^p |x_{ij} - x_{hj}| \quad i = 1, \dots, n; \quad h = 1, \dots, n \quad (2.8)$$

- b. Menghitung  $P_{ih}$  untuk membuat perkiraan awal pusat kelompok:

$$p_{ih} = \frac{d_{ih}}{\sum_{h=1}^n d_{ih}} \quad i = 1, \dots, n; \quad h = 1, \dots, n \quad (2.9)$$

- c. Menghitung  $\sum_{i=1}^n p_{ih}$  ( $h = 1, \dots, n$ ) pada masing-masing obyek dan diurutkan dari urutan paling kecil sampai yang paling besar. Pilih  $k$  obyek yang memiliki nilai minimum pada *initial medoid* kelompok.
- d. Tempatkan masing-masing obyek pada *medoid* yang paling dekat.
- e. Hitung nilai jumlah jarak optimal dari semua obyek pada *medoidnya*.

2. Mendapatkan *medoid* baru

Mempertimbangkan mengganti pasangan obyek  $(i, h)$ , di mana  $i \in$  obyek terpilih sebagai *medoid* dan  $h \in$  obyek data yang tidak terpilih sebagai *medoid*. *Swap* disimbolkan dengan  $i \leftrightarrow h$ . Lihat  $d(x_i, x_h)$  menjadi ukuran jarak antara dua obyek  $i$  dan  $h$ . Sekarang mempertimbangkan obyek tidak terpilih lain yaitu  $j$ .

Hitung  $T_{ih}$  "Total swap contribution" untuk  $i \leftrightarrow h$  sebagai berikut:

$$T_{ih} = \sum_j C_{jih} \quad (2.10)$$

Di mana  $C_{jih}$  adalah kontribusi untuk  $i \leftrightarrow h$  dari obyek  $j$  yang digambarkan di bawah ini:

Menurut Ng, R dan J.Han (1994) ada empat kemungkinan untuk mempertimbangkan ketika menghitung  $C_{jih}$ :

- a. Jika  $j$  masuk pada kelompok yang diwakili *medoid*  $i$ . Maka pertimbangkan jarak antara obyek  $j$  dan obyek  $h$ . jika  $j$  lebih



jauh dari  $h$  dibandingkan jarak  $j$  ke *medoid*  $j,2$ , kontribusi dari obyek  $j$  untuk *swap* adalah:

$$C_{jih} = d(x_j, x_{j,2}) - d(x_j, x_i) \quad (2.11)$$

Hasil pada  $i \leftrightarrow h$  maka obyek  $j$  sekarang menjadi kelompok  $j,2$ , seperti pada kasus 1. Selain itu jika  $j$  lebih dekat ke *medoid*  $h$  dibanding jarak  $j$  ke *medoid*  $j,2$ , maka kontribusi dari  $j$  untuk *swap* adalah:

$$C_{jih} = d(x_j, x_h) - d(x_j, x_{j,2}) \quad (2.12)$$

Hasil pada  $i \leftrightarrow h$  maka obyek  $j$  sekarang menjadi kelompok  $h$ , seperti pada kasus 2.

- b. Jika  $j$  masuk dalam kelompok  $j,2$ , dimana  $j,2 \neq i$ , periksa jarak antara obyek  $j$  dan obyek  $h$ . jika  $j$  lebih jauh dari  $h$  dibanding  $j$  ke *medoid*  $j,2$ , maka kontribusi dari obyek  $j$  *swap* adalah:

$$C_{jih} = 0 \quad (2.13)$$

Hasil pada  $i \leftrightarrow h$  maka obyek  $j$  tetap pada kelompok  $j,2$ , seperti pada kasus 3. Selain itu jika  $j$  lebih dekat  $h$  dibandingkan  $j$  ke  $j,2$ , kontribusi dari  $j$  untuk *swap* adalah:

$$C_{jih} = d(x_j, x_h) - d(x_j, x_{j,2}) \quad (2.14)$$

Hasil pada  $i \leftrightarrow h$  maka obyek  $j$  sekarang menjadi kelompok  $h$ , seperti pada kasus 4.

3. Lihat  $(i,h) = \text{argmin}_{i,h} T_{ih}$ . Jika  $T_{ih} < 0$  maka ganti  $i \leftrightarrow h$ . Sekarang obyek  $h \in$  obyek terpilih sebagai *medoid* dan  $i \in$  obyek tidak terpilih sebagai *medoid*. Selanjutnya kembali kelangkah dua, jika  $T_{ih} > 0$  masuk ke langkah empat.
4. Menempatkan masing-masing obyek pada *medoid* baru yang terdekat.

## 2.5. Metode Perbandingan *Adjusted Rand Index*

Dalam membandingkan hasil pengelompokan terhadap kriteria eksternal, diperlukan suatu pengukuran. Karena algoritma *K-mean* dan *partitional around medoid* termasuk dalam *hard clustering* dimana setiap objek harus termasuk ke kelompok tertentu, pengukuran antara dua partisi dapat digunakan, yaitu *Adjusted Rand index* (ARI). Nilai dari ARI berada anatar 0 dan 1, ketika

pengelompokan dapat memisahkan dengan sempurna maka nilai dari Rand index adalah 1 (Yeung dan Ruzzo, 2001).

*Adjusted Rand Index* yang diusulkan oleh Hubert dan Arabie, mengasumsikan Distribusi Hipergeometri sebagai model acak. Misalkan  $U$  adalah ukuran eksternal dan  $V$  adalah hasil pengelompokan, pemisahan  $U$  dan  $V$  diambil secara acak dengan banyak objek di dalam kelas dan kelompok sudah ditetapkan. Untuk  $n_{ij}$  menjadi angka objek pada kelas  $u_i$  dan kelompok  $v_j$  sedangkan  $n_i$  dan  $v_j$  masing-masing adalah angka objek dalam kelas kelas  $u_i$  dan kelompok  $v_j$ . Ilustrasi dari notasi tersebut adalah sebagai berikut:

Tabel 2.1 Ilustrasi notasi hasil pengelompokan

kelas/kelompok	$k_1$	$k_2$	...	$k_c$	jumlah
$u_1$	$n_{11}$	$n_{12}$	...	$n_{1c}$	$n_{1.}$
$u_2$	$n_{21}$	$n_{22}$	...	$n_{2c}$	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
$u_r$	$n_{r1}$	$n_{r2}$	...	$n_{rc}$	$n_{r.}$
jumlah	$n_{.1}$	$n_{.2}$	...	$n_{.c}$	$n_{..} = n$

Di bawah ini adalah rumus dari *Adjusted Rand Index* yang dapat ditulis sebagai berikut:

$$\frac{\sum_{i,j} \binom{n_{ij}}{2} - \left[ \sum_i \binom{n_i}{2} \sum_j \binom{n_j}{2} \right] / \binom{n}{2}}{\frac{1}{2} \left[ \sum_i \binom{n_i}{2} + \sum_j \binom{n_j}{2} \right] - \left[ \sum_i \binom{n_i}{2} \sum_j \binom{n_j}{2} \right] / \binom{n}{2}} \quad (2.17)$$

(Yeung, 2001)

### 2.6. Validasi Kelompok

Setelah obyek dimasukkan ke dalam kelompoknya masing-masing, maka selanjutnya dilakukan uji perbedaan dengan menggunakan analisis ANOVA (*Analysis of Variance*). Prosedur yang digunakan dalam analisis ANOVA ini adalah prosedur *one way ANOVA*. Alat uji ini untuk menguji apakah dua populasi atau lebih yang *independent*, memiliki rata-rata yang dianggap sama atau tidak sama. Teknik ANOVA akan menguji *variabilitas* dari *observasi* masing-masing kelompok dan *variabilitas* antar rata-rata kelompok.

Melalui kedua *variabilitas* tersebut, akan dapat ditarik kesimpulan mengenai rata-rata populasi (Ridho, 2006).

Uji statistik yang digunakan untuk menguji hipotesis nol bahwa semua kelompok mempunyai mean populasi yang sama adalah Uji F. Nilai F diperoleh dari rata-rata jumlah kuadrat (*mean square*) antar kelompok yang dibagi dengan rata-rata jumlah kuadrat dalam kelompok dengan rumus:

$$F = \frac{\hat{S}_B^2}{\hat{S}_W^2} \quad (2.16)$$

dengan derajat bebas  $a-1$  dan  $a(b-1)$

di mana:  $\hat{S}_B^2$  = ragam antar kelompok

$\hat{S}_W^2$  = ragam dalam kelompok

Hipotesis yang digunakan dalam pengujian ANOVA adalah:

$H_0$  : Seluruh kelompok dari rata-rata populasi adalah sama

$H_1$  : Seluruh kelompok dari rata-rata populasi adalah berbeda

Dasar dari pengambilan keputusan adalah:

- Jika  $F_{hitung} \geq F_{tabel} 0,05$ , maka  $H_0$  ditolak

- Jika  $F_{hitung} < F_{tabel} 0,05$ , maka  $H_0$  diterima

(Johnson dan Wichern, 2002)

Untuk menghitung nilai ragam dari semua data pada tiap kelompok dapat dihitung dengan menggunakan persamaan berikut:

$$S_k^2 = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (x_i - \bar{x}_i)^2 \quad (2.17)$$

di mana:  $S_i^2$  = ragam pada kelompok ke-k

$k$  = 1,2,...,c dimana  $c$  = jumlah kelompok

$n_k$  = jumlah data pada kelompok ke-k

$x_i$  = data ke-i pada suatu kelompok

$\bar{x}_i$  = rata-rata dari data pada suatu kelompok

Seadangkan ragam antar kelompok dapat dihitung menggunakan persamaan berikut:

$$S_B^2 = \frac{1}{k-1} \sum_{k=1}^c n_k (\bar{x}_k - \bar{x})^2 \quad (2.18)$$

di mana:  $S_B^2$  = ragam antar kelompok

$\bar{x}_k$  = rata-rata kelompok ke-k  
 $\bar{x}$  = rata-rata dari  $\bar{x}_k$

Ragam di dalam kelompok dapat dihitung dengan menggunakan persamaan berikut:

$$S_w^2 = \frac{1}{n-k} \sum_{k=1}^c (n_k - 1) S_k^2 \quad (2.19)$$

di mana:  $S_w^2$  = ragam dalam kelompok  
 $n$  = jumlah semua data

(Ridho, 2006)

## 2.7. Analisis Deskriptif

Metode statistika merupakan prosedur-prosedur yang digunakan dalam pengumpulan, penyajian, analisis, dan penafsiran data (Yitnosumarto, 1990). Dalam metode statistika terdapat beberapa prosedur yang digunakan dalam pengumpulan, penyajian, analisis, dan penafsiran data. Kita dapat mengelompokkan metode-metode tersebut ke dalam dua kelompok besar, yaitu statistika deskriptif dan statistika inferensia. Statistika deskriptif adalah metode-metode yang berkaitan dengan pengumpulan dan penyajian suatu gugus data sehingga memberikan informasi yang berguna (Walpole, 1995). Dalam statistika deskriptif kita tidak tertarik untuk membuat kesimpulan yang melibatkan analisis di dalamnya. Kita hanya sekedar menyederhanakan dan menata data untuk memperoleh gambaran secara keseluruhan dari peubah atau karakteristik yang dipelajari atau diamati (Yitnosumarto, 1990).

