PENGKLASIFIKASIAN DATA
MENGGUNAKAN METODE ANALISIS DISKRIMINAN KUADRATIK
DENGAN *EXPECTED COST OF MISCLASSIFICATION* (ECM)
MINIMUM

(Skripsi)

Oleh

SEPRIA HERDYANSAH

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2016

ABSTRACT

DATA CLASSIFICATION
USING QUADRATIC DISCRIMINANT ANALYSIS
WITH EXPECTED COST OF MISCLASSIFICATION (ECM) MINIMUM

By

SEPRIA HERDYANSAH

Discriminant analysis is multivariate analysis method that purpose grouping an object to which one of some population that different based on observations of object characters. The purposes of the research were reviewing theoretically quadratic discriminant analysis model that minimize Expected Cost of Misclassification (ECM). Then, applied on two data population that generated using the R program. Based on these studies, showed that classification rule on quadratic discriminant analysis with Expected Cost of Misclassification (ECM) minimum influenced by probability density function ratio, cost of misclassification ratio, and prior probability ratio. ECM is minimized if R_1 that contains $x_1, \dots , x_p$ such that the integrand is negative.

Keywords : quadratic discriminant analysis, classification, expected cost of misclassification

ABSTRAK

PENGKLASIFIKASIAN DATA MENGGUNAKAN METODE ANALISIS DISKRIMINAN KUADRATIK DENGAN *EXPECTED COST OF MISCLASSIFICATION* (ECM) MINIMUM

Oleh

SEPRIA HERDYANSAH

Analisis diskriminan merupakan metode analisis multivariat yang bertujuan mengelompokkan suatu individu ke salah satu dari beberapa populasi berbeda yang ada berdasarkan pengamatan pada beberapa karakter individu. Penelitian ini bertujuan untuk mengkaji secara teori model analisis diskriminan kuadratik dengan *Expected Cost of Misclassification (ECM)* minimum, kemudian diterapkan pada dua populasi data simulasi yang dibangkitkan menggunakan *software* R. Berdasarkan kajian tersebut diperoleh bahwa kaidah klasifikasi pada Analisis Diskriminan Kuadratik dengan *Expected Cost of Misclassification* (ECM) minimum bergantung pada rasio fungsi kepekatan peluang, rasio biaya kesalahan klasifikasi, dan rasio peluang prior. ECM dikatakan minimum jika R_1 yang memuat $x_1, \dots, x_p$ sedemikian sehingga fungsi dari integralnya bernilai negatif.

Kata Kunci : analisis diskriminan kuadratik, klasifikasi, *minimize expected cost of misclassification*

PENGKLASIFIKASIAN DATA
MENGGUNAKAN METODE ANALISIS DISKRIMINAN KUADRATIK
DENGAN *EXPECTED COST OF MISCLASSIFICATION* (ECM)
MINIMUM

Oleh

SEPRIA HERDYANSAH

Skripsi

Sebagai salah satu syarat untuk mencapai gelar
SARJANA SAINS

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2016

Judul Skripsi : **PENGKLASIFIKASIAN DATA MENGGUNAKAN METODE ANALISIS DISKRIMINAN KUADRATIK DENGAN *EXPECTED COST OF MISCLASSIFICATION* (ECM) MINIMUM**

Nama Mahasiswa : Sepria Herdyansah

Nomor Pokok Mahasiswa : 1117031049

Jurusan : Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

1. Komisi Pembimbing

Widiarti, S.Si., M.Si.	**Drs. Rudi Ruswandi, M.Si.**
NIP 19800502 200501 2 003	NIP 19560208 198902 1 001

2. Ketua Jurusan Matematika

Drs. Tiryono Ruby, M.Sc., Ph.D.
NIP 19620704 198803 1 002

MENGESAHKAN

1. Tim Penguji

Ketua	: **Widiarti, S.Si., M.Si.**	....
Sekretaris	: **Drs. Rudi Ruswandi, M.Si.**	....
Penguji Bukan Pembimbing	: **Dian Kurniasari, S.Si., M.Sc.**	....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Prof. Suharso, Ph.D.
NIP 19690530 199512 1 001

Tanggal Lulus Ujian Skripsi : **25 Januari 2016**

PERNYATAAN SKRIPSI MAHASISWA

Saya yang bertandatangan di bawah ini:

Nama	: **Sepria Herdyansah**
Nomor Pokok Mahawiswa	: **1117031049**
Jurusan	: **Matematika**

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri dan semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah karya penulisan ilmiah Universitas Lampung.

