



**ANALISIS DISKRIMINAN UNTUK VALIDASI *CLUSTER*
PADA STUDI KASUS PENGELOMPOKAN KECAMATAN DI
KABUPATEN JEMBER BERDASARKAN STATUS
KEMISKINAN**

SKRIPSI

Oleh:

**Fikriana Nur Istiqomah
NIM 1318101010**

**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER
2019**



**ANALISIS DISKRIMINAN UNTUK VALIDASI *CLUSTER*
PADA STUDI KASUS PENGELOMPOKAN KECAMATAN DI
KABUPATEN JEMBER BERDASARKAN STATUS
KEMISKINAN**

SKRIPSI

diajukan guna melengkapi tugas akhir dan memenuhi salah satu syarat untuk menyelesaikan Program Studi Matematika (S1) dan mencapai gelar Sarjana Sains

Oleh:

**Fikriana Nur Istiqomah
NIM 1318101010**

**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER
2019**

PERSEMBAHAN

Dengan segala rasa syukur kepada Allah SWT., saya persembahkan skripsi ini kepada :

1. Kedua Orangtua saya, Bapak Sukam dan Ibu Nurika serta Adik saya yang bernama Aulia Indah Sukam yang tidak henti-hentinya memberikan saya dukungan, nasihat, motivasi dan selalu mendoakan atas terselesaikannya skripsi ini;
2. Bapak Prof. I Made Tirta M.Sc., Ph.D., dan Ibu Dian Anggraeni S.Si., M.Si., yang dengan sabar dan tulus ikhlas membimbing sehingga penulisan skripsi ini dapat terselesaikan;
3. Seluruh guru dan almamater dari TK Anggrek Tenggarong, SDN 003 Tenggarong, MTs Negeri Tenggarong dan SMAN 1 Tenggarong serta Universitas Jember tercinta atas kesempatannya untuk menimba Ilmu dan pengalaman yang diberikan selama ini;
4. Keluarga Besar Mbah Taya di Talangsari, dan Keluarga Besar Mbah Marsatun di Tanggul, yang telah mendukung dan memberikan doa, kasih sayang, serta motivasi;
5. Sahabat-sahabat yang inshaAllah terbaik sampai Jannah, Jihan, Yora, dan Eka, yang selalu memberikan dorongan dan motivasi terbaik dalam mengerjakan skripsi ini;
6. Teman-teman ATLAS (Matematika Angkatan 2013) FMIPA Matematika yang telah memberikan rasa kekeluargaan yang tak terlupakan;
7. Almamater tercinta Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

MOTTO

Jadikanlah sabar dan Shalat sebagai penolongmu, dan sesungguhnya yang demikian itu sungguh berat kecuali bagi orang-orang yang khusyu'. (Q.S. Al-Baqarah : 45)¹

Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya. Dia mendapat (pahala) dari (kebajikan) yang dikerjakannya dan dia mendapatkan (siksa) dari (kejahatan) yang diperbuatnya... (Q.S. Al-Baqarah : 286)¹

Jangan pernah lelah untuk berjuang, berhenti untuk membandingkan. Karena takkan berguna, menutup pintu syukur, menghalang Ikhlas untuk mencari jalan, menuju perasaan sehingga dapat menerima semua dengan hati yang lapang, seperti sebuah lahan kosong yang ditanami dengan rasa syukur dan keikhlasan.
(Alyssa Soebandono)²

-
1. Departemen Agama Republik Indonesia. 1998. Al-Qur'an dan Terjemahan. Semarang : PT Kumudasmoro Grafindo.
 2. Dikutip dari media sosial Instagram, IG:Alyssa Soebandono

PERNYATAAN

Saya yang bertandatangan di bawah ini :

Nama : Fikriana Nur Istiqomah

NIM : 131810101010

menyatakan dengan sesungguhnya bahwa skripsi yang berjudul “Analisis Diskriminan Untuk Validasi *Cluster* Pada Studi Kasus Pengelompokan Kecamatan Di Kabupaten Jember Berdasarkan Status Kemiskinan” adalah benar-benar hasil karya sendiri kecuali untuk kutipan yang sudah saya sebutkan sumbernya, belum pernah diajukan pada institusi manapun, dan bukan karya jiplakan. Saya bertanggungjawab atas keabsahan dan kebenaran isinya sesuai dengan sikap ilmiah yang harus dijunjung tinggi.

Demikian pernyataan ini saya buat dengan sebenarnya, tanpa ada ttekanan dan paksaan dari pihak manapun serta bersedia mendapat sanksi akademik jika ternyata di kemudian hari pernyataan ini tidak benar.

Jember, Februari 2019

Yang menyatakan

Fikriana Nur Istiqomah

NIM 131810101010

SKRIPSI

**ANALISIS DISKRIMINAN UNTUK VALIDASI *CLUSTER*
PADA STUDI KASUS PENGELOMPOKAN KECAMATAN DI
KABUPATEN JEMBER BERDASARKAN STATUS
KEMISKINAN**

Oleh

Fikriana Nur Istiqomah

131810101010

Pembimbing

Dosen Pembimbing Utama : Prof. I Made Tirta, M.Sc., Ph.D

Dosen Pembimbing Anggota : Dian Anggraeni, S.Si., M.Si

PENGESAHAN

Skripsi berjudul “Analisis Diskriminan Untuk Validasi *Cluster* Pada Studi Kasus Pengelompokan Kecamatan Di Kabupaten Jember Berdasarkan Status Kemiskinan” telah diuji dan disahkan pada:

Hari, tanggal :

Tempat : Fakultas MIPA Universitas Jember

Tim Penguji :

Dosen Pembimbing Utama,

Dosen Pembimbing Anggota,

Prof. I Made Tirta, M.Sc., Ph.D

Dian Anggraeni, S.Si., M.Si

NIP. 195912201985031002

NIP. 198202162006042002

Dosen Penguji I,

Dosen Penguji II,

Dr. Alfian Futuhul Hadi, S.Si., M.Si

Ika Hesti Agustin, S.Si., M.Si

NIP. 197407192000121001

NIP. 198408012008012006

Mengesahkan

Dekan,

Drs. Sujito, Ph.D

NIP. 196102041987111001

RINGKASAN

Analisis Diskriminan Untuk Validasi Cluster Pada Studi Kasus Pengelompokan Kecamatan Di Kabupaten Jember Berdasarkan Status Kemiskinan; Fikriana Nur Istiqomah; 131810101010; 2018; 81 halaman; Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Analisis *cluster* merupakan salah satu analisis dari teknik analisis multivariat yang bertujuan untuk meng-*cluster*-kan data pada observasi ataupun pada variabel-variabel ke dalam *cluster* sedemikian rupa sehingga untuk masing-masing *cluster* bersifat homogen sesuai dengan faktor yang digunakan untuk melakukan analisis *cluster*. Pada analisis *cluster*, metode-metode pengelompokan yang digunakan tidak ada yang secara khusus menjelaskan bahwa ada metode yang tepat digunakan dalam proses pengelompokan data, karena masing-masing metode pengelompokan mempunyai kelebihan dan kelemahan. Pengelompokan yang baik adalah kelompok yang mempunyai kesamaan yang tinggi antar individu dalam satu kelompok dan kesamaan yang rendah antar kelompoknya. Untuk mendapatkan metode pengelompokan yang tepat penggunaannya dalam pengelompokan data serta memenuhi kriteria pengelompokan yang baik, diperlukan adanya validasi kelompok (*validasi cluster*) sehingga nantinya dihasilkan pengelompokan data yang tepat (benar).

Berdasarkan hal tersebut, penelitian ini bertujuan mengetahui prosedur penentuan metode analisis *cluster* yang terbaik dari kedua metode analisis *cluster* yaitu metode hirarki, dan *kmeans*; Mengetahui hasil dari metode analisis *cluster* dan hasil dari validasi analisis *cluster* terhadap pengelompokan menggunakan data “iris” untuk ketiga metode validasi yang digunakan yaitu validasi internal, validasi stabilitas, dan menggunakan uji analisis diskriminan dan dapat diaplikasikan terhadap pengelompokan kecamatan-kecamatan di Kabupaten Jember berdasarkan status kemiskinan dan mengetahui bahwa uji analisis diskriminan tepat digunakan sebagai salah satu metode untuk memvalidasi hasil solusi dari analisis *cluster* untuk data secara umum. Adapun hasil validasi metode pengelompokan dan

jumlah pengelompokannya terhadap pengelompokan data simulasi yang ada di web Pusat Belajar dan Laboratorium Statistika Virtual yaitu data “iris” dan data Kecamatan di Kabupaten Jember berdasarkan peubah status kemiskinannya mendapatkan hasil pengelompokan yang berbeda-beda disetiap metode yang digunakan. Data “iris” digunakan sebagai salah satu acuan pengelompokan dalam analisis data ini, karena data “iris” mempunyai klasifikasi variabel kelompok yang asli sehingga dengan lebih mudah dilihat ketepatan pengelompokan analisis data yang dihasilkan dari beberapa metode yang digunakan dengan kelompok yang telah terdefinisi sebelumnya. Tidak hanya satu metode yang tepat dalam pengelompokannya, namun ada beberapa metode. Hasil pengelompokan data “iris”, pada proses validasi menggunakan validasi internal dan stabilitas, dan juga uji analisis diskriminan adalah menggunakan metode hirarki tautan *single linkage* dengan pembentukan kelompok optimal sebanyak 3 kelompok. Pada data “iris” tautan *single linkage* digunakan untuk membuat model diskriminannya karena merupakan metode pengelompokan yang optimal untuk semua metode validasi yang akan diaplikasikan selanjutnya pada data Kecamatan di Kabupaten Jember berdasarkan peubah status kemiskinannya. Persentase ketepatan pengklasifikasian objek pada data “iris” sebesar 100% (*hit ratio* = 100 %) dengan tingkat kesalahan. Hasil pengelompokan data kecamatan di Kabupaten Jember berdasarkan peubah status kemiskinannya pada proses validasi menggunakan validasi internal dan stabilitas, dan juga uji analisis diskriminan adalah berdasarkan tujuh variabel yang digunakan yaitu kepadatan penduduk (Jiwa/km²) (X_1), banyaknya sekolah (SD/MI (Negeri & Swasta) (X_2), SLTP/MTs (Negeri & Swasta), SMA/MA (Negeri & Swasta), banyaknya fasilitas kesehatan (Rumah sakit, Puskesmas) (X_3), banyaknya Rumah Tangga Pengguna PLN (X_4), banyaknya Rumah Tangga yang mempunyai Telepon (X_5), banyaknya Industri berbadan hukum (Industri Kecil, Industri Sedang, dan Industri Besar) (X_6) dan Produksi Pertanian (Ton) (X_7) menggunakan metode hirarki tautan *single linkage* (berdasarkan acuan metode pengelompokan yang dilakukan pada data “iris) dan *complete linkage* (berdasarkan metode yang paling banyak dihasilkan sebagai metode pengelompokan paling baik dari semua metode validasi) dengan pembentukan

kelompok optimal sebanyak 3 kelompok. Analisis diskriminan linier merupakan analisis yang digunakan untuk mengklasifikasikan objek dari data kemiskinan Kecamatan Kabupaten Jember kedalam kelompok kecamatan dengan ekonomi menengah kebawah. Persentase ketepatan pengklasifikasian objek pada data Kecamatan di Kabupaten Jember berdasarkan peubah status kemiskinannya adalah 100% (*hit ratio* = 100%) dengan tingkat kesalahan klasifikasinya adalah 0 menggunakan tautan *single linkage* dan *complete linkage*.



PRAKATA

Puji syukur kehadirat Allah SWT., yang telah melimpahkan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Analisis Diskriminan Untuk Validasi *Cluster* Pada Studi Kasus Pengelompokan Kecamatan Di Kabupaten Jember Berdasarkan Status Kemiskinan”. Skripsi ini disusun untuk memenuhi salah satu syarat untuk menyelesaikan pendidikan Strata Satu (S1) pada Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Penyusunan skripsi ini tidak lepas dari bantuan berbagai pihak. Oleh karena itu, penulis menyampaikan terimakasih kepada:

1. Kusbudiono, S.Si., M.Si., dan Ahmad Kamsyakawuni, S.Si., M.Kom., selaku Ketua Jurusan dan Wakil Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember;
2. Prof. I Made Tirta, M.Sc., Ph.D., selaku Dosen Pembimbing Utama, Dian Anggraeni, S.Si., M.Si, selaku Dosen Pembimbing Anggota, Dr. Alfian Futuhul Hadi, S.Si., M.Si., selaku Dosen Penguji I, dan Ika Hesti Agustin, S.Si., M.Si., selaku Dosen Penguji II yang telah meluangkan waktu, pikiran, dan perhatian dalam penulisan skripsi ini;
3. Prof. I Made Tirta, M.Sc., Ph.D., selaku Dosen Pembimbing Akademik yang telah membimbing selama penulis menjadi mahasiswa;
4. Dosen dan karyawan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember;
5. Semua pihak yang tidak dapat disebutkan satu persatu.

Penulis juga menerima kritikk dan saran dari semua pihak demi kesempurnaan skripsi ini. Akhirnya penulis berharap, semoga skripsi ini dapat bermanfaat.

Jember, Februari 2019

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PERSEMBAHAN	ii
HALAMAN MOTTO	iii
HALAMAN PERNYATAAN	iv
HALAMAN PEMBIMBINGAN	v
HALAMAN PENGESAHAN	vi
RINGKASAN	vii
PRAKATA	x
DAFTAR ISI	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN	xv
BAB 1. PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	5
1.4 Batasan Masalah	5
1.5 Manfaat Penelitian	5
BAB 2. TINJAUAN PUSTAKA	
2.1 Analisis Multivariat	6
2.2 Analisis Cluster	7
2.2.1 Metode Hirarki	7
2.2.2 Metode Non Hirarki	9
2.3 Teknik Pengukuran Jarak	10
2.3.1 Ukuran Asosiasi	10
2.3.2 Ukuran Korelasi	10
2.3.3 Ukuran Kedekatan	11
2.4 Validasi Cluster	12

2.4.1	Validasi Internal	12
2.4.2	Validasi Stabilitas	14
2.5	Analisis Diskriminan	16
2.5.1	Asumsi Analisis Diskriminan	16
2.5.2	Model Analisis Diskriminan	17
2.5.3	Uji Validasi dari Analisis Diskriminan	18
2.6	Web Interaktif R-Shiny	22
BAB 3.	METODE PENELITIAN	
3.1	Obyek Penelitian	25
3.2	Langkah-langkah Penelitian	26
3.3	Pengolahan Data dengan Menggunakan Program R	29
BAB 4.	HASIL DAN PEMBAHASAN	
4.1	Hasil dan Pembahasan Validasi Cluster Data “iris”	33
4.2	Hasil dan Pembahasan Validasi Cluster Data Kemiskinan Kecamatan di Kabupaten Jember	36
BAB 5.	KESIMPULAN DAN SARAN	
5.1	Kesimpulan	40
5.2	Saran	43
DAFTAR PUSTAKA	44
LAMPIRAN	48

DAFTAR TABEL

	Halaman
Tabel 2.1 Analisis <i>Cluster</i>	12
Tabel 2.2 Nilai Validasi	15
Tabel 2.3 Klasifikasi Tiga Kelompok	19
Tabel 4.1 Perbandingan Hasil Validasi Keseluruhan Data “iris”	34
Tabel 4.2 Keباikan model analisis diskriminan untuk metode analisis diskriminan linier data “iris”	35
Tabel 4.3 Perbandingan Hasil Validasi Keseluruhan Data Kemiskinan Kecamatan di Kabupaten Jember	36
Tabel 4.4 Ketepatan Klasifikasi Masing-masing Kelompok Pada Metode <i>Single Linkage</i> data Kemiskinan Kecamatan Kabupaten Jember	38
Tabel 4.5 Ketepatan Klasifikasi Masing-masing Kelompok Pada Metode <i>Complete Linkage</i> data Kemiskinan Kecamatan Kabupaten Jember.....	38

DAFTAR GAMBAR

	Halaman
Gambar 2.1 Struktur Komponen R-Shiny	23
Gambar 3.1 Diagram Alur Penelitian	28



DAFTAR LAMPIRAN

	Halaman
A. Data Kemiskinan Kecamatan di Kabupaten Jember	48
B. Hasil Validasi Solusi <i>Cluster</i>	50
B.1 Tampilan Hasil	50
B.2 Perhitungan Hasil Validasi Menggunakan Uji Analisis Diskriminan	61
C. Data “iris”	63
D. Hasil Validasi Solusi <i>Cluster</i>	67
D.1 Tampilan Hasil	65
D.2 Perhitungan Hasil Validasi Menggunakan Uji Analisis Diskriminan	76

BAB 1. PENDAHULUAN

1.1 Latar Belakang

Analisis Multivariat adalah salah satu jenis analisis statistika yang dapat digunakan untuk menganalisis data dengan variabel *independent* dan/atau variabel *dependent*-nya lebih dari satu. Analisis multivariat dapat dibagi menjadi beberapa analisis di antaranya adalah analisis komponen utama, analisis *cluster*, analisis diskriminan dan analisis faktor (Supranto,2004). Analisis *cluster* bertujuan untuk mengelompokkan objek ke dalam suatu kelompok sehingga objek yang berada dalam satu kelompok akan memiliki kesamaan yang tinggi daripada objek yang berada di kelompok lain (Sharma,1996).

Menurut Shidqon (2012) pada analisis *cluster*, tidak ada metode-metode pengelompokan yang secara khusus dapat menjelaskan tentang ketepatan sebuah metode digunakan dalam proses pengelompokan data. Masing-masing metode pengelompokan memiliki kelebihan dan kelemahan, diantaranya adalah sensitif terhadap adanya *outlier*, kesulitan menangani variasi bentuk dan ukuran, serta memisahkan kelompok yang besar. Pengelompokan yang baik yaitu memiliki kesamaan individu dalam satu kelompok yang tinggi dan kesamaan antar kelompok yang rendah serta tidak adanya *outlier*. Untuk mendapatkan metode pengelompokan yang tepat penggunaannya dalam pengelompokan data dan memenuhi kriteria pengelompokan yang baik, dibutuhkan proses validasi *cluster* atau validasi pengelompokan, sehingga nantinya dari pada pengelompokan data tersebut akan didapatkan hasil yang lebih baik. Menurut Ninda (2017), validasi *cluster* adalah prosedur yang mengevaluasi hasil analisis *cluster* secara kuantitatif dan objektif. Menurut Brock (2008) validasi *cluster* memiliki 3 macam metode yang digunakan sebagai acuan untuk menentukan jumlah *cluster* optimal yaitu validasi internal, validasi stabilitas, dan validasi biologis. Dalam penelitian ini, saya menggunakan validasi internal, validasi stabilitas dan juga menggunakan uji analisis diskriminan sebagai salah satu metode untuk memvalidasi hasil solusi *cluster*.