Analisis deskriptif adalah analisis yang dilakukan dengan mengamati perilaku tiap kelas yang telah tercapai, dimana tiap kelas tersebut pasti memiliki suatu perilaku tertentu yang membedakan dirinya dengan kelompok yang lainnya (Ridho, 2006)

## BAB III METODE PENELITIAN

### 3.1. Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang terdiri dari tiga data seperti disajikan pada Tabel 3.1

Tabel 3.1 Variabel yang terlibat untuk setiap penelitian

No	Sumber Data	Data	Variabel
1	Skripsi Syarifuddin. H (2003)	Rasio Keuangan BPR di Daerah Tingkat II Badung Propinsi Bali	$X_1$ = Rasio KAP $X_2$ = Rasio Rentabilitas $X_3$ = Rasio Likuiditas
2	Skripsi Dian A Zuroida (2004)	Faktor-faktor yang mempengaruhi penetasan telur burung walet	$X_1$ = Berat telur $X_2$ = Panjang telur $X_3$ = Lebar telur $X_4$ = Suhu pengeraman
3	<a href="ftp://ftp.ics.uci.edu/pub/machine-learning-databases">ftp://ftp.ics.uci.edu/ pub/machine- learning-databases</a>	Protein localization sites	$X_1$ = MCG $X_2$ = GUH $X_3$ = AAC $X_4$ = ALM 1 $X_5$ = ALM 2

### 3.2. Metode Penelitian

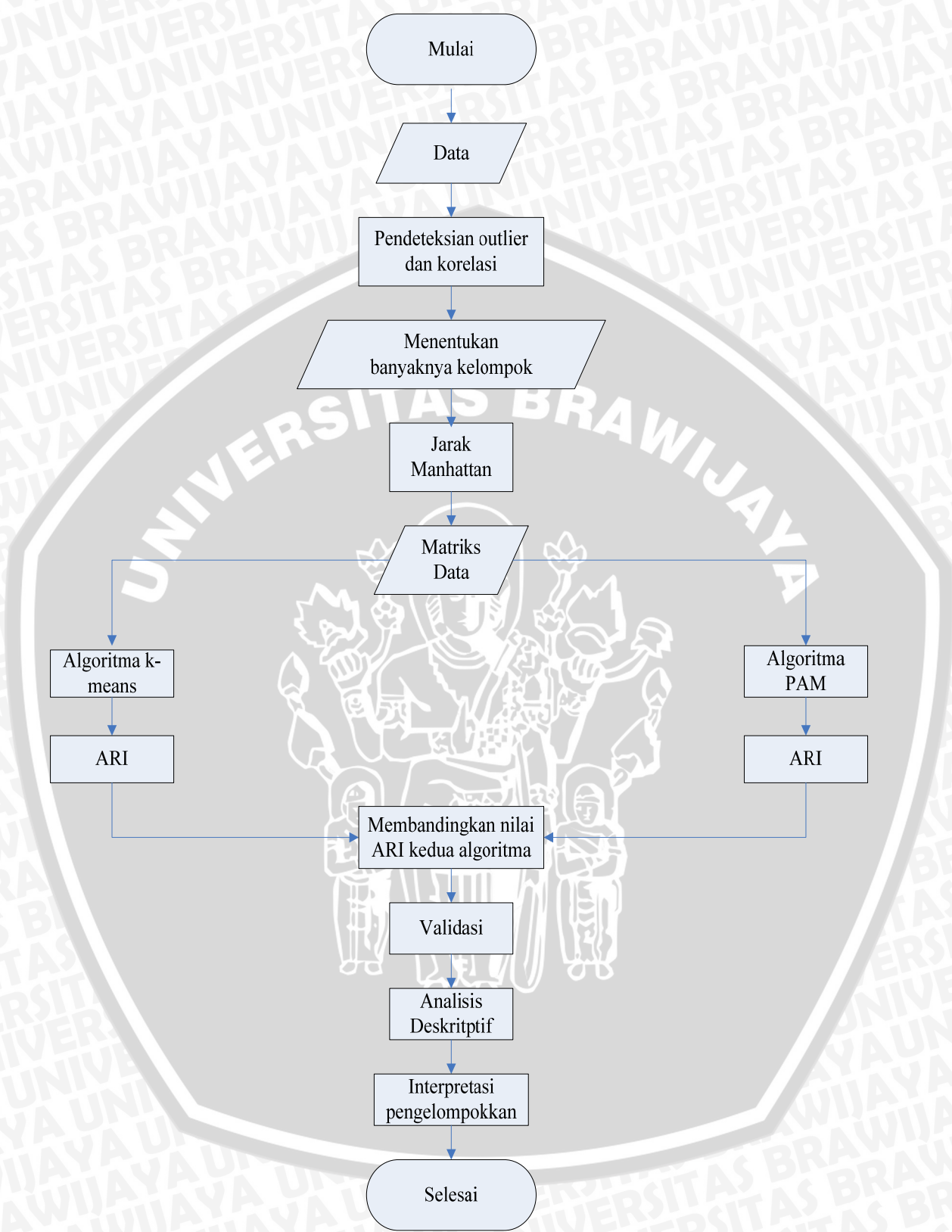
Langkah-langkah analisis pada penelitian ini adalah sebagai berikut:

1. Pendeteksian *outlier* menggunakan persamaan 2.1.
2. Pendeteksian korelasi menggunakan persamaan 2.2.
3. Menentukan banyaknya kelompok yang akan dibentuk.
4. Membentuk matriks ukuran jarak manhattan menggunakan persamaan 2.4.
5. Mencari *centroid* menggunakan algoritma *K-means* dan *partitioning around medoid*.

6. Melakukan alokasi pengelompokan data.
7. Menghitung nilai *Adjusted Rand Index* (ARI) kedua algoritma berdasarkan persamaan 2.15.
8. Menentukan *centroid* terbaik dengan membandingkan nilai dari *Adjusted Rand Index* kedua algoritma.
9. Melakukan validasi terhadap kelompok yang sudah terbentuk dengan persamaan 2.16.
10. Analisis Deskriptif untuk masing-masing kelompok yaitu menghitung jumlah obyek dan menghitung rata-rata tiap variabel..
11. Melakukan interpretasi karakteristik pengelompokan optimal yang terbentuk antara algoritma *K-means* dan *partitional around medoid*.

Bantuan paket program statistika yang digunakan untuk perhitungan dalam analisis ini adalah SPSS, SYSTAT dan S-PLUS. Diagram langkah-langkah perbandingan antara algoritma *k-means* dan *partitioning around medoid* disajikan pada Gambar 3.1.





Gambar 3.1 Langkah-langkah Analisis





## BAB IV HASIL DAN PEMBAHASAN

### 4.1. Deskripsi Data

#### 4.1.1. Data 1

Data I merupakan data sekunder yang digunakan untuk melakukan kajian kinerja pada Bank Perkreditan Rakyat di Daerah Tingkat II Badung propinsi Bali. Penelitian ini bertujuan untuk mengetahui apakah status bank dalam keadaan baik atau tidak. Pada penelitian ini digunakan 72 BPR yang berada di Daerah Badung dan dipilih secara acak..

Adapun variabel yang diamati meliputi:

$X_1$  = Rasio KAP (Kualitas Aktiva Produktif) yaitu rasio yang digunakan untuk menilai kondisi aktiva produktif dan kecukupan manajemen resiko kredit.

$$\text{Rasio KAP} = \frac{\text{Aktiva produktif yang diklasifikasikan}}{\text{Aktifa produktif}} \times 100\%$$

$X_2$  = Rasio Rentabilitas atau ROA (Return on Asset) yaitu rasio yang digunakan untuk mengukur tingkat efisiensi usaha dan probabilitas yang dicapai oleh perusahaan yang bersangkutan

$$\text{ROA} = \frac{\text{Laba sebelum pajak dalam 12 bulan terakhir}}{\text{Total aktiva}} \times 100\%$$

$X_3$  = Rasio Likuiditas atau LDR ( Loan to Deposit) yaitu rasio yang digunakan untuk mengetahui kemampuan perusahaan untuk memenuhi kewajiban sewaktu-waktu.

$$\text{LDR} = \frac{\text{Kredit}}{\text{Dana yang diterima}} \times 100\%$$

Setelah didapatkan nilai dari masing-masing variabel kemudian dibandingkan nilainya dengan Tabel 4.1

Tabel 4.1 Faktor-faktor penilaian, bobot dan batasan predikat

Faktor yang dinilai	Bobot	Batats kredit predikit tingkat kesehatan faktor	
		Sehat	Tidak Sehat
KAP	50%	50% - 33%	0% < 35%
Rentabilitas	25%	25% - 20%	0% < 20%
Likuiditas	25%	25% - 20%	0% < 20%

#### 4.1.2. Data II

Usaha untuk pengembangbiakan burung walet belum mendapatkan perhatian yang serius baik dari peneliti maupun dari para peternak burung walet. Hal ini ditunjukkan dengan sangat terbatasnya penelitian tentang burung walet. Pembibitan yang tidak didasarkan teori dan program yang tepat maka pengembakbiakan tidak akan berjalan dengan baik.

Data yang digunakan pada penelitian ini bertujuan untuk mengetahui faktor-faktor yang mempengaruhi penetasan telur burung walet dengan menggunakan analisis regresi. Adapun variabel-variabel yang akan diteliti ada empat yaitu berat telur, panjang telur, lebar telur dan suhu pengeraman. Selain keempat variabel tersebut diusahakan kondisinya dalam keadaan homogen. Dimana penetasan dilakukan di dalam mesin tetas dengan suhu tertentu, dan kelembapan serta sirkulasi udara yang dikondisikan homogen. Telur yang digunakan pada penelitian ini sebanyak 60 butir dari 60 sarang burung walet yang ada. Karena musim kawin dari burung walet pada saat musim hujan maka penelitian ini dilakukan pada musim hujan hal ini dikarenakan pada saat musim hujan populasi serangga tengah melimpah sebagai sumber makanan burung walet.

#### 4.1.3. Data III

*Protein localization prediction* melibatkan perhitungan prediksi di mana suatu protein terletak di suatu sel. Identifikasi mengenai lokasi protein merupakan suatu komponen yang sangat penting karena dapat membantu identifikasi target.

Lokasi protein dalam sel terutama ditentukan oleh urutan *asam amino*. Sistem ini dapat menghasilkan lokasi urutan protein dengan ketelitian yang baik. Adapun faktor-faktor yang telah ditentukan sesuai aturan yaitu:

- 1.MCG yaitu hasil dari analisis diskriminan untuk pengenalan urutan isyarat yang dihasilkan oleh metode McGeoch's.
- 2.GVH yaitu suatu skor yang dihasilkan oleh metode Von Jeine's untuk pengenalan urutan isyarat dengan menggunakan weight matriks.
- 3.AAC yaitu nilai *asam amino* yang terkandung dalam membran luar.