Bandar Lampung, 25 Januari 2016

Yang Menyatakan,

METERAI TEMPEL
409D3ADF986066237
6000 ENAM RIBU RUPIAH

Sepria Herdyansah
NPM. 1117031049

RIWAYAT HIDUP

Penulis dilahirkan di Banjar Baru, Way Kanan pada 5 September 1992, merupakan anak kedua dari tiga bersaudara, dari Bapak Sukisman, S.Pd. dan Ibu Mujiati.

Penulis menempuh pendidikan taman kanak-kanak di TK Dharma Wanita Baradatu pada tahun 1998-1999, sekolah dasar diselesaikan di SDN Banjar Baru pada tahun 2005, lalu pendidikan selanjutnya di SMPN 1 Baradatu pada tahun 2008, dan pendidikan menengah atas di SMAN 9 Bandar Lampung pada tahun 2011.

Pada tahun 2011 penulis terdaftar sebagai mahasiswa S1 Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung (FMIPA Unila). Selama menjadi mahasiswa, penulis aktif di beberapa organisasi kampus seperti Rohani Islam FMIPA Unila 2012/2013, Himpunan Mahasiswa Jurusan Matematika FMIPA Unila 2012/2013 sebagai anggota Bidang Eksternal, UKMF Natural FMIPA Unila 2012/2013 sebagai Kepala Biro Usaha dan sebagai Pimpinan Umum UKMF Natural pada periode 2013/2014 serta Badan Eksekutif Mahasiswa FMIPA Unila 2014/2015 sebagai Kepala Departemen Media dan Informasi.

Pada tahun 2014 penulis melakukan Kerja Praktik (KP) di Kantor Pengelolaan Kekayaan Negara dan Lelang (KPKNL) dan pada tahun 2015 melaksanakan Kuliah Kerja Nyata (KKN) di Desa Marga Mulya Kecamatan Kelumbayan Barat Kabupaten Tanggamus, Lampung.

PERSEMBAHAN

Dengan mengucap puji dan syukur kehadirat Allah SWT kupersembahkan karya kecilku ini untuk:

Bapak dan Ibu tersayang yang telah menjadi motivasi terbesarku selama ini

Dosen Pembimbing dan Penguji yang sangat berjasa, seluruh sahabat-sahabatku dan Almamaterku Universitas Lampung

SANWACANA

Puji syukur kehadirat Allah SWT atas rahmat dan karunia-Nya sehingga skripsi ini dapat diselesaikan.

Skripsi dengan judul “Pengklasifikasian Data Menggunakan Analisis Diskriminan Kuadratik dengan *Expected Cost of Misclassification* (ECM) Minimum” disusun sebagai salah satu syarat memperoleh gelar Sarjana Sains (S.Si.) di Universitas Lampung.

Dengan ketulusan hati penulis ingin mengucapkan terima kasih banyak kepada:

1. Ibu yang tak pernah berhenti berdoa untuk kesuksesanku dan tak hentinya menasehati untuk terus bermunajat kepada-Nya. Ayah yang selalu mendukung dan sabar menanti kelulusanku, serta kakakku Dwi Arso Munandar, S.Si. dan adikku Rina Diana Sari, Amd.KL.
2. Widiarti, S.Si., M.Si. selaku Dosen Pembimbing I, terima kasih untuk bimbingan dan kesedian waktunya selama penyusunan skripsi ini.
3. Drs. Rudi Ruswandi, M.Si. selaku Dosen Pembimbing II, terima kasih untuk bantuan dan masukannya selama penyusunan skripsi.
4. Dian Kurniasari, S.Si., M.Sc. selaku Dosen Penguji, terima kasih atas kesediannya untuk menguji, memberikan saran dan kritik yang membangun dalam penyelesaian skripsi ini.

5. Ibu Dra. Dorrah Aziz, M.Si. selaku Pembimbing Akademik, terima kasih atas bimbingan dan nasihatnya selama ini.
6. Bapak Drs. Tiryono Ruby, M.Sc., Ph.D. selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Bapak Prof. Suharso, Ph.D. selaku Dekan FMIPA Universitas Lampung.
8. Seluruh dosen dan karyawan Jurusan Matematika.
9. Bunda Lusiana, Pak Drajat, dan staf TU Matematika lainnya.
10. Teman seatap dan seperjuangan Dias, Asmawi, Helmi, Sigit, Kak Udin, Wahyu, dan Irul. Terima kasih atas kebersamaanya.
11. Erick, Jordian, Nova, Meri, Wesly, Reno, Bram, dan sahabat matematika 2011 lainnya.
12. Mbak Reny dan Kak Ridho yang selalu memberikan semangat.
13. Keluarga Besar UKMF Natural terima kasih atas ilmu, kekeluargaan, dan kebersamaan sedari awal perkuliahan sampai saat ini. Natural itu kita, kita itu Natural. *We are family*. Salam Pers.
14. Keluarga Besar BEM FMIPA 2014/2015 yang telah menyelipkan kenangan berharga pada masa akhir perkuliahan ini.
15. Himatika FMIPA Universitas Lampung.
16. Semua pihak yang telah membantu dalam penyusunan skripsi ini.
17. Almamater tercinta Universitas Lampung.