Analisis diskriminan adalah salah satu analisis multivariat yang termasuk *Dependence Method*, yaitu terdapat variabel independen dan dependen sehingga ada hasil variabel yang hasilnya tergantung dari data variabel independen, dan mempunyai nilai *hit ratio*, *cmax* dan *Press's Q* yang berfungsi untuk menunjukkan pengelompokan yang dilakukan adalah optimal (baik). Pada penelitian ini, solusi *cluster* menjadi variabel dependen yang digunakan untuk mendapatkan pengelompokan yang optimal. Adapun pengolahan data dilakukan dengan menggunakan akses web online *Virtual Statistics Laboratory* Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember dengan alamat <http://statslab-rshiny.fmipa.unej.ac.id/RProg/Multiv/>.

R-Shiny merupakan salah satu versi R berbasis web, yang mengalami beberapa perubahan atau mengalami suatu revolusi. Revolusi R terjadi saat tim R-Studio meluncurkan dua program pendukung R yaitu :

1. Rstudio (bersifat online dan sudah dilengkapi dengan GUI editor), namun pemanfaatan R masih menggunakan pendekatan CLI. Program ini cocok untuk para pengembang statistika dan mahasiswa jurusan statistika yang bekerja dalam laboratorium terpusat. Namun program ini tetap belum memenuhi kebutuhan para pengguna statistika yang kemampuan pemrogramannya tidak terlalu kuat.
2. S-Shiny, yang merupakan *interface* (yang dilengkapi dengan server *interface*) yang memungkinkan orang membuat laman web (*web pages*) interaktif sehingga kemampuan R yang pada dasarnya bersifat CLI bisa diakses melalui menu web secara GUI web.

Sejak diluncurkannya R-Shiny oleh Rstudio & Inc (2013,2014), banyak pengguna R membuat contoh aplikasi menggunakan R yang sebagian besar merupakan visualisasi dan simulasi. Perkembangan dari dimanfaatkannya R-Shiny adalah dibentuknya Pusat Belajar dan Laboratorium Statistika Virtual yang bersifat web dan didalamnya berisi berbagai catatan kuliah dan tutorial tentang analisis data menggunakan statistika, serta sebagai pusat analisis data (interaktif) online (Tirta; 2014a, 2014b, 2015). Menurut Tirta (2016) *Virtual Statistics Laboratory* bermanfaat baik bagi mahasiswa maupun dosen / peneliti untuk

berbagai keperluan, misalnya: untuk belajar statistika; berlatih analisis data; dan analisis data langsung. Serta mempunyai kelebihan dan kekurangan, Salah satu kekurangan dari analisis data online yang ada ialah ketergantungan pada keberadaan jaringan internet yang memadai dan juga keterbatasan jumlah pengguna secara serempak (server masih menggunakan versi gratis, bukan yang profesional) (Tirta, 2016). Sedangkan kelebihan dari *Virtual Statistics Laboratory* diantaranya adalah:

1. Tidak perlu lagi mengunduh maupun memasang program pada komputer;
2. Dapat diakses melalui berbagai jenis komputer dan gadget, dengan berbagai platform;
3. Disertai dengan paparan teori dan langkah-langkah yang dapat dipraktikkan untuk data yang berbeda-beda;
4. Mencakup hampir sebagian besar jenis analisis data.

Penelitian analisis *cluster* telah dilakukan oleh beberapa peneliti diantaranya Sitepu (2011) yang menggunakan analisis *cluster* terhadap tingkat pencemaran udara pada sektor industri, dengan hasil validasi diuji menggunakan parsial F yang akan menjadi variabel pembeda dalam peng-cluster-an, Dewi (2014) yang menggunakan aplikasi metode-metode agglomerative dalam analisis cluster pada data tingkat polusi udara untuk hasil validasi dengan plot RMSSTD (*Root Mean Square Standard Deviation*) dan CD (*Cluster Distance*) sehingga hasil plot dari jumlah cluster yang digunakan menghasilkan jumlah *cluster* yang valid (dapat dipercaya) dan dapat dipilih 5 sampai 6 *cluster* dan Shidqon (2012) yang menggunakan metode validasi internal untuk validasi *cluster* pada pengelompokan Kecamatan di Kabupaten Jember berdasarkan status kemiskinan. Akan tetapi setiap data dan metode-metode yg digunakan pada hasil validasi memiliki karakteristik tersendiri sehingga memungkinkan adanya pemilihan data dan penentuan metode hasil validasi yang berbeda.

Analisis diskriminan untuk validasi *cluster* digunakan sebagai pengklasifikasian suatu individu atau objek ke dalam suatu kelompok yang telah ditentukan sebelumnya berdasarkan variabel-variabel bebasnya (Dillon dan Goldstein, 1984). Pada penelitian ini, digunakan data simulasi yang ada di web

Pusat Belajar dan Laboratorium Statistika Virtual yaitu data “iris” dan data Kecamatan di Kabupaten Jember berdasarkan peubah status kemiskinannya. Data “iris” digunakan sebagai salah satu acuan pengelompokan dalam analisis data ini, karena data “iris” mempunyai klasifikasi variabel kelompok yang asli sehingga dengan lebih mudah dilihat ketepatan pengelompokan analisis data yang dihasilkan dari beberapa metode yang digunakan dengan kelompok yang telah terdefinisi sebelumnya. Berdasarkan hasil pengelompokan dengan metode analisis data yang tepat digunakan dalam data “iris” sebelumnya diaplikasikan pada pengelompokan data Kecamatan di Kabupaten Jember berdasarkan peubah status kemiskinannya yang juga dapat di kelompokkan ke dalam beberapa *cluster* kecamatan yang mempunyai sifat relatif sama atau kemiripan sifat berdasarkan peubah status kemiskinannya. Kemiskinan didefinisikan sebagai konsekuensi dari suatu struktur masyarakat dengan penduduk yang padat, terbatasnya sumber daya, tingkat kesehatan yang rendah dan kesempatan pendidikan yang tidak merata (Lewis, 1975, dalam Ernayanti & Ita Novita, 1996).

1.2 Rumusan Masalah

Berdasarkan latar belakang diperoleh rumusan masalah sebagai berikut :

1. Bagaimana prosedur penentuan metode analisis *cluster* yang terbaik dari kedua metode analisis *cluster* yaitu metode hirarki, dan *kmeans*?
2. Bagaimana hasil dari metode analisis *cluster* dan hasil dari validasi analisis *cluster* terhadap pengelompokan menggunakan data “iris” untuk ketiga metode validasi yang digunakan yaitu validasi internal, validasi stabilitas, dan menggunakan uji analisis diskriminan?
3. Bagaimana hasil dari validasi analisis *cluster* dan hasil dari metode analisis *cluster* yang didapatkan pada hasil validasi terhadap pengelompokan kecamatan-kecamatan di Kabupaten Jember berdasarkan status kemiskinan untuk ketiga metode validasi yang digunakan yaitu validasi internal, validasi stabilitas, dan menggunakan uji analisis diskriminan?
4. Apakah uji analisis diskriminan tepat digunakan sebagai salah satu metode memvalidasi hasil solusi dari analisis *cluster* untuk data secara umum?

1.3 Tujuan Penelitian

Tujuan yang ingin dicapai dari penulisan ini adalah :

1. Mengetahui prosedur penentuan metode analisis *cluster* yang terbaik dari kedua metode analisis *cluster* yaitu metode hirarki, dan *kmeans*.
2. Mengetahui hasil dari metode analisis *cluster* dan hasil dari validasi analisis *cluster* terhadap pengelompokan menggunakan data “iris” untuk ketiga metode validasi yang digunakan yaitu validasi internal, validasi stabilitas, dan menggunakan uji analisis diskriminan.
3. Mengetahui hasil dari validasi analisis *cluster* dan hasil dari metode analisis *cluster* yang didapatkan pada hasil validasi terhadap pengelompokan kecamatan-kecamatan di Kabupaten Jember berdasarkan status kemiskinan untuk ketiga metode validasi yang digunakan yaitu validasi internal, validasi stabilitas, dan menggunakan uji analisis diskriminan.
4. Mengetahui bahwa uji analisis diskriminan tepat digunakan sebagai salah satu metode untuk memvalidasi hasil solusi dari analisis *cluster* untuk data secara umum.

1.4 Batasan Masalah

Agar tidak memperluas pembahasan dan mendapatkan hasil yang maksimal, penulisan penelitian ini dibatasi pada hal berikut, metode analisis diskriminan yang digunakan untuk memvalidasi *Cluster* yaitu metode Analisis Diskriminan Linear.

1.5 Manfaat Penelitian

Manfaat yang diperoleh pada penelitian ini yaitu dapat mengelompokkan sebuah data untuk data secara umum menggunakan analisis diskriminan sebagai salah satu metode untuk validasi *cluster*.

BAB 2. TINJAUAN PUSTAKA

2.1 Analisis Multivariat

Analisis multivariat adalah analisis statistik yang digunakan untuk menganalisis data yang terdiri dari beberapa variabel dan masing-masing variabel tersebut saling berkorelasi satu sama lain. Analisis multivariat dapat dibagi menjadi dua, yaitu analisis dependensi dan analisis interdependensi. Ciri-cirinya adalah pada analisis dependensi ada satu atau beberapa variabel yang berfungsi sebagai variabel bebas dan variabel terikat, seperti analisis regresi linier berganda, analisis diskriminan, analisis logit, dan analisis korelasi kanonik. Sebaliknya ciri-ciri pada analisis interdependensi adalah semua variabel yang digunakan bersifat independen, seperti analisis faktor, analisis *cluster*, dan *multidimensional scaling* (Sarwono, 2007).

Menurut Yusuf (2003) bahwa perbedaan analisis multivariat dengan analisis-analisis statistik yang lain adalah bahwa jumlah variabel terikat pada analisis statistik lain, seperti analisis regresi berganda yang terdiri hanya satu variabel misalnya (Y) tetapi pada analisis multivariat, variabel terikat dapat berjumlah lebih dari satu (misalnya Y_1, Y_2, \dots, Y_n). Menurut Fadli (2011:15), dalam analisis multivariat data dapat dinyatakan ke dalam bentuk matriks dimana terdapat n obyek dan p variabel. Misalkan pada suatu pengamatan terdapat p variabel yaitu $X = (X_1, X_2, \dots, X_p)$ dari bnyaknya n obyek, sehingga data tersebut dapat dinyatakan dalam bentuk matriks seperti di bawah ini:

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$$

2.2 Analisis Cluster

Analisis *cluster* merupakan salah satu analisis dari teknik analisis multivariat yang bertujuan untuk meng-*cluster*-kan data pada observasi ataupun pada variabel-variabel ke dalam *cluster* sedemikian rupa sehingga untuk masing-masing *cluster* bersifat homogen sesuai dengan faktor yang digunakan untuk melakukan analisis *cluster* dengan dasar peng-*cluster*-an adalah kesamaan skor nilai yang dianalisis. Data pada ukuran kesamaan tersebut dapat dianalisis dengan analisis *cluster* maka selanjutnya dapat ditentukan siapa saja yang masuk pada tiap-tiap *cluster* (Gudono,2011). Analisis *cluster* terbagi menjadi dua metode yaitu metode hirarki dan metode non hirarki.

Pada metode hirarki jumlah kelompok yang akan diperoleh belum diketahui, sedangkan dalam metode nonhirarki diasumsikan ada k kelompok terlebih dahulu. Metode hirarki dibagi menjadi dua, yaitu metode *agglomerative* (pemusatan) dan metode *divisive* (penyebaran). Metode-metode yang termasuk dalam *agglomerative* adalah *Single Linkage Method*, *Complete Linkage Method*, *Avarage Linkage Method*, *Ward's Method*, *Centroid Method* dan *Median Method* (Everitt,1998). Sedangkan pada metode yang termasuk metode nonhirarki adalah *k-means* dan *Mixture Modelling*.

2.2.1 Metode Hirarki

Metode hirarki dimulai dengan mengelompokkan untuk dua atau lebih obyek yang memiliki kesamaan paling dekat. Setelah itu, proses dilanjutkan ke obyek lain yang mempunyai kedekatan kedua. Seperti itu seterusnya sampai *cluster* akan membentuk serupa “pohon” hirarki (tingkatan) yang jelas antar obyek 1 yaitu pada tingkatan memiliki kesamaan paling dekat sampai pada tingkatan yang memiliki kesamaan paling jauh (tidak sama). Alat bantu pada analisis *cluster* disini adalah berupa dendogram yang digunakan untuk memperjelas proses metode hirarki tersebut.

Metode hirarki dibedakan menjadi dua metode peng-*cluster*-an, yaitu :

1. Metode Aglomeratif

Pada metode ini, pendekatan *cluster* dimulai dengan n *cluster* sehingga masing-masing *cluster* memiliki tepat satu obyek, selanjutnya ditentukan dua *cluster* terdekat dan digabungkan menjadi satu *cluster* baru. Proses tersebut dapat diulangi sampai diperoleh satu *cluster* yang memuat semua himpunan data, yang selalu diikuti dengan perbaikan matriks jarak dan hasilnya disajikan dalam bentuk dendogram. Metode Aglomeratif ada beberapa macam, yaitu :

- 1) Metode *Single Linkage*, untuk menentukan jaraknya dengan metode ini dapat dilakukan dengan melihat jarak antar dua *cluster* yang ada, kemudian memilih jarak paling dekat (*nearest neighbour rule*).
- 2) Metode *Complete Linkage (farthest-neighbour method)*, jarak antar *cluster* ditentukan oleh jarak terjauh (*farthest-neighbour*) antara dua obyek dalam *cluster* yang berbeda.

$$d_{(IJ)K} = \max\{d_{IJ}, d_{JK}\} \quad (2.1)$$

Dimana d_{IJ} dan d_{JK} masing-masing adalah jarak antara anggota yang paling jauh dari *cluster* I dan J serta *cluster* J dan K (Johnson & Wichern, 1992).

- 3) Metode *Average Linkage*, menganggap jarak antara dua *cluster* sebagai jarak rata-rata antar semua anggota dalam satu *cluster* dengan semua anggota *cluster* lain.

$$d_{(IJ)K} = \frac{\sum a \sum b d_{ab}}{N_{IJ}N_K} \quad (2.2)$$

dengan,

d_{ab} : jarak antara obyek a pada *cluster* (IJ) dan obyek b pada *cluster* K.

N_{IJ} : jumlah item pada *cluster* (IJ).

N_K : jumlah item pada *cluster* (IJ) dan K.

2.2.2. Metode Nonhirarki

Metode non hirarki (partisi) dimulai dengan proses penentuan jumlah *cluster* terlebih dahulu, dan yang digunakan adalah metode *k-means*. Metode *kmeans* digunakan sebagai metode alternatif analisis *cluster* untuk data dengan ukuran yang besar karena mempunyai kecepatan yang lebih tinggi dibandingkan dengan metode hirarki. Mac Queen menyarankan bahwa penggunaan *kmeans* untuk menjelaskan algoritma dalam penentuan suatu obyek ke dalam *cluster* tertentu berdasarkan rata-rata terdekat. Proses peng-*cluster*-an dengan *kmeans* adalah :

1. Menentukan besarnya *k*, yaitu banyaknya *cluster* dan menentukan *centroid* (rata-rata) di tiap *cluster*;
2. Menghitung jarak tiap obyek dengan masing-masing *centroid*;
3. Menghitung kembali rata-rata (*centroid*) untuk *cluster* yang terbentuk;
4. Mengulangi langkah 2 sampai tidak ada lagi pemindahan obyek antar *cluster*.

Metode *kmean* ini bertujuan untuk meminimumkan *error* akibat partisi *n* obyek ke dalam *k cluster* dan *error* partisi dapat disebut sebagai fungsi obyektif.

Misalkan $X = \{x_i\}, i = 1, 2, \dots, n$ merupakan titik-titik dalam ruang berdimensi n (R^n) dan titik tersebut dikelompokkan ke dalam i *cluster*, $C_i, i = 1, 2, \dots, K$. Misalkan c_i centroid dari *cluster* C_i sehingga jumlah kuadrat antara c_i dan titik dalam *cluster* yaitu x_i , didefinisikan sebagai :

$$J(c_k) = \sum_{x_i \in c_i} (c_i - x_i)^2 \quad (2.3)$$

Prinsip dasar dari metode *k-means* adalah meminimumkan jumlah kuadrat *error* dari seluruh i *cluster*, yaitu :

$$SSE = \sum_{i=1}^k \sum_{x_i \in c_i} (c_i - x_i)^2 \quad (2.4)$$

2.3 Teknik Pengukuran Jarak

Pada analisis *cluster* mempunyai tiga ukuran untuk mengukur kesamaan antar obyek, yaitu ukuran asosiasi, ukuran korelasi, dan ukuran kedekatan.

2.3.1 Ukuran Asosiasi

Ukuran asosiasi biasanya digunakan untuk mengukur data berskala non metrik (nominal atau ordinal), dengan mengambil bentuk-bentuk dari koefisien korelasi pada tiap obyeknya, dengan cara memutlakan korelasi-korelasi yang bernilai negatif (Simamora, 2005).

2.3.2 Ukuran Korelasi

Ukuran korelasi biasanya digunakan untuk mengukur data skala matriks, namun ukuran ini jarang digunakan karena titik beratnya pada nilai suatu pola tertentu, padahal disini titik berat analisis *cluster* terletak pada besarnya obyek. Kesamaan antar obyek dapat diketahui dari koefisien korelasi antar pasangan obyek yang diukur dengan menggunakan beberapa variabel. Berikut beberapa jenis dari metode ukuran korelasi :

Pearson correlation distance adalah ukuran korelasi yang mengukur tingkat hubungan linier antara dua profil.

$$d_{cor}(x, y) = 1 - \frac{\sum_{i=1}^n (x_i - \bar{x}) - (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2.5)$$

Eisen cosine correlation distance (Eisen et al., 1998), khusus korelasi pearson dengan \bar{x} dan \bar{y} yang keduanya diganti dengan nol.

$$d_{eisen}(x, y) = 1 - \frac{|\sum_{i=1}^n x_i y_i|}{\sqrt{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2}} \quad (2.6)$$

Spearman correlation distance, menghitung korelasi antara rank x dan peringkat dari variabel y.

$$d_{spear}(x, y) = 1 - \frac{\sum_{i=1}^n (x'_i - \bar{x}') - (y'_i - \bar{y}')}{\sqrt{\sum_{i=1}^n (x'_i - \bar{x}')^2 \sum_{i=1}^n (y'_i - \bar{y}')^2}} \quad (2.7)$$

dengan , $x'_i = \text{rank}(x_i)$ dan $y'_i = \text{rank}(y_i)$.

