PERBANDINGAN METODE GAP-STATISTIK DAN INDEKS VALIDITAS GABUNGAN (IVG) DALAM MENENTUKAN BANYAKNYA *CLUSTER* OPTIMAL PADA ANALISIS *CLUSTER* MENGGUNAKAN DATA EKSPRESI GEN

SKRIPSI

oleh:
FITRISIANA
0410950020-95

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

2009

PERBANDINGAN METODE GAP-STATISTIK DAN INDEKS VALIDITAS GABUNGAN (IVG) DALAM MENENTUKAN BANYAKNYA *CLUSTER* OPTIMAL PADA ANALISIS *CLUSTER* MENGGUNAKAN DATA EKSPRESI GEN

SKRIPSI

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

oleh:
FITRISIANA
0410950020-95

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

LEMBAR PENGESAHAN SKRIPSI

PERBANDINGAN METODE GAP-STATISTIK DAN INDEKS VALIDITAS GABUNGAN (IVG) DALAM MENENTUKAN BANYAKNYA *CLUSTER* OPTIMAL PADA ANALISIS *CLUSTER* MENGGUNAKAN DATA EKSPRESI GEN

oleh:
FITRISIANA
0410950020-95

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

Pembimbing I	Pembimbing II
Suci Astutik, SSi., MSi	**Adji Achmad R.F ,SSi.,MSc.**
NIP. 132 233 148	**132 311 764**

Mengetahui,
Ketua Jurusan Matematika
Fakultas MIPA Universitas Brawijaya,

Dr. Agus Suryanto, M.Sc.
NIP. 132 126 049

LEMBAR PERNYATAAN

Saya yang bertanda tangan di bawah ini :

Nama : Fitrisiana
NIM : 0410950020-95
Jurusan : Matematika
Penulisan skripsi berjudul : Perbandingan Metode Gap-Statistik Dan Indeks Validitas Gabungan (IVG) dalam Menentukan Banyaknya *Cluster* Optimal pada Analisis *Cluster* Menggunakan Data Ekspresi Gen

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 ini.**
2. **Apabila di kemudian hari ternyata skripsi yang saya tulis terbukti hasil jiplakan, maka saya akan bersedia menanggung segala risiko.**

Demikian pernyataan ini saya buat dengan segala kesadaran.

Malang, 16 Maret 2009
Yang menyatakan,

Fitrisiana
NIM. 0410950020-95

ABSTRAK

Data ekspresi gen memiliki jumlah relatif besar sehingga bukan suatu hal yang mudah untuk menentukan berapa jumlah *cluster* (kelompok) yang terbentuk. Analisis *cluster* merupakan salah satu komponen statistik yang biasa digunakan dalam data ekspresi gen. Setelah analisis *cluster* dilakukan, perlu diketahui berapa *cluster* optimal dalam suatu himpunan data. Penelitian ini bertujuan untuk menentukan jumlah *cluster* optimal dengan metode Gap Statistik dan Indeks Validitas Gabungan (IVG) pada data ekspresi gen dan menentukan metode jumlah *cluster* optimal terbaik antara keduanya. Indeks Validitas Gabungan (IVG) pada penelitian ini terdiri atas Indeks *Dunn*, Indeks *Davies-Bouldin*, Indeks C, Indeks *Silhoutte* dan Indeks *Goodman-Kruskal*. Kelima indeks tersebut akan diranking menurut kriteria masing-masing. Jumlah *cluster* optimal diperoleh pada rata-rata ranking yang paling tinggi. Metode Gap Statistik memiliki ciri khusus yaitu membangkitkan data dengan menggunakan Distribusi Uniform (0,1). Indeks Validitas gabungan (IVG) dan metode Gap Statistik akan diterapkan pada dua Data. Data 1 merupakan penelitian tentang gen mutan ragi mec 1 dan dun 1 dengan pemberian *MMS* (*Methylating-agent Methylmethane Sulfonate*), Data 2 merupakan data sporulasi gen jamur. Pada data 1, Indeks Validitas Gabungan (IVG) menghasilkan 2 *cluster* optimal sedangkan metode Gap Statistik menghasilkan 5 *cluster optimal*. Kedua indeks tersebut akan dibandingkan berdasarkan nilai CTM ***(****Cluster Tightness Measure)* yang digunakan untuk mengukur kebaikan hasil analisis *cluster*. Berdasarkan nilai CTM Gap Statistik lebih baik digunakan jika diterapkan pada Data 1. Sedangkan pada Data 2, Indeks Validitas Gabungan (IVG) dan metode Gap Statistik memiliki kemampuan yang sama dalam menentukan banyaknya *cluster* optimal yaitu membentuk 2 *cluster*. Kedua metode dalam menentukan banyaknya *cluster* optimal tersebut menghasilkan komposisi anggota yang sama sehingga tidak perlu membandingkan keduanya berdasarkan nilai CTM.

Kata kunci: Analisis *Cluster, Cluster* Optimal, Indeks Validitas, Gap Statistik, Data Ekspresi Gen

ABSTRACT

Gene expression data have amount big data relative so that its not easy to estimating the number of cluster from a gathering of gene expression data. One of statistical component which commonly use is cluster analysis. After analysis of cluster done, it is important to know how many optimal cluster in a data gathering. This research aim is to determine the number of optimal cluster with Gap statistic method and Index Validity Composite at gene expression data and determine the best method the number of cluster optimal among both. Index Validity Composite consist of Dunn's index, Davies-Bouldin Index, C-Index, Silhoutte Index, and Goodman Kruskal Index. Fifth of index will be rank according to each criterion. Amount of optimal cluster obtained at mean with ranking highest. While Gap Statistic method, in determining the number of cluster optimal, owning special characteristic that is generate the data by using distribution of Uniform (0,1). Both of this optimal cluster will be applied for the two Data. Data 1 representing about the experiment of yeast response to MMS and data 2 representing expression of fungus gene sporulation. At data 1 Index Validity composite is produce 2 optimal cluster, whereas Gap Statistic method is produce 5 optimal cluster. Both of index give different result, hence will be compared according to value of CTM. CTM represent a value that used to measure the goodness of result cluster. Formed to be group to be told goodness if owning small value of CTM. Method Gap-Statistic have smaller value of CTM in comparison with Index Validity Composite. So that Gap Statistic will be better to used if applied at Data 1. While for Data 2, Index Validity Composite and Statistical Gap method have same ability at determining the number of cluster optimal that is forming 2 cluster. The both method of the number cluster optimal is yield the same member composition so that needn't to compare both according to value of CTM.

Keywords: Cluster analysis, optimal cluster, validity index, Gap Statistic, gene expression data

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul Perbandingan Metode Gap-Statistik Dan Indeks Validitas Gabungan (IVG) dalam Menentukan Banyaknya *Cluster* Optimal pada Analisis Cluster Menggunakan Data Ekspresi Gen sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains dalam bidang Statistika. Oleh karena itu, penulis mengucapkan rasa hormat dan terima kasih kepada :

1. Ibu Suci Astustik, Ssi.,Msi. selaku Dosen Pembimbing I atas kesabaran dan pengarahan kepada penulis selama penyusunan skripsi ini.
2. Bapak Adji Achmad Rinaldo Fernandes, Ssi., Msc. selaku Dosen Pembimbing II atas kesabaran dan arahan yang telah diberikan kepada penulis selama penyusunan skripsi ini.
3. Ibu Ani Budi Astuti, M.Si, Ir. Soepraptini, MSc dan Eni Sumarminingsih, SSi., MM. selaku Dosen Penguji.
4. Bapak dan Ibu Dosen Statistika atas didikan selama kuliah hingga penulis bisa menyelesaikan kuliah.
5. Keluarga Besarku, Bapak, Ibu, Kak Rina dan Kak Rostrina yang selama ini telah memberikan perhatian, dukungan, motivasi serta kasih dan sayang selama ini
6. Semua pihak yang tidak dapat penulis sebutkan satu per satu yang telah banyak membantu dan memberikan dorongan selama penulisan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan mengingat keterbatasan kemampuan penulis. Untuk itu, dengan segala kerendahan hati penulis mengharap kritik dan saran. Akhirnya penulis berharap semoga skripsi ini dapat bermanfaat bagi pembaca.

Malang, 16 Maret 2009

Penulis

DAFTAR ISI

Halaman

Halaman

DAFTAR GAMBAR

Halaman

DAFTAR TABEL

Halaman

DAFTAR LAMPIRAN

Halaman

BAB I
PENDAHULUAN

1.1 Latar Belakang

Data tentang ekspresi gen semakin banyak seiring perkembangan teknologi *microarray*. *Microarray* adalah bagian baru di bidang teknologi yang digunakan untuk mengamati ekspresi gen. Ekspresi gen merupakan perbedaan pola penyediaan protein yang bervariasi antar obyek di mana gen merupakan rantaian DNA yang menurunkan sebuah protein dan memiliki fungsi tertentu di dalam sebuah sel penyusun tubuh (Meiyanto, 2001). Gen tersebut ditranskripsikan (disalin) menjadi mRNA (*messenger*RNA) yang selanjutnya mRNA ditranslasikan menjadi protein. Data ekspresi gen ini memiliki jumlah data yang relatif besar sehingga bukan suatu hal yang mudah untuk menentukan berapa jumlah *cluster* (*cluster*) dari suatu himpunan data ekspresi gen. Salah satu metode statistik yang cukup umum digunakan dalam menentukan banyaknya *cluster* pada data ekspresi gen adalah dengan menggunakan analisis *cluster*.

Analisis *cluster* merupakan suatu metode pengelompokkan di mana data yang akan di*cluster*kan belum membentuk *cluster* sehingga pengelompokkan yang akan dilakukan bertujuan agar data yang terdapat di dalam *cluster* yang sama relatif lebih homogen daripada data yang berada pada *cluster* yang berbeda (Iman, 2008). Tetapi, setelah analisis *cluster* dilakukan, perlu diketahui berapa *cluster* optimal dalam suatu himpunan data. *Cluster* optimal merupakan *cluster* yang padat antar obyek dalam *cluster* dan terisolasi dari *cluster* lain dengan baik . Tujuan dari *cluster* optimal adalah mendapatkan hasil pengelompokkan di mana bisa memberikan informasi sebenarnya mengenai data tersebut.

Wulandari (2006) melakukan penelitian tentang indeks-indeks *cluster* optimal (Indeks Validitas) di antaranya Indeks *Dunn* (ID), Indeks *Davies-Bouldin* (IDB), Indeks C (IC), Indeks *Silhouette-Rousseauw* (ISR) dan Indeks *Goodman-Kruskal* (IGK). Penelitian tersebut melakukan Indeks Validitas Gabungan (IVG) dengan mengkombinasikan indeks-indeks validitas tersebut yang kemudian dapat dipilih jumlah *cluster* optimalnya pada saat indeks tersebut berkombinasi paling banyak.

Tibshirani, *et al* (2000) mengusulkan metode Gap Statistik dalam menentukan jumlah *cluster* optimal pada sejumlah data. Untuk itu

pada penelitian ini akan membandingkan indeks *cluster* optimal yang paling baik antara metode Gap Statistik dan Indeks Validitas Gabungan (IVG) jika diterapkan pada data ekspresi gen dengan metode pengelompokkan hirarki. Kedua indeks *cluster* optimal tersebut akan dibandingkan berdasarkan nilai *Cluster Tightness Measure* (CTM).

CTM merupakan suatu nilai yang digunakan untuk mengukur kebaikan hasil analisis *cluster* berdasarkan simpangan baku tiap peubah pada tiap *cluster*. Indeks validitas yang menghasilkan CTM terkecil merupakan indeks validitas terbaik.

1.2 Rumusan Masalah

1. Berapa banyaknya *cluster* optimal yang dihasilkan metode Gap Statistik dan Indeks Validitas Gabungan (IVG) pada data ekspresi gen?

2. Metode penentuan banyaknya *cluster* optimal apa yang terbaik antara Gap Statistik dan Indeks Validitas Gabungan (IVG) pada data ekspresi gen berdasarkan nilai CTM?

1.3 Tujuan

1. Menentukan banyaknya *cluster* optimal dengan metode Gap Statistik dan Indeks Validitas Gabungan (IVG) pada data ekspresi gen.

2. Membandingkan metode penentuan banyaknya *cluster* optimal yang terbaik antara metode Gap Statistik dan Indeks Validitas Gabungan (IVG) pada data ekspresi gen berdasarkan nilai CTM.

1.4 Batasan Masalah

Permasalahan dibatasi sampai dengan pengelompokkan hirarki pada analisis *cluster* dalam menentukan banyaknya *cluster* optimal menggunakan metode Gap Statistik dan Indeks Validitas Gabungan (IVG) .

1.5 Manfaat Penelitian

Manfaat penelitian ini adalah menerapkan metode Gap Statistik dan Indeks Validitas Gabungan (IVG) untuk mendapatkan *cluster* optimal dari analisis *cluster* pada data *microarray* ekspresi gen agar diperoleh informasi yang sebenarnya dari data tersebut.