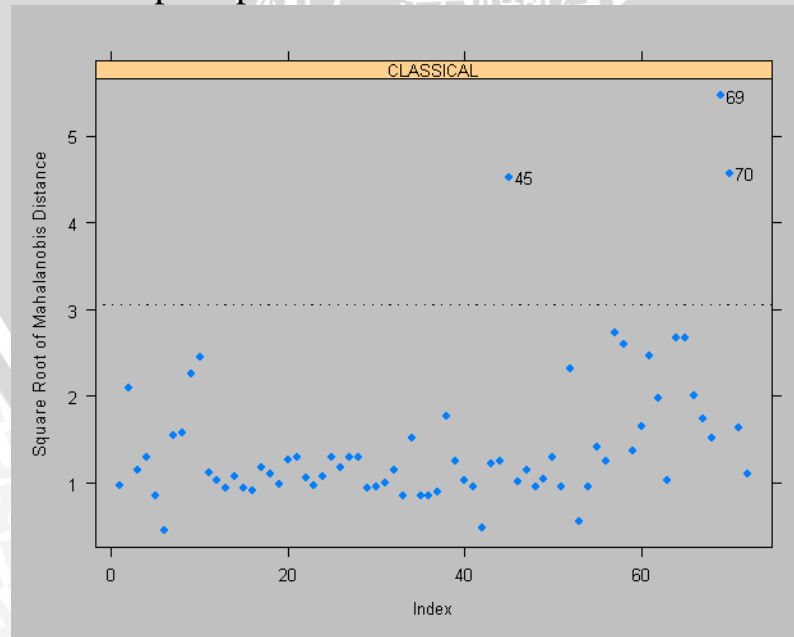
4. ALM 1 yaitu nilai output dari program ALOM untuk identifikasi disekeliling daerah *membran* untuk keseluruhan urutan isyarat.

5. ALM 2 yaitu nilai output dari program ALOM untuk identifikasi yang tidak terdapat disekeliling daerah *membran* dengan metode Von Heijne's.

Untuk melakukan klasifikasi dalam penelitian ini digunakan bentuk *bayesian network* di mana pada penelitian ini menggunakan sel *Ecoli* dan dilaksanakan di dalam laboratorium yang *steril* agar tidak terkontaminasi dengan lingkungan luar. Dalam penelitian ini digunakan 30 sel *Ecoli*.

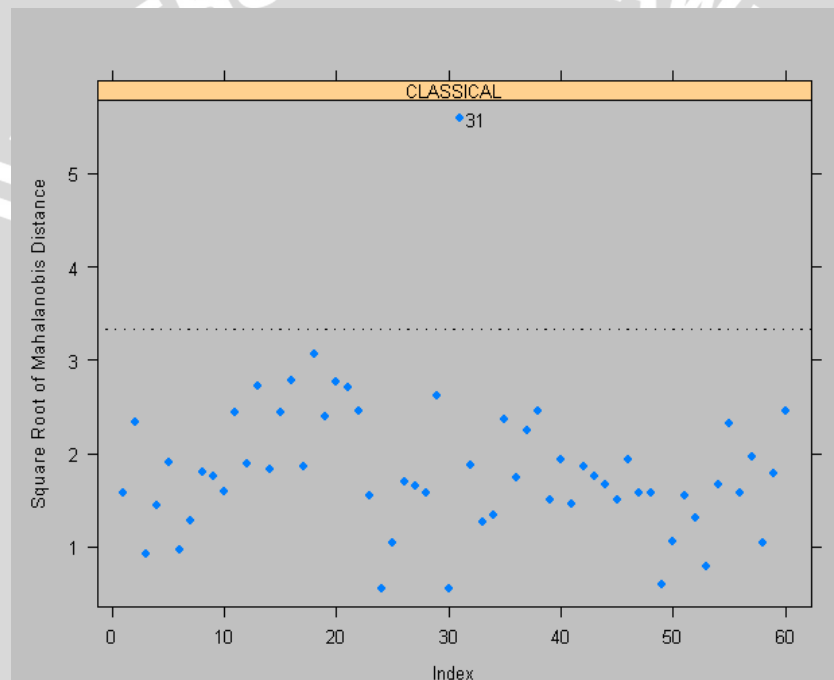
#### 4.2. Pendeteksian *Outlier*

*Outlier* merupakan suatu keganjilan dan menandakan suatu titik pengamatan yang sama sekali berbeda dengan pengamatan lain (Marazzi, 1993). Mengabaikan *outlier* dari suatu himpunan data bukanlah tindakan yang tepat, karena *outlier* dapat memberikan pengaruh terhadap hasil analisis secara drastis. Deteksi *outlier* pada data dapat dilihat dari nilai jarak *Mahalanobis* yang dibandingkan dengan nilai  $\chi^2_{p,(1-\alpha)}$  dan dapat juga di gambarkan pada sebuah grafik di mana titik-titik yang berada di luar batas diidentifikasi sebagai *outlier*. Seperti pada Gambar 4.1 dibawah ini:



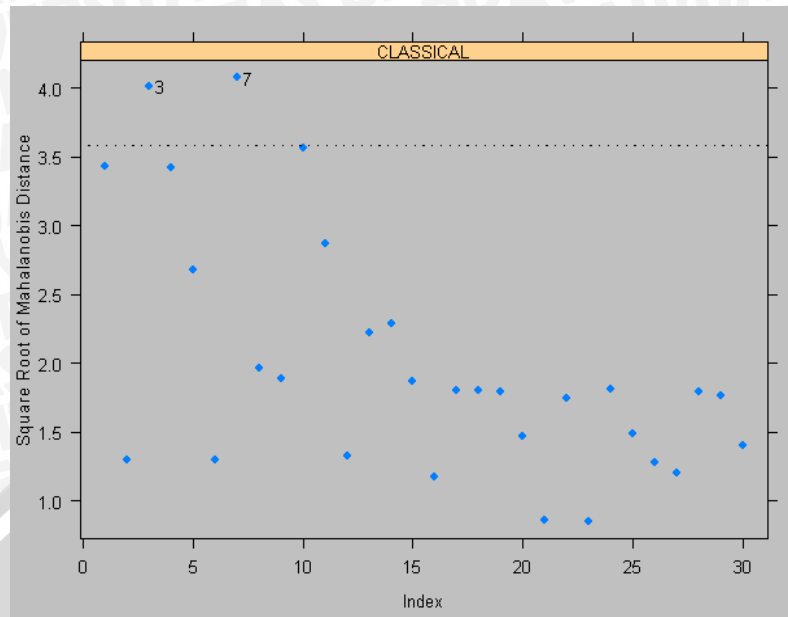
Gambar 4.1 Plot Jarak *Mahalanobis* Data I

Pada Gambar 4.1 hasil analisis deteksi *outlier* untuk Data I terlihat bahwa terdapat 3 titik berada di luar batas. Hal ini mengindikasikan bahwa terdapat 3 *outlier* dari 72 data pengamatan. Untuk jarak *Mahalanobis* setiap pengamatan dapat dilihat pada Lampiran 4 di mana pengamatan yang memiliki nilai lebih besar dari  $\chi^2_{3,0.975} = 9,348$  dideteksi sebagai *outlier*. Sedangkan pada Gambar 4.2 hasil analisis deteksi *outlier* Data II terlihat bahwa 1 titik berada di luar batas. Jadi terdapat 1 *outlier* dari 60 data pengamatan. Pada Lampiran 4 dapat dilihat Jarak *Mahalanobis* untuk setiap pengamatan di mana pengamatan yang memiliki nilai lebih besar dari  $\chi^2_{4,0.975} = 11,14$  dideteksi sebagai *outlier*.



Gambar 4.2 Plot Jarak *Mahalanobis* Data II

Dari hasil plot jarak *Mahalanobis* pada Data III terdapat 2 obyek yang berada diluar batas garis Hal ini mengindikasikan bahwa terdapat 2 *outlier* dari 30 data pengamatan. Jarak *Mahalanobis* untuk setiap pengamatan dapat dilihat pada Lampiran 4 di mana pengamatan yang memiliki nilai lebih besar dari  $\chi^2_{5,0.975} = 12,83$  dideteksi sebagai *outlier*. Plot Jarak *Mahalanobis* di tunjukkan pada Gambar 4.3 di bawah ini:



Gambar 4.3 Plot Jarak *Mahalanobis* Data III

### 4.3. Pendeteksian Korelasi

Uji korelasi pada data perlu dilakukan untuk menguji apakah terdapat korelasi atau tidak. Penggunaan jarak *Manhattan* pada algoritma *k-means* dan PAM mensyaratkan tidak adanya korelasi pada data. Hipotesis nol ( $H_0$ ) menyatakan bahwa tidak terdapat korelasi yang signifikan antar antar peubah yang diuji, sedangkan Hipotesis tandingan ( $H_1$ ) menyatakan bahwa terdapat korelasi yang signifikan antar peubah yang diuji. Besar nilai  $\alpha$  yang digunakan sebagai pembanding bagi *p-value* adalah 5 %. Sehingga bila *p-value*  $\geq \alpha$  menghasilkan kesimpulan terima  $H_0$  dan bila *p-value*  $< \alpha$  menghasilkan kesimpulan tolak  $H_0$ . Dari hasil output SPSS yang terdapat pada lampiran 5, didapatkan nilai matrik korelasi pearson dan *p-value* untuk 72 obyek pada Data I seperti di bawah ini:

Tabel 4.2 Nilai korelasi antar peubah dan *P-value* pada Data I

	KAP ( $X_1$ )	Rentabilitas ( $X_2$ )	Likuiditas ( $X_3$ )
KAP ( $X_1$ )	1,000 (0,000)	-0,021 (0,863)	0,146 (0,221)
Rentabilitas ( $X_2$ )	-0,021 (0,863)	1,000 (0,000)	-0,184 (0,123)
Likuiditas ( $X_3$ )	0,146 (0,221)	-0,184 (0,123)	1,000 (0,000)

Keterangan : Nilai dalam kurung merupakan nilai *p-value*

Berdasarkan Tabel 4.2 menunjukkan bahwa *p-value* untuk peubah Rasio KAP ( $X_1$ ) pada peubah Rasio Rentabilitas ( $X_2$ ) dan Rasio Likuiditas ( $X_3$ ) masing-masing memiliki nilai 0,863 dan 0,221. *P-value* tersebut lebih besar dari nilai  $\alpha$  sebesar 5 % sehingga menunjukkan tidak ada korelasi yang signifikan antara peubah KAP ( $X_1$ ) dengan peubah yang lain. Begitu juga pada peubah Rasio Rentabilitas ( $X_2$ ) pada peubah Rasio Likuiditas ( $X_3$ ) yang memiliki *p-value* sebesar 0,123 yang lebih besar dari nilai  $\alpha$  sebesar 5%, sehingga menunjukkan tidak ada korelasi untuk semua peubah. Jadi dapat disimpulkan bahwa tidak terdapat korelasi antar peubah pada Data I, sehingga langsung bisa dilakukan analisis kelompok.