Kendall correlation distance, mengukur korespondensi antara rank x dan variabel y . Jumlah total kemungkinan pasangan x dan pengamatan y adalah $n(n-1)/2$, dimana n adalah ukuran x dan y . Jika x dan y berkorelasi, maka pasangan x dan y memiliki urutan peringkat relatif yang sama.

$$d_{kend}(x, y) = 1 - \frac{n_c - n_d}{\frac{1}{2}n(n-1)} \quad (2.8)$$

Dengan, n_c : jumlah pasangan sesuai

n_d : jumlah pasangan tidak sesuai

n : ukuran x dan y

Dari penjelasan sebelumnya, terdapat catatan bahwa :

- Analisis korelasi pearson adalah metode yang paling umum digunakan dan dikenal sebagai korelasi parametrik yang bergantung pada distribusi data.
- Korelasi Kendall dan korelasi Spearman adalah non-parametrik dan biasanya digunakan untuk melakukan analisis korelasi berbasis peringkat.

2.3.3 Ukuran Kedekatan

1) *Euclidean distance* adalah metrika yang sering digunakan untuk menghitung kesamaan dua vektor. *Euclidean distance* menghitung akar dari kuadrat perbedaan dua vektor. Rumus *Euclidean distance* adalah sebagai berikut (Putra, 2010) :

$$d_{ij} = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \quad (2.9)$$

dengan,

d_{ij} = distance antara obyek ke- i dan obyek ke- j

$i = i_1, i_2, i_3, \dots, i_p$

$j = j_1, j_2, j_3, \dots, j_p$

k = merepresentasikan nilai atribut

n = dimensi data

x = obyek data

Semakin besar jarak antara dua vektor, maka tingkat kesamaan atau kemiripannya kecil. Sebaliknya, semakin kecil jarak antar dua vektor, maka tingkat kesamaan atau kemiripannya besar.

Tabel 2.1 Analisis *Cluster*

No.	Jenis <i>Cluster</i>	Jarak	Tautan	Korelasi
1.	Hirarki (pengelompokan untuk dua/lebih objek dengan kesamaan paling dekat) <ul style="list-style-type: none"> • Aglomeratif 	<i>Euclid</i>	<ul style="list-style-type: none"> • <i>Single Linkage</i> • <i>Complete Linkage</i> • <i>Average Linkage</i> 	<ul style="list-style-type: none"> • <i>Pearson correlation distance</i> • <i>Eisen cosine correlation distance</i> • <i>Spearman correlation distance</i> • <i>Kendall correlation distance</i>
2.	Non Hirarki <ul style="list-style-type: none"> • <i>K-Means</i> 	<i>Euclid</i>	Tidak memiliki tautan.	

2.4 Validasi *Cluster*

Setiap kelompok yang dibentuk pada tahapan sebelumnya, masing-masing kelompok tersebut memiliki seperangkat ukuran karakteristik, yang diantaranya berupa nilai indeks validitas *cluster* (Brock *et al*, 2008). Nilai indeks validasi *cluster* tersebut digunakan sebagai acuan dalam menentukan kelompok optimal.

2.4.1 Validasi Internal

Validasi internal ialah validasi kelompok yang menggunakan informasi internal pada data untuk menilai kualitas pengelompokan berdasarkan sifat strukturalnya dan dapat mencerminkan kepadatan, hubungan dan pemisahan kelompok partisi yang masing-masing menunjukkan kegunaannya. Seperti kepadatan, berhubungan dengan mengevaluasi homogenitas kelompok, biasanya terlihat dari varian *intra-cluster*-nya. Hubungan menunjukkan penempatan beberapa data pengamatan dalam sebuah kelompok, dimana data tersebut sebagai tetangga terdekat sehingga mempunyai nilai kepadatan yang diukur dengan konektivitas. Pemisahan *cluster* partisi menunjukkan tingkat pengukuran jarak antar dua kelompok (menggunakan jarak antar *centroid*). *Silhouette* dan *Dunn*

index adalah bentuk pengukuran yang mengkombinasikan keduanya. Validasi internal meliputi :

a. Konektivitas (*Connectivity*)

Konektivitas menunjukkan tingkat hubungan kelompok, ditentukan dengan jumlah tetangga terdekat. Dalam pengukuran ini seluruh objek pengamatan berada dalam kelompok masing-masing yang nantinya akan dikaitkan satu sama lain dengan jarak yang paling minimum. Nilai konektivitas dihitung menggunakan persamaan (2.10).

$$K = \sum_{i=1}^N \sum_{j=1}^L x_{i,nn_{i(j)}} \quad (2.10)$$

Dengan, K: konektivitas; N: jumlah data; L: jumlah tetangga; $nn_{i(j)}$: ukuran jarak tetangga terdekat data ke-i, dimana, $nn_{i(j)} : \begin{cases} 0, i \text{ dan } j \text{ satu cluster} \\ 1, j \text{ lainnya} \end{cases}$.

b. Nilai rata-rata *Silhouette*

Nilai rata-rata *Silhouette* merupakan rata-rata nilai *Silhouette* setiap data. Nilai *Silhouette* mengukur tingkat kepercayaan pada proses *clustering* dari setiap data observasi. Nilai *Silhouette* dihitung menggunakan persamaan (2.11).

$$S(i) = \frac{b_i - a_i}{\max(b_i, a_i)} \quad (2.11)$$

dengan,

$S(i)$: nilai rata-rata *Silhouette*;

b_i : jarak rata-rata antara data ke-i dan objek data pada kelompok berbeda (jarak terdekat);

a_i : jarak rata-rata antara data ke-i dengan seluruh objek data lain dalam kelompok yang sama.

c. *Dunn Index*

Dunn index merupakan rasio dari jarak terdekat antara objek data di *cluster* berbeda terhadap jarak terjauh pada intra-*cluster*. Nilai indeks *Dunn* dapat dihitung menggunakan persamaan (2.12).

$$D(C) = \frac{\min_{C_k, C_1 \in C, C_k \neq C_1} \left(\min_{i \in C_k, j \in C_1} \text{dist}(i, j) \right)}{\max_{C_m \in C} \text{diam}(C_m)} \quad (2.12)$$

dengan,

$D(C)$: *dunn Index*;

$diam(C_m)$: jarak maksimum antar data dalam *cluster* ke- m .

Pengelompokan dianggap bersifat *compact* dan *well separated* jika jarak antar kelompok besar dan diameter kelompok kecil, nilai *Dunn index* berada dalam selang nilai 0 sampai ∞ , sehingga jika *Dunn index* memiliki nilai semakin besar maka kelompok semakin baik (Halkidi *et al.*, 2001).

Menurut Shidqon (2012) Semakin minimal nilai pada *connectivity* maka masing-masing individu yang dimasukkan pada kelompok yang sama semakin baik, kemudian semakin mendekati 1 nilai *silhouette* maka masing-masing individu memiliki tingkat kesamaan dan keterpisahan semakin tinggi dalam satu kelompok dan semakin maksimal nilai *dunn index* maka keterpisahan masing-masing kelompok makin jauh. Namun, jika setiap pengukuran memiliki hasil validasi yang berbeda, maka hasil pengukuran *silhouette* dan *dunn index* yang lebih diutamakan, karena keduanya merupakan hasil kombinasi dari pengukuran kepadatan dan keterpisahan antar kelompok yang mencirikan pengelompokan yang optimal.

2.4.2 Validasi Stabilitas

Validasi stabilitas mengevaluasi kekonsistenan hasil pengelompokan dengan membandingkannya pada perolehan kelompok dengan menghapus satu kolom pada suatu waktu. Menurut Bhuyan *et al* (2017) Validasi stabilitas meliputi:

a. *Average Proportion Of Non-overlap* (APN)

Average Proportion Of Non-overlap (APN) merupakan proporsi rata-rata data pada kelompok berbeda dari dataset dan *clustering* berdasarkan dataset dengan menghilangkan satu kolom (variabel) dimana APN mempunyai interval $[0,1]$ dengan nilai mendekati nol yang diperoleh dari hasil pengelompokan sangatlah konsisten. Ukuran *average proportion of nonoverlap* (APN) didefinisikan pada persamaan (2.13).

$$APN(K) = \frac{1}{MN} \sum_{i=1}^N \sum_{l=1}^M \left(1 - \frac{n(C^{i,l} \cap C^{i,0})}{n(C^{i,0})} \right) \quad (2.13)$$

dengan,

M : jumlah variabel

N : jumlah data

$C^{i,l}$: *cluster* dari data ke- i dengan menghapus kolom ke- l

$C^{i,0}$: *cluster* dari data ke- i tanpa menghapus kolom

b. *Figure Of Merit* (FOM)

Figure Of Merit (FOM) menunjukkan rata-rata varian intra-*cluster* dataset dengan menghapus satu kolom (variabel). Perkiraan kesalahan rata-rata dari FOM menggunakan prediksi berdasarkan rata-rata kelompok. Nilai FOM bervariasi interval $[0, \infty]$ dengan nilai yang lebih kecil yang sesuai dengan jumlah *Cluster* yang optimal. Ukuran *figure of merit* (FOM) didefinisikan pada persamaan (2.14).

$$FOM(l, K) = \sqrt{\frac{1}{N} \sum_{k=1}^K \sum_{C_k(l)} \text{dist}(x_{i,l}, \bar{x}_{C_k(l)})} \quad (2.14)$$

dengan,

K : jumlah *cluster*;

$x_{i,l}$: nilai data ke- i pada kolom ke- l pada *cluster* $C_k(l)$;

$\bar{x}_{C_k(l)}$: rata-rata *cluster* $C_k(l)$.

Tabel 2.2 Nilai validasi

Validasi	Komponen	Nilai	Kriteria
Internal	Konektivitas	$[0, +\infty]$	*
	Nilai rata-rata Silhouette	$[-1, +1]$	**
	Indeks Dunn	$[0, +\infty]$	**
Stabilitas	<i>average proportion non-overlap</i> (APN)	$[0, 1]$	*
	<i>figure of merit</i> (FOM)	$[0, +\infty]$	*

Keterangan : * : minimum

** : maksimum

2.5 Analisis Diskriminan

Analisis diskriminan adalah salah satu analisis multivariat yang termasuk *Dependence Method*, yaitu terdapat variabel independen dan dependen sehingga ada hasil variabel yang hasilnya tergantung dari data variabel independen. Prinsip dari analisis diskriminan adalah membuat model baru yang bisa secara *significant* menunjukkan perbedaan (diskriminasi) antar isi variabel dependen (Santoso, 2002:147).

Menurut (Supranto, 2004:77), tujuan analisis diskriminan adalah sebagai berikut:

1. Membuat suatu fungsi diskriminan atau kombinasi linier, dari prediktor atau variabel bebas yang bisa mendiskriminasi atau membedakan kategori variabel tak bebas, artinya mampu membedakan suatu objek (responden) masuk ke dalam kelompok/ kategori yang mana.
2. Menguji apakah ada perbedaan signifikan antara kategori/kelompok, dikaitkan dengan variabel bebas atau prediktor.
3. Menentukan prediktor/ variabel bebas yang mana yang memberikan sumbangan terbesar terhadap terjadinya perbedaan antar kelompok.
4. Mengelompokkan objek/ kasus/ responden ke dalam suatu kelompok/ kategori didasarkan pada nilai variabel bebas.
5. Mengevaluasi keakuratan klasifikasi (*the accuracy of classification*).
6. Melakukan interpretasi terhadap fungsi diskriminan.
7. Melakukan uji validasi fungsi diskriminan.

2.5.1 Asumsi Analisis Diskriminan

Asumsi penting yang harus dipenuhi agar model analisis diskriminan dapat digunakan adalah sebagai berikut:

1. *Multivariate Normality*, atau variabel independen seharusnya berdistribusi normal. Jika tidak berdistribusi normal, hal ini akan menyebabkan masalah pada ketepatan fungsi (model) diskriminan. Regresi logistik bisa dijadikan alternatif metode jika memang data tidak berdistribusi normal.
2. Matriks kovarians dari semua variabel independen seharusnya sama.

3. Tidak ada korelasi antar variabel independen. Jika dua variabel independen mempunyai korelasi yang kuat, maka dikatakan terjadi multikolinieritas.
4. Tidak adanya data yang sangat ekstrim (outlier) pada variabel independen.

2.5.2 Model Analisis Diskriminan

Model dasar analisis diskriminan hampir sama dengan regresi berganda. Perbedaan adalah pada variabel dependen yang digunakan untuk regresi berganda dilambangkan dengan Y maka dalam analisis diskriminan dilambangkan dengan D yang berupa data metrik (interval atau rasio) dan variabel independen (X_1, X_2, \dots, X_n) berupa data nonmetrik (nominal atau ordinal) (Johnson *et al*, 2007). Model analisis diskriminan adalah sebuah persamaan yang menunjukkan suatu kombinasi linear dari berbagai variabel independen yaitu :

$$D_i = b_0 + b_1X_{i1} + b_2X_{i2} + b_3X_{i3} + \dots + b_jX_{ij} \quad (2.15)$$

dengan,

D_i : nilai (skor) diskriminan dari responden.

i : 1, 2, ..., g.

X_{ij} : prediktor atau variabel independent (atribut) ke-j dari responden ke-i.

b_j : koefisien atau timbangan diskriminasi dari variabel atau atribut ke-j.

j : 1, 2, ..., p.

Yang diestimasi disini adalah koefisien dari b_k , sehingga 'D' setiap kelompok akan berbeda. Hal tersebut terjadi pada saat rasio jumlah kuadrat antar kelompok (*betweengroup sum of square*) terhadap jumlah kuadrat dalam kelompok (*within-group sum of square*) untuk skor diskriminan mencapai maksimum, sehingga dari nilai 'D' itulah keanggotaan sebuah data diprediksi.

Fungsi ini dibentuk dengan memaksimalkan jarak antar kelas, sehingga memiliki kemampuan untuk membedakan antar kelas. Setiap variabel dalam fungsi analisis diskriminan ditujukan untuk kelompok yang memiliki skor diskriminan tertinggi. Sehingga untuk memastikan solusi optimal, kesalahan klasifikasi, yang disebabkan oleh pemberian variabel pada kelompok yang salah juga dapat dipertimbangkan. Salah satu caranya ialah membangun matriks

klasifikasi untuk menentukan akurasi prediktif fungsi diskriminan dengan ukuran matriks $n \times n$.

Prinsip dari analisis diskriminan adalah membuat model baru yang bisa secara *significant* menunjukkan perbedaan (diskriminan) antar isi dari variabel *dependent*, seperti yang telah dijelaskan sebelumnya. Pada penelitian ini, pilihan yang tepat menggunakan analisis diskriminan sebagai salah satu metode validasi dari hasil solusi analisis *cluster* yang telah didapat. Dimana hasil solusi *cluster* yang diperoleh dari dua metode analisis *cluster* yaitu metode hirarki dan metode *Kmeans* digunakan sebagai variabel dependen untuk mendapatkan hasil uji analisis diskriminan untuk validasi *cluster* yang diinginkan melalui persamaan (2.15) diatas. Kelompok yang dihasilkan dari uji analisis diskriminan tersebut dapat terlihat bahwa kelompok tersebut baik atau tidak, karena uji analisis diskriminannya disini berfungsi sebagai alat yang membantu analisis *cluster* melakukan pengecekan terhadap kelompok-kelompok yang dihasilkan apakah sudah optimal (baik). Proses pengecekan tersebut dilakukan dengan melakukan uji validasi dari analisis diskriminan. Sehingga, diketahui nilai validitas dari masing-masing *cluster* yang terbentuk.

2.5.3 Uji Validasi dari Analisis Diskriminan

Setelah Fungsi diskriminan terbentuk, langkah selanjutnya adalah menilai validitas dari analisis diskriminan. Hasil klasifikasi dapat didasarkan pada analisis sampel ataupun validitas sampel yang dapat dilakukan dengan cara sebagai berikut:

a) Kesalahan Klasifikasi (*Misclassification*) dan APER (*Apparent Error Rate*)

Misclassification merupakan kesalahan dari pengklasifikasian suatu observasi baru ke dalam suatu kelompok (Sugiarto, 2000). Pada setiap prosedur klasifikasi dinilai dengan cara menghitung tingkat kesalahan atau probabilitas kesalahan klasifikasi. Salah satu metode yang dapat digunakan untuk menghitung kesalahan klasifikasi adalah APER (*Apparent Error Rate*) yaitu perhitungan hasil klasifikasi yang dilakukan dengan menggunakan matriks konfusi (*confusion matrix*). Untuk sebanyak n_1 observasi kelompok pertama (π_1), sebanyak n_2

observasi kelompok kedua (π_2) sampai sebanyak π_k observasi kelompok ke-k (π_k). Pada penelitian ini menggunakan analisis diskriminan lebih dari 2 kelompok yaitu 3 kelompok ($k=3$), persentase kesalahan klasifikasi dapat dihitung dari matriks yang menunjukkan nilai sebenarnya atau aktual dan nilai prediksi dari setiap kelompok.