BAB II
TINJAUAN PUSTAKA

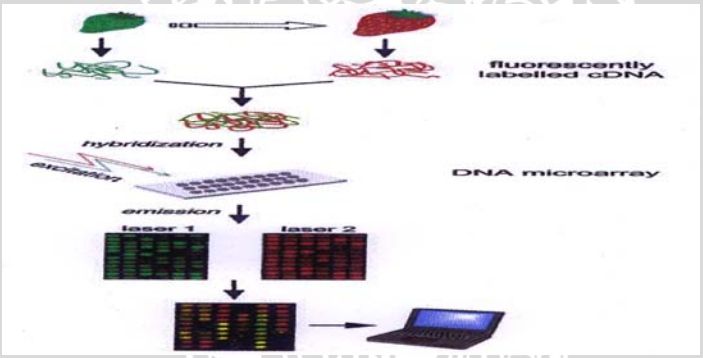
2.1 DNA *Microarray*

DNA adalah materi genetik yang membawa informasi. Menurut Speed (2008), kode protein gen ditentukan untuk mengungkapkan sebuah sel ketika kode protein tersebut menyampaikan pesan RNA (mRNA). Di dalam sel manusia, DNA dapat ditemukan di dalam inti sel dan di dalam mitokondria. DNA *microarray* merupakan teknologi percobaan untuk mempelajari gen dengan mengidentifikasi dan menghitung tingkat ekspresi gen semua jenis organisme. Dengan adanya DNA *microarray* akan bisa diketahui gen-gen apa saja yang aktif terhadap perlakuan tersebut.

Data *microarray* juga merupakan jenis data yang dipakai dalam bioinformatika. Bioinformatika merupakan teknologi pengumpulan, penyimpanan, analisis, interpretasi dan aplikasi dari data biologi molekul. Dengan bioinformatika, data yang dihasilkan dari proyek genom dapat disimpan teratur dalam waktu singkat dengan tingkat akurasi tinggi. Genom merupakan materi genetik yang merupakan suatu kumpulan gen-gen dari suatu makhluk hidup.

2.1.1 Proses Microarray

Gambar 2.1 merupakan proses untuk mendapatkan data *microarray* (Aharoni dan Vorst, 2001).



Gambar 2.1. Proses Mendapatkan Data *Microarray*

Pada Gambar 2.1 tingkat mRNA buah hijau (kontrol) dibandingkan dengan tingkat buah merah (perlakuan) dari buah yang diteliti. Proses tentang DNA *microarray* dimulai dari mRNA yang diisolasi dari obyek dan dikembalikan terlebih dahulu dalam bentuk DNA menggunakan reaksi *reverse transcription*. Setelah reaksi *reverse transcription*, DNA yang komplementer (cDNA) akan didapatkan dan dilabeli dengan warna *fluorescent* berbeda (hijau dan merah). Selanjutnya cDNA yang telah dilabeli tersebut dicampur dan dihibridisasi pada cDNA *microarray*. Setelah dihibridisasi, *microarray* discan dengan menggunakan sinar laser untuk memancarkan warna *fluorescent.* Jumlah mRNA relatif tiap gen adalah warna hijau dan merah yang direfleksikan dengan rasio banyaknya warna hijau dan merah yang diukur sebagai warna *fluorescent.* Apabila ekspresi gen buah hijau lebih banyak dari buah merah, maka warna yang terbentuk adalah kuning begitupun sebaliknya akan terbentuk warna merah. Jika kedua gen tidak berekspresi, maka warna gelap akan terbentuk (Draghici, 2003).

2.1.2 Data *Microarray* Ekspresi Gen

Data *microarray* ekspresi gen adalah data berupa rasio banyaknya warna merah dan hijau atau sebaliknya. Bahkan beberapa peneliti menggunakan log rasio untuk tiap jenis gen. Untuk kemudahan para peneliti biasanya menggunakan logaritma basis 10. Penggunaan basis berbeda dalam logaritma tidak akan mengubah hasil analisis (Speed, 2008).

Dengan adanya ekspresi dari gen-gen tersebut, maka akan bisa mengetahui dan meningkatkan informasi tentang data ekspresi gen (D'haeseleer, *et al*. 1999). Salah satu langkah dalam analisis ekspresi gen adalah dengan mendeteksi tingkat kesamaan ekspresi dari gen tersebut (Bolshakova dan Azuaje, 2003). Analisis yang paling sering digunakan dalam data ekspresi gen adalah analisis *cluster*. Hal ini disebabkan karena kekuatan teknik analisis *cluster* dalam pengelompokkan gen akan dapat menerangkan gen yang karakternya tidak diketahui (Speed, 2008). Dengan analisis *cluster* juga diharapkan dapat mengelompokkan gen-gen sesuai dengan karakter gen masing-masing

2.2 Analisis *Cluster*

Menurut Mattjik, *et al* (2002), analisis *cluster* adalah suatu metode dalam analisis peubah ganda yang bertujuan untuk mengelompokkan n (obyek) satuan pengamatan ke dalam *k cluster* berdasar peubah, sehingga unit-unit pengamatan dalam satu *cluster* mempunyai ciri-ciri yang lebih homogen dibandingkan unit pengamatan dalam *cluster* lain. Salah satu metode pengelompokkan dalam analisis *cluster* adalah metode pengelompokkan hirarki.

Sebelum melakukan analisis *cluster*, hal pertama yang akan dilakukan adalah membentuk matriks data $\vec{X}_{nxp}$ yaitu matriks pengukuran n obyek dengan p variabel.

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

dimana :
$\mathbf{X}_{ik}$ adalah data pada obyek ke-i dan variabel ke-k
i = 1, 2, 3,...n
k = 1, 2, 3, ...p

Setelah membentuk matriks data $\vec{X}_{nxp}$, akan dibentuk matriks proksimitas. Adapun format dari matriks *proksimitas* yaitu :

$$d(x_a,x_b) = \begin{bmatrix} d_{(x_1,x_1)} & d_{(x_1,x_2)} & \dots & d_{(x_1,x_n)} \\ d_{(x_2,x_1)} & d_{(x_2,x_2)} & \dots & d_{(x_2,x_n)} \\ . & & & \\ . & & & \\ . & & & \\ d_{(x_n,x_1)} & d_{(x_n,x_2)} & \dots & d_{(x_n,x_n)} \end{bmatrix}$$

di mana :

a, b = 1,...,n
n = banyaknya obyek
$d(x_a,x_b)$ = jarak antara obyek ke-*a* dan obyek ke-*b* di mana $a \neq b$

Metode pengelompokkan hirarki digunakan jika banyak *cluster* yang dikehendaki tidak diketahui. Pada dasarnya, ada dua tipe metode hirarki yaitu *agglomerative* dan *divisive* (Hair, *et al*, 1998). Proses *agglomerative* (penggabungan) adalah masing-masing obyek dianggap satu *cluster* kemudian antar *cluster* yang jaraknya berdekatan bergabung menjadi satu *cluster*. Sedangkan proses *divisive* (pemecahan) pada awalnya semua obyek berada dalam satu *cluster*. Kemudian sifat obyek yang paling beda (jarak yang paling jauh) dipisahkan dan membentuk satu *cluster* yang lain. Proses *divisive* berlanjut sampai semua obyek tersebut masing-masing membentuk satu *cluster*. Dengan demikian proses pengelompokkannya dilakukan secara bertingkat atau bertahap.

2.3 Jarak *Euclid* (*Euclidean Distance)*

Jarak *euclid* merupakan salah satu konsep jarak dalam analisis *cluster* (Hair, *et al*,1998). Jarak ini merupakan salah satu ukuran kedekatan yang cukup sering digunakan.

Rumus jarak *euclid* adalah :

$$d(x_a,x_b) = \sqrt{\sum_{j=1}^{p}(x_{aj} - x_{bj})^2} \qquad (2.1)$$

di mana : $d(x_a,x_b)$ = jarak antara obyek ke-a dan ke-b
x_{aj} = obyek ke-*a* untuk nilai peubah ke-*j*
x_{bj} = obyek ke-*b* untuk nilai peubah ke-*j*

2.3.1 Jarak *Intercluster* (*Intercluster Distance*)

Menurut Bolshakova (2001), ada enam metode *intercluster distance*.Metode *intercluster distance* adalah ukuran jarak antara obyek pada *cluster* yang berbeda, dituliskan dengan $d(c_1,c_2)$. Jarak-jarak tersebut adalah sebagai berikut :

1. *Average Linkage*

Average linkage yaitu rata-rata jarak seluruh obyek suatu *cluster* terhadap seluruh obyek pada *cluster* lainnya. Langkah pertama adalah menemukan jarak terdekat antar obyek, kemudian menggabungkan obyek-obyek tersebut dengan menggunakan jarak rata-rata antar tiap pasangan obyek yang mungkin. Jarak *average linkage* dirumuskan dengan :

$$d(c_1 c_2) = \frac{1}{n_1 n_2} \sum_{\substack{x_a \in c_1 \\ x_b \in c_2}} d(x_a, x_b) \qquad (2.2)$$

Menurut Hair, *et al* (1998) *average linkage* lebih baik karena kriteria penggabungan *cluster* tidak berdasarkan jarak ekstrim melainkan rata-rata jarak dari semua obyek dalam suatu *cluster*.

2. C*entroid Linkage*

Jarak antara pusat dua *cluster* dirumuskan dengan :

$$d(c_1,c_2)=d(vc_1,vc_2) \qquad (2.3)$$

di mana :

$$vc_1 = \frac{1}{n} \sum_{x_a \in c_1} x_a$$

$$vc_2 = \frac{1}{n} \sum_{x_b \in c_2} x_b$$

vc_1 adalah pusat pada *cluster* 1 dan vc_2 adalah pusat pada *cluster* 2.

3. C*omplete Linkage*

Jarak antara obyek paling jauh dari dua *cluster* berbeda dirumuskan dengan :

$$d(c_1, c_2) = \max\left\{ d(x_a , x_b \right\} \qquad (2.4)$$
$$\substack{x_a \in c_1 \\ x_b \in c_2}$$

4. *Single Linkage*

Jarak antara obyek paling dekat dari dua *cluster* berbeda dirumuskan dengan :

$$d(c_1, c_2) = \min\left\{ d(\underset{x_a \in c_1}{x_a}, \underset{x_b \in c_2}{x_b}) \right\} \quad (2.5)$$

5. *Average To Centroid Linkage*

Jarak antara pusat dari suatu *cluster* dan semua obyek dari *cluster* berbeda dirumuskan dengan **:**

$$d(c_{1,} c_2) = \frac{1}{n_1 + n_2}\left(\sum_{x_a \in c_1} d(x_a, vc_2) + \sum_{x_b \in c_2} d(x_b, vc_1) \right) \quad (2.6)$$

di mana merupakan obyek yang berada pada *cluster* 1 dan *y* merupakan obyek yang berada pada *cluster* 2.

6. *Hausdorff Metric*

Jarak maksimal dari obyek pada suatu *cluster* dengan obyek yang paling dekat pada *cluster* yang lain dirumuskan dengan :

$$d(c_1,c_2)=\max\left\{ d_1(c_1 c_2), d_2(c_2 c_1) \right\} \quad (2.7)$$

di mana :

$$d_1(c_1, c_2) = \max_{x_a \in c_1}\left\{ \min_{x_b \in c_2}\left\{ d(x_a, x_b) \right\} \right\}$$

$$d_2(c_2, c_1) = \max_{x_b \in c_2}\left\{ \min_{x_a \in c_1}\left\{ d(x_a, x_b) \right\} \right\}$$

2.3.2 Jarak *Intracluster* (*Intracluster Distance)*

Jarak *intracluster* merupakan ukuran jarak antar obyek pada *cluster* yang sama. Terdapat tiga macam *intracluster* yang dituliskan dengan d(c) yaitu :

1. C*omplete Diameter*

Jarak ini menunjukkan jarak antara obyek yang paling jauh pada *cluster* yang sama. Jarak *intracluster* ini dirumuskan dengan :

$$d(c) = \max_{x_a, x_b \in c}\left\{ d(x_a, x_b) \right\} \quad (2.8)$$

2. *Average Diameter*

Jarak ini menunjukkan rata-rata jarak antara semua obyek dari *cluster* yang sama. Jarak *intracluster* ini dirumuskan dengan :

$$d(c) = \frac{1}{n+(n-1)} \sum_{\substack{x_a, x_b \in C \\ x_a \neq x_b}} d(x_a, x_b) \qquad (2.9)$$

3. C*entroid Diameter*

Jarak ini menunjukkan dua kali rata-rata jarak antara semua obyek dan pusat *cluster* . Jarak *intracluster* ini dirumuskan dengan:

$$\text{d(c)} = 2\left[\frac{\sum_{x \in C} d(x,v)}{n}\right] \qquad (2.10)$$

$$\text{v} = \frac{1}{n}\sum_{x \in C} x$$

Metode jarak *intercluster* dan *intracluster* yang sudah dijelaskan tersebut akan digunakan pada Indeks Validitas *Dunn* dan Indeks Validitas *Davies-Bouldin* dengan mengkombinasikan 6 metode jarak *intercluster* dan 3 metode jarak *intracluster* tersebut.

2.4 Indeks Validitas *Cluster*

Validitas *cluster* merupakan suatu nilai untuk mengevaluasi hasil dari analisis *cluster* sehingga dihasilkan *cluster* optimal. Berikut ini merupakan indeks-indeks validitas *cluster* untuk mendapatkan *cluster* optimal. Indeks tersebut adalah Indeks *Dunn* (ID), Indeks *Davies-Bouldin* (IDB), Indeks C (IC), Indeks *Silhoutte-Rousseauw* (ISR) dan Indeks *Goodman-Kruskal* (IGK). Indeks-indeks ini akan digabung agar mendapatkan Indeks Validitas Gabungan (IVG).