Untuk Data II seperti yang terdapat pada Lampiran 5, didapatkan nilai matrik korelasi pearson untuk 60 obyek pada data II seperti pada Tabel 4.3 di bawah ini :

Tabel 4.3 Nilai korelasi antar peubah dan *P-value* pada Data II

	$X_1$	$X_2$	$X_3$	$X_4$
$X_1$	1,000 (0,000)	0,139 (0,289)	-0,080 (0,542)	0,241 (0,063)
$X_2$	0,139 (0,289)	1,000 (0,000)	0,015 (0,912)	0,134 (0,306)
$X_3$	-0,080 (0,542)	0,015 (0,912)	1,000 (0,000)	0,210 (0,107)
$X_4$	0,241 (0,063)	0,134 (0,306)	0,210 (0,107)	1,000 (0,000)

Keterangan : Nilai dalam kurung merupakan nilai *p-value*

Berdasarkan Tabel 4.3 di atas dapat dilihat bahwa untuk semua peubah pada Data II yaitu peubah  $x_1$  (berat telur),  $x_2$  (panjang telur),  $x_3$  (lebar telur),  $x_4$  (suhu pengeraman) mempunyai *p-value* yang lebih besar dari  $\alpha$  sebesar 5 % sehingga menunjukkan tidak ada korelasi yang signifikan antar semua peubah pada Data II. Sedangkan untuk Data III seperti yang terdapat pada lampiran 5, didapatkan nilai matrik korelasi pearson untuk 30 obyek pada data III seperti pada Tabel 4.4 di bawah. Pada Data III dapat disimpulkan bahwa tidak terdapat korelasi yang signifikan, hal ini dapat dilihat dari Tabel 4.4 dimana nilai *p-value* untuk peubah  $x_1$  (MCG),  $x_2$  (GVH),  $x_3$  (AAC),

$x_4$  (ALM 1),  $x_5$  (ALM 2) semuanya bernilai lebih besar dari  $\alpha$  sebesar 5 %.

Tabel 4.4 Nilai korelasi antar peubah dan *P-value* pada Data III

	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$
$X_1$	1,000 (0,000)	0,309 (0,097)	-0,031 (0,870)	0,004 (0,985)	0,090 (0,635)
$X_2$	0,309 (0,097)	1,000 (0,000)	0,109 (0,566)	0,302 (0,104)	-0,094 (0,621)
$X_3$	-0,031 (0,870)	0,109 (0,566)	1,000 (0,000)	-0,316 (0,089)	0,212 (0,216)
$X_4$	0,004 (0,985)	0,302 (0,104)	-0,316 (0,089)	1,000 (0,000)	0,115 (0,545)
$X_5$	0,090 (0,635)	-0,094 (0,621)	0,212 (0,216)	0,115 (0,545)	1,000 (0,000)

Keterangan : Nilai dalam kurung merupakan nilai *p-value*

#### 4.4. Analisis Kelompok Non-hierarki

Pada analisis kelompok non-hierarki banyaknya kelompok ditentukan sebelum dilakukan analisis ( Hair *et al*, 1998). Pada Data I dan Data II telah ditentukan sebanyak 2 kelompok yang ingin dibentuk, sedangkan pada Data III banyaknya kelompok yang ingin dibentuk sebanyak 3 kelompok Hasil pengelompokan dengan algoritma *k-means* dan *partitioning around medoid* pada Lampiran 6 dan 7.

##### 4.4.1. Algoritma *K-means*

Dari hasil analisis kelompok diperoleh hasil yaitu banyaknya anggota kelompok dan *centroid* (titik pusat) kelompok pada Data I. Setelah didapatkan *centroid* selanjutnya menguji masing-masing komponen di dalam populasi data dan menandai komponen tersebut ke salah satu pusat kelompok yang telah didefinisikan tergantung dari jarak minimum antar komponen dengan tiap-tiap *centroid* kelompok (Matteua, 2008). Berdasarkan Lampiran 6 didapatkan hasil dari analisis algoritma *k-means* untuk kelompok 1 beranggotakan 56 obyek sedangkan kelompok 2 beranggotakan 16 obyek. Serta didapatkan nilai dari *centroid* untuk masing-masing kelompok seperti pada Tabel 4.5 di bawah ini

Tabel 4.5 *Centroid* kelompok algoritma *k-means* pada Data I

variabel	Kelompok	
	1	2
X <sub>1</sub>	26,65	12,08
X <sub>2</sub>	7,40	7,66
X <sub>3</sub>	9,95	9,66

Selanjutnya untuk hasil analisis kelompok pada Data II dengan metode algoritma *k-means* didapatkan nilai *centroid* seperti pada tabel 4.6.

Tabel 4.6 *Centroid* kelompok algoritma *k-means* pada Data II

variabel	Kelompok	
	1	2
X <sub>1</sub>	-0,62	0,95
X <sub>2</sub>	-0,25	0,49
X <sub>3</sub>	-0,06	0,02
X <sub>4</sub>	-0,46	0,63

Pengalokasian objek pada masing-masing kelompok dapat dilihat pada Lampiran 6 di mana pada Data II didapatkan hasil untuk kelompok 1 beranggotakan 37 obyek dan untuk kelompok 2 terdiri dari 23 obyek. Sedangkan hasil analisis kelompok pada data III dengan metode algoritma *k-means* didapatkan nilai *centroid* seperti pada Tabel 4.7. Sedangkan pengalokasian objek pada masing-masing kelompok yaitu dengan mendapatkan jarak minimum setiap obyek dengan tiap-tiap *centroid* kelompok dapat dilihat pada Lampiran 6. Di mana untuk kelompok 1 beranggotakan 17 obyek, kelompok 2 beranggotakan 5 objek sedangkan untuk kelompok 3 terdiri dari 8 obyek.

Tabel 4.7 *Centroid* kelompok algoritma *k-means* pada Data III

Variabel	Kelompok		
	1	2	3
X <sub>1</sub>	0,38	0,27	0,67
X <sub>2</sub>	0,41	0,51	0,62
X <sub>3</sub>	0,47	0,33	0,50
X <sub>4</sub>	0,35	0,76	0,52
X <sub>5</sub>	0,44	0,33	0,45



#### 4.4.2. Algoritma *Partitioning Around Medoid*

Dari hasil analisis pada Data I didapatkan hasil yaitu banyaknya anggota kelompok dan *centroid* (titik pusat) kelompok. Sama halnya dengan algoritma *k-means*, pada algoritma PAM setelah didapatkan *centroid* selanjutnya menandai tiap komponen tersebut ke salah satu pusat kelompok yang telah didefinisikan tergantung dari jarak minimum antar komponen dengan tiap-tiap *centroid* kelompok. Pada Data I hasil dari algoritma PAM dapat dilihat pada tabel 4.8 di bawah ini:

Tabel 4.8 *Centroid* kelompok algoritma PAM pada Data I

Variabel	Kelompok	
	1	2
$X_1$	27,48	14,21
$X_2$	9,17	6,86
$X_3$	10,00	10,00

Pengalokasian masing-masing obyek ke dalam kelompok dapat dilihat pada Lampiran 7. Berdasarkan Lampiran 7 didapatkan hasil dari analisis PAM untuk kelompok 1 beranggotakan 54 obyek sedangkan kelompok 2 beranggotakan 18 obyek. Pada Data II diperoleh hasil banyaknya anggota kelompok dan *centroid* kelompok seperti pada Tabel 4.9 yang menyajikan nilai *centorid* yang dihasilkan dengan algoritma PAM .

Tabel 4.9 *Centroid* kelompok algoritma PAM pada Data II

variabel	Kelompok	
	1	2
$X_1$	-0,49	1,21
$X_2$	-0,25	0,81
$X_3$	0,45	0,45
$X_4$	-0,66	0,50

Anggota masing-masing kelompok dapat dilihat pada Lampiran 7. Untuk kelompok 1 beranggotakan 38 sedangkan untuk kelompok 2 beranggotakan 22 obyek. Untuk Data III diperoleh hasil banyaknya anggota kelompok dan *centroid* kelompok. Tabel 4.10 menyajikan nilai *centorid* yang dihasilkan dengan algoritma PAM.

Tabel 4.10 *Centroid* kelompok algoritma PAM pada Data III

Variabel	Kelompok		
	1	2	3
$X_1$	0,30	0,41	0,70
$X_2$	0,51	0,43	0,61
$X_3$	0,42	0,45	0,56
$X_4$	0,61	0,31	0,52
$X_5$	0,34	0,41	0,43

Sedangkan pengalokasian obyek pada masing-masing kelompok yaitu dengan mendapatkan jarak minimum setiap obyek dengan tiap-tiap *centroid* kelompok dapat dilihat pada Lampiran 7. Untuk kelompok 1 beranggotakan 5 obyek, kelompok 2 beranggotakan 7 obyek dan untuk kelompok 3 beranggotakan 18 obyek.

#### 4.5. Uji Perbandingan *Adjusted Rand Index*

Dalam membandingkan hasil pengelompokan terhadap kriteria eksternal, diperlukan suatu pengukuran. Karena algoritma *K-mean* dan *partitional around medoid* termasuk dalam *hard clustering* dimana setiap objek harus termasuk ke kelompok tertentu, pengukuran antara dua partisi dapat digunakan yaitu *Adjusted Rand index* (ARI). Nilai dari *Adjusted Rand index* berada anatar 0 dan 1, ketika pengelompokan dapat memisahkan dengan sempurna maka nilai dari *Adjusted Rand index* adalah 1 (Yeung dan Ruzzo, 2001). Nilai ARI untuk semua Data pada algoritma *k-means* dan PAM adalah sebagai berikut:

Tabel 4.11 Perbandingan Nilai ARI untuk semua Data

Data	<i>K-means</i>	PAM
I	0,7680	0,8805
II	0,7340	0,8049
III	0,7292	0,8384

Berdasarkan Tabel 4.11 nilai *Adjusted Rand Index* yang dihasilkan dari semua Data pada algoritma *k-means* lebih kecil dibandingkan nilai *Adjusted Rand Index* yang dihasilkan oleh algoritma PAM, sehingga dapat disimpulkan bahwa *centroid* yang baik untuk Data I adalah dengan menggunakan algoritma PAM.

#### 4.6. Validasi Kelompok

Tahap validasi menyangkut usaha analisis untuk mendapat keyakinan bahwa kelompok yang terbentuk benar-benar berbeda (Ridho, 2006). Setelah obyek dimasukkan ke dalam kelompoknya masing-masing, maka selanjutnya dilakukan uji perbedaan dengan menggunakan analisis ANOVA. Hipotesis nol ( $H_0$ ) menyatakan bahwa tidak terdapat perbedaan yang signifikan antar kelompok untuk tiap-tiap nilai rata-rata peubah yang diuji, sedangkan Hipotesis tandingan ( $H_1$ ) menyatakan bahwa terdapat perbedaan yang signifikan antar kelompok untuk tiap-tiap nilai rata-rata peubah yang diuji. Besar nilai  $\alpha$  yang digunakan sebagai pembanding bagi *p-value* adalah 5 %. Sehingga bila *p-value*  $\geq \alpha$  menghasilkan kesimpulan terima  $H_0$  dan bila *p-value*  $< \alpha$  menghasilkan kesimpulan tolak  $H_0$ . *P-value* untuk pengujian beda masing-masing peubah pada pembentukan kelompok Data I terdapat pada Tabel 4.12 dan hasil dari ANOVA dengan bantuan SPSS 11 terdapat pada Lampiran 8.