Tabel 2.3 Klasifikasi Tiga kelompok

Aktual	Prediksi			Total
	π_1	π_2	π_3	
π_1	n_{11}	n_{12}	n_{13}	$n_{11} + n_{12} + n_{13}$
π_2	n_{21}	n_{22}	n_{23}	$n_{21} + n_{22} + n_{23}$
π_3	n_{31}	n_{32}	n_{33}	$n_{31} + n_{32} + n_{33}$
	N =			
Total	$n_{11} + n_{21} + n_{31}$	$n_{12} + n_{22} + n_{32}$	$n_{13} + n_{23} + n_{33}$	$n_{11} + n_{12} + n_{13} + n_{21} + n_{22} + n_{23} + n_{31} + n_{32} + n_{33}$

dengan,

n_{11} = jumlah objek dari π_1 yang tepat diklasifikasikan sebagai π_1

n_{12} = jumlah objek dari π_1 yang tepat diklasifikasikan sebagai π_2

n_{13} = jumlah objek dari π_1 yang tepat diklasifikasikan sebagai π_3

n_{21} = jumlah objek dari π_2 yang tepat diklasifikasikan sebagai π_1

n_{22} = jumlah objek dari π_2 yang tepat diklasifikasikan sebagai π_2

n_{23} = jumlah objek dari π_2 yang tepat diklasifikasikan sebagai π_3

n_{31} = jumlah objek dari π_3 yang tepat diklasifikasikan sebagai π_1

n_{32} = jumlah objek dari π_3 yang tepat diklasifikasikan sebagai π_2

n_{33} = jumlah objek dari π_3 yang tepat diklasifikasikan sebagai π_3

$$\text{Maka APER} = \frac{n_{12} + n_{21} + n_{32}}{N} \text{ dengan Akurasi} = (1 - \text{APER}) \times 100 \% \quad (2.16)$$

b) *Hit Ratio*

Setelah semua observasi diprediksi keanggotaannya, kita dapat menghitung *hit ratio*. *Hit Ratio* merupakan nilai yang dapat menjawab:

“Berapa persen objek yang dapat diklasifikasi secara tepat dari jumlah total objek?” *Hit Ratio* adalah persentase kasus atau responden yang kelompoknya

dapat diprediksi secara tepat. Fattah (2008) menyatakan bahwa jika model yang dihasilkan dapat memberi tingkat akurasi (*hit ratio*) lebih besar sama dengan 50% maka model tersebut dianggap baik, dan sebaliknya jika model yang dihasilkan memberi tingkat akurasi (*hit ratio*) kurang dari 50%, maka model tersebut dianggap kurang baik. Rumus *hit ratio* yang digunakan adalah:

$$\text{Hit Ratio} = \frac{n_{\text{benar}}}{N} \times 100\% \quad (2.17)$$

dengan,

$n_{\text{ benar}}$: jumlah sampel yang diklasifikasi secara tepat = $n_{11} + n_{22} + n_{33} + \dots + n_{kp}$

N : jumlah data (sampel) secara keseluruhan ,

$$N = n_{11} + n_{12} + \dots + n_{k1} + n_{21} + n_{22} + \dots + n_{k2} + n_{31} + n_{32} + \dots + n_{k3} + \dots + n_{kp}$$

c) Perhitungan Nilai C_{max}

Pengklasifikasian data sampel dari fungsi diskriminan dapat di uji keakuratannya dengan melakukan perhitungan untuk mengetahui *chane classification* (peluang klasifikasi) berdasarkan pada jumlah observasi, kemudian dibandingkan dengan nilai *hit ratio*, seperti yang telah dilakukan sebelum perhitungan nilai C_{max} ini. *Chane Classification* terdiri dari maximum *chane criterion* (C_{max}) dan Proportional *chane classification* (C_{pro}). Rumus C_{max} yang digunakan adalah sebagai berikut :

$$C_{\text{max}} = \frac{n_{\text{max}}}{N} \times 100\% \quad (2.18)$$

dengan,

$n_{\text{ max}}$: jumlah sampel yang terbesar pada salah satu kelompok / nilai elemen yang terbesar dalam diagonal matriks hasil analisis diskriminan.

N : jumlah data (sampel) secara keseluruhan.

d) Akurasi statistik

Klasifikasi yang dilakukan dengan fungsi diskriminan dapat diuji keakuratannya dengan menggunakan *Press's Q Statistik*. Menurut Hair, *et al* (2006) ukuran statistik uji *Press's Q* ini membandingkan jumlah kasus yang

diklasifikasikan secara tepat dengan ukuran sampel dan jumlah grup. Nilai yang diperoleh dari perhitungan kemudian dibandingkan dengan nilai kritis (*critical value*) yang diambil dari tabel Chi-Square berderajat bebas 1. Dengan kriteria keputusannya adalah jika *Press's Q* < X^2 , maka hasil analisis diskriminan dinyatakan tidak stabil, dan begitu pula sebaliknya. *Press's Q Statistik* dapat ditulis :

$$Press's Q = \frac{[N-(nK)]^2}{N(K-1)} \sim \chi^2_{(1)} \quad (2.19)$$

dengan,

N : ukuran sampel (banyaknya sampel)

n : jumlah sampel yang diklasifikasi secara tepat = $n_{11} + n_{22} + n_{33} \dots + n_{kp}$

K : jumlah kelompok

Hit ratio dan *Cmax*, perhitungan akurasi statistik (*Press's Q*) serta perhitungan dari Kesalahan Klasifikasi (*Misclassification*) dan *APER* (*Apparent Error Rate*) telah cukup untuk digunakan sebagai perhitungan uji keakuratan model diskriminan yang dihasilkan karena mampu memprediksi total jumlah kasus yang diprediksi secara tepat.

Discriminant Analysis (DA) umumnya digunakan untuk menyelesaikan masalah klasifikasi, dengan jenis metode yang digunakan dalam analisis diskriminan dan diperkenalkan oleh R.Fisher adalah *Linier Diskriminan Analysis* (LDA) (Hastie, *et al*: 2007). LDA digunakan sebagai cara menemukan proyeksi linear dari data yang akan memaksimalkan jarak data yang memiliki kesamaan, dan menjadi dasar klasifikasi dari data yang ada. Sebagai contoh pengaplikasian dari kombinasi linear ini adalah data makanan *junkfood*. Makanan yang termasuk dalam jenis ini adalah burger, pizza, dan kebab. Masing-masing makanan ini dibedakan dari kelasnya dan memiliki nilai jumlah N_i , sebanyak n dimensi. Dari data yang ada tersebut, maka sampel dari data tersebut dapat dinotasikan $\{X_1, X_2, \dots, X_{N_i}\}$, yang ditujukan kepada kelas yang ada. Setelah itu, data tersebut dikumpulkan menjadi matriks X yang tiap kolomnya mempresentasikan satu sampel data. Kemudian, dicari hasil transformasi dari X ke Y dengan

menggunakan proyeksi sampel dalam matriks X ke dalam hyperplane dengan dimensi sebesar K-1 (Anthasenna, 2014).

Fungsi diskriminan versi sampel biasanya digunakan dalam LDA, ditunjukkan dengan rumus :

$$f_i = \mu_i C^{-1} x_k^T - \frac{1}{2} \mu_i C^{-1} \mu_i^T + \ln(P_i) \quad (2.20)$$

Dengan,

- f_i : fungsi diskriminan kelas ke-i
- μ_i : rata-rata nilai setiap kelas dari masing-masing matriks Xi
- C^{-1} : invers dari grup matriks kovarian
- x_k^T : transpos dari matriks data uji
- P_i : peluang munculnya kelas ke-i
- μ_i^T : transpos rata-rata kelas ke-i

Tujuan LDA adalah untuk mengklasifikasi objek ke salah satu dari dua atau lebih kelompok berdasarkan serangkaian fitur yang menggambarkan kelas atau kelompok. Secara umum, dalam menetapkan objek untuk salah satu dari sejumlah kelompok yang telah ditentukan didasarkan dari pengamatan yang dilakukan pada objek tersebut.

2.6 Web Interaktif dengan R-Shiny

R-Shiny merupakan salah satu versi R berbasis web, Tirta (2014) menyatakan bahwa komponen program Shiny dibedakan menjadi dua kelompok yaitu :

1. *User Interface*, memiliki manfaat untuk :
 - a. Panel kontrol, adalah panel yang berfungsi untuk mengontrol *input* berupa data, variabel, model, tergantung kompleksitas modul. Tampilan pada panel kontrol dapat berupa *slider, radiobutton, check-box*, dan lain-lain;
 - b. Pemasukan permintaan nilai *input* (data dengan berbagai jenis variabel yang diperlukan, pemilihan model, jenis dan kriteria uji statistika);

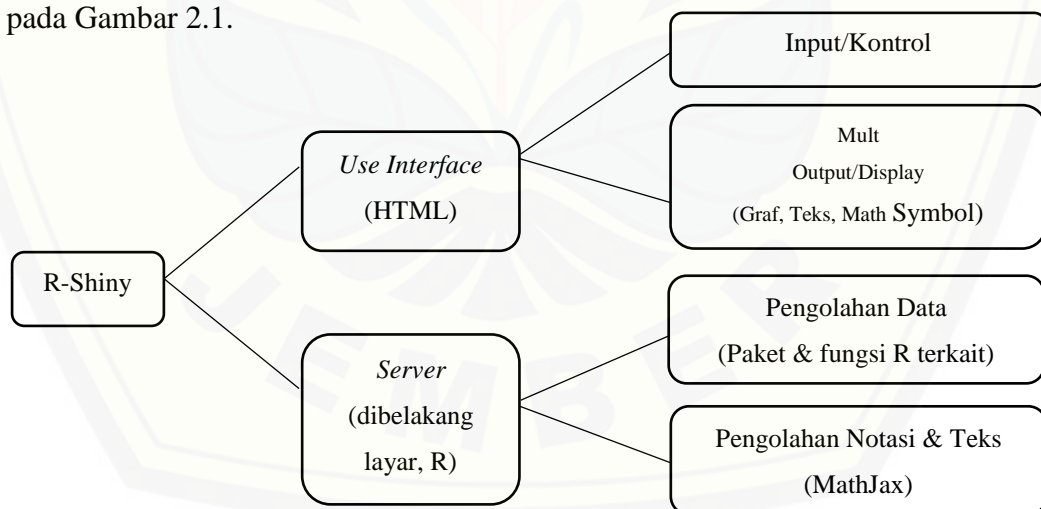
c. Penyajian *output* terkait hasil analisis/uji yang hasil *output*- nya dapat berupa :

- 1) Grafik (histogram, diagram pencar, dan lain-lain)
- 2) Bentuk angka/teks, berbentuk asli (verbatim) maupun berbentuk tabel;
- 3) Teks khusus dengan notasi matematika dengan format LaTeX.

R-shiny menyediakan berbagai format *output* seperti *plotOutput*, *textOutput*, *verbatimTextOutput*, *tabelOutput*, dan lain-lain untuk mengakomodasikan berbagai luaran diatas. Bagian *user interface* dapat disajikan pada file khusus *ui.r* dapat juga disajikan penuh melalui file HTML : *index.html*.

2. *Server*, bagian server ini merupakan otak dari program yang bertugas melakukan simulasi, berbagai analisis data sesuai pilihan penggunaan dan selanjutnya mengirim hasilnya ke bagian output. Server didukung oleh berbagai prosedur dan analisis data yang pada umumnya telah tersedia di berbagai paket program dan disimpan dalam file *server.r*.

Struktur umum komponen pemrograman dengan R-Shiny dapat dilihat pada Gambar 2.1.



Gambar 2.1 Struktur Komponen R-Shiny

Secara umum, penyajian file server ada dua macam yaitu :

- 1) Membuat file *server.r* dan *ui.r* secara terpisah lalu diletakkan pada satu direktori dengan nama sesuai topik, dan selanjutnya nama direktori dipanggil. Direktori : *shiny1*

File :

```
-server.r  
/subdir www/  
-index.html
```

- 2) Rumus-rumus matematika dalam tampilan R-Shiny, dapat dilakukan dengan memanfaatkan MathJax dengan sintaks mirip LaTeX.

```
withMathJax(  
  helpText("Teks biasa $$simbol matematika $$").
```



BAB 3. METODE PENELITIAN

3.1 Objek Penelitian

Data yang digunakan dalam penelitian ini adalah data simulasi yang ada di web Pusat Belajar dan Laboratorium Statistika Virtual yaitu data “iris” (Data dapat dilihat pada lampira C) dan data sekunder skripsi Muhda Shidqon (2012). Sumber data berasal dari hasil sensus Badan Pusat Statistik (BPS) Kabupaten Jember tahun 2010 (Data dapat dilihat pada lampiran A). Pengelompokan ini dilakukan berdasarkan faktor-faktor yang mempengaruhi kemiskinan pada suatu kecamatan.

Adapun pengolahan data menggunakan akses web online *Virtual Statistics Laboratory* Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember dengan alamat <http://statslab-rshiny.fmipa.unej.ac.id/RProg/Multiv/>. Dalam penelitian ini, peneliti memilih objek pengamatan Kecamatan di Kabupaten Jember yang terdiri dari 31 kecamatan, yaitu :

- | | | | |
|---------------|-----------------|-----------------|-------------|
| 1. Kencong | 11. Ajung | 21. Sukorambi | 31. Patrang |
| 2. Gumuk Mas | 12. Rambipuji | 22. Arjasa | |
| 3. Puger | 13. Balung | 23. Pakusari | |
| 4. Wuluhan | 14. Umbulsari | 24. Kalisat | |
| 5. Ambulu | 15. Semboro | 25. Ledokombo | |
| 6. Tempurejo | 16. Jombang | 26. Sumberjambe | |
| 7. Silo | 17. Sumberbaru | 27. Sukowono | |
| 8. Mayang | 18. Tanggul | 28. Jelbuk | |
| 9. Mumbulsari | 19. Bangsalsari | 29. Kaliwates | |
| 10. Jenggawah | 20. Panti | 30. Sumpalsari | |

Dalam penelitian ini, pengamatan pada data kemiskinan Kecamatan Kabupaten Jember didasarkan pada peubah status kemiskinan di 31 kecamatan yang tersedia di Badan Pusat Statistik (BPS) Kabupaten Jember, yaitu :

X_1 = Kepadatan Penduduk (Jiwa/km²).

X_2 = Banyaknya sekolah (SD/MI (Negeri & Swasta), SLTP/MTs (Negeri & Swasta), SMA/MA (Negeri & Swasta).

X_3 = Banyaknya fasilitas kesehatan (Rumah sakit, Puskesmas).

X_4 = Banyaknya Rumah Tangga Pengguna PLN.

X_5 = Banyaknya Rumah Tangga yang Mempunyai Telepon.

X_6 = Banyaknya Industri berbadan hukum (Industri Kecil, Industri Sedang, dan Industri Besar).

X_7 = Produksi Pertanian (Ton).

3.2 Langkah-langkah Penelitian

Adapun langkah-langkah dalam penelitian menggunakan Analisis diskriminan untuk validasi *cluster* yaitu :

- a. *Input* data pada web Pusat Belajar dan Laboratorium Statistika Virtual bagian Analisis Multivariat secara online.
- b. Menentukan prosedur pengelompokan

Pada langkah menentukan prosedur pengelompokan ini menggunakan metode hirarki dan metode *kmeans*.

1. Metode hirarki

Pengelompokan data dengan metode hirarki ini menggunakan tautan *single linkage*, *complete linkage* dan *average linkage*, sehingga masing-masing objek dalam data akan membentuk kelompok dan jumlah kelompok.

2. Metode *kmeans*

Pengelompokan data dengan metode *kmeans* ini menggunakan algoritma *Mac Queen*.

- c. Melakukan validasi kelompok

Pada langkah ini, akan dilakukan validasi terhadap hasil pengelompokan untuk mendapatkan hasil pengelompokan yang optimal.

1. Validasi internal

Pada langkah sebelumnya, dipilih metode hirarki dengan tautan *single linkage*, *complete linkage* dan *average linkage* yang dilakukan denganurut. Namun, untuk setiap tautan yang dipilih langsung dilanjutkan pada langkah melakukan validasi kelompok dengan validasi internal karena setiap tautan yang dipilih dan dilakukan validasi kemungkinan didapatkan nilai pada validasi internal yaitu nilai *connectivity*, *silhouette*, dan *dunn index* yang berbeda-beda atau mungkin ada yang sama. Begitu juga pada metode *Kmeans* menggunakan algoritma *Mac Queen*.

2. Validasi Stabilitas

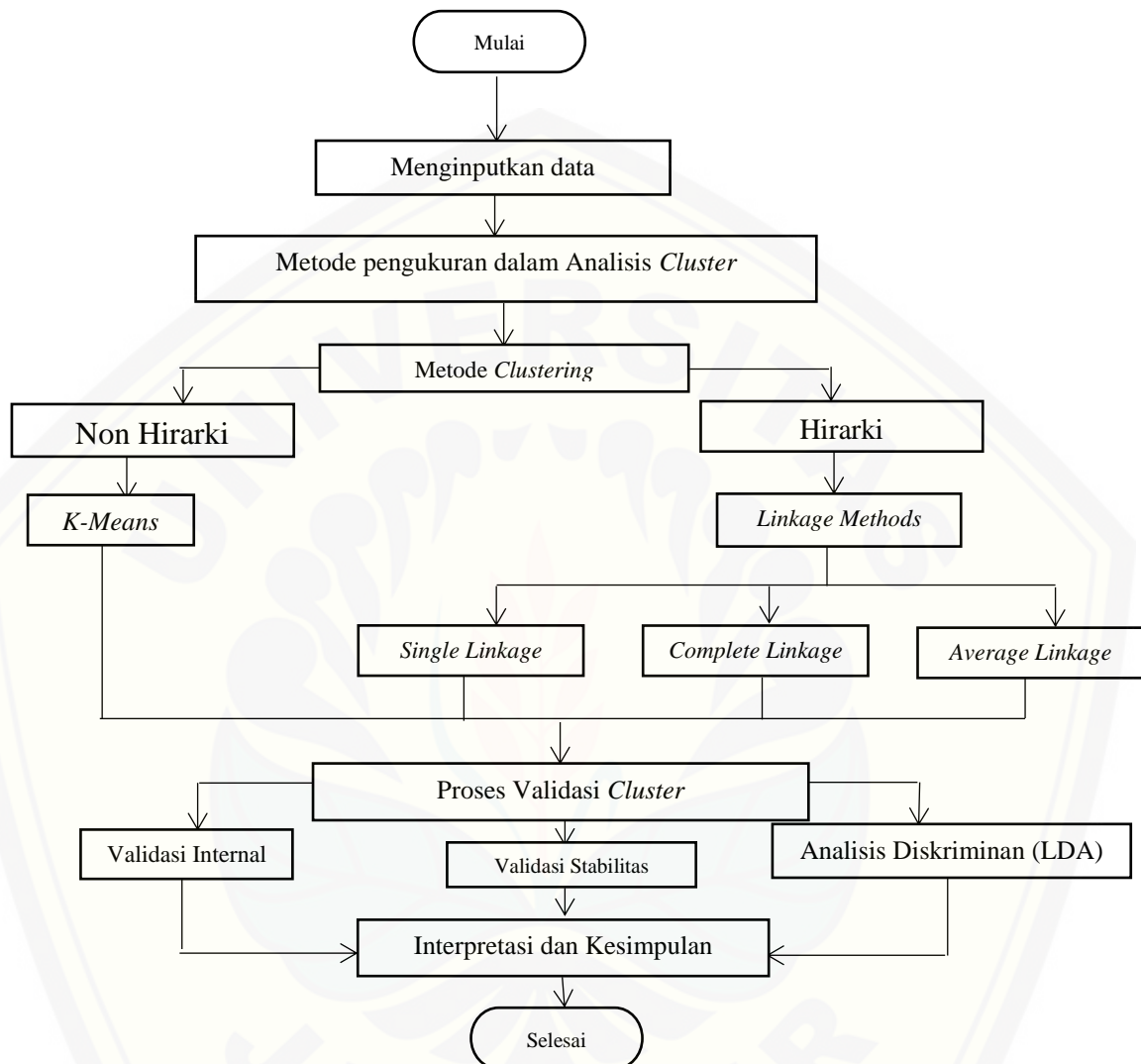
Pada langkah sebelumnya, dipilih metode hirarki dengan tautan *single linkage*, *complete linkage* dan *average linkage* yang dilakukan denganurut. Namun, untuk setiap tautan yang dipilih langsung dilanjutkan pada langkah melakukan validasi kelompok dengan validasi stabilitas karena setiap tautan yang dipilih dan dilakukan validasi kemungkinan didapatkan nilai pada validasi stabilitas yaitu nilai *APN*, dan *FOM* yang berbeda-beda atau mungkin ada yang sama. Begitu juga pada metode *Kmeans* menggunakan algoritma *Mac Queen*.