2.4.1 Indeks *Dunn* (ID)

Indeks *Dunn* (ID) bisa dihitung dengan rumus 2.11 :

$$ID = \min_{\substack{1 \le i \le n \\ 1 \le j \le n \\ i \neq j}} \left\{ \frac{d(c_i, c_j)}{\max_{1 \le k \le n} \{d'(c_k)\}} \right\} \qquad (2.11)$$

di mana :

d (c_i ,c_j) = jarak *intercluster* pada *cluster* ke-*i* dan ke-*j*

d'(c_k) = jarak *intracluster* pada *cluster* ke-k

Tujuan utama indeks ini adalah untuk memaksimalkan jarak *intercluster* dan meminimumkan jarak *intracluster*. Sehingga nilai terbesar dari D diambil sebagai jumlah *cluster* optimal. Jarak *intercluster* dan *intracluster* untuk Indeks *Dunn* (ID) ini menggunakan persamaan 2.2 sampai dengan persamaan 2.10.

2.4.2 Indeks *Davies-Bouldin* (IDB)

Rumus Indeks *Davies-Bouldin* (IDB) sebagai berikut :

$$\text{IDB}= \frac{1}{n}\sum_{i=1}^{n}\max\left\{\frac{d'(c_i)+d'(c_j)}{d(c_i,c_j)}\right\} \qquad (2.12)$$

di mana :

n = jumlah *cluster*

d(c_i,c_j) = jarak *intercluster* pada *cluster* ke-*i* dan ke-*j*

d'(c_i) = jarak *intracluster* pada *cluster* ke-*i*

d'(c_j) = jarak *intracluster* pada *cluster* ke-*j*

Seperti Indeks *Dunn* (ID), Indeks *Davies-Bouldin* (IDB) juga bertujuan untuk mengidentifikasikan *cluster* sebagai *cluster* yang padat dan terpisah dengan *cluster* lain dengan baik. Karenanya nilai Indeks *Davies Bouldin* (IDB) kecil jika *cluster- cluster* tersebut padat dan jauh satu sama lain. Sebagai konsekuensi, indeks *Davies-Bouldin* akan mempunyai suatu nilai kecil untuk suatu *cluster* yang baik (Su,2003). Jarak *intercluster* dan *intracluster* yang digunakan adalah persamaan 2.2 sampai dengan 2.10.

2.4.3 Indeks *C* (IC)

Indeks ini dapat dirumuskan sebagai berikut:

$$C = \frac{S - S_{\min}}{S_{\max} - S_{\min}} \qquad (2.13)$$

di mana :

S = jumlah jarak pada semua pasangan obyek amatan dari *cluster* yang sama

dengan *l* merupakan jumlah pasangan tersebut,

S_{min} = jumlah dari *l* dengan jarak terkecil jika semua pasangan obyek amatan dari *cluster* yang sama

S_{max} = jumlah dari *l* dengan jarak terbesar dari semua pasangan.

Nilai C yang kecil menandai *cluster* yang baik (Bolshakova, 2001).

2.4.4 Indeks *Silhouette-Rousseauw* (ISR)

Untuk mendapatkan hasil *Silhouette* digunakan rumus sebagai berikut :

$$S(i) = \frac{(h(i) - k(i))}{\max\{k(i), h(i)\}} \quad (2.14)$$

di mana :

k(i) = rata-rata perbedaan (rata-rata jarak) dari i-obyek dengan semua obyek lain di dalam *cluster* yang sama

h(i) = nilai minimum dari rata-rata perbedaan (rata-rata jarak) dari i-obyek dengan semua obyek pada *cluster* lain (di *cluster* terdekat).

Nilai S(i) berada di antara -1 dan 1. Jika nilai *Silhouette* dekat dengan 1, berarti obyek tersebut telah di*cluster*kan dengan baik. Jika nilai *silhouette* dekat dengan 0 ini berarti bahwa obyek dapat ditempatkan pada *cluster* terdekat. Jika nilai *silhouette* dekat dengan -1 maka diartikan bahwa pengelompokkannya salah. Nilai yang paling besar dari keseluruhan rata-rata *Silhoutte* (Global *Silhoutte)* menandai jumlah *cluster* terbaik dan diambil sebagai *cluster* optimal. Rumusan Global *Silhoutte* sebagai berikut:

$$GS_u = \frac{1}{c}\sum_{j=1}^{c} S_j \quad (2.15)$$

di mana :

GS_u = *Global Silhouette* pada pembagi-u

S_j = *Silhouette Cluster* ke-j

C = Banyaknya *Cluster* (Bolshakova and Azuaje, 2003)

2.4.5 Indeks *Goodman-Kruskal* (IGK)

Misalkan ada 4 pasang obyek yaitu q, r, s dan t dengan *d* merupakan jarak antara q dan r atau s dan t. Empat pasang obyek tersebut dikatakan konkordan jika memenuhi kondisi yaitu: d(q,r)<d(s,t), di mana q dan r berada pada *cluster* yang sama dan s dan t pada *cluster* yang berbeda. Sebaliknya, empat pasang obyek dikatakan diskordan jika memenuhi kondisi yaitu : d(q,r)<d(s,t),di mana q dan r berada pada *cluster* yang berbeda dan s dan t pada *cluster* yang sama

Indeks IGK dihitung dari nilai hasil perhitungan nilai pasangan konkordan dan diskordan dengan rumus :

$$IGK = \frac{S_c - S_d}{S_c + S_d} \tag{2.16}$$

di mana :

S_c = Jumlah pasangan konkordan
S_d = Jumlah pasangan diskordan
IGK = Nilai Indeks *Goodman Kruskal*

Nilai-nilai Indeks *Goodman Kruskal* (IGK) yang besar menunjukkan *cluster* yang baik (Bolshakova, 2001). *Cluster* yang baik harus memiliki dua hal yaitu homogenitas yang tinggi antar anggota dalam *cluster* (*withincluster*) dan heterogenitas (perbedaan) yang tinggi antar *cluster* yang satu dengan yang lainnya.

2.5 Indeks Validitas Gabungan (IVG)

Pouwels (1998) memberikan alternatif untuk memilih jumlah *cluster* optimal dengan mengkombinasikan indeks validitas *cluster* yang kemudian dapat dipilih jumlah *cluster* optimalnya pada saat indeks tersebut berkombinasi paling banyak. Langkah dalam IVG ini adalah dengan menghitung kelima indeks validitas lalu memberi ranking tiap jumlah *cluster* yang mungkin pada masing-masing indeks. Jumlah *cluster* optimal diperoleh pada rata-rata ranking yang paling tinggi. Dari penelitian Mufidah (2004), indeks validitas yang

berbeda akan menghasilkan *cluster* optimal yang berbeda dan menganjurkan untuk menggunakan Indeks Validitas Gabungan. Tetapi menurut Tibshirani,*et al* (2001) metode Gap Statistik dapat digunakan dalam menentukan banyaknya *cluster* optimal pada suatu himpunan data dengan membangkitkan data referensi distribusi uniform U (0,1)

2.6 Gap Statistik

Menurut Arima *et al* (2008), gap statistik merupakan metode untuk menduga *cluster* optimal pada analisis *cluster*. Secara detil dapat dijelaskan sebagai berikut : Misalkan X_{ik} dengan i = 1,2,….,n dan j=1,2,….,p terdiri dari himpunan data p (peubah) pada observasi *independent* n (obyek) , kemudian data di *cluster*kan menjadi k *cluster*, yaitu $C_1,C_2,\ldots,C_k$ dengan C_r menandakan indikasi pengamatan pada *cluster r*, dan n_r adalah banyaknya observasi (anggota) *cluster* ke-r. Kemudian didefinisikan sebagai berikut:

$$D_r = \sum_{x_a, x_b} d(x_a, x_b) \qquad (2.17)$$

di mana :

D_r = jarak *euclid* data observasi

$d(x_a,x_b)$ = jarak antara obyek ke-*a* dan obyek ke-*b* di mana $a \neq b$

Kemudian menghitung W_k :

$$W_k = \sum_{r=1}^{k} \frac{1}{2_{n_r}} D_r \qquad (2.18)$$

dengan W_k adalah jumlah kuadrat dalam *cluster*.

Menurut Tibshirani, *et al* (2000), prosedur Gap Statistik sebagai berikut :

1. Mengelompokkan data observasi dengan menggunakan metode *average linkage* (hirarki) dengan banyaknya *cluster* adalah K dan besarnya K berjalan dari $1 \leq k \leq n$ di mana n menyatakan banyaknya obyek yang di*cluster*kan.
2. Menentukan K optimum metode Gap Statistik dengan langkah-langkah :
   a. Mendapatkan nilai W_k untuk tiap-tiap K (banyaknya *cluster*), kemudian menghitung Log W_k

b. Melakukan resampling (dari data simulasi) dengan pengembalian sebanyak B kali dengan distribusi uniform U(0.1).
c. Mendapatkan nilai W^*_{kb} dari poin (b) di mana b=1,2,...,B dan k = 1,2,...,K
d. Menghitung standar deviasinya (sd_k) dan mendapatkan :

$$Sd_k = \left[\left(\frac{1}{B}\right)\sum_b \{Log(W^*_{kb}) - \ell\}^2\right]^{\frac{1}{2}} \quad (2.19)$$

di mana :

$$\ell = (1/B)\sum_b \log(W * kb)$$

e. Menghitung S_k dengan rumus :

$$s_k = sd_k\sqrt{(1+1/B)} \quad (2.20)$$

f. Menghitung *cluster* optimal dengan Gap Statistik

$$Gap(k) = \left[\frac{1}{B}\right]\sum_b \log(W^*_{kb}) - Log(W_k) \quad (2.21)$$

di mana B merupakan resampling dengan distribusi uniform U(0,1) dan B = n (jumlah observasi)

g. Banyaknya *cluster* (k) optimum ditentukan melalui nilai k terkecil sedemikian hingga :

$$Gap(k) \geq Gap(k+1)-s_{k+1} \quad (2.22)$$

Indeks Validitas Gabungan (IVG) dan metode Gap Statistik merupakan dua metode pengelompokkan optimal yang dibahas dalam penelitian ini. Untuk mengetahui kebaikan dari masing-masing *cluster* optimal yang terbentuk, keduanya akan dibandingkan berdasarkan nilai *Cluster Tightness Measures (CTM)*.

2.7 *Cluster Tightness Measure* (CTM)

Dalam rangka mengukur suatu kebaikan dari suatu *cluster*, Epps dan Ambikairajah (2008) menganjurkan suatu rumus pengukuran berdasarkan pada simpangan baku dari tiap peubah. Rumus tersebut adalah :

$$\text{CTM} = \frac{1}{A}\sum_{a=1}^{A}\left[\frac{1}{p}\sum_{j=1}^{p}\frac{\sigma_j^a}{\sigma_j^n}\right] \qquad (2.23)$$

di mana :

σ_j^a = simpangan baku pada *cluster* ke-a untuk peubah ke-j

σ_j^n = simpangan baku seluruh data untuk peubah ke-j

p = banyaknya peubah

A = banyaknya *cluster*

Cluster yang terbentuk dikatakan baik jika memiliki CTM kecil (Epps dan Ambikairajah, 2006).

BAB III
METODOLOGI PENELITIAN

3.1 Data Penelitian

Data adalah log rasio banyaknya warna merah dengan hijau yang terbentuk pada waktu tertentu. Cara memperoleh data ekspresi gen adalah sebagai berikut : cDNA yang telah disiapkan dari 2 sampel terlebih dahulu dilabel dengan *fluorescent* (bahan celup) yaitu *Cyanine*-3 (Cy 3 ^{TM}) yang berwarna hijau untuk gen kontrol dan *Cyanine*-5 (Cy 5 ^{TM}) yang berwarna merah untuk gen yang diberi perlakuan. Setelah pelabelan dilakukan hibridisasi dan kedua sampel dicampur pada cDNA *microarray.* Setelah proses hibridisasi akan dilakukan pengukuran ekspresi dengan pemancaran gelombang warna dari sinar laser. Dengan menggunakan bantuan *software, image file fluorescent* dapat diubah menjadi nilai numerik. Dari alat *microarray* tersebut, akan dihitung banyaknya bintik-bintik warna merah dan hijau dan kemudian dirasiokan. Untuk bintik-bintik yang kosong dan lebih lemah dari rata-ratanya,akan dihilangkan dengan cara melakukan proses filtering. Setelah proses filtering tersebut, data ditransformasi dengan logaritma dengan menggunakan bantuan software. Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari *http://www-genome.stanford* yaitu :

1. Data 1
   Data percobaan gen ragi mutan *mec 1* dan mutan *dun 1* yang sengaja dirusak oleh *MMS* (*Methylating-agent Methylmethane Sulfonate*) di mana *MMS* merupakan bahan kimia yang merusak DNA. Sebanyak 159 gen mutan *mec 1* diamati pada saat 5 , 15, 30, 45, 60, 90 dan 120 menit sedangkan untuk *dun 1* diamati pada saat 30, 90 dan 120 menit. Penelitian ini untuk melihat apakah setelah pemberian *MMS* ekspresi gen menjadi on (aktif) atau off (pasif) dalam pengamatan waktu tertentu. Jika berekspresi positif, berarti gen tersebut on (aktif) yang menandakan *MMS* tersebut bekerja. Sebaliknya jika off (pasif) berarti *MMS* tersebut tidak bekerja. Data merupakan log rasio banyaknya warna merah dengan hijau yang terbentuk pada waktu pengamatan tersebut. Warna hijau merupakan warna pelabelan cDNA untuk kontrol sedangkan warna hijau untuk

gen yang diberikan perlakuan. Data ini merupakan penelitian dari Gasch *et al* (2001)**.**

2. Data 2
   Data terdiri dari 457 gen jamur yang sedang bersporulasi dan diamati pada waktu berbeda yaitu 0.5 , 2 , 5 , 7 , 9 dan 11.5 jam. Penelitian ini bertujuan untuk melihat pada saat kapan gen tersebut on (aktif) atau off (pasif) dalam melakukan sporulasi. Data merupakan log rasio banyaknya warna merah dengan hijau yang terbentuk pada waktu tersebut. Data ini adalah penelitian dari Eichenberger *et al* (2003).