Terima kasih, semoga skripsi ini dapat bermanfaat bagi banyak pihak.

Bandar Lampung, Januari 2016
Penulis

Sepria Herdyansah

DAFTAR ISI

Halaman

DAFTAR TABEL .... **xiv**

DAFTAR GAMBAR .... **xv**

I. PENDAHULUAN **1**

1.1. Latar Belakang 1
1.2. Perumusan Masalah 4
1.3. Tujuan Penelitian 4
1.4. Manfaat Penelitian 5

II. TINJAUAN PUSTAKA **6**

2.1 Konsep Matriks 6
2.2 Analisis Peubah Ganda 8
2.3 Distribusi Normal Multivariat 9
2.4 Parameter Distribusi Normal Multivariat 10
2.4.1 Vektor Rataan 10
2.4.2 Matriks Varian Kovarian 11
2.5 Metode Kemungkinan Maksimum Likelihood 12
2.6 Analisis Diskriminan 13
2.6.1 Analisis Diskriminan Kuadratik 16
2.7 Asumsi Analisis Diskriminan Kuadratik 18
2.7.1 Uji Distribusi Normal Multivariat 18
2.7.2 Uji Homogenitas Matriks Varian Kovarian 19

III. METODOLOGI PENELITIAN **21**

3.1 Waktu dan Tempat Penelitian 21
3.2 Data Penelitian 21
3.3 Metode Penelitian 22

IV. HASIL DAN PEMBAHASAN **27**

4.1 Fungsi Diskriminan Kuadratik 27
4.2 Analisis Diskriminan Kuadratik 29

4.3 Pendugaan Parameter **μ** dan $\sum$ 34
4.4 Aplikasi Analisis Diskriminan Kuadratik pada Data Simulasi 38
4.5 Uji Asumsi Analisis Diskriminan 38
4.5.1 Uji Normal Multivariat 39
4.5.2 Uji Homogenitas Matriks Varian Kovarian 40
4.6 Nilai Dugaan Vektor Nilai Tengah dan Matriks Varian Kovarian 40
4.7 Analisis Diskriminan Kuadratik.. 42
4.7.1 Analisis Diskriminan Kuadratik untuk $p_1 = p_2$ dan $C(2|1) = C(1|2)$ 42
4.7.2 Analisis Diskriminan Kuadratik untuk $p_1 < p_2$ dan $C(2|1) < C(1|2)$ 44
4.7.3 Analisis Diskriminan Kuadratik untuk $p_1 > p_2$ dan $C(2|1) > C(1|2)$ 45

V. KESIMPULAN 48

DAFTAR PUSTAKA 49

LAMPIRAN.. 50

Program R 51
Tabel 6-9 55

DAFTAR TABEL

Tabel Halaman

1. Struktur Data pada Analisis Diskriminan 22
2. Data Penelitian 38
3. Hasil Klassifikasi untuk $p_1 = p_2$ dan $C(2|1) = C(1|2)$ 43
4. Hasil Klassifikasi untuk $p_1 < p_2$ dan $C(2|1) < C(1|2)$ 45
5. Hasil Klassifikasi untuk $p_1 > p_2$ dan $C(2|1) > C(1|2)$ 46
6. Data Awal Bangkitan 55
7. Data Hasil Klassifikasi untuk $p_1 = p_2$ dan $C(2|1) = C(1|2)$.......... 59
8. Data Hasil Klassifikasi untuk $p_1 < p_2$ dan $C(2|1) < C(1|2)$.......... 61
9. Data Hasil Klassifikasi untuk $p_1 > p_2$ dan $C(2|1) > C(1|2)$.......... 63