3. Uji analisis diskriminan

Pada tahapan validasi *cluster* dengan uji analisis diskriminan, ada beberapa tahapan yang dilakukan; dimulai dengan melakukan pengelompokan dengan metode hirarki tautan *single linkage*, *complete linkage* dan *average linkage* secara berurutan. Selanjutnya, dilanjutkan dengan pengelompokan metode *kmeans*. Hasil yang diperoleh dari pengelompokan metode hirarki dan *kmeans* digunakan sebagai variabel *dependent* dalam uji analisis diskriminan untuk validasi *cluster*. Hasil pengelompokan metode hirarki dan *kmeans* pada validasi *cluster* menggunakan uji analisis diskriminan masing-masing adalah tidak sama (berbeda).

- d. Interpretasi hasil dan mengambil kesimpulan.

Skema dari langkah-langkah penelitian dapat dilihat pada Gambar 3.1 dibawah ini.



Gambar 3.1 Diagram Alur Penelitian

Pada langkah menginputkan data, peneliti menggunakan data “iris” sebagai data simulasi untuk analisis data ini, dan selanjutnya hasil dari pengelompokan dengan metode analisis data yang tepat digunakan dalam data “iris” diaplikasikan pada pengelompokan data Kecamatan di Kabupaten Jember berdasarkan peubah status kemiskinannya.

3.3 Pengolahan Data dengan Menggunakan Program R

Pengolahan data dalam penelitian ini menggunakan program R untuk mendapatkan hasil validasi kelompok, jumlah kelompok dan bentuk kelompoknya. Program statistika R merupakan paket *opensource*, yang dapat diperoleh secara gratis dari situs <http://www.r-project.org/>. Dalam melakukan proses validasi kelompok ini menggunakan peket-paket yang telah tersedia dalam program R, yaitu paket LDA dan paket `clvalid`.

Linier Discriminant Analysis (LDA)

```
lda(x, ...)

## S3 method for class 'formula'
lda(formula, data, ..., subset, na.action)

## Default S3 method:
lda(x, grouping, prior = proportions, tol = 1.0e-4,
     method, CV = FALSE, nu, ...)

## S3 method for class 'data.frame'
lda(x, ...)

## S3 method for class 'matrix'
lda(x, grouping, ..., subset, na.action)
```

Keterangan :

- Formula** : rumus dari grup $\sim x_1 + x_2 + \dots$. Yaitu, respons adalah faktor pengelompokan dan sisi kanan menentukan (non-faktor) discriminator.
- data** : data dari variabel mana yang ditentukan dalam rumus secara khusus akan diambil.
- x** : (diperlukan jika tidak ada rumus yang diberikan sebagai argumen utama) matriks atau data atau Matriks yang berisi variabel penjelas.
- Grouping** : (diperlukan jika tidak ada argumen prinsip dasar yang diberikan.) faktor yang menentukan kelas untuk setiap observasi.
- prior** : Probabilitas prior dari keanggotaan kelas. Jika tidak ditentukan, proporsi kelas untuk set pelatihan digunakan. Jika ada, probabilitas harus ditentukan dalam urutan tingkat faktor.

<code>tol</code>	: Toleransi untuk memutuskan apakah matriks itu tunggal ia akan menolak variabel dan kombinasi linier dari variabel unit-variens yang variannya kurang dari tol^2 .
<code>Subset</code>	: Vektor indeks yang menentukan kasus yang akan digunakan dalam sampel pelatihan.
<code>na.action</code>	: Suatu fungsi untuk menentukan tindakan yang akan diambil jika NAs ditemukan. Tindakan standar untuk prosedur gagal. Alternatifnya adalah <code>na.omit</code> , yang mengarah pada penolakan kasus dengan nilai yang hilang pada variabel yang diperlukan.
<code>method</code>	: "momen" untuk penaksir standar dari mean dan varians, "mle" untuk MLE, "mve" untuk menggunakan <code>cov.mve</code> , atau "t" untuk estimasi kuat berdasarkan distribusi t.
<code>CV</code>	: Jika benar, kembalikan hasil (kelas dan probabilitas posterior) untuk validasi silang keluar-satu. Perhatikan bahwa jika perkiraan sebelumnya, proporsi dalam seluruh dataset digunakan.
<code>nu</code>	: derajat kebebasan untuk metode = "t"
<code>...</code>	: argumen dilewatkan ke atau dari metode lain.

Nilai (value) dari LDA, jika `CV = TRUE` nilai kembalian adalah daftar dengan kelas komponen, klasifikasi MAP (faktor), dan posterior, probabilitas posterior untuk kelas. Kalau tidak, itu adalah objek kelas "lda" yang berisi komponen-komponen berikut:

<code>Prior</code>	:probabilitas sebelumnya digunakan.
<code>Means</code>	:mean kelompok
<code>Scaling</code>	:matriks yang mengubah pengamatan menjadi fungsi diskriminan, dinormalisasi sehingga dalam kelompok matriks kovariansi adalah bulat.
<code>Svd</code>	:nilai-nilai singular, yang memberikan rasio antara standar deviasi antara dan dalam kelompok pada variabel diskriminan linier. Kotak mereka adalah F-statistik kanonis.

`N` : Jumlah observasi yang digunakan.

`Call` : Panggilan fungsi (yang ditetapkan).

Validasi *cluster* telah tersedia dalam paket `clValid`. Berikut fungsi dalam paket `clValid` yang digunakan dalam mengestimasi jumlah *cluster*.

```
clValid(obj, nClust, clMethods = "...", validation =
"...")
```

Berikut fungsi dalam paket `clvalid` yang digunakan untuk validasi internal, dan validasi stabilitas :

- Validasi internal :

- `connectivity(distance = NULL, clusters, Data = NULL, neighbSize = 10, method = "euclidean")`

- `dunn(distance = NULL, clusters, Data = NULL, method = "euclidean")`

- Validasi stabilitas :

```
stability(mat, Dist=NULL, del, cluster, clusterDel,
method="euclidean")
```

keterangan:

`Obj` : data yang telah diproses sebelumnya.

`nClust` : jumlah kelompok yang akan dievaluasi.

`clMethods` : metode pengelompokan yang akan digunakan.

`Validation` : jenis-jenis validasi yang akan digunakan. Pilihan yang tersedia adalah internal dan stabilitas

`Maxitems` : jumlah maksimum objek (baris dalam matriks) yang dapat dikelompokkan.

`Metric` : pemilihan pengukuran jarak.

`Method` : perbaikan matriks jarak.

`neighbSize` : jumlah ukuran ketetanggan, digunakan pada pengukuran *connectivity*.

`cluster` : sebuah vektor integer yang menunjukkan partisi cluster berdasarkan semua data.

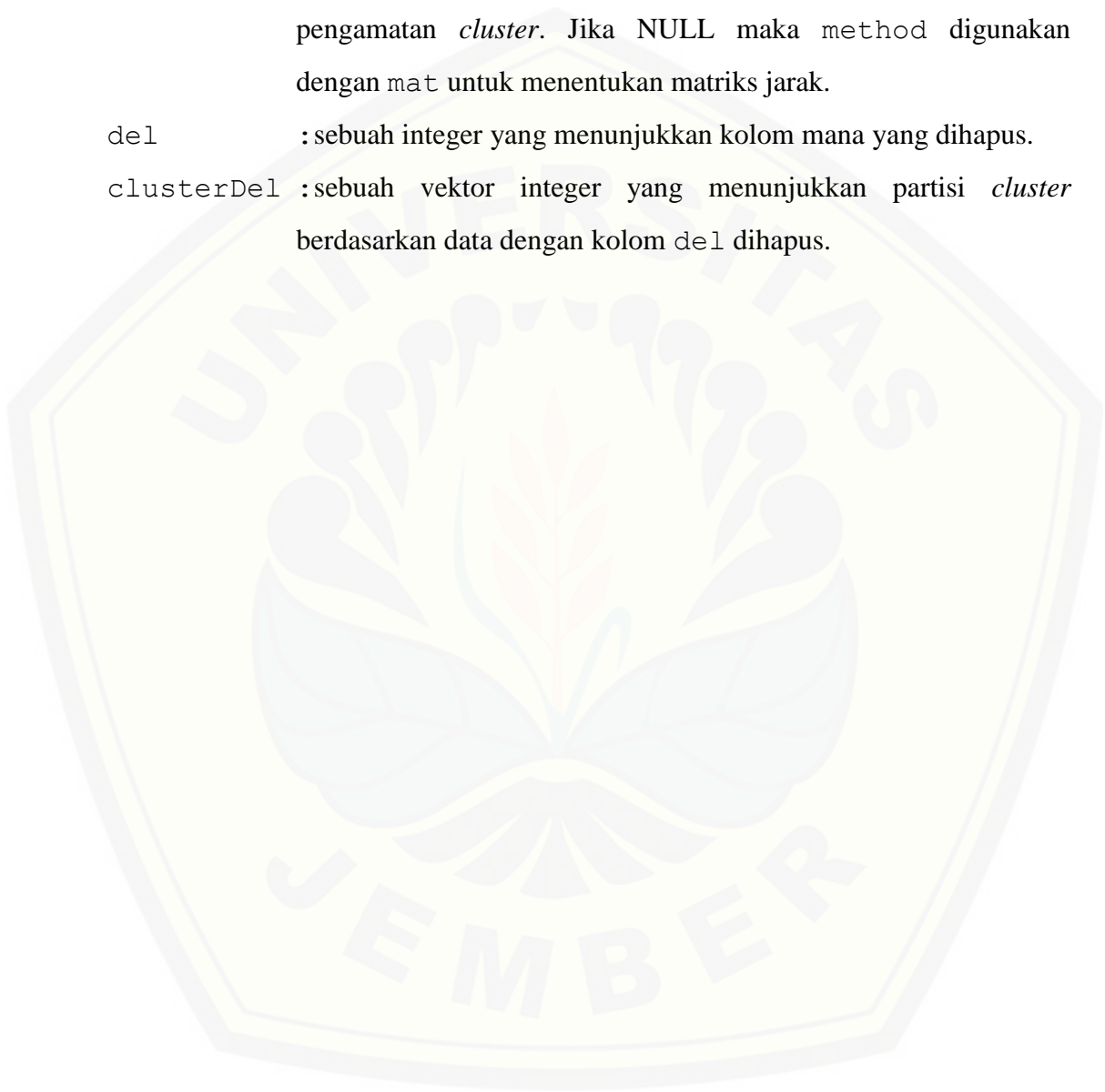
`Data` : matriks data dari pengamatan berkelompok diperlukan jika jarak adalah NULL.

`mat` : matriks data dari pengamatan berkelompok.

`Dist` : matriks jarak (sebagai matriks atau objek `Dist`) dari pengamatan *cluster*. Jika NULL maka method digunakan dengan `mat` untuk menentukan matriks jarak.

`del` : sebuah integer yang menunjukkan kolom mana yang dihapus.

`clusterDel` : sebuah vektor integer yang menunjukkan partisi *cluster* berdasarkan data dengan kolom `del` dihapus.



BAB 5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan Analisis Diskriminan untuk Validasi *Cluster* pada bab 4 berbasis online menggunakan *Virtual Statistics Laboratory* web dan aplikasinya pada data “iris” dan data Kecamatan Kabupaten Jember berdasarkan peubah status kemiskinan tahun 2010 dapat diambil kesimpulan sebagai berikut:

1. Prosedur penentuan metode analisis *cluster* yang terbaik adalah dengan menggunakan uji analisis diskriminan untuk validasi *cluster* yang berdasarkan pada nilai *hit ratio*, *cmax*, dan *Press's Q*. Ketiga nilai tersebut sudah cukup digunakan sebagai perhitungan uji keakuratan model diskriminan yang dihasilkan, karena mampu memprediksi sejumlah kasus yang diprediksi secara tepat dari total jumlah kasus. Pautan pada metode hirarki dan algoritma pada metode *kmeans* menghasilkan nilai yang berbeda-beda disetiap validasi yang digunakan, sehingga tidak menutup kemungkinan tidak hanya 1 pautan/algoritma yang terbaik pengelompokannya tetapi ada beberapa yang ternyata baik hasil pengelompokannya.
2. Hasil validasi *cluster* dengan metode hirarki dengan tautannya *single linkage* merupakan pengelompokan dengan hasil yang optimal untuk metode validasi internal, validasi stabilitas dan validasi *cluster* menggunakan uji analisis diskriminan. Pada validasi internal dan stabilitas yang telah digunakan sebelumnya mendapatkan hasil jumlah kelompok optimalnya sebanyak 2 kelompok, sehingga pada validasi *cluster* dengan menggunakan uji analisis diskriminan pengelompokan yang dilakukan harus membentuk kelompok optimal sebanyak 3 kelompok sesuai dengan jumlah kelompok pada variabel *species*. Nilai peluang awal (*prior probabilities*) pengklasifikasian objek ke dalam 3 kelompok tautan *single linkage* adalah sama yaitu sebesar 0,33; 0,65; dan 0,013. Berdasarkan hal tersebut, pengelompokan metode hirarki dengan

menggunakan tautan *single linkage* adalah pengelompokan yang optimal dan paling baik untuk seluruh metode validasi yang digunakan pada data “iris” ini, serta untuk semua variabel yang digunakan dalam pengelompokan ini tidak ada yang hasil pengelompokan keanggotaannya sesuai atau sama dengan pengelompokan keanggotaan yang asli pada variabel *species*, tetapi hanya pada metode hirarki hasil pengelompokan keanggotaannya hampir sesuai dengan pengelompokan keanggotaan daripada variabel *species* yang asli untuk tautan *average linkage*.

3. Hasil validasi *cluster* pada pengelompokan data kemiskinan Kecamatan Kabupaten Jember metode hirarki dengan tautannya *single linkage* dan *complete linkage* digunakan pada data ini karena tautan *single linkage* dapat kita gunakan pada data kemiskinan Kecamatan Kabupaten Jember sebagai tautan yang digunakan pada data “iris” sebelumnya yang dijadikan acuan dan Tautan *complete linkage* dapat kita gunakan sebagai pautan yang memenuhi kriteria pengelompokan yang optimal untuk semua metode validasi yang digunakan pada pembuatan model analisis diskriminannya. Berdasarkan hal tersebut, pengelompokan metode hirarki dengan menggunakan tautan *single linkage* dan *complete linkage* merupakan pengelompokan yang optimal dan paling baik untuk seluruh metode validasi yang digunakan pada data pengelompokan Kecamatan di Kabupaten Jember berdasarkan peubah status kemiskinannya. Nilai peluang awal, model persamaan analisis diskriminan dan kebaikan model serta ketepatan klasifikasi masing-masing kelompok pada tautan *single linkage* dan *complete linkage* dapat dilihat pada subbab 4.2 yang telah dijelaskan sebelumnya.
4. Uji analisis diskriminan sebagai salah satu metode untuk memvalidasi hasil solusi dari analisis *cluster* adalah tepat digunakan untuk analisis data pengelompokan secara umum. Analisis diskriminan disini berfungsi sebagai salah satu alat yang membantu analisis *cluster* melakukan pengecekan terhadap kelompok-kelompok yang dihasilkan apakah sudah optimal (baik), daripada metode validasi internal dan metode validasi stabilitas. Validasi *cluster* menggunakan uji analisis dikriminan lebih mudah dilakukan dalam

proses perhitungannya dan lebih spesifik pada pengklasifikasian anggota dalam setiap kelompoknya serta dapat juga melihat kesalahan klasifikasi (*missclassification*) dari nilai APER dan persentase nilai akurasi secara langsung tanpa harus mempertimbangkan nilai-nilai lainnya seperti pada validasi internal dan validasi stabilitas sehingga diketahui kelompok yang dihasilkan tersebut adalah tepat pengklasifikasian keanggotaan pengelompokannya.

5.2 Saran

Penelitian selanjutnya dapat mengembangkan metode algoritma pada metode *Kmeans* lainnya, seperti *Hartigan-Wong* dan *Llyod* ataupun mencoba validasi *cluster* yang lainnya, yaitu validasi biologis pada data yang juga bersifat biologis untuk melihat apakah tautan *single linkage* tetap sebagai tautan yang paling baik untuk data “iris” dengan data lainnya dan apakah analisis diskriminan juga cocok untuk memvalidasi hasil *clustering* pada data yang bersifat biologis.