Untuk data 1 dan data 2 dapat dilihat pada Lampiran 1 dan 2.

3.2 Metode Penelitian

3.2.1 Analisis *Cluster*

Langkah-langkah dalam analisis *cluster* adalah :

1. Membuat matriks proksimitas dengan menggunakan jarak *euclid* sesuai persamaan 2.1
2. Setelah terbentuk matriks proksimitas, kemudian mencari nilai minimum dari jarak antara 2 obyek kemudian digabungkan dalam satu *cluster*.
3. Menghitung kedekatan (jarak) antara *cluster* satu dengan satu atau beberapa obyek di luar *cluster* yang telah terbentuk sebelumnya menggunakan metode *average linkage* sesuai persamaan 2.2
4. Menyusun kembali matriks proksimitas yang baru setelah terbentuk *cluster* yang pertama
5. Mengulangi langkah 3 dan 4 sampai semua obyek berada dalam satu *cluster*
6. Menentukan *cluster* optimal menggunakan Indeks Validitas Gabungan (IVG) dan Gap Statistik

3.2.2 Indeks Validitas Gabungan (IVG) dan Gap Statistik

1. Indeks Validitas Gabungan (IVG)
   Langkah-langkah yang dilakukan adalah **:**

a. Menghitung Indeks *Dunn* (ID) dengan persamaan (2.11), Indeks *Davies-Bouldin* (IDB) dengan persamaan (2.12), Indeks C (IC) dengan persamaan (2.13) , Indeks *Silhouette-Rousseauw* (ISR) dengan persamaan (2.14) dan Indeks *Goodman-Kruskal* (IGK) dengan persamaan (2.15).
b. Memilih *cluster* optimal seperti yang dijelaskan pada sub bab 2.6.

2. Gap Statistik

Langkah-langkah yang dilakukan :

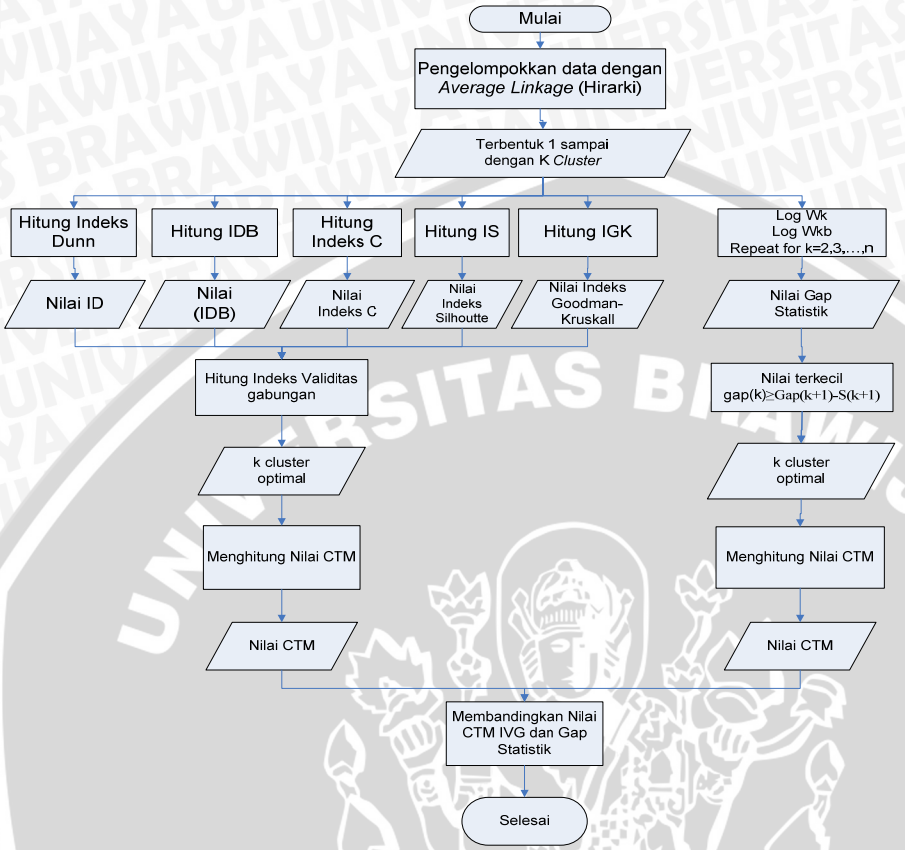
a. *Cluster* data observasi seperti yang sudah dijelaskan pada sub bab 2.6.
b. Hitung D_r dengan persamaan (2.17)
c. Hitung W_k, untuk k=1,2,…,K dengan persamaan (2.18)
d. Bangkitkan data referensi (B) : dengan distribusi uniform U(0,1) sebagai resampling sebanyak (n)
e. *Cluster* data referensi (B)
f. Hitung D_r untuk data referensi dengan rumus pada persamaan (2.17)
g. Hitung W^*_{kb}, di mana b=1,2,…,B dan k=1,2,…K dengan rumus pada persamaan (2.18)
h. Hitung log W^*_{kb}
i. Hitung Standar deviasi dengan persamaan (2.19)
j. Hitung S_k dengan persamaan (2.20)
k. Hitung *cluster* optimal dengan Gap Statistik menggunakan persamaan (2.21)
l. Cari nilai k terkecil sesuai dengan persamaan (2.22)

3.2.3 Membandingkan Indeks Validitas Gabungan (IVG) dan Gap Statistik

Setelah melakukan proses *cluster* optimal dengan menggunakan Indeks Validitas Gabungan (IVG) dan Gap Statistik, akan dilanjutkan dengan membandingkan hasil keduanya berdasarkan nilai CTM. Nilai CTM dapat dihitung dengan menggunakan persamaan (2.23). Dalam semua proses perhitungan yang dilakukan, digunakan bantuan perangkat lunak *Package Clustersim* pada *software* R 2.8.0 dan *software Machaon*.

3.3 Diagram Alir

Diagram alir dari Indeks Validitas Gabungan (IVG) dan Metode Gap Statistik adalah sebagai berikut :



Gambar 3.3. Diagram Alir Indeks Validitas Gabungan (IVG) dan Metode Gap Statistik

BAB IV
HASIL DAN PEMBAHASAN

4.1 Analisis *cluster*

Dalam pengelompokkan hirarki pada analisis *cluster*, terlebih dahulu dilakukan pembentukan matriks proksimitas yaitu matriks kedekatan obyek berdasarkan jarak *euclid.* Hasil proses analisis *cluster* pada data 1 dan data 2 dalah sebagai berikut :

1. Data 1

Matriks proksimitas untuk data 1 ditampilkan pada Lampiran 3. Matriks tersebut digunakan untuk proses pengelompokkan gen-gen dari mutan ragi *mec 1* dan *dun 1* dengan menggunakan metode *average linkage.* Proses pengelompokkan dimulai dengan mencari jarak terdekat antara obyek-obyek tersebut. Misalkan dari matriks proximitas pada Lampiran 3, gen ke-5 dan ke- 6 memiliki jarak terdekat yaitu 0.193 maka kedua gen tersebut bergabung menjadi satu *cluster*. Kemudian, proses pengelompokkan dilanjutkan kembali setelah *cluster* pertama telah terbentuk dengan menyusun matriks proksimitas yang baru. Demikian seterusnya sampai seluruh gen berada dalam satu *cluster.*

2. Data 2

Matriks proksimitas untuk data 2 ditampilkan pada Lampiran 4. Matriks tersebut digunakan untuk proses pengelompokkan gen-gen dari sporulasi jamur dengan menggunakan metode *average linkage.* Pada Lampiran 4, gen ke-3 dan ke- 457 memiliki jarak terdekat 1.152 maka kedua gen tersebut bergabung menjadi satu *cluster*. Kemudian, proses pengelompokkan dilanjutkan kembali dengan menyusun matriks proksimitas yang baru setelah *cluster* pertama telah terbentuk. Demikian seterusnya sampai seluruh gen berada dalam satu *cluster.*

Berdasarkan tahap pengelompokkan ini, belum diketahui berapa banyaknya *cluster* yang paling baik untuk dibentuk. Oleh karena itu, langkah selanjutnya adalah membentuk *cluster* optimal dengan menggunakan Indeks Validitas Gabungan (IVG).

4.2 Indeks Validitas Gabungan (IVG)

Indeks-indeks yang digunakan dalam Indeks Validitas Gabungan (IVG) pada penelitian ini adalah Indeks *Dunn* (ID), Indeks *Davies-Bouldin* (IDB), Indeks *C* (C), Indeks *Silhoutte-Rousseauw* (ISR) dan Indeks *Goodman-Kruskal* (IGK).

1. Data 1

Pada awalnya, untuk menentukan banyaknya *cluster* optimal dengan Indeks Validitas Gabungan (IVG), harus dihitung terlebih dahulu masing-masing nilai indeks pada tiap-tiap *cluster*. Adapun hasil kelima nilai indeks tersebut dapat dilihat pada Tabel 4.1.

Tabel 4.1 Indeks Validitas Gabungan Berdasarkan Banyaknya *Cluster* padaData 1

	Banyaknya *Cluster*						
Indeks	2	3	4	5	6	7	8
ID	**2.467**	2.0125	1.0817	1.043	0.823	0.865	0.76
IDB	**0.632**	1.431	1.7446	1.878	1.733	1.496	1.71
ISR	**0.67**	0.578	0.471	0.391	0.337	0.342	0.26
IC	**0.162**	0.089	0.1	0.059	0.057	0.057	0.07
IGK	**0.949**	0.929	0.855	0.874	0.876	0.81	0.81

Nilai Indeks *Dunn* (ID) dan Indeks *Davies-Bouldin* (IDB) pada Tabel 4.1 merupakan nilai rata-rata dari kombinasi 6 jarak *intercluster* dan 3 jarak *intracluster*. Sedangkan, untuk nilai Indeks *Silhoutte-Rousseauw* (ISR) yang ada pada tabel tersebut merupakan nilai *Gsu*-nya. Untuk selengkapnya dapat dilihat pada Lampiran 5 dan 6.

Pengelompokkan untuk Indeks Validitas Gabungan (IVG) ini dibentuk hanya sampai dengan 9 *cluster* saja. Hal ini disebabkan karena pembentukan *cluster* tersebut sudah bisa dianggap mewakili proses dalam mendapatkan *cluster* optimal. Sebagai contoh dapat dilihat nilai Indeks *Dunn* (ID) pada tabel tersebut. Sebagaimana yang telah diketahui, nilai Indeks *Dunn* (ID) yang besar berarti menunjukkan *cluster* yang baik. Saat terbentuk 2 *cluster,* nilai

Indeks *Dunn* (ID) besar, tetapi setelah lebih dari 2 *cluster*, nilai indeks akan semakin mengecil.

Contoh lainnya adalah Indeks *Davies-Bouldin* (IDB). Dari tabel tersebut dapat dilihat bahwa semakin besar banyaknya *cluster,* maka nilai indeks pun semakin besar. Seperti yang sudah diketahui, Indeks *Davies Bouldin* (IDB) akan memiliki suatu nilai yang kecil untuk *cluster* yang baik. Oleh karena itu 9 *cluster* sudah bisa dianggap mewakili dalam mendapatkan *cluster* optimal. Hal ini juga berlaku pada ketiga indeks lainnya.

Setelah dilakukan pembentukan Tabel 4.1, kelima indeks akan diranking menurut kriteria masing-masing indeks. Untuk Indeks *Dunn* (ID), Indeks *Silhoutte* (IS) dan Indeks *Goodman-Kruskal* (IGK), nilai yang paling besar menunjukkan *cluster* yang paling baik, maka akan diberi ranking yang paling tinggi. Sedangkan, untuk Indeks *Davies-Bouldin* (IDB) dan Indeks *C* (IC), nilai yang paling kecil menunjukkan *cluster* yang paling baik, sehingga nilai terkecil yang akan diberi ranking paling tinggi. Adapun hasil ranking yang didapatkan dari 5 indeks tersebut disajikan pada Tabel 4.2

Tabel 4.2 Ranking Berdasarkan Indeks Validitas Gabungan (IVG) pada Data 1

	Banyaknya *Cluster*							
Indeks	2	3	4	5	6	7	8	9
ID	8	7	6	5	3	4	1	2
IDB	8	7	2	1	3	6	4	5
ISR	8	7	6	5	3	4	2	1
IC	1	3	2	8	6.5	6.5	4	5
IGK	8	7	4	5	6.5	6.5	1.5	1.5
Rata-rata	**6.6**	6.2	4	4.8	4.4	5.4	2.5	2.9

Berdasarkan Tabel 4.2, Indeks *Dunn* (ID), Indeks *Davies-Bouldin* (IDB), Indeks *Silhoutte* (S) dan Indeks *Goodman-Kruskal* (IGK), menghasilkan 2 *cluster* optimal. Sedangkan Indeks *C* (IC) memberikan hasil berbeda dibandingkan ke-empat indeks lainnya. Indeks *C* (IC) memberikan 2 nilai *cluster* optimal yaitu 6 dan 7 *cluster*. Setelah ranking tersebut dirata-rata, seperti ke-empat indeks tersebut, 2 *cluster* merupakan *cluster* optimal. Jadi, berdasarkan Indeks Validitas Gabungan (IVG), *cluster* optimal yang terbentuk pada data 1 adalah sebanyak 2 *cluster*.