DAFTAR GAMBAR

Gambar Halaman

1. Diagram Alir Pengklasifikasian Data (2 kelompok) Menggunakan Analisis Diskriminan Kuadratik 26

2. Grafik QQ Plot Normal Multivariat.. 39

I. PENDAHULUAN

1.1 Latar Belakang

Menurut Kamus Besar Bahasa Indonesia (KBBI), klasifikasi merupakan penyusunan bersistem dalam suatu kelompok berdasarkan aturan yang telah ditetapkan, maksudnya menunjuk kepada sebuah metode untuk menyusun data menurut aturan tertentu. Klasifikasi juga bisa diartikan sebagai suatu cara pengelompokan yang didasarkan pada ciri-ciri tertentu. Klasifikasi ini diperlukan karena bisa mempermudah dalam membedakan suatu objek antar kelompok berdasarkan ciri-cirinya, sehingga apabila ingin mencari suatu objek dari suatu kelompok akan lebih mudah mencari asal kelompok dari objek tersebut dan ketika sudah diketahui ciri atau peubah penjelas dari suatu objek baru maka akan lebih mudah mengelompokkan objek tersebut. Seperti halnya ketika di perpustakaan, semua buku diletakkan sesuai kategori bidang atau jenisnya. Dalam statistika, untuk mempermudah pengelompokan suatu objek atau untuk membedakan suatu kelompok dengan kelompok lainnya, maka digunakan suatu teknik pengklasifikasian data.

Pengklasifikasian data merupakan salah satu hal penting dalam permasalahan peubah ganda. Dalam peubah ganda, pengklasifikasian digunakan untuk menentukan suatu objek bisa masuk menjadi anggota salah satu dari beberapa kelompok (populasi) yang ada berdasarkan peubah penjelas yang berhasil dikumpulkan. Salah satu teknik peubah ganda yang digunakan dalam pengklasifikasian data ini adalah analisis diskriminan.

Analisis diskriminan merupakan salah satu teknik statistik yang digunakan pada hubungan dependensi (hubungan antar variabel yang sudah bisa dibedakan antara peubah respon dan peubah penjelas). Lebih spesifik lagi, analisis diskriminan digunakan pada kasus dengan peubah respon berupa data kualitatif dan peubah penjelas berupa data kuantitatif

Peubah penjelas dalam analisis diskriminan berupa data berskala interval atau rasio yang digunakan sebagai pertimbangan untuk pengklasifikasian suatu objek baru ke dalam suatu populasi. Peubah penjelas ini berkontribusi untuk memisahkan kelompok dan mendapatkan hasil pengelompokan yang optimal dengan menggunakan peubah penjelas yang dapat menjelaskan gambaran dari masing-masing kelompok. Misalkan dalam bidang pendidikan, peubah penjelas (X_p) yang berupa data kuantitatif ialah nilai raport dari beberapa mata pelajaran ketika di Sekolah Menengah Atas (SMA) dan kategori keberhasilan atau kegagalan studi mahasiswa ketika di perguruan tinggi adalah sebagai data kualitatifnya (Y_k).

Analisis diskriminan bertujuan untuk mengetahui peubah-peubah penjelas yang mampu membedakan suatu kelompok. Kemudian, dapat dipergunakan sebagai kriteria pengelompokan suatu objek baru ke dalam salah satu kelompok dari beberapa kelompok yang ada.

Terdapat beberapa kasus analisis diskriminan yang diketahui. Setiap kasus analisis diskriminan memiliki penggunaan yang berbeda dalam menganalisis data. Analisis diskriminan linear digunakan jika data berdistribusi normal multivariat dan setiap kelompoknya memiliki matriks varian kovarian yang homogen. Analisis diskriminan fisher digunakan jika data tidak berdistribusi normal multivariat tetapi matriks varian kovariannya homogen dalam setiap kelompoknya. Analisis diskriminan nonparametrik digunakan jika data tidak berdistribusi normal multivariat dan matriks varian kovariannya tidak homogen setiap kelompoknya. Sedangkan analisis diskriminan kuadratik digunakan jika data berdistribusi normal multivariat tetapi matriks varian kovariannya tidak homogen dalam setiap kelompoknya.