DAFTAR PUSTAKA

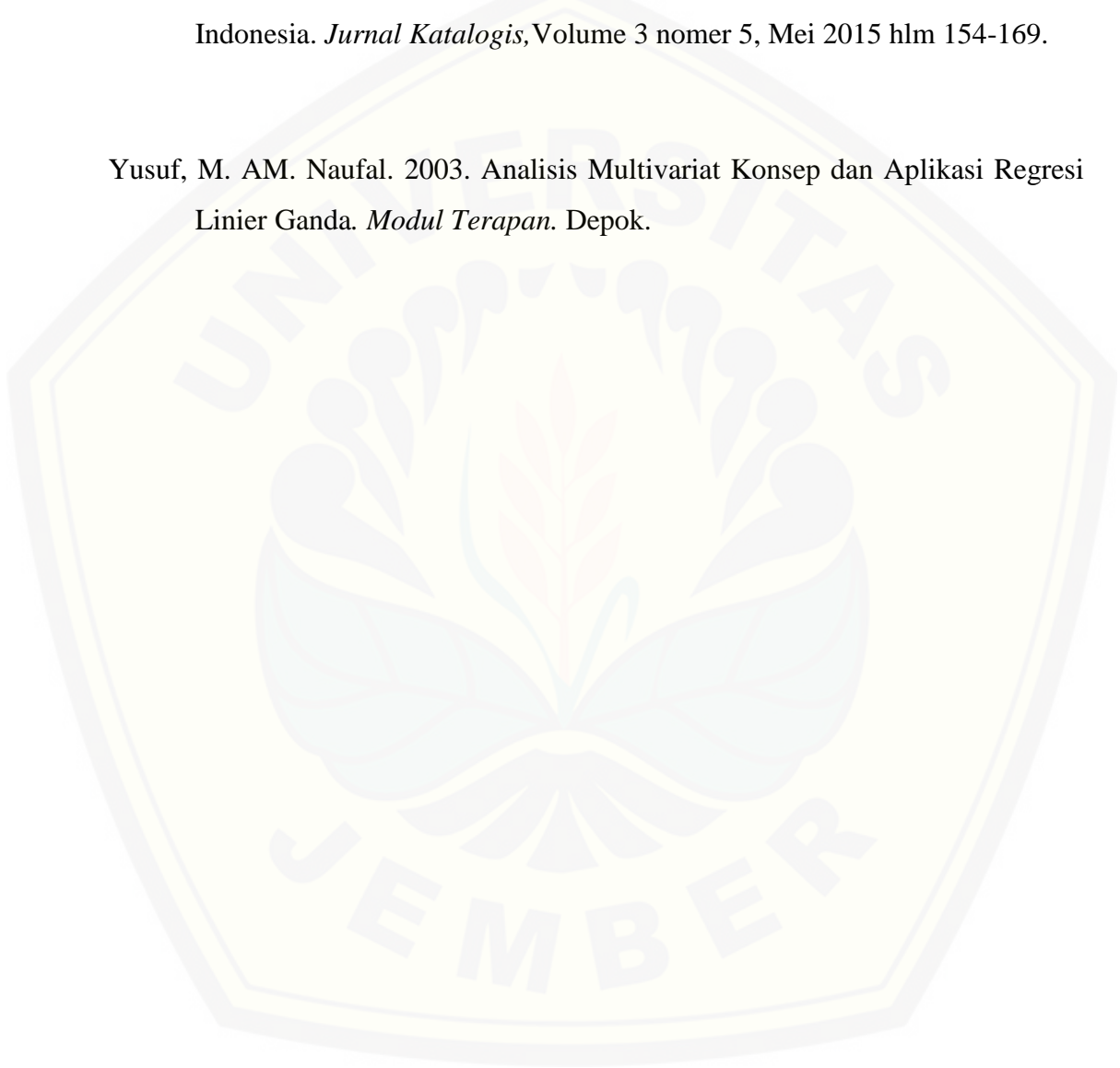
- Anthasenna, I.D.G., 2014. Sistem Identifikasi Genre Musik dengan Metode Ekstraksi Fitur FFT dan Metode Klasifikasi Linear Discriminant Analysis Beserta Similarity Measure. *Skripsi*. Malang: Universitas Brawijaya.
- Avan. 2014. Penggunaan Analisis Diskriminan Linier Robust Menggunakan Metode Fast MCD Pada Data yang Mengandung Pencilan. *Skripsi*. Jember: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.
- Brock, G., Vasyl, P., Susmita, D., dan Somnath, D. 2008. cValid: *An R Package for Cluster Validation*. *Journal of Statistical Software*, **25**(4): 1-22.
- Dillon, W.R. dan Goldstein, M. 1984. *Multivariate Analysis Methods and Application*. John Wiley&Sons.Inc. New York.
- Ernayanti, dan Ita, N. 1996. *Budaya Kemiskinan di Desa Tertinggal di Yogyakarta* (Kasus desa Karang Tengah, Kecamatan Imogiri Kabupaten Bantul). Direktorat Sejarah dan Nilai Tradisional Bagian Proyek Pengkajian dan Pembinaan Kebudayaan Masa Kini, Jakarta.
- Everitt, B.S., dan Dun G. 1998. *Applied Multivariate Data Analysis*. London: Edward Arnold.
- Fadhli. (2011). Analisis *Cluster* Untuk Pemetaan Mutu Pendidikan di Aceh. *Tesis*. PPs-UGM.
- Fisher , R. A. 1938. *The Statistical Utilization of Multiple Measurements*. *Annals of Eugenics*,[s.1.], 8, 376-386.

- Gudono. (2011). *Analisis Data Multivariate Edisi Pertama*. Yogyakarta: BPFPE.
- Hair, J.F. (Jr), Anderson, R.E., Tathan, R.L., dan Nack, W.C. (1998). *Multivariate Data Analysis* (5th Ed). New Jersey, USA : Prentice-Hall International, Inc.
- Hair, J. F. 2006. *MultiVariate Data Analysis*. Edisi 5. Jakarta: Gramedia Pustaka Utama.
- Halkidi, M., Batistakis, Y., & Vazirgiannis, M. 2001. *Cluster Validity Method : Part I*. <http://u.cs.biu.ac.il/~louzouy/courses/seminar/asses2.pdf>
[5 November 2011]
- Hastie, T., Guo, Y., dan Tibshirani, R. 2006. *Regularized Linear discriminant Analysis and its Application in Microarrays*. *Biostatistics*, Volume 8, Issue 1,1 January 2007, Pages 86-100, <http://doi.org/10.1093/biostatistics/kxj035>
.[07 April 2006].
- Johnson, R. A., dan Wichern, D. W. 2007. *Applied Multivariate Statistical Analysis*. Sixth Ed. , Prentice Hall. New Jersey.
- K. Das, A dan K. Bhuyan, P. 2017. *Self-Organizing Tree Algorithm (SOTA) Clustering for Defining Level of Service (LOS) Criteria of Urban Streets*. *Research Article*. pp. <http://doi.org/10.3311/PPtr.9911> .
- N.B. Carvalho, R. Minim, V. P., Nascimento, M., T. R. Vidigal, M. C., M. Ferreira, M. A., A. Goncalves, A. C., dan A. Minim, L. 2015. *A discriminant function for validation of the cluster analysis and behavioral prediction of the coffee market*. *Jurnal Food Research International* 77 (2015) 400-407

- N. Halim, N., dan Widodo, E. 2017. *Clustering Dampak Gempa Bumi di Indonesia Menggunakan Kohonen Self Organizing Maps. Prosiding SI MaNIs (Seminar Nasional Integrasi atematika dan Nilai Islami, Vol.1, No.1, Juli 2017, Hal. 188-194.*
- Nurmaleni. 2015. Perbandingan Metode Klasifikasi Antara Analisis Diskriminan Verteks dan Diskriminan Fisher. *Tesis*. Bogor: Sekolah Pascasarjana Intitut Pertanian Bogor.
- Rachmatin, D dan Sawitri, K. 2014. *Perbandigan Antara Metode Agglomeratif, Metode Divisif, dan Metode K-Means dalam Analisis Klaster*. Seminar Nasional Matematika UNPAR.
- RStudio and Inc. 2013. *shiny: Web Application Frame work for R. Rpackageversion 0.8.1*. <http://CRAN.R-project.org/package=shiny>
- RStudio and Inc.2014. Shiny Widget Gallery. <http://shiny.rstudio.com/gallery/widgetgallery.html> [September 2014]
- Santosa, S. 2002. *Buku Latihan SPSS Statistik Multivariat*. Jakarta: PT Elex Media Komptindo.
- Sarwono, J. 2006. *Metodelogi Penelitian Kuantitatif dan Kualitatif*. Jogjakarta : Graha Ilmu.
- Sharma, S. (1996). *Applied Multivariate Technique*. Canada: John Wiley & Sons.
- Shidqon, M. 2012. Validasi Metode Kelompok Pada Pengelompokan Kecamatan Di Kabupaten Jember Berdasarkan Peubah Status Kemiskinan. *Skripsi*. Jember: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

- Simamora, B. 2005. *Analisis Multivariat Pemasaran*. Jakarta: PT Gramedia Pustaka Umum.
- Sitepu, R., Irmeilyana, dan Gultom, B. 2011. Analisis *Cluster* terhadap Tingkat Pencemaran Udara Pada Sektor Industri di Sumatera Selatan. *Jurnal Penelitian Sains*. 4(34).
- Sugiarto, D.S. 2000. *Metode Statistika*. Jakarta : Gramedia Pustaka Utama.
- Supranto, J. 2004. *Analisis Multivariat Arti dan Interpretasi*. PT. Rineka Cipta Jakarta.
- Tazkiyah, O. 2016. *Self Organizing Map (SOM) clustering* berbasis web interaktif dengan r-shiny. *Skripsi*. Jember: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.
- Tirta IM. 2014a “Aktivitas Laboratorium Statistika Virtual Menggunakan Rshiny. *Prosiding Seminar Nasional Matematika Universitas Udayana* 235-244
- Tirta, IM. 2014b. Pengembangan E-module Statistika yang Terintegrasi dan Dinamik dengan R-Shiny dan MathJax. *Prosiding Seminar Nasional Matematika Universitas Jember*.
- Tirta, I.M. 2014. Pengembangan E-Modul Statistik Terintegrasi dan Dinamik dengan R-shiny dan mathJax. *Prosiding Seminar Nasional Matematika, Universitas Jember*. 223-232.

- Tirta, IM. 2015. Pengembangan Analisis Respon Item Interaktif Online Menggunakan R untuk Respon Dikotomus dengan Model Logistik (1-PI, 2-PI 3-PI). *Prosiding Seminar Nasional Pendidikan* 420-427.
- Yuliana. 2015. Analisis Kebangkrutan pada Perusahaan Batubara Di Bursa Efek Indonesia. *Jurnal Katalogis*, Volume 3 nomer 5, Mei 2015 hlm 154-169.
- Yusuf, M. AM. Naufal. 2003. Analisis Multivariat Konsep dan Aplikasi Regresi Linier Ganda. *Modul Terapan*. Depok.



LAMPIRAN

A. DATA KEMISKINAN KECAMATAN DI KABUPATEN JEMBER

Kecamatan	X1	X2	X3	X4	X5	X6	X7
Kencong	1112.32	88	4	152	12702	25	11903
Gumuk Mas	950.60	110	4	76	10512	11	36135
Puger	1557.65	96	4	112	9836	27	50747
Wuluhan	1285.40	93	4	231	10123	52	30377
Ambulu	955.47	130	5	487	13271	27	38941
Tempurejo	93.96	90	5	101	8438	47	58797
Silo	338.63	69	4	62	7271	26	21152
Mayang	863.67	82	3	72	8934	33	27584
Mumbulsari	667.74	59	2	59	9867	28	6578
Jenggawah	1229.68	62	3	312	13288	28	156
Ajung	117.24	46	3	154	14211	32	37295
Rambipuji	1394.00	61	3	631	15456	26	1788
Balung	1530.21	109	4	392	14999	19	31862
Umbulsari	1075.59	87	4	156	12234	53	4512
Semboro	1161.20	50	3	129	13344	51	26513
Jombang	988.93	34	2	43	11453	36	597
Sumberbaru	609.21	87	3	57	10356	18	43029
Tanggul	711.83	95	4	214	8621	51	50163
Bangsalsari	647.48	109	3	71	9724	21	39377
Panti	614.20	76	2	64	8611	32	38464
Sukorambi	900.49	45	2	77	10465	19	17237
Arjasa	895.03	29	2	89	10189	15	1771

Digital Repository Universitas Jember

Pakusari	1311.00	22	2	594	11765	8	18790
Kalisat	1342.50	66	2	223	12578	22	23901
Ledokombo	365.28	49	2	46	7667	46	38023
Sumberjambe	869.46	40	2	156	7519	34	17344
Sukowono	1260.00	39	2	289	14018	17	6735
Jelbuk	726.48	38	4	450	10433	14	13560
Kaliwates	3581.42	63	6	1874	26724	210	4304
Sumbersari	2994.45	77	4	1557	24787	213	16711
Patrang	2477.64	105	2	1027	22558	80	14638



B. Hasil Validasi Solusi *Cluster* Data Kemiskinan Kecamatan Kabupaten Jember

B.1 Tampilan Hasil

X1	X2	X3	X4
Min. : 93.96	Min. : 22.00	Min. : 2.000	Min. : 43.0
1st Qu.: 689.78	1st Qu.: 47.50	1st Qu.: 2.000	1st Qu.: 74.0
Median : 955.47	Median : 69.00	Median : 3.000	Median : 154.0
Mean : 1117.06	Mean : 71.16	Mean : 3.194	Mean : 321.2
3rd Qu.: 1298.20	3rd Qu.: 91.50	3rd Qu.: 4.000	3rd Qu.: 352.0
Max. : 3581.42	Max. : 130.00	Max. : 6.000	Max. : 1874.0

X5	X6	X7	Kecamatan
Min. : 7271	Min. : 8.00	Min. : 156	Ajung : 1
1st Qu.: 9780	1st Qu.: 20.00	1st Qu.: 9319	Ambulu : 1
Median : 10512	Median : 28.00	Median : 21152	Arjasa : 1
Mean : 12321	Mean : 42.61	Mean : 23516	Balung : 1
3rd Qu.: 13316	3rd Qu.: 46.50	3rd Qu.: 37659	Bangsalsari: 1
Max. : 26724	Max. : 213.00	Max. : 58797	Guduk Mas : 1
			(Other) : 25

Metode hirarki

Validasi internal

```
Call: hclust(d = dist(datasetInput0()[, input$xc], method = input$distc)
, method = input$method)
```

Cluster method : single

Distance : euclidean

Number of objects: 31

Score	Method	Clusters
Connectivity 9.6734127	hierarchical	3
Dunn 0.1222677	hierarchical	3
Silhouette 0.8547908	hierarchical	3

Call:

```
hclust(d = dist(datasetInput0()[, input$xc], method = input$distc),  
method = input$method)
```

Cluster method : complete

Distance : euclidean

Number of objects: 31

	Score	Method	Clusters
Connectivity	9.6734127	hierarchical	3
Dunn	0.1222677	hierarchical	3
Silhouette	0.8547908	hierarchical	3

Call:

```
hclust(d = dist(datasetInput0()[, input$xc], method = input$distc),  
method = input$method)
```

Cluster method : average

Distance : euclidean

Number of objects: 31

	Score	Method	Clusters
Connectivity	11.38571429	hierarchical	3
Dunn	0.03176821	hierarchical	3
Silhouette	0.81150690	hierarchical	3

Validasi Stabilitas

Call:

```
hclust(d = dist(datasetInput0()[, input$xc], method = input$distc),  
method = input$method)
```

Cluster method : single

Distance : euclidean

Number of objects: 31

	Score	Method	Clusters
APN	7.572965e-02	hierarchical	3
FOM	2.968325e+03	hierarchical	3

```
Call: hclust(d = dist(datasetInput0()[, input$xc], method = input$distc)  
, method = input$method)
```

Cluster method : complete

```

Distance      : euclidean
Number of objects: 31
      Score      Method Clusters
APN 2.285714e-01 hierarchical      3
FOM 2.852140e+03 hierarchical      3
Call:
hclust(d = dist(datasetInput0()[, input$xc], method = input$distc),
method = input$method)
Cluster method : average
Distance      : euclidean
Number of objects: 31
      Score      Method Clusters
APN 4.321403e-02 hierarchical      3
FOM 2.923033e+03 hierarchical      3

```

Metode Kmeans

Validasi internal

```

Jenis      Algoritma  NKlas
[1,] "K-Means" "MacQueen" "3"   "9" "10" "12"
K-means clustering with 3 clusters of sizes 10, 12, 9

Cluster means:
      X1      X2      X3      X4      X5      X6
X7
1  662.292 88.80000 3.500000 138.2000 10124.70 31.20000
43097.10
2 1533.607 61.83333 3.166667 553.2500 15309.08 62.08333
6937.75
3 1066.951 64.00000 2.888889 215.1111 10777.56 29.33333
23862.22

Clustering vector:
 [1] 2 1 1 3 1 1 3 3 2 2 1 2 3 2 3 2 1 1 1 1 3 2 3 3 1 3 2 2 2
 2 2

Within cluster sum of squares by cluster:
 [1] 557734516 779025606 303537017
 (between_SS / total_SS = 81.7 %)

Available components:

 [1] "cluster"      "centers"      "totss"        "withinss"
 "tot.withinss"

```

```
[6] "betweenss"      "size"          "iter"          "ifault"
      Score Method Clusters
Connectivity 14.7892857 kmeans      3
Dunn          0.1581622 kmeans      3
Silhouette   0.4633359 kmeans      3
```

Validasi Stabilitas

```
Jenis      Algoritma      NKlas
[1,] "K-Means" "Mac Queen" "3"  "12" "9" "10"
K-means clustering with 3 clusters of sizes 4, 12, 15
```

Cluster means:

```
      X1      X2      X3      X4      X5      X6
X7
1  743.1625 92.00000 4.000000 121.0000  9312.75 35.75000
50684.000
2  847.6567 82.41667 3.250000 167.2500 10937.08 31.00000
32468.667
3 1432.2820 56.60000 2.933333 497.7333 14230.53 53.73333
9108.267
```

Clustering vector:

```
[1] 3 2 1 2 2 1 2 2 3 3 2 3 2 3 2 3 1 1 2 2 3 3 3 2 2 3 3 3 3
3 3
```

Within cluster sum of squares by cluster:

```
[1] 128418882 538091625 1143544899
(between_SS / total_SS = 79.8 %)
```

Available components:

```
[1] "cluster"      "centers"      "totss"      "withinss"
"tot.withinss"
[6] "betweenss"    "size"         "iter"         "ifault"
```

Score Method Clusters

```
APN 1.313364e-01 kmeans      3
FOM 3.004107e+03 kmeans      3
```

LDA

Metode Hirarki, pautan *single linkage*

Kecamatan	X1	X2	X3
Ajung	: 1 Min. : 93.96	Min. : 22.00	Min. : 2.000
Ambulu	: 1 1st Qu.: 689.78	1st Qu.: 47.50	1st Qu.: 2.000
Arjasa	: 1 Median : 955.47	Median : 69.00	Median : 3.000
Balung	: 1 Mean : 1117.06	Mean : 71.16	Mean : 3.194
Bangsalsari	: 1 3rd Qu.: 1298.20	3rd Qu.: 91.50	3rd Qu.: 4.000

```

Gumuk Mas : 1 Max. :3581.42 Max. :130.00 Max. :6.000
(Other) :25

      X4          X5          X6          X7          KMean
Min.   : 43.0    Min.   : 7271    Min.   : 8.00    Min.   : 156    1: 8
1st Qu.: 74.0    1st Qu.: 9780    1st Qu.: 20.00   1st Qu.: 9319   2:12
Median : 154.0   Median :10512    Median : 28.00   Median :21152   3:11
Mean   : 321.2   Mean    :12321    Mean    : 42.61   Mean    :23516
3rd Qu.: 352.0   3rd Qu.:13316    3rd Qu.: 46.50   3rd Qu.:37659
Max.   :1874.0   Max.    :26724    Max.    :213.00   Max.    :58797

HC
1:28
2: 1
3: 2

datasetInputLDA(): IMPOR plus HC, KMean
Call:
lda(form, data = datasetInputLDA())

Prior probabilities of groups:
      1          2          3
0.90322581 0.03225806 0.06451613

Group means:
      X1          X2          X3          X4          X5          X6          X7
1  913.4018 70.03571 3.107143 196.3929 10995.89 29.21429 24761.82
2 3581.4200 63.000000 6.000000 1874.0000 26724.00 210.00000 4304.00
3 2736.0450 91.000000 3.000000 1292.0000 23672.50 146.50000 15674.50

Coefficients of linear discriminants:
      LD1          LD2
X1 -5.033063e-05 -1.408860e-03
X2 2.244709e-02 4.361567e-02
X3 -7.886385e-01 -1.433798e+00
X4 3.649612e-03 1.124436e-03

```

```
X5 1.247667e-04 1.072117e-04
X6 3.539871e-02 7.066491e-03
X7 1.520670e-06 -3.495610e-06
```