2. Data 2

Sama seperti data gen mutan ragi (Data 1), setelah dilakukan pengelompokkan hirarki, akan dicari *cluster* optimal yang terbentuk berdasarkan nilai Indeks Validitas Gabungan (IVG). Nilai-nilai indeks yang dihasilkan dari data sporulasi gen jamur ini dapat dilihat pada Tabel 4.3

Tabel 4.3 Nilai Indeks Validitas Gabungan (IVG) pada Data 2

	Banyaknya *Cluster*							
Indeks	2	3	4	5	6	7	8	9
ID	**1.674**	1.316	1	1.042	0.87	0.751	0.703	0.705
IDB	2.96	2.794	3.45	3.11	2.733	2.813	2.646	**2.525**
ISR	**0.591**	0.479	0.4	0.311	0.282	0.299	0.283	0.254
IC	0.032	**0.031**	0.06	0.047	0.047	0.048	0.048	0.045
IGK	0.919	0.88	**0.923**	**0.923**	0.912	0.915	0.917	0.921

Dari tabel 4.3, Indeks *Dunn* (ID) dan Indeks *Silhoutte-Rousseauw* (ISR) menghasilkan 2 *cluster* optimal, Indeks *Davies Bouldin* (IDB) menghasilkan 9 *cluster* optimal, Indeks *C* (IC) menghasilkan 3 *cluster* optimal, sedangkan Indeks *Goodman-Kruskal* (IGK) memberikan 2 *cluster* optimal yaitu 4 dan 5 *cluster*. Kemudian, nilai Indeks tersebut akan diranking sesuai dengan kriteria masing-masing. Ranking indeks tersebut dapat dilihat pada Tabel 4.4

Tabel 4.4 Ranking Berdasarkan Indeks Validitas Gabungan (IVG) pada Data 2

	Banyaknya *Cluster*							
Indeks	2	3	4	5	6	7	8	9
ID	8	7	5	6	4	3	1	2
IDB	3	5	1	2	6	4	7	8
ISR	8	7	6	5	2	4	3	1
IC	7	8	1	4.5	4.5	2.5	2.5	6
IGK	5	1	7.5	7.5	2	3	4	6
Rata-rata	**6.2**	5.6	4.1	5	3.7	3.3	3.5	4.6

Dari Tabel 4.4 didapatkan rata-rata ranking tertinggi adalah 6.2. Hal ini menunjukkan bahwa *cluster* optimal yang terbentuk

berdasarkan rata-rata ranking adalah 2 *cluster*. Jadi, berdasarkan Indeks Validitas Gabungan (IVG), *cluster* optimal yang terbentuk pada data 2 adalah sebanyak 2 *cluster*.

Proses pengelompokkan *cluster* optimal dengan Indeks Validitas Gabungan (IVG) sudah dilakukan. Selanjutnya akan dilihat hasil dari proses pengelompokkan *cluster* optimal dengan menggunakan metode Gap Statistik.

4.3 Gap Statistik

Dalam Gap Statistik, besarnya *cluster* optimal ditentukan berdasarkan *cluster* terkecil sedemikian hingga Gap(k) ≥Gap(k+1)-s_{k+1} dan juga bisa dilihat dari nilai *diffu.*

1. Data 1

Hasil proses pengelompokkan *cluster* optimal pada data 1 dapat dilihat pada Tabel 4.5 yang menampilkan nilai Gap Statistik dari data gen mutan ragi *mec* 1 dan *dun* 1.

Tabel 4.5 Nilai Gap Statistik dari Data Gen Mutan Ragi *Mec* 1 dan *Dun* 1

k	Gap(k)	S(k)	Gap(k)-S(k)	*Diffu*
2	1.386115	0.036651	1.3494645	-0.4846938
3	1.935364	0.064555	1.8708088	-0.352066
4	2.352393	0.064959	2.28743	-0.126391
5	2.533597	0.054813	2.478784	0.057004
6	2.530233	0.05364	2.476593	0.07294341

Berdasarkan Tabel 4.5 tersebut, dapat diketahui bahwa nilai Gap(2) adalah 1.386115 dan masih memiliki nilai yang lebih kecil daripada nilai Gap(3)-$s_{(3)}$ yaitu 1.8707088. Begitu pula untuk nilai Gap(3) masih lebih kecil daripada nilai Gap(4)-$s_{(4)}$. Hal yang sama berlaku pada nilai Gap(4) yang bernilai 2.352393, masih lebih kecil daripada nilai Gap(5)-$s_{(5)}$ yaitu 2.478784. Tetapi saat banyaknya *cluster* yang terbentuk adalah 5, nilai Gap yang didapatkan adalah 2.533597 dan nilai Gap(6)-$s_{(6)}$ adalah 2.476593. Nilai Gap(5) sudah lebih besar daripada nilai Gap(6)-$s_{(6)}$, maka berdasarkan hasil tersebut dapat disimpulkan 5 *cluster* merupakan *cluster* optimal untuk data gen mutan ragi *mec* 1 dan *dun* 1. *Syntax*

yang digunakan beserta *output* dalam proses ini dapat dilihat pada Lampiran 7.

Selain proses tersebut, dalam mendapatkan *cluster* optimal berdasarkan metode Gap Statistik dapat juga dilihat dari selisih Gap(k)-Gap(k+1)-s_{k+1} yang disebut sebagai nilai *diffu.* Nilai *diffu* ini merupakan nilai yang perlu dalam mendapatkan banyaknya *cluster* optimal menggunakan metode gap statistik di mana *cluster* optimal didapatkan jika nilai *diffu* ≥ 0. Dari Tabel 4.3, dapat diketahui bahwa saat terbentuk 2 *cluster* nilai *diffu* yang didapatkan untuk Gap(2)-[(Gap(3)-$s_{(3)}$)] adalah -0.4846938. Nilai *diffu* tersebut masih kurang dari nilai 0. Hal ini berlaku sampai terbentuk 4 *cluster* dengan nilai *diffu* untuk Gap(4)-[Gap(5)-$s_{(5)}$)] adalah -0.126391. Tetapi pada saat terbentuk 5 *cluster* nilai *diffu* sudah lebih besar dari 0. Nilai *diffu* untuk Gap(5)-[Gap(6)-$s_{(6)}$)] adalah 0.057004. Agar lebih mudah hasil ini dapat dilihat pada Lampiran 8 di mana nilai *diffu* dan banyaknya *cluster* ditampilkan secara grafis.

Dari grafik pada Lampiran 8 tersebut, saat terbentuk 2 sampai 4 *cluster* nilai *diffu* masih dibawah 0. Tetapi pada saat terbentuk 5 *cluster* nilai *diffu* sudah lebih besar dari 0.

2. Data 2

Hasil proses pengelompokkan *cluster* optimal pada data 2 dapat dilihat pada Tabel 4.6 yang menampilkan nilai Gap Statistik dari data gen sporulasi jamur.

Tabel 4.6 Nilai Gap Statistik dari Data Gen Sporulasi Jamur

k	Gap(k)	S(k)	Gap(k)-S(k)	*Diffu*
2	1.726859	0.0665698	1.6602892	0.1368185
3	1.644764	0.0547235	1.5900405	-0.03548813
4	1.708369	0.0281169	1.68025213	-0.002762581
5	1.750971	0.0398394	1.711131581	0.07005663
6	1.716084	-0.1406159	1.8566991	-0.1057289

Dari Tabel 4.6 tersebut, dapat diketahui bahwa nilai Gap(2) adalah 1.726859 dan nilai Gap(3)-$s_{(3)}$ adalah 1.5900405. Nilai Gap(2) sudah lebih besar daripada nilai Gap(3)-$s_{(3)}$, maka berdasarkan hasil tersebut dapat disimpulkan 2 *cluster* merupakan *cluster* optimal

untuk data sporulasi gen jamur. *Syntax* yang digunakan beserta *output* dalam proses ini dapat dilihat pada Lampiran 11.

Sedangkan untuk nilai *diffu,* pada Tabel 4.6, dapat diketahui bahwa saat terbentuk 2 *cluster* nilai *diffu* yang didapatkan untuk Gap(2)-[(Gap(3)-s($_3$)] adalah 0.1368185. Nilai *diffu* tersebut sudah lebih besar dari nilai 0 maka *cluster* optimal yang terbentuk adalah 2 *cluster.* Untuk lebih mudah, hasil ini dapat dilihat pada Lampiran 12 di mana nilai *diffu* dan banyaknya *cluster* ditampilkan secara grafis.Dari grafik pada Lampiran 12 tersebut, saat terbentuk 2 *cluster,* nilai *diffu* sudah lebih besar dari nilai 0.

Proses untuk mendapatkan *cluster* optimal menggunakan Indeks Validitas Gabungan (IVG) dan Gap Statistik telah dilakukan, selanjutnya akan mengukur kebaikan hasil kedua indeks tersebut berdasarkan nilai CTM.

4.4 *Cluster Tightness Measures (CTM)*

1. Data 1

Cluster optimal Indeks Validitas Gabungan (IVG) pada data 1 menghasilkan 2 *cluster* sedangkan metode Gap Statistik menghasilkan 5 *cluster.* Keduanya menghasilkan *cluster* optimal yang berbeda sehingga perlu diketahui *cluster* optimal mana yang terbaik antara kedua metode tersebut berdasarkan nilai CTM. Nilai CTM yang didapatkan dari Indeks Validitas Gabungan (IVG) dan Gap Statistik untuk data 1 dapat dilihat pada Tabel 4.7.

Tabel 4.7 Nilai CTM untuk IVG dan Gap-Statistik pada Data 1

Penentuan Banyaknya *Cluster* Optimal	CTM
IVG	0.48
Gap Statistik	0.44

Berdasarkan Tabel 4.7 di atas, Indeks Validitas Gabungan (IVG) menghasilkan nilai CTM sebesar 0.48 sedangkan metode Gap Statistik menghasilkan nilai CTM sebesar 0.44. Metode Gap Statistik menghasilkan nilai CTM yang lebih kecil dibandingkan Indeks Validitas Gabungan (IVG). Berarti metode Gap Statistik lebih baik dibandingkan dengan Indeks Validitas Gabungan (IVG) jika diterapkan pada data 1. Dari hasil nilai CTM tersebut dapat

disimpulkan bahwa banyaknya *cluster* yang tepat dalam pengelompokkan 159 gen mutan ragi *mec 1* dan *dun 1* adalah sebanyak 5 *cluster*.

Karena pembentukan *cluster* yang paling baik adalah sebanyak 5 *cluster*, maka untuk melihat karakteristik gen tersebut akan berdasarkan pengelompokkan sebanyak 5 *cluster*. Karakteristik gen mutan ragi *mec 1* dan *dun 1* tersebut dapat dilihat pada Tabel 4.8.

Tabel 4.8 Rata-rata *Cluster* dari Log Rasio Banyaknya warna merah dan Hijau pada Data 1

		Peubah									
Cluster	n	M5m	15m	30m	45m	60m	90m	120m	D30m	90m	120m
1	109	0.14	0.13	0.04	0.17	0.2	0.16	0.18	0.097	0.24	0.214
2	15	0.16	0.04	-0.06	-0.25	-0.43	-0.8	-1	-0.09	-0.95	-1.28
3	22	0.19	0.6	1.03	1.36	1.27	1.14	1.21	1.44	1.63	1.544
4	12	-0.24	-1.2	-1.67	-2.25	-2.47	-2.38	-2	-2.12	-2.24	-2.02
5	1	0.03	1.28	2.59	3.75	3.28	4.03	3.83	2.26	4.35	4.2

Dari Tabel 4.8, pada *cluster* 1, gen mutan ragi *mec 1* memiliki rata-rata tingkat ekspresi paling besar pada menit ke-60 sedangkan untuk mutan *dun 1* memiliki rata-rata tingkat ekspresi paling besar pada menit ke-90. Selain itu diketahui bahwa *cluster* 1 merupakan *cluster* dengan jumlah gen terbanyak dan semuanya memiliki rata-rata tingkat ekspresi positif. *Cluster* 1 juga memiliki n (banyaknya gen) sebanyak 109 yang berarti bahwa *cluster* 1 memiliki jumlah anggota terbanyak dibandingkan ke-4 *cluster* lainnya.

Untuk *cluster* 2, gen mutan ragi *mec* 1 di menit ke-5 memiliki rata-rata tingkat ekspresi positif, kemudian tingkat ekspresi tersebut menurun pada saat menit ke-15. Setelah menit ke-30, rata-rata tingkat ekspresi gen mutan *mec 1* tersebut menjadi negatif sampai pengamatan terakhir dengan rata-rata tingkat ekspresi paling negatif terjadi di menit ke-60. Sedangkan untuk gen mutan *dun 1* pada *cluster* 2, mulai dari 30 menit sampai waktu pengamatan terakhir, rata-rata tingkat ekspresinya negatif. Berarti *cluster* 2 terdiri atas gen-gen mutan *mec 1* dan *dun 1* yang memiliki rata-rata tingkat ekspresi negatif.