Tabel 4.12 Uji signifikansi peubah antar kelompok Data I

	$X_1$	$X_2$	$X_3$
<i>p-value</i>	0,000	0,556	0,075

Berdasarkan Tabel 4.12 *p-value* untuk peubah Rasio Rentabilitas ( $X_2$ ) dan Rasio Likuiditas ( $X_4$ ) sebesar 0,556 dan 0,075 yang lebih besar dari nilai  $\alpha$  sebesar 5 %, sehingga disimpulkan terima  $H_0$  yang berarti nilai rata-rata untuk kedua peubah tersebut sama pada tiap-tiap kelompok. Sedangkan *p-value* untuk peubah Rasio KAP ( $X_1$ ) sebesar 0,000 yang lebih kecil dari nilai  $\alpha$  sebesar 5 % yang dapat disimpulkan tolak  $H_0$  yang berarti nilai rata-rata peubah KAP pada tiap-tiap kelompok mempunyai perbedaan yang signifikan. Jadi pembentukan 2 kelompok pada Data 1 benar-benar berbeda. Untuk Data II hasil nilai dari ANOVA untuk pengujian beda masing-masing peubah pada pembentukan kelompok dapat dilihat pada Tabel 4.13 di mana nilai *p-value* dari peubah Berat telur ( $X_1$ ), Panjang telur ( $X_2$ ) dan Suhu pengeraman ( $X_4$ ) masing-masing sebesar 0,000; 0,004 dan 0,000 yang lebih kecil dari nilai  $\alpha$  sebesar 5 %, sehingga dapat disimpulkan tolak  $H_0$  yang berarti rata-rata untuk ketiga peubah tersebut berbeda untuk untuk tiap-tiap kelompok. Sedangkan untuk *p-value*  $X_3$  atau peubah Lebar telur sebesar 0,569

yang lebih besar dari nilai  $\alpha$  sebesar 5 % yang dapat disimpulkan terima  $H_0$  yang berarti rata peubah  $X_3$  atau lebar telur untuk tiap-tiap kelompok sama. Sehingga kelompok yang terbentuk, dibedakan oleh peubah berat telur, panjang telur dan suhu pengeraman

Tabel 4.13 Uji signifikansi peubah antar kelompok Data II

	$X_1$	$X_2$	$X_3$	$X_4$
<i>P-value</i>	0,000	0,004	0,569	0,000

Untuk Data III hasil nilai dari ANOVA untuk pengujian beda masing-masing peubah pada pembentukan kelompok dapat dilihat pada Tabel 4.14 dan hasil pengujian hipotesis terdapat pada lampiran 8.

Tabel 4.14 Uji signifikansi peubah antar kelompok Data III

	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$
<i>p-value</i>	0,000	0,000	0,017	0,000	0,121

Nilai *p-value* dari peubah MCG ( $X_1$ ), GVH ( $X_2$ ), AAC ( $X_3$ ) dan ALM 1 ( $X_4$ ) mempunyai nilai yang lebih kecil dari nilai  $\alpha$  sebesar 5 % , sehingga disimpulkan tolak  $H_0$  yang berarti nilai rata-rata untuk keempat peubah tersebut berbeda pada tiap-tiap kelompok, sedangkan untuk peubah ALM 2 ( $X_5$ ) mempunyai nilai *p-value* yang lebih besar dari nilai  $\alpha$  sebesar 5 % yang dapat disimpulkan terima  $H_0$  yang berarti kedua nilai rata-rata produksi untuk kedua peubah tersebut berbeda pada tiap-tiap kelompok, sehingga kelompok-kelompok yang terbentuk benar-benar berbeda antara kelompok yang satu dengan kelompok yang lainnya, dimana perbedaan tersebut ditentukan oleh peubah  $X_1$ ,  $X_2$ ,  $X_3$  dan  $X_4$ .

#### 4.7. Interpretasi Karakteristik Pengelompokan

Interpretasi karakteristik pengelompokan dengan menggunakan analisis deskriptif yang dilakukan dengan mengamati perilaku tiap kelompok yang telah dihasilkan, di mana tiap kelompok tersebut pasti memiliki suatu perilaku tertentu yang membedakan kelompok tersebut dengan kelompok yang lainnya. Tabel 4.15 menyajikan rata-rata kelompok yang terbentuk untuk Data I berdasarkan hasil analisis deskriptif dengan SPSS yang terdapat pada Lampiran 9. Nilai rata-rata untuk tiap-tiap peubah sebagai berikut:

Tabel 4.15 Rata-rata masing-masing kelompok pada Data I

Kelompok	N	$X_1$	$X_2$	$X_3$
Kelompok 1	54	29,9	7,58	9,95
Kelompok 2	18	13,00	7,09	9,70

Berdasarkan Tabel 4.15 kelompok 1 yang merupakan kelompok Bank dengan status baik beranggotakan objek data sebanyak 54 objek sedangkan kelompok 2 yang merupakan kelompok Bank dengan status tidak baik beranggotakan 18 objek. Pada kelompok Bank dengan status baik Nilai Rata-rata peubah Rasio KAP ( $X_1$ ), Rasio Rentabilitas ( $X_2$ ) dan Rasio likuiditas ( $X_3$ ) lebih besar dibandingkan dengan Nilai Rata-rata ketiga peubah pada kelompok Bank dengan status tidak baik. Tabel 4.16 menyajikan rata-rata kelompok yang terbentuk untuk Data II berdasarkan hasil analisis deskriptif dengan SPSS yang terdapat pada Lampiran 9 sehingga dapat dilakukan interpretasi untuk masing-masing kelompok yang sudah terbentuk.

Tabel 4.16 Rata-rata masing-masing kelompok pada Data II

Kelompok	N	$X_1$	$X_2$	$X_3$	$X_4$
Kelompok 1	38	2,05	2,00	1,36	33,37
Kelompok 2	22	2,86	2,07	1,38	35,27

Pada Data II obyek-obyek yang ada dikelompokkan menjadi dua kelompok di mana kelompok 1 terdiri dari 38 obyek sedangkan kelompok 2 terdiri dari 22 obyek. Rata-rata untuk peubah panjang telur ( $X_2$ ) dan lebar telur ( $X_3$ ) pada kelompok 1 hampir sama dengan nilai rata-rata kedua peubah pada kelompok 2. Sedangkan rata-rata nilai peubah berat telur ( $X_1$ ) dan suhu pengeraman ( $X_4$ ) lebih tinggi pada kelompok 1 yaitu kelompok telur yang menetas dibandingkan dengan kelompok 2 yaitu kelompok telur yang tidak menetas. Tabel 4.16 menyajikan rata-rata peubah untuk masing masing kelompok, sedangkan hasil analisis deskriptif dengan SPSS dapat dilihat pada Lampiran 9

Tabel 4.17 Rata-rata masing-masing kelompok pada Data III

Kelompok	N	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$
Kelompok 1	5	0,27	0,51	0,33	0,76	0,33
Kelompok 2	18	0,39	0,42	0,47	0,36	0,44
Kelompok 3	7	0,68	0,64	0,50	0,52	0,46

Tabel 4.17 menyatakan hasil dari analisis kelompok pada Data III di mana kelompok 1 beranggotakan 5 obyek, kelompok 2 beranggotakan 18 objek sedangkan kelompok 3 beranggotakan 7 obyek. Pada kelompok 3 yang merupakan kelompok Perisplasma mempunyai nilai rata-rata yang paling tinggi untuk semua peubah dibandingkan kelompok lain kecuali pada peubah  $X_4$  (ALM 1). Sedangkan untuk kelompok 1 yang merupakan kelompok Inner Membran mempunyai nilai rata-rata terkecil dibandingkan dengan kelompok lain pada peubah  $X_1$  (MCG),  $X_3$  (AAC) dan  $X_5$  (ALM 2).

#### 4.8. Pembahasan

Dalam membandingkan hasil pengelompokan terhadap kriteria eksternal, diperlukan suatu pengukuran. Karena algoritma *K-mean* dan *partitional around medoid* termasuk dalam *hard clustering* dimana setiap objek harus masuk ke dalam satu kelompok tertentu. Maka pengukuran antara dua partisi dapat menggunakan metode perbandingan *Adjusted Rand index*. Nilai dari *Adjusted Rand index* berada anatar 0 dan 1, ketika pengelompokan dapat memisahkan dengan sempurna maka nilai dari Rand index adalah 1 (Yeung dan Ruzzo, 2001). Nilai ARI untuk ketiga data pada algoritma *k-means* dan *PAM* adalah sebagai berikut:

Tabel 4.18 Perbandingan Nilai ARI untuk semua data

Data	<i>K-means</i>	PAM	<i>Outlier</i>
I	0,7680	0,8805	3
II	0,7340	0,8049	1
III	0,7292	0,8384	2

Dari Tabel 4.18 dapat dilihat nilai ARI yang dihasilkan dari ketiga data pada algoritma *k-means* lebih kecil dibandingkan nilai *Adjusted Rand index* yang dihasilkan oleh algoritma PAM, sehingga dapat disimpulkan bahwa algoritma yang menghasilkan *centroid* yang baik untuk data yang mengandung *outlier* adalah dengan menggunakan algoritma PAM.

Berdasarkan *outlier* yang terdapat pada data yaitu untuk Data I terdapat tiga *outlier*, Data II terdapat satu *outlier* sedangkan pada Data III terdapat dua *outlier* dengan membandingkan dengan nilai *Adjusted Rand Index* pada Tabel 4.18 di mana semakin banyak *outlier* dan semakin besar jarak *outlier* maka nilai *Adjusted Rand*

*Index yang dihasilkan oleh algoritma  $k$ -means dan algoritma PAM perbedaannya semakin besar.*







## BAB V KESIMPULAN DAN SARAN

### 5.1. Kesimpulan

Berdasarkan hasil dan pembahasan dapat ditarik kesimpulan sebagai berikut:

1. Semakin banyak *outlier* dan semakin besar jarak *outlier* pada data maka semakin besar perbedaan nilai dari metode perbandingan *Adjusted Rand index* (ARI) yang dihasilkan oleh algoritma *k-means* dan algoritma *partitioning around medoid* (PAM).
2. Berdasarkan metode perbandingan *Adjusted Rand index* (ARI) maka dapat dikatakan bahwa algoritma *partitioning around medoid* (PAM) menghasilkan *centroid* yang dapat menghasilkan pengelompokan yang lebih baik dibandingkan dengan algoritma *k-means* pada data yang mengandung *outlier*.

### 5.2. Saran

Pada penelitian selanjutnya, dapat dibahas metode kelompok non-hirarki yang lain seperti algoritma *Gustafson-Kessel* dan *Gath-Geva*.