Proportion of trace:

```
LD1 LD2
0.9447 0.0553
```

datasetInputLDA(): IMPOR plus HC, KMean

Sebaran Kelompok Awal dan hasil LDA. Total Sampel: 31

Validasi LDA

Hit Ratio: 100 %, Cmax: 90.32 %, PRESS'Q: 62

```
LDA
HC 1 2 3
1 28 0 0
2 0 1 0
3 0 0 2
```

datasetInputLDA(): IMPOR plus HC, KMean

Prosentase Sebaran Kelompok Awal dan hasil LDA:

```
LDA
HC 1 2 3
1 90.32 0.00 0.00
2 0.00 3.23 0.00
3 0.00 0.00 6.45
```

Metode Hirarki, pautan *complete linkage*

Kecamatan	X1	X2	X3
Ajung	: 1 Min. : 93.96	Min. : 22.00	Min. : 2.000
Ambulu	: 1 1st Qu.: 689.78	1st Qu.: 47.50	1st Qu.: 2.000
Arjasa	: 1 Median : 955.47	Median : 69.00	Median : 3.000
Balung	: 1 Mean : 1117.06	Mean : 71.16	Mean : 3.194
Bangsalsari	: 1 3rd Qu.: 1298.20	3rd Qu.: 91.50	3rd Qu.: 4.000
Gumuk Mas	: 1 Max. : 3581.42	Max. : 130.00	Max. : 6.000


```
(Other)      :25
      X4          X5          X6          X7          KMean
Min.   : 43.0    Min.   : 7271    Min.   :  8.00    Min.   :  156    1:  4
1st Qu.: 74.0    1st Qu.: 9780    1st Qu.: 20.00    1st Qu.: 9319    2:15
Median : 154.0   Median :10512    Median : 28.00    Median :21152    3:12
Mean   : 321.2   Mean    :12321    Mean    : 42.61    Mean    :23516
3rd Qu.: 352.0   3rd Qu.:13316    3rd Qu.: 46.50    3rd Qu.:37659
Max.   :1874.0   Max.    :26724    Max.    :213.00    Max.    :58797
```

HC

1:16

2:12

3: 3

datasetInputLDA(): IMPOR plus HC, KMean

Call:

```
lda(form, data = datasetInputLDA())
```

Prior probabilities of groups:

```
      1      2      3
0.51612903 0.38709677 0.09677419
```

Group means:

```
      X1      X2      X3      X4      X5      X6      X7
1 1363.9288 57.37500 3.000000 470.5000 13795.56 52.00000 9861.00
2  870.2050 83.91667 3.166667 166.8333 11194.17 30.33333 34291.75
3  787.8133 93.66667 4.333333 142.3333  8965.00 41.66667 53235.67
```

Coefficients of linear discriminants:

```
      LD1      LD2
X1  0.0016762628  1.486650e-03
X2 -0.0112894866 -2.830492e-02
X3  0.4273345188  8.145533e-01
X4 -0.0048378699  9.294488e-04
```

```

X5  0.0001456018 -3.343122e-04
X6  0.0010246224  1.303436e-03
X7  0.0001842413  8.107912e-06
Proportion of trace:
      LD1      LD2
0.9769 0.0231
datasetInputLDA(): IMPOR plus HC, KMean
Sebaran Kelompok Awal dan hasil LDA. Total Sampel: 31
Validasi LDA
Hit Ratio: 100 %,      Cmax: 51.61 %,  PRESS'Q: 62
      LDA
HC   1  2  3
1  16  0  0
2   0 12  0
3   0  0  3
datasetInputLDA(): IMPOR plus HC, KMean
Prosentase Sebaran Kelompok Awal dan hasil LDA:
      LDA
HC      1      2      3
1  51.61  0.00  0.00
2   0.00 38.71  0.00
3   0.00  0.00  9.68
    
```

Metode Hirarki, pautan *average linkage*

Kecamatan	X1	X2	X3
Ajung	: 1 Min. : 93.96	Min. : 22.00	Min. : 2.000
Ambulu	: 1 1st Qu.: 689.78	1st Qu.: 47.50	1st Qu.: 2.000
Arjasa	: 1 Median : 955.47	Median : 69.00	Median : 3.000
Balung	: 1 Mean : 1117.06	Mean : 71.16	Mean : 3.194
Bangsalsari	: 1 3rd Qu.: 1298.20	3rd Qu.: 91.50	3rd Qu.: 4.000
Gumuk Mas	: 1 Max. : 3581.42	Max. : 130.00	Max. : 6.000
(Other)	: 25		

	X4	X5	X6	X7	KMean
Min.	: 43.0	Min. : 7271	Min. : 8.00	Min. : 156	1: 9
1st Qu.:	74.0	1st Qu.: 9780	1st Qu.: 20.00	1st Qu.: 9319	2:12
Median :	154.0	Median :10512	Median : 28.00	Median :21152	3:10
Mean :	321.2	Mean :12321	Mean : 42.61	Mean :23516	
3rd Qu.:	352.0	3rd Qu.:13316	3rd Qu.: 46.50	3rd Qu.:37659	
Max. :	1874.0	Max. :26724	Max. :213.00	Max. :58797	

HC

1:16

2:12

3: 3

datasetInputLDA(): IMPOR plus HC, KMean

Call:

lda(form, data = datasetInputLDA())

Prior probabilities of groups:

	1	2	3
	0.51612903	0.38709677	0.09677419

Group means:

	X1	X2	X3	X4	X5	X6	X7
1	1363.9288	57.37500	3.000000	470.5000	13795.56	52.00000	9861.00
2	870.2050	83.91667	3.166667	166.8333	11194.17	30.33333	34291.75
3	787.8133	93.66667	4.333333	142.3333	8965.00	41.66667	53235.67

Coefficients of linear discriminants:

	LD1	LD2
X1	0.0016762628	1.486650e-03
X2	-0.0112894866	-2.830492e-02
X3	0.4273345188	8.145533e-01
X4	-0.0048378699	9.294488e-04
X5	0.0001456018	-3.343122e-04

```

X6 0.0010246224 1.303436e-03
X7 0.0001842413 8.107912e-06
Proportion of trace:
  LD1  LD2
0.9769 0.0231
datasetInputLDA(): IMPOR plus HC, KMean
Sebaran Kelompok Awal dan hasil LDA. Total Sampel: 31
Validasi LDA
Hit Ratio: 100 %,      Cmax: 51.61 %,  PRESS'Q: 62
  LDA
HC  1  2  3
1 16  0  0
2  0 12  0
3  0  0  3
datasetInputLDA(): IMPOR plus HC, KMean
Prosentase Sebaran Kelompok Awal dan hasil LDA:
  LDA
HC    1    2    3
1 51.61  0.00  0.00
2  0.00 38.71  0.00
3  0.00  0.00  9.68
    
```

Metode Kmeans, Algoritma *Mac Queen*

Kecamatan	X1	X2	X3	X4	X5	X6	X7	KMean
Ajung	: 1	Min. : 93.96	Min. : 22.00	Min. : 2.000				
Ambulu	: 1	1st Qu.: 689.78	1st Qu.: 47.50	1st Qu.: 2.000				
Arjasa	: 1	Median : 955.47	Median : 69.00	Median : 3.000				
Balung	: 1	Mean : 1117.06	Mean : 71.16	Mean : 3.194				
Bangsalsari	: 1	3rd Qu.: 1298.20	3rd Qu.: 91.50	3rd Qu.: 4.000				
Gumuk Mas	: 1	Max. : 3581.42	Max. : 130.00	Max. : 6.000				
(Other)	: 25							
	X4	X5	X6	X7				
	Min. : 43.0	Min. : 7271	Min. : 8.00	Min. : 156				1:11

1st Qu.: 74.0	1st Qu.: 9780	1st Qu.: 20.00	1st Qu.: 9319	2:12
Median : 154.0	Median :10512	Median : 28.00	Median :21152	3: 8
Mean : 321.2	Mean :12321	Mean : 42.61	Mean :23516	
3rd Qu.: 352.0	3rd Qu.:13316	3rd Qu.: 46.50	3rd Qu.:37659	
Max. :1874.0	Max. :26724	Max. :213.00	Max. :58797	

HC

1:16

2:12

3: 3

datasetInputLDA(): IMPOR plus HC, KMean

Call:

lda(form, data = datasetInputLDA())

Prior probabilities of groups:

	1	2	3
	0.3548387	0.3870968	0.2580645

Group means:

	X1	X2	X3	X4	X5	X6	X7
1	1281.6218	62.00000	2.909091	409.0000	12941.45	47.72727	19030.273
2	786.5442	90.83333	3.583333	167.0833	10530.75	31.91667	41100.833
3	1386.5487	54.25000	3.000000	431.6250	14153.62	51.62500	3305.125

Coefficients of linear discriminants:

	LD1	LD2
X1	-4.471094e-04	-1.051042e-03
X2	1.624842e-02	1.222395e-02
X3	-2.124700e-01	6.820866e-01
X4	1.372539e-03	-3.797591e-03
X5	2.670332e-05	2.791535e-04
X6	-8.182529e-03	8.417840e-03
X7	1.644316e-04	-3.924934e-05

Proportion of trace:

	LD1	LD2
	0.986	0.014

```
datasetInputLDA(): IMPOR plus HC, KMean
Sebaran Kelompok Awal dan hasil LDA. Total Sampel: 31
Validasi LDA
Hit Ratio: 93.55 %,      Cmax: 35.48 %,  PRESS'Q: 50.58
```

```
      LDA
KMean  1  2  3
      1 10  0  1
      2  1 11  0
      3  0  0  8
```

```
datasetInputLDA(): IMPOR plus HC, KMean
Prosentase Sebaran Kelompok Awal dan hasil LDA:
```

```
      LDA
KMean   1    2    3
      1 32.26  0.00  3.23
      2  3.23 35.48  0.00
      3  0.00  0.00 25.81
```

B.2 Perhitungan Hasil Validasi Menggunakan Uji Analisis Diskriminan

- Metode Hirarki
 - pautan *single linkage*

Sebaran Kelompok Awal dan hasil LDA:

```
      LDA
HC    1  2  3
      1 28  0  0
      2  0  1  0
      3  0  0  2
```

$$\text{Hit Ratio} = \frac{n \text{ benar}}{N} \times 100\% = \frac{31}{31} \times 100\% = 100\%$$

$$C_{\max} = \frac{n \text{ max}}{N} \times 100\% = \frac{28}{31} \times 100\% = 90,32\%$$

$$\text{Press's } Q = \frac{[N - (nK)]^2}{N(K - 1)} = \frac{[31 - (31 \times 3)]^2}{31 \times (3 - 1)} = \frac{3844}{62} = 62$$

- pautan *complete linkage*

Sebaran Kelompok Awal dan hasil LDA:

```
      LDA
HC    1  2  3
      1 16  0  0
      2  0 12  0
      3  0  0  3
```

$$\text{Hit Ratio} = \frac{n \text{ benar}}{N} \times 100\% = \frac{31}{31} \times 100\% = 100\%$$

$$C_{\max} = \frac{n \text{ max}}{N} \times 100\% = \frac{16}{31} \times 100\% = 51,61\%$$

$$\text{Press's } Q = \frac{[N - (nK)]^2}{N(K - 1)} = \frac{[31 - (31 \times 3)]^2}{31 \times (3 - 1)} = \frac{3844}{62} = 62$$

o pautan *average linkage*

Sebaran Kelompok Awal dan hasil LDA:

		LDA		
HC		1	2	3
	1	16	0	0
	2	0	12	0
	3	0	0	3

$$\text{Hit Ratio} = \frac{n \text{ benar}}{N} \times 100\% = \frac{31}{31} \times 100\% = 100\%$$

$$C_{\max} = \frac{n \text{ max}}{N} \times 100\% = \frac{16}{31} \times 100\% = 51,61\%$$

$$\text{Press's } Q = \frac{[N - (nK)]^2}{N(K - 1)} = \frac{[31 - (31 \times 3)]^2}{31 \times (3 - 1)} = \frac{3844}{62} = 62$$

• Metode *Kmeans* algoritma *Mac Queen*

Sebaran Kelompok Awal dan hasil LDA:

		LDA		
KMean		1	2	3
	1	14	1	0
	2	1	12	0
	3	0	1	2

$$\text{Hit Ratio} = \frac{n \text{ benar}}{N} \times 100\% = \frac{28}{31} \times 100\% = 90,32\%$$

$$C_{\max} = \frac{n \text{ max}}{N} \times 100\% = \frac{14}{31} \times 100\% = 45,16\%$$

$$\text{Press's } Q = \frac{[N - (nK)]^2}{N(K - 1)} = \frac{[31 - (28 \times 3)]^2}{31 \times (3 - 1)} = \frac{2809}{62} = 45,3$$

$$\text{APER} = \frac{n_{12} + n_{21} + n_{32}}{N} = \frac{1+1+1}{31} = \frac{3}{31} = 0,09677$$

$$\text{Akurasi} = (1 - \text{APER}) \times 100\% = (1 - 0,09677) \times 100\% = 90,32\%$$

C. DATA “iris”

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	Setosa
4.9	3.0	1.4	0.2	Setosa
4.7	3.2	1.3	0.2	Setosa
4.6	3.1	1.5	0.2	Setosa
5.0	3.6	1.4	0.2	Setosa
5.4	3.9	1.7	0.4	Setosa
4.6	3.4	1.4	0.3	Setosa
5.0	3.4	1.5	0.2	Setosa
4.4	2.9	1.4	0.2	Setosa
4.9	3.1	1.5	0.1	Setosa
5.4	3.7	1.5	0.2	Setosa
4.8	3.4	1.6	0.2	Setosa
4.8	3.0	1.4	0.1	Setosa
4.3	3.0	1.1	0.1	Setosa
5.8	4.0	1.2	0.2	Setosa
5.7	4.4	1.5	0.4	Setosa
5.4	3.9	1.3	0.4	Setosa
5.1	3.5	1.4	0.3	Setosa
5.7	3.8	1.7	0.3	Setosa
5.1	3.8	1.5	0.3	Setosa
5.4	3.4	1.7	0.2	Setosa
5.1	3.7	1.5	0.4	Setosa
4.6	3.6	1.0	0.2	Setosa
5.1	3.3	1.7	0.5	Setosa
4.8	3.4	1.9	0.2	Setosa
5.0	3.0	1.6	0.2	Setosa
5.0	3.4	1.6	0.4	Setosa
5.2	3.5	1.5	0.2	Setosa
5.2	3.4	1.4	0.2	Setosa
4.7	3.2	1.6	0.2	Setosa
4.8	3.1	1.6	0.2	Setosa
5.4	3.4	1.5	0.4	Setosa
5.2	4.1	1.5	0.1	Setosa
5.5	4.2	1.4	0.2	Setosa

4.9	3.1	1.5	0.2	Setosa
5.0	3.2	1.2	0.2	Setosa
5.5	3.5	1.3	0.2	Setosa
4.9	3.6	1.4	0.1	Setosa
4.4	3.0	1.3	0.2	Setosa
5.1	3.4	1.5	0.2	Setosa
5.0	3.5	1.3	0.3	Setosa
4.5	2.3	1.3	0.3	Setosa
4.4	3.2	1.3	0.2	Setosa
5.0	3.5	1.6	0.6	Setosa
5.1	3.8	1.9	0.4	Setosa
4.8	3.0	1.4	0.3	Setosa
5.1	3.8	1.6	0.2	Setosa
4.6	3.2	1.4	0.2	Setosa
5.3	3.7	1.5	0.2	Setosa
5.0	3.3	1.4	0.2	Setosa
7.0	3.2	4.7	1.4	Versicolor
6.4	3.2	4.5	1.5	Versicolor
6.9	3.1	4.9	1.5	Versicolor
5.5	2.3	4.0	1.3	Versicolor
6.5	2.8	4.6	1.5	Versicolor
5.7	2.8	4.5	1.3	Versicolor
6.3	3.3	4.7	1.6	Versicolor
4.9	2.4	3.3	1.0	Versicolor
6.6	2.9	4.6	1.3	Versicolor
5.2	2.7	3.9	1.4	Versicolor
5.0	2.0	3.5	1.0	Versicolor
5.9	3.0	4.2	1.5	Versicolor
6.0	2.2	4.0	1.0	Versicolor
6.1	2.9	4.7	1.4	Versicolor
5.6	2.9	3.6	1.3	Versicolor
6.7	3.1	4.4	1.4	Versicolor
5.6	3.0	4.5	1.5	Versicolor
5.8	2.7	4.1	1.0	Versicolor
6.2	2.2	4.5	1.5	Versicolor
5.6	2.5	3.9	1.1	Versicolor

5.9	3.2	4.8	1.8	Versicolor
6.1	2.8	4.0	1.3	Versicolor
6.3	2.5	4.9	1.5	Versicolor
6.1	2.8	4.7	1.2	Versicolor
6.4	2.9	4.3	1.3	Versicolor
6.6	3.0	4.4	1.4	Versicolor
6.8	2.8	4.8	1.4	Versicolor
6.7	3.0	5.0	1.7	Versicolor
6.0	2.9	4.5	1.5	Versicolor
5.7	2.6	3.5	1.0	Versicolor
5.5	2.4	3.8	1.1	Versicolor
5.5	2.4	3.7	1.0	Versicolor
5.8	2.7	3.9	1.2	Versicolor
6.0	2.7	5.1	1.6	Versicolor
5.4	3.0	4.5	1.5	Versicolor
6.0	3.4	4.5	1.6	Versicolor
6.7	3.1	4.7	1.5	Versicolor
6.3	2.3	4.4	1.3	Versicolor
5.6	3.0	4.1	1.3	Versicolor
5.5	2.5	4.0	1.3	Versicolor
5.5	2.6	4.4	1.2	Versicolor
6.1	3.0	4.6	1.4	Versicolor
5.8	2.6	4.0	1.2	Versicolor
5.0	2.3	3.3	1.0	Versicolor
5.6	2.7	4.2	1.3	Versicolor
5.7	3.0	4.2	1.2	Versicolor
5.7	2.9	4.2	1.3	Versicolor
6.2	2.9	4.3	1.3	Versicolor
5.1	2.5	3.0	1.1	Versicolor
5.7	2.8	4.1	1.3	Versicolor
6.3	3.3	6.0	2.5	Virginica
5.8	2.7	5.1	1.9	Virginica
7.1	3.0	5.9	2.1	Virginica
6.3	2.9	5.6	1.8	Virginica
6.5	3.0	5.8	2.2	Virginica
7.6	3.0	6.6	2.1	Virginica