Cluster 3 sama dengan *cluster* 1 yaitu di semua waktu pengamatan gen-gen tersebut memiliki rata-rata tingkat ekspresi positif. Perbedaannya adalah gen-gen yang berada pada *cluster* 3 rata-rata tingkat ekspresinya lebih positif dibandingkan dengan gen-gen yang berada pada *cluster* 1. Pada *cluster* 3, untuk mutan *dun 1* rata-rata tingkat ekspresi paling positif terjadi pada menit ke-45, sedangkan gen mutan *dun 1* rata-rata tingkat ekspresi paling positif terjadi pada menit ke-90.

Sebaliknya untuk *cluster* 4, baik untuk mutan *mec 1* maupun *dun 1*, di semua waktu pengamatan memiliki rata-rata tingkat ekspresi negatif. Hampir sama dengan *cluster 2* , *cluster* 4 terdiri atas gen-gen yang memiliki rata-rata tingkat ekspresi negatif. Perbedaan karakteristik kedua *cluster* tersebut yaitu anggota gen pada *cluster* 4 memiliki rata-rata tingkat ekspresi lebih negatif dibandingkan anggota gen *cluster* 2. Selain itu, untuk *cluster* 4, rata-rata tingkat ekspresi negatif gen tersebut sudah terjadi mulai dari waktu pengamatan pertama sampai waktu pengamatan terakhir.

Pada *cluster* 5, hanya memiliki anggota 1 gen.Untuk mutan *mec 1* dan *dun 1,* rata-rata tingkat ekspresi paling positif sama-sama terjadi pada menit ke-90. *Cluster 5* memiliki gen-gen yang mempunyai nilai rata-rata tingkat ekspresi paling positif dari semua *cluster*. Ini menunjukkan bahwa gen yang berada pada *cluster* 5 adalah gen yang paling terpengaruh jika diberi perlakuan pemberian *MMS*.

Gen yang menunjukkan rata-rata tingkat ekspresi positif mengindikasikan pemberian *MMS* berpengaruh pada kedua gen mutan tersebut. Semakin positif maka semakin besar pengaruh *MMS* tersebut. Sebaliknya, semakin negatif maka pemberian *MMS* tersebut tidak berpengaruh terhadap gen mutan tersebut.

2. Data 2

Pada data 2, jika Indeks Validitas Gabungan (IVG) dan metode Gap-Statistik diterapkan, keduanya memiliki kemampuan yang sama dalam menghasilkan *cluster* optimal yaitu 2 *cluster*. Keduanya menghasilkan komposisi anggota yang sama baik untuk *cluster 1* maupun *cluster 2* sehingga tidak perlu membandingkan keduanya berdasarkan nilai CTM. Maka disimpulkan bahwa banyaknya *cluster* yang tepat dalam pengelompokkan 457 gen sporulasi jamur adalah sebanyak 2 *cluster*

Untuk melihat karakteristik dari gen sporulasi jamur dapat dilihat pada Tabel 4.9.

Tabel 4.9 Rata-Rata *Cluster* dari Log Rasio Banyaknya Warna Merah dan Hijau pada Data 2

		Peubah					
C	n	t=0.5	t=2	t=5	t=7	t=9	t=11.5
1	239	-0.55	-0.9	-1.52	-2.43	-1.97	-2
2	218	1.7	1.33	1.595	1.883	1.36	1.12

Untuk *cluster* 1, terdiri atas gen-gen yang memiliki nilai rata-rata tingkat ekspresi negatif. Hal ini mengindikasikan bahwa gen-gen yang menjadi anggota *cluster* 1 tidak aktif melakukan sporulasi dengan rata-rata tingkat ekspresi paling negatif terjadi pada menit ke-7. Sedangkan pada *cluster* 2, terdiri atas gen-gen yang memiliki nilai rata-rata tingkat ekspresi positif. Berarti gen-gen yang berada pada *cluster* 2, aktif dalam melakukan sporulasi (proses pembentukan spora) di mana rata-rata tingkat ekspresi paling positif terjadi pada menit ke-7. Kedua *cluster* ini menunjukkan bahwa gen-gen tersebut akan berekspresi kuat pada menit ke-7. Sehingga informasi yang dapat diambil dari 457 gen ini adalah gen-gen tersebut paling aktif atau pasif dalam melakukan pembentukan spora akan terjadi pada menit ke-7.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari penelitian ini dapat diambil kesimpulan sebagai berikut

1. Pada data 1, dengan metode Gap Statistik jumlah *cluster* optimal yang didapatkan adalah 5 sedangkan untuk Indeks validitas Gabungan (IVG) didapatkan 2 *cluster* optimal. Dan untuk data 2, metode Gap Statistik dan Indeks Validitas Gabungan (IVG) memiliki kemampuan yang sama dalam pengelompokkan data ekspresi gen yaitu membentuk 2 *cluster* .

2. Berdasarkan nilai CTM, metode Gap Statistik lebih baik daripada Indeks Validitas Gabungan (IVG) jika diterapkan pada data 1. Sedangkan untuk Data 2, Metode Gap Statistik dan Indeks Validitas Gabungan (IVG) berkemampuan sama dengan menghasilkan *cluster* optimal sebanyak 2 dengan komposisi anggota yang sama.

5.2 Saran

1. Disarankan pada peneliti lain, untuk menerapkan metode gap-statistik pada data yang bukan berjenis *microarray* untuk mengetahui apakah metode gap-statistik juga cukup baik untuk data tersebut.

2. Metode Gap-Statistik perlu ditindaklanjuti yaitu dengan menggunakan metode *weighted* Gap-Statistik yang juga diterapkan pada data *microarray*.

DAFTAR PUSTAKA

Aharoni, A and O. Vorst. 2001. **DNA Microarrays for Functional Plant Genomics**. *http://www.springerlink.com/.* Tanggal akses: 27 Desember 2008.

Arima,C., K.Hakamada, M.Okamoto and T.Hanai .2008. **Validity Index for Fuzzy K-Means clustering Using the Gap statistic Method**. *http://www.jsbi.org.* Tanggal Akses : 15 Juli 2008.

Bolshakova,N. 2001. **Cluster Validity Algorithms. Department of Computer Sciences**. *http://www.cs.tcd.ie.* Trinity College Dublin, Ireland. Tanggal Akses : 20 November 2008

Bolshakova,N and F.Azuaje.2003. **Cluster Validation for Gene Expression data.***http://bioinformatics.oupjournals.org/* Tanggal Akses : 20 November 2008.

D'haeseleer,P.,S.Liang, and R.Somogyi.1999.**Gene Expression Data Analysis and Modeling.** *www.citeseer.ist.psu.edu***.**Tanggal Akses : 15 juli 2008.

Draghici, S. 2003. **Data Analysis Tools For DNA Microarray**. Chapman and Hall, Inc. New York.

Eichenberger, P., S.T Jensen, E.M Conlon, C.V Ooij and J. Silvaggi. 2003. **The Identification of Additional Sporulation Genes**. *http://www-genome.stanford*. Tanggal Akses : 23 Desember 2008.

Epps, J. and E. Ambikairajah. 2008. **Visualisation of Reduced-Dimension Microarray Data Using Gaussian Mixture Model.** *http://crpit.com/confpapers/.* Tanggal Akses: 12 November 2008.

Gasch, A.P., M. Huang, S. Metzner, D. Botstein, S.J. Elledge and P.O Brown. **Genomic Expression Responses to DNA-**

damaging Agents and the Regulatory Role of the Yeast ATR Homolog Mec1p. *http://www-genome.stanford*. Tanggal Akses : 15 Desember 2008.

Hair, J.F., R.E. Anderson, R.L. Tatham and W.C. Black. 1998. **Multivariate Data Analysis with Readings**. *Fifth Edition*. Prentice-Hall, Inc. New Jersey.

Mattjik,A.A, M.Sumertajaya, H.Wijayantoo, Indahwati, A.kurnia dan B.Sartono. 2002. **Aplikasi Analisis Peubah Ganda.** Jurusan Statistika FMIPA IPB. Bogor.

Meiyanto, E. 2001. **Farmakogenomik; Sebuah Paradigma Baru Dalam Sistem Pengobatan.** *http://wwwstd.ryu.titech.ac.jp/* Tanggal Akses : 24 Oktober 2008

Mufidah, R. 2004. **Analisis Kelompok Optimal dengan Kombinasi Indeks Validitas pada Analisis Kelompok Hirarki.** Fakultas MIPA Universitas Brawijaya. Malang. Skripsi. Tidak dipublikasikan.

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

Speed,T. 2008. **Statistics and Gene Expression Analysis**. *www.proba.jussieu.fr*. Tanggal Akses : 20 Juli 2008.

Su, Mc. 2003. **A new Index of Validity Cluster Validity**. *http://www.ctlseer.nj.nec.com*. Tanggal Akses : 10 November 2008.

Tibshirani, R., G.Wattner and T.Hastie, 2002. **Estimating the number of clusters in a dataset via the Gap Statistics**.University,Stanford.*www.stat.stanford.edu* Tanggal Akses : 15 Mei 2008

Wulandari, T. 2006. **Penerapan Analisis Cluster Hirarki pada Data Berskala Campuran dengan Menggunakan**

Analisis Komponen Utama Nonlinier (Studi Kasus di Sentra Industri Tempe Kota Malang). Fakultas MIPA Universitas Brawijaya, Malang. Skripsi. Tidak dipublikasikan

Lampiran 1. Data Log Rasio Banyaknya Warna Merah dengan Hijau pada Respon *Mec 1* dan *Dun 1* terhadap *MMS*

		Waktu Pengamatan									
1	Gen	Mec 1 5m	15m	30m	45m	60m	90m	120m	Dun 1 30m	90m	120m
2	'GENE1163X'	0.91	0.77	0.3	0.44	0.19	0.11	0.11	0.33	-0.03	-0.2
3	'GENE1140X'	0.49	0.66	0.1	0.31	0.28	0.6	0.49	-0.14	0.06	-0.18
4	'GENE1202X'	0.51	0.29	-0.23	0.01	0.14	0.14	0.18	-0.45	-0.27	-0.54
5	'GENE1206X'	0.42	0.34	-0.36	0.15	0.16	0.11	0.16	0.04	0.18	0.28
6	'GENE1194X'	0.42	0.43	0.17	0.4	0.53	0.29	0.27	0.18	0.12	-0.07
7	'GENE1187X'	0.71	0.61	0.01	0.45	0.4	0.35	0.27	0.18	-0.04	-0.12
8	'GENE1189X'	0.24	0.33	-0.09	0.12	0.19	0.12	0.03	0.08	0.2	0.32
9	'GENE1110X'	0.49	0.18	-0.09	-0.14	-0.15	-0.18	-0.06	0.06	0.21	-0.03
10	'GENE1235X'	0.57	0.72	0.36	0.72	0.81	0.61	0.62	0.32	0.36	0.3
11	'GENE1031X'	0.44	0.38	0.38	0.45	0.43	0.32	0.28	0.44	0.89	0.82
12	'GENE1089X'	0.41	0.38	0.23	0.64	-0.2	0.43	0.53	0.16	0.18	-0.04
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
155	'GENE1866X'	0.28	0.12	0.01	0.12	-0.04	-0.14	-0.07	-0.14	-0.14	-0.3
156	'GENE3082X'	0.23	0.23	0.19	0.33	0.4	-0.23	-0.56	0.24	-0.4	-0.42
157	'GENE2918X'	-0.22	-0.97	-1.36	-2.64	-2.74	-2.56	-2.4	-2.56	-2.74	-2.47
158	'GENE2908X'	-0.6	-1.36	-1.22	-1.94	-2.4	-1.84	-1.64	-2.32	-2.47	-2.32
159	'GENE2958X'	-0.43	-1.6	-1.32	-2.12	-2.18	-2.12	-1.6	-2	-1.69	-1.25

Lampiran 2. Data Log Rasio Banyaknya Warna Merah dengan Warna Hijau pada Sporulasi Gen Jamur