**DAFTAR PUSTAKA**

- Afifi, A.A. dan V. Clark. 1990. **Computer-Aided multivariate Analysis, Second Edition**. Champman and Hall. New York.
- Agusta, Y. 2007. **Penerapan K-Means, Permasalahan dan Metode Terkait**. STMIK STIKOM. Bali.
- Anderberg, M.R. 1973. **Cluster Analysis for Application**. Academic Press. New York.
- Anonymous. 2009. **Multivariate Outlier Detection**. Canada. <http://www.med.monash.edu.au/pysh/research/rda/outliers.htm>  
Tanggal akses: 10 Maret 2009.
- Atom. 2008. **Winner Statistik**. <http://winnerstatistik.com/2008/02/analisisgerombolcluster-analysis.html>. Tanggal akses: 17 April 2009.
- De Hoon, M. 2002. **Distance or Similarity Measure**. Human Genome Center. University of Tokyo. Japan. <http://bonsai.ims.Utokyo.oc.jp/~mdehoon/software/cluster/manual/Distance.html>. Tanggal akses: 15 April 2008.
- Epps, J dan Ambikairajah. 2008. **Visualitation of Reduced-Dimension Microarray data Using Gaussian Mixture Model**. <http://crpit.com/confpapers/CRPITV45Epps.pdf>. Tanggal Akses: 29 April 2009.
- Garg, S dan R.C. Jain. 2006. **Variations of k-means Algorithm: A Study for High-dimensional Large Data Set**. <http://ica.ign.fr/leicestur/paper/blnv2ICAworkshop.pdf>. Tanggal Akses: 30 April 2009.
- Giudici, P. 2007. **Applied Data Mining, Statistical Methods for Business and Industry**. John Wiley & Sons, Inc. New York.

Hines, W.W. dan D.C. Montgomery. 1990. **Probability and Statistics in Engineering and Management Science, Third Edition**. John Wiley & Sons, Inc. Canada.

Hair, Jr. J.F, R.E Anderson, R.L Tataham, and W.C Black. 1998. **Multivariate Data Analysis**. Prentice-Hall International. New Jersey.

Huan, L. 2003. **Introduction to Bioinformatics Microarray**. Electrical Engineering and Computer Science. <http://people.eecs.ku.edu/~jhuan>. Tanggal Akses: 29 April.

Johnson, R.A. dan D.W. Wichern, 2002. **Applied Multivariate Statistical Analysis**. Fifth edition. Prentice-Hall, Inc. New Jersey.

Kirana, D.A. 2008. **Pengelompokan Terbaik Berdasarkan Algoritma K-Means dan Fuzzy C-Means pada Analisis Kelompok Non-hierarki**. Skripsi Program Studi Statistika Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Brawijaya Malang (tidak dipublikasikan).

Laboratorium Data Mining Jurusan Teknik Industri Fakultas Teknologi Industri Universitas Islam Indonesia. 2009. **Clustering**.[http://www.adadata.id.or.id/dl\\_jump.php?id=49&Login=31c097c3fb5534f53825c99e05a4cbe5](http://www.adadata.id.or.id/dl_jump.php?id=49&Login=31c097c3fb5534f53825c99e05a4cbe5). Tanggal akses: 15 April 2009.

Likes, A., N Vlassis and J.J Verbeek. 2002. **The Global K-means Clustering Algorithm**. University of Iounina. Grece.

Marazzi, A. 1993. **Algorithms, Routines and S functions for Robust Statistics**. California: Wadworth, Inc.

Matteua. 2008. **A Tutorial on Clustering Algorithms, Fuzzy C-means Clustering**.<http://home.dei.polimit/matteucc/clustering/tutorialhtml/cmeans.html>. Tanggal akses: 15 April 2009.

Mattjik, A.I.M. Sumartajaya, H.Wijayanto, Indahwati dan B.Sartono. 2002. **Aplikasi Analisis Peubah Ganda**. Depdiknas Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Pertanian Bogor. Bogor.

Mercer, D.P. and L. College. 2003. **Clustering Large Datasets**. <http://ieeexplore.ieu.org/ie15/4089190/4089191/04089238.pdf>.  
Tanggal akses: 17 April 2009

Nakai, K. 1996. **Protein Localization Sites**. Osaka university. <ftp://ftp.ics.uci.edu/pub/machine-learning-databases>.  
Tanggal akses 19 juni 2009.

Park, H, J. Lee and C. Jun. 2005. **A K-means-like Algorithm for K-medoids Clustering and Its Performance**. Department of Industrial and Management Engineering, POSTECH San 31 Hyoja-dong. Korea.

Raymond and J. Han. 1994. **Efficient and Effective Clustering Methods for Spatial Data Mining**. Proceedings of the 20th International Conference on Very Large Databases. Santiago. Chile. <http://cs.sfu.ca/cc/459/han/papers/ng94.pdf>.  
Tanggal akses: 30 April 2009.

Ridho B, A. 2006. **Clustering Workshop Data mining**. <http://www.Let.rug.nl/nerbonne/teach/Tema-Stats-methsemina/cluster-prokic.2006.pdf>. Tanggal Akses: 19 Juni 2009.

Rousseeuw P, J. and B.C. Van Zomeren. 1990. **Unmasking multivariate outliers and leverage points**, *Journal of the American Statistical Association*. <http://www.comp.nus.edu.sg/tmcl/papers/CIKM2004/i12-niu.pdf>.

Sharma, S. 1996. **Applied Multivariate Techniques**. John Wiley & Sons, Inc. New York.

Syarifuddin. 2003. **Pemilihan Model Terbaik pada Analisis Diskriminan Dengan Metode Stepwise**. Skripsi Program Studi

Statistika Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Brawijaya Malang (tidak dipublikasikan).

Timm, N.H. 2002. **Applied Multivariate Analysis**. University of Pittsburgh. USA.

Toledo, M.G. 2005. **A Comparison in Cluster Validation Techniques**. University of Puerto Rico Mayaguez Campus. <http://grad.UPTM.edu/tesis/gonzalestoledo.pdf>. Tanggal akses: 19 Juni 2009.

Walpole, R.E. 1995. **Pengantar Statistika, Edisi ke-3**. Terjemah Sumantri, B. PT.Garmedia Pustaka Utama. Jakarta.

Walpole, R.E. dan R.H. Myers. 1995. **Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan, Edisi Keempat**. Terjemahan R.K. Sembering. Penerbit ITB. Bandung.

Wolkenhauer, O. 1998. **Fuzzy Clustering**. [www.csc.umist.ac.uk/people/wolkenhauer.html](http://www.csc.umist.ac.uk/people/wolkenhauer.html). Tanggal Akses 28 Juli 2009.

Yeung, K.Y and W.L. Ruzzo. 2001. **Details of the Adjusted Rand Index and Clustering Algorithms**. <http://faculty.washington.edu/kayee/pca/supp.pdf>. Tanggal akses 19 juni 2009.

Yitnosumarto, S. 1990. **Dasar-dasar Statistika**. PT. Raja Grafindo Persada. Jakarta.

Zuroida, D.A. 2004. **Pendeteksian *Outlier* dan Obyek Berpengaruh Pada Analisis Regresi Logistik**. Skripsi Program Studi Statistika Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Brawijaya Malang (tidak dipublikasikan).

**Lampiran 1. Data I yatiu Rasio keuangan BPR di Daerah Tingkat II Badung Propinsi Bali**

No	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	Status Bank
1	30,00	8,03	10,00	Baik
2	25,55	1,29	10,00	Baik
3	30,00	5,42	10,00	Baik
4	27,51	3,95	10,00	Baik
5	27,48	9,17	10,00	Baik
6	22,83	6,21	10,00	Baik
7	30,00	3,75	10,00	Baik
8	30,00	3,63	10,00	Baik
9	29,66	1,26	10,00	Baik
10	25,82	0,24	10,00	Baik
...	....	....	....	....
...	....	....	....	....
...	....	....	....	....
...	....	....	....	....
60	12,32	7,92	10,00	Tidak
61	20,25	0,28	10,00	Tidak
62	15,06	2,81	10,00	Tidak
63	19,59	4,90	10,00	Tidak
64	5,10	6,56	10,00	Tidak
65	5,10	6,54	10,00	Tidak
66	11,33	10,00	10,00	Tidak
67	11,57	7,39	10,00	Tidak
68	13,75	8,81	10,00	Tidak
69	11,83	10,00	7,03	Tidak
70	15,36	10,00	7,48	Tidak
71	16,69	3,56	10,00	Tidak
72	16,19	6,75	10,00	Tidak

Keterangan:

X<sub>1</sub> = Rasio KAP

X<sub>2</sub> = Rasio Rentabilitas

X<sub>3</sub> = Rasio Likuiditas

**Lampiran 2. Data II yaitu Faktor-faktor yang mempengaruhi penetasan telur burung walet**

No	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	Hasil
1	1,9	1,90	1,40	33	Menetas
2	2,0	2,20	1,30	33	Menetas
3	2,1	2,00	1,40	33	Menetas
4	3,0	2,10	1,40	35	Menetas
5	1,8	1,90	1,35	35	Menetas
6	2,5	2,10	1,40	35	Menetas
7	2,0	2,10	1,40	35	Menetas
8	2,8	1,90	1,40	35	Menetas
9	3,2	2,10	1,40	35	Menetas
9	3,2	2,10	1,40	35	Menetas
10	3,1	2,10	1,40	35	Menetas
11	3,6	2,10	1,40	35	Menetas
12	2,8	2,00	1,40	37	Menetas
...	...	...	...	...	.....
...	...	...	...	...	.....
...	...	...	...	...	.....
51	2,0	2,10	1,30	34	Tidak Menetas
52	2,8	2,00	1,30	34	Tidak Menetas
53	2,0	2,00	1,40	34	Tidak Menetas
54	1,9	2,10	1,30	33	Tidak Menetas
55	3,0	2,10	1,40	32	Tidak Menetas
56	2,0	1,95	1,40	32	Tidak Menetas
57	1,7	2,10	1,40	32	Tidak Menetas
58	2,1	2,10	1,40	34	Tidak Menetas
59	1,8	2,15	1,40	34	Tidak Menetas
60	1,6	2,20	1,40	34	Tidak Menetas

**Keterangan**

- X<sub>1</sub> = Berat telur (gr)
- X<sub>2</sub> = Panjang telur (cm)
- X<sub>3</sub> = Lebar telur (cm)
- X<sub>4</sub> = Suhu pengermanan (C)



## Lampiran 2. (Lanjutan)

No	ZX <sub>1</sub>	ZX <sub>2</sub>	ZX <sub>3</sub>	ZX <sub>4</sub>	Hasil
1	-0,87	-1,3	0,45	-0,66	Menetas
2	-0,68	1,86	-1,06	-0,66	Menetas
3	-0,49	-0,25	0,45	-0,66	Menetas
4	1,214	0,81	0,45	0,50	Menetas
5	-1,06	-1,30	-0,31	0,50	Menetas
6	0,27	0,81	0,45	0,50	Menetas
7	-0,68	0,81	0,45	0,50	Menetas
8	0,84	-1,30	0,45	0,50	Menetas
9	1,59	0,81	0,45	0,50	Menetas
10	1,40	0,81	0,45	0,50	Menetas
11	2,35	0,81	0,45	0,50	Menetas
12	0,84	-0,25	0,45	1,67	Menetas
...	...	...	...	...	.....
...	...	...	...	...	.....
...	...	...	...	...	.....
51	-0,68	0,81	-1,06	-0,08	Tidak Menetas
52	0,84	-0,25	-1,06	-0,08	Tidak Menetas
53	-0,68	-0,25	0,45	-0,08	Tidak Menetas
54	-0,87	0,81	-1,06	-0,66	Tidak Menetas
55	1,21	0,81	0,45	-1,25	Tidak Menetas
56	-0,68	-0,77	0,45	-1,25	Tidak Menetas
57	-1,25	0,81	0,45	-1,25	Tidak Menetas
58	-0,49	0,81	0,45	-0,08	Tidak Menetas
59	-1,06	1,33	0,45	-0,08	Tidak Menetas
60	-1,44	1,86	0,45	-0,08	Tidak Menetas