Kaidah klasifikasi akan menjadi lebih rumit ketika matriks varian kovarian populasinya tidak homogen. Dengan begitu, matriks varian kovariannya maupun vektor mean berbeda antara satu dengan yang lain. Ketika densitas normal multivariat mempunyai matriks varian kovarian yang berbeda, maka hubungan dalam rasio densitas yang menyangkut $|\sum_{\boldsymbol{i}}|^{\frac{1}{2}}$ tidak dihilangkan seperti yang dilakukan ketika $\sum_{\mathbf{1}} = \sum_{\mathbf{2}}$ dan bentuk kuadratik dalam eksponen tidak digabungkan agar memberikan hasil yang lebih sederhana. Kemudian, dalam skema klasifikasi sering kali dinilai dari peluang kesalahan tapi mengabaikan biaya kesalahan klasifikasi $c(i|j)$ sehingga dapat menimbulkan masalah. Oleh karena itu, rata-rata biaya kesalahan klasifikasi atau *Expexted Cost of Misclassification* (ECM) dan total peluang kesalahan klasifikasi atau *Total Probability of Misclassfication* (TPM) haruslah minimum. ECM dan TPM minimum bergantung pada rasio densitas $f_1(x)/f_2(x)$ atau sebanding dengan logaritma natural dari rasio densitas $ln[f_1(x)/f_2(x)]$.

Berdasarkan uraian tersebut, penulis tertarik untuk mengkaji analisis diskriminan kuadratik dengan kaidah klasifikasi ECM minimum yang melibatkan rasio biaya kesalahan klasifikasi, rasio fungsi kepekatan peluang, dan rasio peluang prior. Kemudian menggunakan *software* R akan dikaji penerapannya pada data simulasi.

1.2 Perumusan Masalah

Dalam penelitian ini, akan dikaji secara teori terkait teknik pengklasifikasian atau pengelompokan suatu data mengunakan metode analisis diskriminan kuadratik dengan kaidah klasifikasi *Expexted Cost of Misclassification* (ECM) minimum.

Analisis diskriminan kuadratik dibatasi pada data berdistribusi normal ganda dan ragam yang tidak homogen. Selanjutnya analisis diskriminan kuadratik ini akan diaplikasikan pada suatu kelompok populasi. Tidak ada batasan jumlah kelompok untuk pengelompokan dalam analisis ini, akan tetapi dalam penelitian ini akan dilakukan pengelompokan hanya pada dua kelompok.

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Mengkaji analisis diskriminan kuadratik dalam pengklasifikasian data.
2. Menerapkan pada contoh data simulasi.
3. Mencari model analisis diskriminan kuadratik pada data simulasi.

1.4 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Memahami lebih dalam mengenai Metode Analisis Diskriminan Kuadratik dalam suatu pengklasifikasian data.
2. Memperoleh model analisis diskriminan kuadratik yang mampu mengklasifikasikan suatu objek ke dalam suatu populasi pada data simulasi.

II. TINJAUAN PUSTAKA

2.1 Konsep Matriks

Definisi Matriks

Matriks adalah susunan persegi panjang dari bilangan-bilangan yang diatur dalam baris dan kolom. Apabila matriks **A** terdiri dari m baris dan n kolom. Entri yang terdapat pada baris i dan kolom j dari matriks **A** dinotasikan menjadi a_{ij}. Secara umum matriks **A** dapat ditulis sebagai berikut:

$$A_{m\times n} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

(Anton dan Rorres, 2010)

Transpose Matriks

Jika **A** adalah sebarang matriks $m \times n$ maka **transpose A**, dinyatakan oleh $\mathbf{A}^{\mathbf{T}}$ dan didefenisikan sebagai matriks $n \times m$ yang kolom pertamanya adalah baris pertama dari **A**, dan seterusnya (Anton dan Rorres, 2010).

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \qquad \boldsymbol{A}^{\boldsymbol{T}} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nm} \end{bmatrix}$$

Trace Matriks

Jika **A** adalah matriks kuadrat, maka ***trace A*** dinyatakan oleh ***tr(A)*** yang didefinisikan sebagai jumlah entri-entri pada diagonal utama. ***Trace A*** tidak terdefinisi jika ***A*** bukan matriks kuadrat (Anton dan Rorres, 2010).

Matriks Simetrik

Matriks kuadrat **A** dikatakan simetrik, jika $\mathbf{A}^{\mathbf{T}} = \mathbf{A}$ (Anton dan Rorres, 2010).

Determinan Matriks

Misalkan **A** adalah matriks kuadrat. Fungsi determinan dinyatakan oleh det, dan kita defenisikan det(**A**) sebagai jumlah semua hasil kali elementer bertanda dari **A**. Jumlah det(**A**) kita namakan determinan **A** yang ditulis $det(\boldsymbol{A}) = \sum \pm a_{1_{j1}} a_{2_{j2}} \dots a_{n_{jn}}$, dimana $\sum$ menunjukkan bahwa suku-suku tersebut harus dijumlahkan terhadap semua permutasi (j_1, j_2, ..., j_n) dan simbol + atau – dapat dipilih dalam masing-masing suku sesuai dengan apakah permutasi itu genap atau ganjil (Anton dan Rorres, 2010).