4.9	2.5	4.5	1.7	Virginica
7.3	2.9	6.3	1.8	Virginica
6.7	2.5	5.8	1.8	Virginica
7.2	3.6	6.1	2.5	Virginica
6.5	3.2	5.1	2.0	Virginica
6.4	2.7	5.3	1.9	Virginica
6.8	3.0	5.5	2.1	Virginica
5.7	2.5	5.0	2.0	Virginica
5.8	2.8	5.1	2.4	Virginica
6.4	3.2	5.3	2.3	Virginica
6.5	3.0	5.5	1.8	Virginica
7.7	3.8	6.7	2.2	Virginica
7.7	2.6	6.9	2.3	Virginica
6.0	2.2	5.0	1.5	Virginica
6.9	3.2	5.7	2.3	Virginica
5.6	2.8	4.9	2.0	Virginica
7.7	2.8	6.7	2.0	Virginica
6.3	2.7	4.9	1.8	Virginica
6.7	3.3	5.7	2.1	Virginica
7.2	3.2	6.0	1.8	Virginica
6.2	2.8	4.8	1.8	Virginica
6.1	3.0	4.9	1.8	Virginica
6.4	2.8	5.6	2.1	Virginica
7.2	3.0	5.8	1.6	Virginica
7.4	2.8	6.1	1.9	Virginica
7.9	3.8	6.4	2.0	Virginica
6.4	2.8	5.6	2.2	Virginica
6.3	2.8	5.1	1.5	Virginica
6.1	2.6	5.6	1.4	Virginica
7.7	3.0	6.1	2.3	Virginica
6.3	3.4	5.6	2.4	Virginica
6.4	3.1	5.5	1.8	Virginica
6.0	3.0	4.8	1.8	Virginica
6.9	3.1	5.4	2.1	Virginica
6.7	3.1	5.6	2.4	Virginica
6.9	3.1	5.1	2.3	Virginica

5.8	2.7	5.1	1.9	Virginica
6.8	3.2	5.9	2.3	Virginica
6.7	3.3	5.7	2.5	Virginica
6.7	3.0	5.2	2.3	Virginica
6.3	2.5	5.0	1.9	Virginica
6.5	3.0	5.2	2.0	Virginica
6.2	3.4	5.4	2.3	Virginica
6.2	3.4	5.4	2.3	Virginica

D. Hasil Validasi Solusi *Cluster Data "iris"*

D.1 Tampilan Hasil

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Min. :4.300	Min. :2.000	Min. :1.000	Min. :0.100
1st Qu.:5.100	1st Qu.:2.800	1st Qu.:1.600	1st Qu.:0.300
Median :5.800	Median :3.000	Median :4.350	Median :1.300
Mean :5.843	Mean :3.057	Mean :3.758	Mean :1.199
3rd Qu.:6.400	3rd Qu.:3.300	3rd Qu.:5.100	3rd Qu.:1.800
Max. :7.900	Max. :4.400	Max. :6.900	Max. :2.500
Species			
setosa :50			
versicolor:50			
virginica :50			

Metode Hirarki

Validasi Internal

```
Call:
hclust(d = dist(datasetInput0()[, input$xc], method =
input$distc), method = input$method)
```

```
Cluster method : single
Distance       : euclidean
Number of objects: 150
```

	Score	Method	Clusters
Connectivity	0.0000000	hierarchical	2
Dunn	0.4356690	hierarchical	2
Silhouette	0.9518406	hierarchical	2

Call:

```
hclust(d = dist(datasetInput0()[, input$xc], method =  
input$distc), method = input$method)
```

Cluster method : complete
Distance : euclidean
Number of objects: 150

	Score	Method	Clusters
Connectivity	0.0000000	hierarchical	2
Dunn	0.4356690	hierarchical	2
Silhouette	0.9518406	hierarchical	2

Call:

```
hclust(d = dist(datasetInput0()[, input$xc], method =  
input$distc), method = input$method)
```

Cluster method : average
Distance : euclidean
Number of objects: 150

	Score	Method	Clusters
Connectivity	0.0000000	hierarchical	2
Dunn	0.4356690	hierarchical	2
Silhouette	0.9518406	hierarchical	2

Validasi Stabilitas

Call:

```
hclust(d = dist(datasetInput0()[, input$xc], method =  
input$distc), method = input$method)
```

Cluster method : single
Distance : euclidean
Number of objects: 150

	Score	Method	Clusters
APN	0.003266667	hierarchical	2
FOM	0.460346133	hierarchical	8

Call:

```
hclust(d = dist(datasetInput0()[, input$xc], method =  
input$distc), method = input$method)
```

Cluster method : complete
Distance : euclidean
Number of objects: 150


```
Sepal.Length 0.4801072 -0.0217617
Sepal.Width 1.6957809 -2.5309705
Petal.Length -1.6804854 -1.5715342
Petal.Width -0.3374967 2.8343891
```

```
Proportion of trace:
  LD1  LD2
0.9805 0.0195
```

```
datasetInputLDA(): iris plus HC, KMean
```

```
Sebaran Kelompok Awal dan hasil LDA. Total Sampel: 150
Validasi LDA
Hit Ratio: 100 %, Cmax: 65.33 %, PRESS'Q: 300
```

```
      LDA
HC    1  2  3
  1  50  0  0
  2   0 98  0
  3   0  0  2
```

```
datasetInputLDA(): iris plus HC, KMean
```

```
Prosentase Sebaran Kelompok Awal dan hasil LDA:
      LDA
HC    1    2    3
  1 33.33  0.00  0.00
  2  0.00 65.33  0.00
  3  0.00  0.00 13.33
```

Metode Hirarki (*Complete Linkage*)

```
Sepal.Length  Sepal.Width  Petal.Length  Petal.Width
Min. :4.300    Min. :2.000    Min. :1.000    Min. :0.100
1st Qu.:5.100  1st Qu.:2.800  1st Qu.:1.600  1st
Qu.:0.300
Median :5.800    Median :3.000    Median :4.350    Median :1.300
Mean :5.843     Mean :3.057     Mean :3.758     Mean :1.199
3rd Qu.:6.400    3rd Qu.:3.300    3rd Qu.:5.100    3rd Qu.:1.800
Max. :7.900     Max. :4.400     Max. :6.900     Max. :2.500
Species      KMean  HC
setosa      :50   1:61  1:50
versicolor:50   2:50  2:72
virginica :50   3:39  3:28
```

```
datasetInputLDA(): iris plus HC, KMean
```

```
Call:
lda(form, data = datasetInputLDA())
```

```
Prior probabilities of groups:
      1      2      3
0.3333333 0.4800000 0.1866667
```

```
Group means:
```



```

Sepal.Length Sepal.Width Petal.Length Petal.Width
Speciesversicolor
1      5.006000    3.428000    1.462000    0.246000
0.0000000
2      6.545833    2.963889    5.273611    1.850000
0.3194444
3      5.532143    2.635714    3.960714    1.228571
0.9642857
Speciesvirginica
1      0.00000000
2      0.68055556
3      0.03571429

```

Coefficients of linear discriminants:

```

                LD1      LD2
Sepal.Length  -0.4012021 -1.6707830
Sepal.Width   -1.6024646 -0.4214812
Petal.Length   2.2148552  0.6628354
Petal.Width    3.0025798 -0.1543882
Speciesversicolor 1.7708843  1.1849368
Speciesvirginica -1.7708843 -1.1849368

```

Proportion of trace:

```

LD1  LD2
0.974 0.026

```

datasetInputLDA(): iris plus HC, KMean

Sebaran Kelompok Awal dan hasil LDA. Total Sampel: 150

Validasi LDA

Hit Ratio: 97.33 %, Cmax: 45.33 %, PRESS'Q: 276.48

```

LDA
HC  1  2  3
1  50  0  0
2   0 68  4
3   0  0 28

```

datasetInputLDA(): iris plus HC, KMean

Prosentase Sebaran Kelompok Awal dan hasil LDA:

```

LDA
HC  1      2      3
1 33.33  0.00  0.00
2  0.00 45.33  2.67
3  0.00  0.00 18.67

```

Metode Hirarki (Average Linkage)

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Min. :4.300	Min. :2.000	Min. :1.000	Min. :0.100
1st Qu.:5.100	1st Qu.:2.800	1st Qu.:1.600	1st Qu.:0.300
Median :5.800	Median :3.000	Median :4.350	Median :1.300
Mean :5.843	Mean :3.057	Mean :3.758	Mean :1.199

```

3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
Max.      :7.900   Max.      :4.400   Max.      :6.900   Max.      :2.500
  Species  KMean  HC
setosa    :50    1:39   1:50
versicolor:50    2:61   2:64
virginica :50    3:50   3:36
  
```

```

datasetInputLDA(): iris plus HC, KMean
Call:
lda(form, data = datasetInputLDA())
  
```

```

Prior probabilities of groups:
      1      2      3
0.3333333 0.4266667 0.2400000
  
```

```

Group means:
  Sepal.Length Sepal.Width Petal.Length Petal.Width
Speciesversicolor
1      5.006000    3.428000    1.462000    0.246000
0.00000
2      5.929688    2.757812    4.410938    1.439062
0.78125
3      6.852778    3.075000    5.786111    2.097222
0.00000
Speciesvirginica
1      0.00000
2      0.21875
3      1.00000
  
```

```

Coefficients of linear discriminants:
              LD1      LD2
Sepal.Length -0.5724339  0.97787172
Sepal.Width  -1.4747773  0.73562399
Petal.Length  2.3623145 -0.09916186
Petal.Width   2.7529262 -0.55678475
Speciesversicolor 1.5355254 -1.75104850
Speciesvirginica -1.5355254  1.75104850
  
```

```

Proportion of trace:
      LD1      LD2
0.9497 0.0503
  
```

```

datasetInputLDA(): iris plus HC, KMean
  
```

```

Sebaran Kelompok Awal dan hasil LDA. Total Sampel: 150
Validasi LDA
Hit Ratio: 91.33 %,      Cmax: 34 %,      PRESS'Q: 227.07
LDA
HC   1   2   3
  1  50   0   0
  2   0  51  13
  3   0   0  36
  
```

```

datasetInputLDA(): iris plus HC, KMean
Prosentase Sebaran Kelompok Awal dan hasl LDA:
LDA
  
```

```

HC      1      2      3
1 33.33  0.00  0.00
2  0.00 34.00  8.67
3  0.00  0.00 24.00
    
```

Metode Kmeans

```

Sepal.Length Sepal.Width Petal.Length Petal.Width
Min. :4.300 Min. :2.000 Min. :1.000 Min.
:0.100
1st Qu.:5.100 1st Qu.:2.800 1st Qu.:1.600 1st
Qu.:0.300
Median :5.800 Median :3.000 Median :4.350 Median
:1.300
Mean :5.843 Mean :3.057 Mean :3.758 Mean
:1.199
3rd Qu.:6.400 3rd Qu.:3.300 3rd Qu.:5.100 3rd
Qu.:1.800
Max. :7.900 Max. :4.400 Max. :6.900 Max.
:2.500
Species KMean HC
setosa :50 1:39 1:50
versicolor:50 2:61 2:64
virginica :50 3:50 3:36
    
```

```

datasetInputLDA(): iris plus HC, KMean
Call:
lda(form, data = datasetInputLDA())
    
```

```

Prior probabilities of groups:
      1      2      3
0.2600000 0.4066667 0.3333333
    
```

```

Group means:
Sepal.Length Sepal.Width Petal.Length Petal.Width
Speciesversicolor
1 6.853846 3.076923 5.715385 2.053846
0.07692308
2 5.883607 2.740984 4.388525 1.434426
0.77049180
3 5.006000 3.428000 1.462000 0.246000
0.00000000
Speciesvirginica HC2 HC3
1 0.9230769 0.07692308 0.9230769
2 0.2295082 1.0000000 0.0000000
3 0.0000000 0.0000000 0.0000000
    
```

```

Coefficients of linear discriminants:
LD1 LD2
Sepal.Length 0.4982313 0.73022915
Sepal.Width 1.4663428 0.05306084
Petal.Length -2.3571313 -0.14240428
Petal.Width -2.7778114 0.24824705
Speciesversicolor -1.5412992 0.24934288
    
```

```
Speciesvirginica 1.5412992 -0.24934288
HC2 -0.5680933 -3.69245877
HC3 0.5680933 3.69245877
```

Proportion of trace:

```
LD1 LD2
0.8087 0.1913
```

datasetInputLDA(): iris plus HC, KMean

Sebaran Kelompok Awal dan hasil LDA. Total Sampel: 150

Validasi LDA

Hit Ratio: 98 %, Cmax: 40.67 %, PRESS'Q: 282.27

```
LDA
KMean 1 2 3
1 36 3 0
2 0 61 0
3 0 0 50
```

datasetInputLDA(): iris plus HC, KMean

Prosentase Sebaran Kelompok Awal dan hasil LDA:

```
LDA
KMean 1 2 3
1 24.00 2.00 0.00
2 0.00 40.67 0.00
3 0.00 0.00 33.33
```

Species

datasetInputLDA(): iris plus HC, KMean

Call:

```
lda(form, data = datasetInputLDA())
```

Prior probabilities of groups:

```
setosa versicolor virginica
0.3333333 0.3333333 0.3333333
```

Group means:

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
setosa	5.006	3.428	1.462	0.246
versicolor	5.936	2.770	4.260	1.326
virginica	6.588	2.974	5.552	2.026

Coefficients of linear discriminants:

```
LD1 LD2
Sepal.Length 0.8293776 0.02410215
Sepal.Width 1.5344731 2.16452123
Petal.Length -2.2012117 -0.93192121
Petal.Width -2.8104603 2.83918785
```

Proportion of trace:

```
LD1 LD2
0.9912 0.0088
```

datasetInputLDA(): iris plus HC, KMean

Sebaran Kelompok Awal dan hasil LDA. Total Sampel: 150

Validasi LDA

Hit Ratio: 98 %, Cmax: 33.33 %, PRESS'Q: 282.27

	LDA		
Species	setosa	versicolor	virginica
setosa	50	0	0
versicolor	0	48	2
virginica	0	1	49

datasetInputLDA(): iris plus HC, KMean

Prosentase Sebaran Kelompok Awal dan hasil LDA:

	LDA		
Species	setosa	versicolor	virginica
setosa	33.33	0.00	0.00
versicolor	0.00	32.00	1.33
virginica	0.00	0.67	32.67

D.2 Perhitungan Hasil Validasi Menggunakan Uji Analisis Diskriminan

- Metode Hirarki

- pautan *single linkage*

Sebaran Kelompok Awal dan hasil LDA:

	LDA		
HC	1	2	3
1	50	0	0
2	0	98	0
3	0	0	2

$$\text{Hit Ratio} = \frac{n \text{ benar}}{N} \times 100\% = \frac{150}{150} \times 100\% = 100\%$$

$$Cmax = \frac{n \text{ max}}{N} \times 100\% = \frac{98}{150} \times 100\% = 65,33\%$$

$$\text{Press's } Q = \frac{[N - (nK)]^2}{N(K - 1)} = \frac{[150 - (150 \times 3)]^2}{150 \times (3 - 1)} = \frac{90000}{300} = 300$$

- pautan *complete linkage*

Sebaran Kelompok Awal dan hasil LDA:

	LDA		
HC	1	2	3
1	50	0	0
2	0	68	4
3	0	0	28

$$\text{Hit Ratio} = \frac{n \text{ benar}}{N} \times 100\% = \frac{146}{150} \times 100\% = 97,33\%$$

$$Cmax = \frac{n \text{ max}}{N} \times 100\% = \frac{68}{150} \times 100\% = 45,33\%$$

$$\text{Press's } Q = \frac{[N - (nK)]^2}{N(K - 1)} = \frac{[150 - (146 \times 3)]^2}{150 \times (3 - 1)} = \frac{82944}{300} = 276,48$$

$$\text{APER} = \frac{n_{12} + n_{21} + n_{32}}{N} = \frac{0 + 4 + 0}{150} = \frac{4}{150} = 0,027$$

$$\text{Akurasi} = (1 - \text{APER}) \times 100\% = (1 - 0,027) \times 100\% = 97,30\%$$

o *pautan average linkage*

Sebaran Kelompok Awal dan hasil LDA:

HC	LDA		
	1	2	3
1	50	0	0
2	0	51	13
3	0	0	36

$$\text{Hit Ratio} = \frac{n \text{ benar}}{N} \times 100\% = \frac{137}{150} \times 100\% = 91,33\%$$

$$C_{\max} = \frac{n \text{ max}}{N} \times 100\% = \frac{51}{150} \times 100\% = 34\%$$

$$\text{Press's } Q = \frac{[N - (nK)]^2}{N(K - 1)} = \frac{[150 - (137 \times 3)]^2}{150 \times (3 - 1)} = \frac{68121}{300} = 227,07$$

$$\text{APER} = \frac{n_{12} + n_{21} + n_{32}}{N} = \frac{0 + 13 + 0}{150} = \frac{13}{150} = 0,087$$

$$\text{Akurasi} = (1 - \text{APER}) \times 100\% = (1 - 0,087) \times 100\% = 91,30\%$$

• *Metode Kmeans algoritma Mac Queen*

Sebaran Kelompok Awal dan hasil LDA:

KMean	LDA		
	1	2	3
1	36	3	0
2	0	61	0
3	0	0	50

$$\text{Hit Ratio} = \frac{n \text{ benar}}{N} \times 100\% = \frac{147}{150} \times 100\% = 98\%$$

$$C_{\max} = \frac{n \text{ max}}{N} \times 100\% = \frac{61}{150} \times 100\% = 40,67\%$$

$$\text{Press's } Q = \frac{[N - (nK)]^2}{N(K - 1)} = \frac{[150 - (147 \times 3)]^2}{150 \times (3 - 1)} = \frac{84681}{300} = 282,27$$

$$APER = \frac{n_{12} + n_{21} + n_{32}}{N} = \frac{3 + 0 + 0}{150} = \frac{3}{150} = 0,02$$

$$Akurasi = (1 - APER) \times 100\% = (1 - 0,02) \times 100\% = 98\%$$

Species

Sebaran Kelompok Awal dan hasil LDA:

Species	LDA		
	setosa	versicolor	virginica
setosa	50	0	0
versicolor	0	48	2
virginica	0	1	49

$$Hit\ Ratio = \frac{n\ \text{benar}}{N} \times 100\% = \frac{147}{150} \times 100\% = 98\%$$

$$Cmax = \frac{n\ \text{max}}{N} \times 100\% = \frac{50}{150} \times 100\% = 33,33\%$$

$$Press's\ Q = \frac{[N - (nK)]^2}{N(K - 1)} = \frac{[150 - (147 \times 3)]^2}{150 \times (3 - 1)} = \frac{84681}{300} = 282,27$$

$$APER = \frac{n_{12} + n_{21} + n_{32}}{N} = \frac{0 + 2 + 1}{150} = \frac{3}{150} = 0,02$$

$$Akurasi = (1 - APER) \times 100\% = (1 - 0,02) \times 100\% = 98\%$$