## Keterangan

ZX<sub>1</sub> = Skor baku Berat telurZX<sub>2</sub> = Skor baku Panjang telurZX<sub>3</sub> = Skor baku Lebar telurZX<sub>4</sub> = Skor baku Suhu pengeraman

### Lampiran 3. Data III yaitu Protein Localization Sites

No	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	Kategori
1	0,06	0,61	0,49	0,92	0,37	Inner membrane
2	0,44	0,52	0,43	0,47	0,54	Inner membrane
3	0,63	0,47	0,51	0,82	0,84	Inner membrane
4	0,36	0,45	0,38	0,79	0,17	Inner membrane
5	0,00	0,51	0,35	0,67	0,44	Inner membrane
6	0,30	0,51	0,42	0,61	0,34	Inner membrane
7	0,61	0,47	0,00	0,80	0,32	Inner membrane
8	0,74	0,49	0,42	0,54	0,36	Perisplasma
9	0,70	0,61	0,56	0,52	0,43	Perisplasma
10	0,66	0,86	0,34	0,41	0,36	Perisplasma
...	...	...	...	...	...	.....
...	...	...	...	...	...	.....
...	...	...	...	...	...	.....
21	0,41	0,48	0,51	0,44	0,51	Cytoplasma
22	0,44	0,28	0,43	0,27	0,37	Cytoplasma
23	0,41	0,43	0,45	0,31	0,41	Cytoplasma
24	0,29	0,47	0,41	0,23	0,34	Cytoplasma
25	0,34	0,55	0,58	0,31	0,41	Cytoplasma
26	0,40	0,46	0,52	0,49	0,56	Cytoplasma
27	0,52	0,44	0,37	0,36	0,42	Cytoplasma
28	0,17	0,52	0,49	0,37	0,46	Cytoplasma
29	0,44	0,35	0,44	0,52	0,59	Cytoplasma
30	0,27	0,42	0,37	0,38	0,43	Cytoplasma

Keterangan:

X<sub>1</sub> = MCG (Skor diskriminan untuk pengenalan urutan isyarat yang dihasilkan oleh metode McGeoch's )

X<sub>2</sub> = GVH (Skor yang dihasilkan oleh metode von Heijne's untuk pengenalan urutan isyarat dengan menggunakan weight matriks)

X<sub>3</sub> = AAC ( Nilai asam amino yang terkandung dalam membran luar)

X<sub>4</sub> = ALM 1 (Nilai program ALOM untuk identifikasi di sekeliling daerah membran)

X<sub>5</sub> = ALM 2 ( Nilai program ALOM untuk identifikasi yang tidak terdapat di sekeliling membran dengan metode Von Heijne's)

## Lampiran 4. Nilai Jarak Mahalanobis

### 1. Data I

Comparison of Mahalanobis Distances:

	1	2	3	4	5	6
Classical	0.9298	4.3741	1.3126	1.6950	0.7330	0.2061
	7	8	9	10	11	12
Classical	2.3867	2.4888	5.1021	6.0011	1.2357	1.0514
	13	14	15	16	17	18
Classical	0.8767	1.1372	0.8835	0.8238	1.3970	1.2036
	19	20	21	22	23	24
Classical	0.9667	1.6066	1.6876	1.1299	0.9460	1.1372
	25	26	27	28	29	30
Classical	1.6876	1.3743	1.6876	1.6876	0.8731	0.9174
	31	32	33	34	35	36
Classical	0.9917	1.3110	0.7152	2.2869	0.7152	0.7152
	37	38	39	40	41	42
Classical	0.8091	3.1415	1.5720	1.0725	0.9168	0.2311
	43	44	45	46	47	48
Classical	1.5009	1.5845	20.4994	1.0424	1.3197	0.9037
	49	50	51	52	53	54
Classical	1.1056	1.6876	0.8982	5.3462	0.3088	0.9235
	55	56	57	58	59	60
Classical	2.0164	1.5627	7.4677	6.7795	1.8934	2.7176
	61	62	63	64	65	66
Classical	6.0482	3.9120	1.0658	7.1624	7.1658	3.9957
	67	68	69	70	71	72
Classical	3.0353	2.3115	29.9080	20.9213	2.6772	1.2211

## Lampiran 4. ( Lanjutan )

### 2. Data II

Comparison of Mahalanobis Distances:

	1	2	3	4	5	6
Classical	2.5122	5.4459	0.8497	2.1015	3.6441	0.9345
	7	8	9	10	11	12
Classical	1.6281	3.2335	3.0965	2.5613	5.9903	3.5607
	13	14	15	16	17	18
Classical	7.4572	3.3260	5.9519	7.7459	3.4727	9.3595
	19	20	21	22	23	24
Classical	5.7236	7.6856	7.3436	6.0023	2.3926	0.3083
	25	26	27	28	29	30
Classical	1.0724	2.8805	2.7266	2.4817	6.8336	0.3083
	31	32	33	34	35	36
Classical	31.2525	3.5415	1.6215	1.7910	5.5736	3.0233
	37	38	39	40	41	42
Classical	5.0618	6.0559	2.2674	3.7353	2.1214	3.4681
	43	44	45	46	47	48
Classical	3.0706	2.7774	2.2648	3.7686	2.4999	2.4739
	49	50	51	52	53	54
Classical	0.3510	1.1222	2.4090	1.7093	0.6321	2.7597
	55	56	57	58	59	60
Classical	5.3862	2.4750	3.8529	1.0847	3.1889	6.0300

## Lampiran 4. ( Lanjutan )

### 3. Data III

Comparison of Mahalanobis Distances:

	1	2	3	4	5	6
Classical	11.7633	1.6793	16.1182	11.6895	7.1558	1.6673
	7	8	9	10	11	12
Classical	16.6246	3.8522	3.5569	12.7276	8.2097	1.7646
	13	14	15	16	17	18
Classical	4.9248	5.2131	3.4998	1.3670	3.2368	3.2471
	19	20	21	22	23	24
Classical	3.2077	2.1518	0.7372	3.0402	0.7197	3.2798
	25	26	27	28	29	30
Classical	2.2159	1.6316	1.4298	3.2087	3.1081	1.9717



## Lambran 5. Matrikas Korelasi Antar Peubah Dengan SPSS

### 1. Data I

Correlations

		X1	X2	X3
X1	Pearson Correlation	1	-,021	,146
	Sig. (2-tailed)	,	,863	,221
	N	72	72	72
X2	Pearson Correlation	-,021	1	-,184
	Sig. (2-tailed)	,863	,	,123
	N	72	72	72
X3	Pearson Correlation	,146	-,184	1
	Sig. (2-tailed)	,221	,123	,
	N	72	72	72

### 2.Data II

Correlations

		X1	X2	X3	X4
X1	Pearson Correlation	1	,139	-,080	,241
	Sig. (2-tailed)	,	,289	,542	,063
	N	60	60	60	60
X2	Pearson Correlation	,139	1	,015	,134
	Sig. (2-tailed)	,289	,	,912	,306
	N	60	60	60	60
X3	Pearson Correlation	-,080	,015	1	,210
	Sig. (2-tailed)	,542	,912	,	,107
	N	60	60	60	60
X4	Pearson Correlation	,241	,134	,210	1
	Sig. (2-tailed)	,063	,306	,107	,
	N	60	60	60	60

## Lampiran 5. (Lanjutan)

### 3.Data III

**Correlations**

		X1	X2	X3	X4	X5
X1	Pearson Correlation	1	,309	-,031	,004	,090
	Sig. (2-tailed)	,	,097	,870	,985	,635
	N	30	30	30	30	30
X2	Pearson Correlation	,309	1	,109	,302	-,094
	Sig. (2-tailed)	,097	,	,566	,104	,621
	N	30	30	30	30	30
X3	Pearson Correlation	-,031	,109	1	-,316	,212
	Sig. (2-tailed)	,870	,566	,	,089	,261
	N	30	30	30	30	30
X4	Pearson Correlation	,004	,302	-,316	1	,115
	Sig. (2-tailed)	,985	,104	,089	,	,545
	N	30	30	30	30	30
X5	Pearson Correlation	,090	-,094	,212	,115	1
	Sig. (2-tailed)	,635	,621	,261	,545	,
	N	30	30	30	30	30



## Lampiran 6. Hasil Output Untuk Algoritma *k-means* Menggunakan SYSTAT

### 1. Data I

Cluster 1 of 2 Contains 56 Cases

Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	Standard Deviation
Case 1	1,341	VAR_1	19,590	26,653	30,000	3,270
Case 2	2,420	VAR_2	0,240	7,399	10,000	3,178
Case 3	1,791	VAR_3	7,640	9,953	10,000	0,317
Case 4	1,451					
Case 5	0,881					
Case 6	1,686					
Case 7	2,348					
Case 8	2,388					
Case 9	3,064					
Case 10	2,680					
Case 11	1,588					
Case 12	0,891					
Case 13	2,117					
Case 14	3,087					
Case 15	1,567					
Case 16	1,402					
Case 17	1,585					
Case 18	1,241					
Case 19	1,137					
Case 20	1,891					
Case 21	1,998					
Case 22	1,085					
Case 23	2,533					
Case 24	3,087					
Case 25	1,998					
Case 26	1,548					
Case 27	1,998					
Case 28	1,998					
Case 29	1,677					
Case 30	2,403					
Case 31	1,050					
Case 32	1,441					
Case 33	1,400					
Case 34	2,073					
Case 35	1,400					
Case 36	1,400					



## Lampiran 6. (Lanjutan)

Case 37	1,326
Case 38	2,293
Case 39	1,961
Case 40	1,578
Case 41	2,400
Case 42	0,143
Case 43	1,904
Case 44	1,861
Case 45	2,754
Case 46	1,515
Case 47	1,768
Case 48	1,304
Case 49	1,585
Case 50	1,998
Case 51	1,457
Case 52	2,558
Case 53	1,977
Case 54	2,433
Case 61	4,523
Case 63	3,203

## Cluster 2 of 2 Contains 16 Cases

Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	Standard Deviation
Case 55	2,264	VAR_1	5,100	12,084	17,850	4,495
Case 56	2,817	VAR_2	2,810	7,657	10,000	2,267
Case 57	2,914	VAR_3	7,030	9,657	10,000	0,941
Case 58	2,715					
Case 59	1,089					
Case 60	0,281					
Case 62	2,722					
Case 64	2,808					
Case 65	2,815					
Case 66	1,147					
Case 67	0,375					
Case 68	1,054					
Case 69	1,741					
Case 70	2,599					
Case 71	3,015					
Case 72	1,785					