	Gen	Waktu Pengamatan						
		t=0	t=0.5	t=2	t=5	t=7	t=9	t=11.5
1	'YAL062W'	-0.33	-2.42	-3.01	-1.43	-0.42	-0.5	-1.48
2	'YAR007C'	0.094	-1.58	-2.42	-2.09	-2.16	-0.7	-0.96
3	'YAL005C'	-0.65	0.238	1.267	1.285	1.857	1.74	2.101
4	'YAL012W'	-0.06	-1.22	0.829	1.657	2.715	2.34	2.299
5	'YAL025C'	0.097	3.553	1.846	1.138	2.275	0.92	0.488
6	'YAL036C'	0.108	2.925	1.783	1.323	1.985	-0.4	0.757
7	'YAL038W'	0.022	2.042	1.689	2.35	2.303	0.68	1.492
8	'YAL054C'	-0.54	-2.44	-1.91	-1.68	-1.46	-0.1	0.114
9	'YAL055W'	0.281	-0.67	-0.59	-0.8	-1.51	-2.2	-2.36
10	'YAL067C'	0.129	-3.28	-0.59	-1.54	-3.38	-1.7	-2.92
11	'YBL009W'	-0.02	-0.46	-1.05	-1.28	-2.64	-2.3	-2.36
12	'YBL010C'	-0.06	-0.05	-0.72	-1.43	-2.36	-1.9	-1.36
13	'YBL042C'	-0.3	2.22	1.06	0.459	-0.65	-1.5	-1.69
14	'YBL067C'	-0.35	-0.4	0.598	1.659	2.29	1.52	1.047
15	'YBL072C'	-0.65	1.75	1.917	2.363	2.407	1.47	0.408
16	'YBL078C'	0.054	-0.82	-1.11	-1.45	-2.07	-2.6	-2.83
17	'YBL099W'	-0.27	1.745	0.309	0.68	1.124	2.05	2.132
18	'YBR025C'	-0.53	2.772	1.943	1.028	0.067	-0.6	-0.85
19	'YBR063C'	-0.23	0.406	-0.44	-2.67	-4.04	-2.1	-3.26
20	'YBR069C'	-0.25	0.157	2.18	0.828	-0.7	-1.1	-1.44
21	'YBR086C'	0.146	-0.03	0.673	1.816	2.073	1.42	0.728
22	'YBR088C'	0.136	-0.82	-2.58	-2.18	-1.69	-0.6	-0.25
23	'YBR181C'	-0.12	2.308	2.002	2.645	2.442	1.68	0.884
24	'YBR184W'	0.205	-0.77	-1.78	-2.95	-3.3	-3.1	-3.16
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
452	'YPR106W'	0.335	-0.53	-0.37	-0.88	-2.19	-1.7	-1.07
453	'YPR110C'	0.208	2.095	1.217	0.454	-0.37	-0.9	-0.58
454	'YPR111W'	0.189	-0.71	-0.67	-0.93	-2	-2.1	-1.82
455	'YPR120C'	0.376	-0.4	-1.49	-2.36	-3.45	-2.8	-2.81
456	'YPR141C'	0.486	0.456	-0.82	-1.78	-2.74	-1.7	-1.93
457	'YPR145W'	0.193	0.109	1.046	1.817	1.855	1.43	2.031

Lampiran 3. Matriks Proximitas Data 1

Case	1	2	3	4	5	6	7	8	9	10	.	.	.	157	158	159
1	0	0,867	1,646	1.307	0,61	0,306	1,289	1,401	1.497	2,409	.	.	.	45,835	39,391	1,397
2	0,867	0	0.998	0.966	0.396	0.32	1	1.65	1.076	2.349	.	.	.	45.479	39,36	1,398
3	1,646	0.998	0	1.164	1.292	1.14	1.373	1.049	3.6	4.693	.	.	.	33.927	28.762	0.495
4	1.307	0.966	1.164	0	0.679	0.739	0	0.514	2.086	1.727	.	.	.	41.986	34.106	0.832
5	0,61	0.396	1.292	0.679	0	0.193	0.559	1.243	0.763	1.013	.	.	.	46.443	39.484	1.068
6	0,306	0.32	1.14	0.739	0.193	0	0.834	1.369	0.941	2.081	.	.	.	46.101	39.564	1.228
7	1,289	1	1.373	0	0.559	0.834	0	0.489	1.88	1.388	.	.	.	42.279	34.359	0.735
8	1,401	1.65	1.049	0.514	1.243	1.369	0.489	0	3.447	2.643	.	.	.	36.482	29.046	0.375
9	1.497	1.076	3.6	2.086	0.763	0.941	1.88	3.447	0	1.116	.	.	.	58.524	50.579	3.509
10	2,409	2.349	4.693	1.727	1.013	2.081	1.388	2.643	1.116	0	.	.	.	57.509	47.367	3.546
.	.	.	.	.	.	.	.	.	.	.	0			.	.	.
.	.	.	.	.	.	.	.	.	.	.		0		.	.	.
.	.	.	.	.	.	.	.	.	.	.			0	.	.	.
157	45,839	45.479	33.927	41.986	46.443	46.101	42.279	36.482	58.524	57.509	.	.	.	0	2,113	33,908
158	39,391	39,36	28.762	34.106	39.484	39.564	34.359	29.046	50.579	47.367	.	.	.	2,113	0	27,855
159	1,397	1,398	0.495	0.832	1.068	1.228	0.735	0.375	3.509	3.546	.	.	.	33,908	27,855	0

Lampiran 4. Matriks Proximitas Data 2

Case	1	2	3	4	5	6	7	8	9	10	.	.	.	455	456	457
1	0	5.007	55.718	57.7	79.1	70.1	73.99	5.097	14.56	19.22	.	.	.	24.32	20.999	57.792
2	5.007	0	60.198	68.2	79.3	69.8	77.47	3.559	10.47	13	.	.	.	12.13	9.293	59.663
3	55.718	60.2	0	3.4	15.4	14.6	6.698	44.36	56.06	89.04	.	.	.	95.51	54.792	1.152
4	57.74	68.23	3.397	0	29.6	28.8	15.44	48.44	68.39	97.3	.	.	.	113.3	81.374	3.54
5	79.117	79.25	15.355	29.6	0	2.38	4.845	73.44	59.56	110.3	.	.	.	96.82	63.309	15.794
6	70.13	69.75	14.555	28.8	2.38	0	3.714	64.25	48.12	96.26	.	.	.	83.5	53.86	13.806
7	73.986	77.47	6.698	15.4	4.85	3.71	0	66.23	60.1	106.2	.	.	.	102.4	68.986	5.515
8	5.097	3.559	44.359	48.4	73.4	64.2	66.23	0	16.72	18.43	.	.	.	25.67	9.063	44.983
9	14.564	10.47	56.057	68.4	59.6	48.1	60.1	16.72	0	11.34	.	.	.	7.79	4.217	53.781
10	19.218	13	89.042	97.3	110	96.3	106.2	18.43	11.34	0	.	.	.	11.09	15.576	87.317
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
455	24.318	12.13	95.588	113	96.8	83.5	102.4	25.67	7.79	11.09	.	.	.	0	4.057	93.997
456	20.999	9.293	64.428	81.4	63.3	53.9	68.99	9.063	4.217	15.58	.	.	.	4.057	0	63.364
457	54.792	59.66	1.152	3.54	15.8	13.8	5.515	44.98	53.78	87.32	.	.	.	94	63.364	0

Lampiran 5. Tabel Nilai Indeks Validitas Dunn (ID) pada Data 1

	Kelompok							
D	2	3	4	5	6	7	8	9
D11	1.3	1.184	0.911	1.087	0.569	0.82	0.499	0.499
D12	5.795	4.964	2.809	2.653	1.96	2.074	1.398	1.517
D13	4.314	3.655	1.909	1.803	1.402	1.536	1.046	1.079
D21	0.709	0.677	0.475	0.587	0.374	0.499	0.433	0.433
D22	3.161	2.839	1.465	1.4732	1.288	1.262	1.214	1.318
D23	2.353	2.09	0.966	0.973	0.922	0.935	0.909	0.937
D31	0.338	0.107	0.144	0.144	0.144	0.113	0.127	0.127
D32	1.509	0.447	0.443	0.35	0.494	0.285	0.357	0.387
D33	1.123	0.329	0.301	0.238	0.354	0.211	0.267	0.275
D41	0.704	0.663	0.444	0.534	0.326	0.372	0.371	0.371
D42	3.139	2.778	1.37	1.303	1.122	0.941	1.041	1.13
D43	2.336	2.045	0.931	0.886	0.803	0.697	0.78	0.804
D51	0.709	0.67	0.457	0.556	0.349	0.438	0.414	0.414
D52	3.161	2.81	1.409	1.357	1.201	1.107	1.16	1.258
D53	2.353	2.069	0.958	0.922	0.859	0.82	0.868	0.895
D61	1.3	1.075	0.725	0.768	0.383	0.641	0.475	0.475
D62	5.795	4.506	2.235	1.874	1.321	1.621	1.332	1.446
D63	4.314	3.317	1.519	1.273	0.945	1.2	0.997	1.028
Rata-rata	**2.46739**	2.0125	1.08172	1.0434	0.8231	0.865	0.7604	0.7996111

Lampiran 6

1. Tabel Nilai Indeks Davies Bouldin (IDB) pada Data 1

	Kelompok							
D	2	3	4	5	6	7	8	9
DB11	0.77	0.945	1.033	1.07	1.166	1.046	1.06	1.068
DB12	0	0.221	0.403	0.423	0.515	0.409	0.489	0.548
DB13	0.232	0.373	0.632	0.663	0.739	0.674	0.761	0.797
DB21	1.411	1.624	2.099	2.127	2.049	1.823	1.97	1.936
DB22	0	0.386	0.81	0.844	0.891	0.734	0.874	0.919
DB23	0.425	0.643	1.23	1.267	1.253	1.126	1.302	1.304
DB31	2.955	8.566	6.92	7.578	6.113	5.359	5.843	5.497
DB32	0	2.453	2.731	3.089	2.55	2.084	2.498	2.444
DB33	0.89	3.518	4.023	4.375	3.51	2.974	3.533	3.398
DB41	1.421	1.654	2.278	2.323	2.288	1.995	2.286	2.244
DB42	0	0.394	0.875	0.918	1.008	0.813	1.019	1.069
DB43	0.428	0.655	1.323	1.373	1.408	1.237	1.503	1.506
DB51	1.411	1.637	2.22	2.246	2.164	1.904	2.117	2.074
DB52	0	0.39	0.851	0.889	0.945	0.77	0.941	0.981
DB53	0.425	0.648	1.288	1.331	1.325	1.177	1.395	1.392
DB61	0.77	1.015	1.357	1.66	1.558	1.355	1.449	1.399
DB62	0	0.243	0.525	0.642	0.715	0.565	0.682	0.699
DB63	0.232	0.403	0.804	0.979	1.003	0.878	1.016	0.999
	0.63167	1.43156	1.74456	1.8776	1.7333	1.496	1.7077	1.6818889

2. Tabel Nilai Indeks Silhoutte-Rousseauw (ISR) pada Data 1

S	Gsu	S1	S2	S3	S4	S5	S6	S7	S8	S9
2	**0.67**	0.668	1							
3	0.578	0.569	0.773	1						
4	0.471	0.497	0.077	0.562	1					
5	0.391	0.4	0.045	0.562	0.491	1				
6	0.337	0.299	0.508	0.28	0.491	1	0.516			
7	0.342	0.299	0.508	1	0.491	1	0.477	0.354		
8	0.264	0.204	0.272	0.333	1	0.464	1	0.477	0.354	
9	0.24	0.189	0.272	0.33	1	0.284	0.534	1	0.477	0.354

Lampiran 7. Syntax *Package clustersim* Program R untuk Metode Gap Statistik pada Data 1

```
library(clusterSim)
data(mecnew)
md<-dist(mecnew,method="euclidean")^2
# nc - number_of_clusters
min_nc=2
max_nc=15
min <- 0
clopt <- NULL
res <- array(0,c(max_nc-min_nc+1,2))
res[,1] <- min_nc:max_nc
found <- FALSE
for (nc in min_nc:max_nc)
{
print(nc)
hc <- hclust(md, method="average")
cl1 <- cutree(hc, k=nc)
cl2 <- cutree(hc, k=nc+1)
cl3 <- cutree(hc, k=nc1)
clall <- cbind(cl1,cl2)
  gap <- index.Gap(mecnew,clall,B=159,method="average")
  res[nc-min_nc+1, 2] <- diffu <- gap$diffu
  if ((res[nc-min_nc+1, 2] >=0) && (!found)){
    nc1 <- nc
    min <- diffu
    clopt <- cl1
    found <- TRUE
  }
print(gap)
}
if (found){
print(paste("Minimal nc where diffu>=0 is",nc1,"for
      diffu=",round(min,4)),quote=FALSE)
}else
{
print("I have not found clustering with diffu>=0", quote=FALSE)
```

Lampiran 7 (Lanjutan)

```
plot(res,type="p",pch=0,xlab="Number of
    clusters",ylab="diffu",xaxt="n")
abline(h=0, untf=FALSE)
axis(1, c(min_nc:max_nc))
```

Output :