## Lampiran 6. (Lanjutan)

### 2. Data II

Cluster 1 of 2 Contains 37 Cases

Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	Standard Deviation
Case 1	0,504	VAR_1	-1,628	-0,619	1,593	0,624
Case 2	0,841	VAR_2	-1,298	-0,246	1,856	0,984
Case 3	0,210	VAR_3	-1,057	-0,062	1,950	0,794
Case 5	0,674	VAR_4	-1,247	-0,460	1,665	0,791
Case 15	0,816					
Case 16	1,079					
Case 21	1,262					
Case 22	0,976					
Case 24	0,348					
Case 26	0,686					
Case 30	0,348					
Case 32	0,673					
Case 33	0,473					
Case 34	0,336					
Case 36	0,710					
Case 37	0,862					
Case 38	0,995					
Case 39	0,524					
Case 40	0,771					
Case 41	0,572					
Case 42	0,741					
Case 43	0,687					
Case 44	0,565					
Case 45	0,386					
Case 46	0,761					
Case 47	0,656					
Case 48	0,535					
Case 49	0,396					
Case 50	0,517					
Case 51	0,622					
Case 53	0,237					
Case 54	0,626					
Case 56	0,471					
Case 57	0,744					
Case 58	0,516					
Case 59	0,726					
Case 60	0,952					

### Lampiran 6. (Lanjutan)

Cluster 2 of 2 Contains 23 Cases

Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	Standard Deviation
Case 4	0,284	VAR_1	-0,870	0,950	2,350	0,772
Case 6	0,389	VAR_2	-1,298	0,485	1,856	0,921
Case 7	0,626	VAR_3	-5,569	0,021	1,198	1,338
Case 8	0,612	VAR_4	-1,247	0,627	1,665	0,895
Case 9	0,379					
Case 10	0,331					
Case 11	0,568					
Case 12	0,577					
Case 13	0,891					
Case 14	0,569					
Case 17	0,664					
Case 18	1,114					
Case 19	0,727					
Case 20	1,064					
Case 23	0,429					
Case 25	0,487					
Case 27	0,580					
Case 28	0,538					
Case 29	0,925					
Case 31	2,008					
Case 35	0,790					
Case 52	0,659					
Case 55	0,721					



## Lampiran 6. (Lanjutan)

### 3. Data III

Cluster 1 of 3 Contains 17 Cases

Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	Standard Deviation
Case 2	0,084	VAR_1	0,070	0,378	0,670	0,150
Case 14	0,107	VAR_2	0,280	0,413	0,550	0,084
Case 15	0,080	VAR_3	0,360	0,465	0,580	0,067
Case 16	0,052	VAR_4	0,230	0,353	0,520	0,090
Case 18	0,093	VAR_5	0,340	0,440	0,590	0,075
Case 19	0,094					
Case 20	0,082					
Case 21	0,060					
Case 22	0,077					
Case 23	0,027					
Case 24	0,084					
Case 25	0,073					
Case 26	0,076					
Case 27	0,058					
Case 28	0,075					
Case 29	0,093					
Case 30	0,049					

Cluster 2 of 3 Contains 5 Cases

Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	Standard Deviation
Case 1	0,134	VAR_1	0,000	0,266	0,610	0,246
Case 4	0,079	VAR_2	0,450	0,510	0,610	0,062
Case 5	0,098	VAR_3	0,000	0,328	0,490	0,191
Case 6	0,057	VAR_4	0,610	0,758	0,920	0,121
Case 7	0,152	VAR_5	0,170	0,328	0,440	0,099

## Lampiran 6. (Lanjutan)

Cluster 3 of 3 Contains 8 Cases

Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	Standard Deviation
Case 3	0,180	VAR_1	0,590	0,666	0,740	0,053
Case 8	0,079	VAR_2	0,470	0,623	0,860	0,146
Case 9	0,025	VAR_3	0,340	0,500	0,600	0,087
Case 10	0,119	VAR_4	0,400	0,515	0,820	0,133
Case 11	0,088	VAR_5	0,310	0,445	0,840	0,170
Case 12	0,042					
Case 13	0,079					
Case 17	0,076					



## Lampiran 7. Hasil Output Untuk Algoritma *Partitioning Around Medoid* Menggunakan S-PLUS

### 1. Data I

```

*** Partitioning Around Medoids ***
Call:
pam(x = menuModelFrame(data = SDF80, variables = "<ALL>",
  subset = NULL, na.rm = T), k = 2, diss = F, metric =
  "manhattan", stand = F, save.x = T, save.diss = T)
Medoids:
numeric matrix: 2 rows, 3 columns.
      V1  V2 V3
5  27.48 9.17 10
59 14.21 6.86 10
Clustering vector:
 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63
1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2
64 65 66 67 68 69 70 71 72
2 2 2 2 2 2 2 2 2
Objective function:
      build      swap
5.736806 5.388194
Metric : manhattan
Number of objects : 72

```



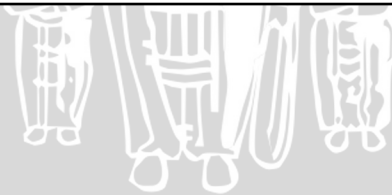
## Lampiran 7. (Lanjutan)

### 2. Data II

```

*** Partitioning Around Medoids ***
Call:
pam(x = menuModelFrame(data = SDF84, variables = "<ALL>",
  subset = NULL, na.rm = T), k = 2, diss = F, metric =
  "manhattan", stand = F, save.x = T, save.diss = T)
Medoids:
numeric matrix: 2 rows, 4 columns.
      V1      V2      V3      V4
3 -0.49 -0.25  0.45 -0.66
4  1.21  0.81  0.45  0.50
Clustering vector:
 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
1  1  1  2  1  2  2  2  2  2  2  2  2  2  1  1  2  2  2  2  1  1  2
24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43
1  2  1  2  2  2  1  2  1  1  1  2  1  1  1  1  1  1  1  1  1  1
44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
1  1  1  1  1  1  1  1  1  1  1  2  1  1  1  1  1
Objective function:
      build      swap
2.533667 2.405333
Metric : manhattan
Number of objects : 60

```



## Lampiran 7. (Lanjutan)

### 3. Data III

```

*** Partitioning Around Medoids ***
Call:
pam(x = menuModelFrame(data = SDF85, variables = "<ALL>",
  subset = NULL, na.rm = T), k = 3, diss = F, metric =
  "manhattan", stand = F, save.x = T, save.diss = T)
Medoids:
numeric matrix: 3 rows, 5 columns.
  V1  V2  V3  V4  V5
6 0.30 0.51 0.42 0.61 0.34
23 0.41 0.43 0.45 0.31 0.41
9 0.70 0.61 0.56 0.52 0.43
Clustering vector:
 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
1 2 3 1 1 1 1 3 3 3 3 3 2 2 2 2 2 2 2 2 2 2
24 25 26 27 28 29 30
 2 2 2 2 2 2 2
Objective function:
build swap
0.425 0.425
Metric : manhattan
Number of objects : 30

```





## Lampiran 8. Uji Perbandingan Peubah Antar Kelompok Dengan SPSS

### 1. Data I

ANOVA						
		Sum of Squares	df	Mean Square	F	Sig.
X1	Between Groups	2626,227	1	2626,227	202,862	,000
	Within Groups	906,212	70	12,946		
	Total	3532,439	71			
X2	Between Groups	3,159	1	3,159	,351	,556
	Within Groups	630,229	70	9,003		
	Total	633,387	71			
X3	Between Groups	,889	1	,889	3,276	,075
	Within Groups	19,003	70	,271		
	Total	19,892	71			

### 2. Data II

ANOVA						
		Sum of Squares	df	Mean Square	F	Sig.
X1	Between Groups	9,164	1	9,164	65,414	,000
	Within Groups	8,126	58	,140		
	Total	17,290	59			
X2	Between Groups	,074	1	,074	8,930	,004
	Within Groups	,479	58	,008		
	Total	,552	59			
X3	Between Groups	,002	1	,002	,328	,569
	Within Groups	,273	58	,005		
	Total	,275	59			
X4	Between Groups	50,528	1	50,528	25,004	,000
	Within Groups	117,206	58	2,021		
	Total	167,733	59			

## Lampiran 8. (Lanjutan)

### 3. Data III

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
X1	Between Groups	,591	2	,295	12,136	,000
	Within Groups	,657	27	,024		
	Total	1,248	29			
X2	Between Groups	,257	2	,129	13,135	,000
	Within Groups	,264	27	,010		
	Total	,521	29			
X3	Between Groups	,096	2	,048	4,719	,017
	Within Groups	,274	27	,010		
	Total	,369	29			
X4	Between Groups	,659	2	,329	28,145	,000
	Within Groups	,316	27	,012		
	Total	,974	29			
X5	Between Groups	,056	2	,028	2,286	,121
	Within Groups	,331	27	,012		
	Total	,388	29			



## Lampiran 9. Analisis Deskriptif Berdasarkan Algoritma *k-means* Dengan SPSS

### 1. Data I

#### a. kelompok 1

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
X1	54	20,04	30,00	26,9026	3,05219
X2	54	,24	10,00	7,5776	3,06165
X3	54	7,64	10,00	9,9517	,32231
Valid N (listwise)	54				

#### b. Kelompok 2

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
X1	18	5,10	20,25	12,9550	4,92574
X2	18	,28	10,00	7,0939	2,80151
X3	18	7,03	10,00	9,6950	,89103
Valid N (listwise)	18				

### 2. Data II

#### a. Kelompok 1

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
X1	38	1,50	3,20	2,0526	,34776
X2	38	1,90	2,20	2,0000	,09227
X3	38	1,30	1,50	1,3645	,05315
X4	38	32,00	37,00	33,3684	1,34408
Valid N (listwise)	38				

#### b. Kelompok 2

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
X1	22	1,90	3,60	2,8636	,41696
X2	22	1,90	2,20	2,0727	,08827
X3	22	1,00	1,45	1,3750	,08964
X4	22	32,00	37,00	35,2727	1,54863
Valid N (listwise)	22				

## Lampiran 9. (Lanjutan)

### 3. Data III

#### a. Kelompok 1

**Descriptive Statistics**

	N	Minimum	Maximum	Mean	Std. Deviation
X1	5	,00	,61	,2660	,24572
X2	5	,45	,61	,5100	,06164
X3	5	,00	,49	,3280	,19071
X4	5	,61	,92	,7580	,12112
X5	5	,17	,44	,3280	,09935
Valid N (listwise)	5				

#### b. Kelompok 2

**Descriptive Statistics**

	N	Minimum	Maximum	Mean	Std. Deviation
X1	18	,07	,67	,3894	,15391
X2	18	,28	,55	,4172	,08365
X3	18	,36	,58	,4678	,06647
X4	18	,23	,52	,3583	,09005
X5	18	,34	,59	,4356	,07548
Valid N (listwise)	18				

#### c. Kelompok 3

**Descriptive Statistics**

	N	Minimum	Maximum	Mean	Std. Deviation
X1	7	,63	,74	,6771	,04608
X2	7	,47	,86	,6414	,14724
X3	7	,34	,60	,4971	,09394
X4	7	,40	,82	,5243	,14105
X5	7	,31	,84	,4571	,18034
Valid N (listwise)	7				