[1] **Minimal nc** where diffu>=0 is **5** for diffu= **0.057004**

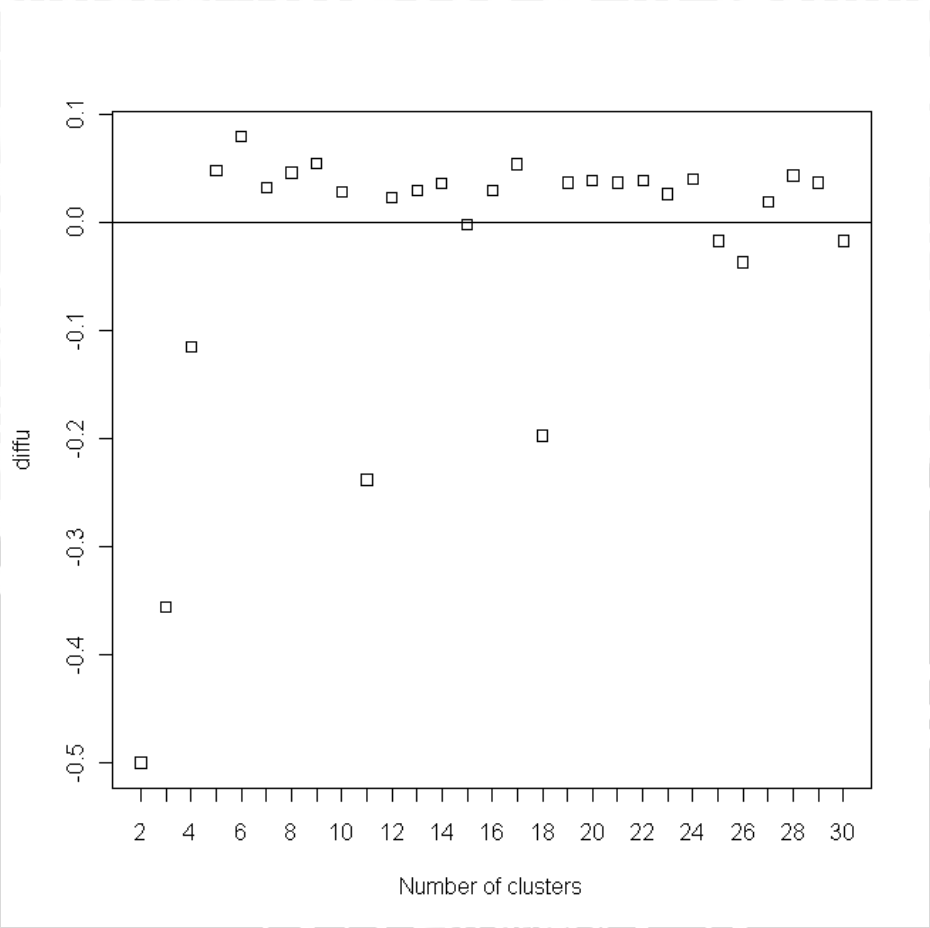
$gap

[1] **2.533597**

```
> cl3
 [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1
1 1 1 1
 [38] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 1 2 2 2 2 2 2 3 1 1 1
4 2 1 1
 [75] 1 1 1 2 2 1 1 1 1 3 3 3 2 1 1 3 1 1 1 1 1 1 1 1 1 3 1 1 3 3 1 1 1
1 1 1 1
[112] 1 1 1 3 1 3 1 1 1 3 3 3 3 3 3 3 3 1 1 3 3 5 3 3 1 1 1 1 1 1 1 1 1
1 1 1 4
[149] 4 4 4 4 4 4 4 4 4 4 1
```

[1] 2 $gap [1] 1.385918 $diffu [1] -0.4960124
[1] 3 $gap [1] 1.932298 $diffu [1] -0.3536842
[1] 4 $gap [1] 2.354661 $diffu [1] -0.1232382
[1] **5** $gap [1] **2.533597** $diffu [1] **0.057004**
[1] 6 $gap [1] 2.538388 $diffu [1] 0.07294341
[1] 7 $gap [1] 2.502563 $diffu [1] 0.02302984
[1] 8 $gap [1] 2.513269 $diffu [1] 0.03478886
[1] 9 $gap [1] 2.519203 $diffu [1] 0.06240753
[1] 10 $gap [1] 2.494840 $diffu [1] 0.02277419
[1] 11 $gap [1] 2.510189 $diffu [1] -0.2338091
[1] 12 $gap [1] 2.783133 $diffu [1] 0.02749675
[1] 13 $gap [1] 2.792119 $diffu [1] 0.02512002
[1] 14 $gap [1] 2.797991 $diffu [1] 0.02726733
[1] 15 $gap[1] 2.813557 $diffu [1] 0.001842853

Lampiran 8. Grafik Banyaknya *cluster* dan Nilai *Diffu* Data 1



Lampiran 9. Tabel Nilai Indeks Validitas Dunn (ID) pada Data 2

Kelompok								
D	2	3	4	5	6	7	8	9
D11	1.402	1.145	0.915	1.032	0.598	0.64	0.508	0.519
D12	4.883	3.559	3.118	3.162	1.832	1.73	1.373	1.373
D13	3.568	2.559	2.277	2.249	1.303	1.238	0.983	0.983
D21	0.658	0.547	0.39	0.463	0.456	0.443	0.418	0.427
D22	2.291	1.699	1.33	1.42	1.396	1.198	1.13	1.13
D23	1.675	1.241	0.971	1.01	0.993	0.858	0.808	0.808
D31	0.086	0.085	0.062	0.077	0.077	0.082	0.082	0.084
D32	0.299	0.265	0.213	0.235	0.235	0.222	0.222	0.222
D33	0.219	0.193	0.156	0.167	0.167	0.159	0.159	0.159
D41	0.616	0.498	0.296	0.362	0.362	0.347	0.349	0.357
D42	2.144	1.548	1.009	1.109	1.109	0.938	0.943	0.943
D43	1.567	1.13	0.737	0.789	0.789	0.671	0.675	0.675
D51	0.637	0.536	0.352	0.419	0.419	0.396	0.396	0.405
D52	2.219	1.666	1.201	1.182	1.282	1.069	1.07	1.07
D53	1.622	1.216	0.877	0.912	0.912	0.765	0.766	0.766
D61	0.89	0.91	0.596	0.668	0.598	0.491	0.491	0.502
D62	3.099	2.83	2.032	2.048	1.832	1.326	1.326	1.326
D63	2.265	2.066	1.484	1.456	1.303	0.949	0.949	0.949
Rata-rata	**1.6744**	1.316	1.0009	1.0422	0.87	0.7512	0.7027	0.7054444

Lampiran 10

1. Tabel Nilai Indeks Davies Bouldin (IDB) pada Data 2

	Kelompok							
DB	2	3	4	5	6	7	8	9
DB11	1.188	1.121	1.322	1.369	1.282	1.398	1.354	1.281
DB12	0.366	0.425	0.496	0.533	0.433	0.498	0.508	0.509
DB13	0.513	0.567	0.691	0.735	0.695	0.774	0.771	0.762
DB21	2.532	2.358	3.082	2.937	2.653	2.847	2.715	2.534
DB22	0.781	0.892	1.09	1.142	0.928	1.03	1.082	0.993
DB23	1.094	1.191	1.506	1.58	1.441	1.558	1.527	1.464
DB31	19.983	16.97	20.113	15.974	13.43	13.279	12.037	11.469
DB32	5.978	6.027	7.085	6.23	5.095	5.102	4.75	4.648
DB33	8.337	8.177	9.788	8.624	7.35	7.335	6.78	6.606
DB41	2.705	2.563	3.765	3.544	3.149	3.486	3.29	3.045
DB42	0.834	0.972	1.334	1.377	1.109	1.27	1.25	1.194
DB43	1.169	1.297	1.846	1.908	1.712	1.905	1.847	1.752
DB51	2.614	2.426	3.315	3.173	2.823	3.125	2.956	2.748
DB52	0.806	0.913	1.173	1.233	0.995	1.138	1.123	1.078
DB53	1.13	1.22	1.621	1.708	1.534	1.708	1.659	1.582
DB61	1.872	1.73	2.068	2.034	1.788	2.172	2.03	1.908
DB62	0.577	0.605	0.751	0.791	1.832	0.803	0.791	0.765
DB63	0.809	0.831	1.048	1.095	0.937	1.202	1.165	1.119
Rata-rata	2.9604	2.794	3.4497	3.1104	2.733	2.8128	2.6464	**2.5253889**

2. Tabel Nilai Indeks Silhoutte-Rousseauw (ISR) pada Data 2

	Kelompok									
S	Gsu	S1	S2	S3	S4	S5	S6	S7	S8	S9
2	**0.591**	0.543	0.62							
3	**0.479**	0.337	0.647	0.633						
4	**0.396**	0.372	0.621	0.166	0.626					
5	**0.311**	0.347	0.385	0.201	0.492	0.607				
6	**0.282**	0.34	0.325	0.196	0.485	0.607	1			
7	**0.299**	0.264	0.325	0.32	0.47	0.061	0.477	1		
8	**0.283**	0.234	0.325	0.276	0.47	0.056	0.615	0.477	1	
9	**0.254**	0.162	0.271	0.276	0.47	0.056	0.784	0.615	0.477	1

Lampiran 11. Syntax *package clustersim* Program R untuk Metode Gap Statistik pada Data 2

```
library(clusterSim)
data(dataspor1)
md<-dist(dataspor1,method="euclidean")^2
# nc - number_of_clusters
min_nc=2
max_nc=30
min <- 0
clopt <- NULL
res <- array(0,c(max_nc-min_nc+1,2))
res[,1] <- min_nc:max_nc
found <- FALSE
for (nc in min_nc:max_nc)
{
print(nc)
hc <- hclust(md, method="average")
cl1 <- cutree(hc, k=nc)
cl2 <- cutree(hc, k=nc+1)
cl3 <- cutree(hc, k=nc1)
clall <- cbind(cl1,cl2)
  gap <- index.Gap(dataspor1,clall,B=457,method="average")
  res[nc-min_nc+1, 2] <- diffu <- gap$diffu
  if ((res[nc-min_nc+1, 2] >=0) && (!found)){
    nc1 <- nc
    min <- diffu
  clopt <- cl1
    found <- TRUE
  }
}
if (found){
print(paste("Minimal nc where diffu>=0 is",nc1,"for
                  diffu=",round(min,4)),quote=FALSE)
}else
{
print("I have not found clustering with diffu>=0", quote=FALSE)
}
```

Lampiran 11 (Lanjutan)

```
plot(res,type="p",pch=0,xlab="Number of
        clusters",ylab="diffu",xaxt="n")
abline(h=0, untf=FALSE)
axis(1, c(min_nc:max_nc))
  clopt <- cl1
   found <- TRUE
 }
print(gap)
}
if (found){
print(paste("Minimal nc where diffu>=0 is",nc1,"for
        diffu=",round(min,4)),quote=FALSE)
}else
{
print("I have not found clustering with diffu>=0", quote=FALSE)
}
plot(res,type="p",pch=0,xlab="Number of
        clusters",ylab="diffu",xaxt="n")
abline(h=0, untf=FALSE)
axis(1, c(min_nc:max_nc))
```

Output :

[1] **Minimal nc** where diffu>=0 is **2** for diffu= **0.1368**

$gap

[1] **1.726859**

```
> cl3
  [1] 1 1 2 2 2 2 2 1 1 1 1 1 2 2 2 1 2 2 1 2 2 1 2 1 2 1 1 2 2 2 2 2 2
2 1 1 1
 [38] 1 1 1 2 1 2 2 1 2 2 1 1 2 1 2 2 2 2 1 2 1 1 1 2 1 2 1 1 1 1 2 1 1
1 2 1 1
 [75] 2 1 2 1 2 2 1 2 1 2 2 2 1 1 2 2 2 2 1 1 1 1 1 1 2 1 2 1 1 1 2 2 1
2 2 1 1
[112] 1 2 2 2 2 2 1 1 1 1 1 1 1 2 1 2 2 2 1 2 1 2 2 1 2 1 1 2 1 1 2 2
1 1 1 2
[149] 1 1 2 1 1 2 2 2 2 2 1 2 1 1 1 1 1 1 1 2 1 2 1 2 1 2 1 1 2 2 1 1 1
1 2 1 2
```

Lampiran 11 (Lanjutan)

[186] 1 1 1 1 2 1 2 1 2 2 2 1 2 1 2 1 1 2 1 1 1 1 1 1 1 2 2 2 1 1 2 2 2 1 2 1 2

[223] 1 1 2 2 2 1 1 1 1 2 1 1 1 1 1 2 2 2 2 1 2 1 1 1 2 2 2 2 2 2 2 1 1 1 2 1 2

[260] 2 1 1 1 1 2 1 2 2 2 2 2 2 2 1 1 2 2 1 2 2 1 1 1 2 2 1 2 1 1 1 1 1 1 1 1 2

[297] 1 2 2 1 1 2 1 1 2 2 2 2 2 1 1 1 1 1 2 2 2 2 1 2 1 1 2 1 2 2 2 2 1 2 2 1 1

[334] 1 2 2 2 1 2 1 2 2 1 2 1 1 2 1 1 2 2 2 2 2 2 1 2 2 1 2 2 1 2 1 2 2 1 1 2 1

[371] 1 1 1 1 2 2 1 1 1 2 2 1 1 1 2 2 2 2 2 2 1 2 1 2 2 1 1 2 1 2 2 1 2 2 1 1 2

[408] 2 1 1 2 1 1 2 1 1 2 2 2 2 1 1 1 1 2 1 2 1 1 2 2 2 2 2 1 1 1 1 1 2 2 1 1 2

[445] 2 1 1 1 1 1 1 1 2 1 1 1 2

[1] **2** $gap [1] **1.726859** $diffu [1] **0.1368185**
[1] 3 $gap [1] 1.644764 $diffu [1] -0.03548813
[1] 4 $gap [1] 1.708369 $diffu [1] -0.002762581
[1] 5 $gap [1] 1.750971 $diffu [1] 0.07005663
[1] 6 $gap [1] 1.716084 $diffu [1] -0.1057289
[1] 7 $gap [1] 1.859776 $diffu [1] 0.05700276
[1] 8 $gap [1] 1.841545 $diffu [1] 0.07450982
[1] 9 $gap [1] 1.797688 $diffu [1] 0.07510283
[1] 10 $gap [1] 1.766236 $diffu [1] 0.02024546
[1] 11 $gap [1] 1.77516 $diffu [1] -0.2055675
[1] 12 $gap [1] 2.012718 $diffu [1] -0.02554728
[1] 13 $gap [1] 2.074349 $diffu [1] 0.04351597
[1] 14 $gap [1] 2.06136 $diffu [1] 0.01394969
[1] 15 $gap [1] 2.082381 $diffu [1] 0.02575659

Lampiran 12. Grafik Banyaknya *Cluster* dan Nilai *Diffu* Data 2