Invers Matriks

Jika **A** adalah sebuah matriks kuadrat dan jika suatu matriks **B** yang berukuran sama dapat dicari sehingga **AB = BA = I**, maka **A** dikatakan dapat balik (*invertible*) dan **B** dinamakan invers (*inverse*) dari **A** (Anton dan Rorres, 2010).

Matriks Definit Positif dan Semidefinit Positif

Misalkan **A** adalah sebuah matriks simetrik berukuran nxn dan $\boldsymbol{x}$ sebarang vektor berukuran $nx1$, maka $\boldsymbol{x}'\boldsymbol{Ax}$ disebut bentuk kuadrat dari **A**. Matriks **A** dikatakan definit positif jika $\boldsymbol{x}'\boldsymbol{Ax} > 0$ dan dikatakan semidefinit positif jika $\boldsymbol{x}'\boldsymbol{Ax} \geq 0$ (Mattjik & Sumertajaya, 2011).

2.2 Analisis Peubah Ganda

Menurut Johnson dan Wichern (2007), analisis peubah ganda digunakan untuk menganalisa data penelitian yang dikumpulkan dari sejumlah objek dengan setiap objek diukur lebih dari satu peubah respon. Secara umum dalam n buah amatan dilakukan pengukuran p peubah. Data tersebut digambarkan sebagai matriks **X** yang berukuran $n \times p$:

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \qquad (2.1)$$

Matriks **X** memuat data yang terdiri dari seluruh data pengamatan terhadap seluruh peubah penjelasnya.

Pengukuran pada baris ke-i yaitu $x_{i1}, x_{i2}, \dots, x_{ip}$ merupakan pengukuran pada individu yang sama, jika disusun sebagai vektor kolom $\boldsymbol{x_i}$ diperoleh:

$$\boldsymbol{x_i} = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix} \qquad (2.2)$$

maka $\boldsymbol{x_i}$ disebut sebagai pengamatan peubah ganda.

2.3 Distribusi Normal Multivariat

Menurut Johnson dan Wichern (2007), kepekatan normal multivariat merupakan generalisasi dari kepekatan normal univariat untuk dimensi ≥ 2.

Pada distribusi normal univariat, peubah acak X dikatakan berdistribusi normal jika fungsi kepekatan peluangnya adalah:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]; -\infty < x < \infty \qquad (2.3)$$

Misalkan $\left(\frac{x-\mu}{\sigma}\right)^2 = (x-\mu)'(\sigma^2)^{-1}(x-\mu)$ (2.4)

adalah ukuran jarak dari x ke μ dalam satuan standar deviasi pada eksponensial dari fungsi kepekatan peluang normal univariat. Jarak ini dapat digeneralisasi untuk suatu vektor pengamatan $\boldsymbol{x}$ berukuran p x l pada beberapa peubah:

$$(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-\mathbf{1}}(\boldsymbol{x}-\boldsymbol{\mu}) \qquad (2.5)$$

Vektor $\boldsymbol{\mu}$ berukuran $p\ x\ l$ merupakan nilai harapan vektor acak $\boldsymbol{x}$ dan matriks Σ berukuran $p\ x\ p$ merupakan matriks varian kovarian. Matriks simetris Σ diasumsikan definit positif, sehingga persamaan (2.5) merupakan jarak kuadrat dari $\boldsymbol{x}$ ke $\boldsymbol{\mu}$.

Kepekatan normal multivariat diperoleh dengan mengganti jarak kuadrat univariat dalam persamaan (2.4) dengan jarak multivariat dalam persamaan (2.5). Sehingga fungsi kepekatan normal *p*-dimensi untuk peubah acak **X** adalah:

$$f(x) = \frac{1}{(2\pi)^{p/2}\,|\Sigma|^{1/2}} exp\left[-\frac{1}{2}\,(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}\,(\boldsymbol{x}-\boldsymbol{\mu})\right] \qquad (2.6)$$

Sehingga dapat ditulis $X \sim N_p\,(\boldsymbol{\mu}, \Sigma)$.

2.4 Parameter Distribusi Normal Multivariat

2.4.1 Vektor Rataan

Misalkan $\boldsymbol{x}$ menggambarkan suatu vektor acak dari p peubah pada suatu unit sampel. Jika ada n pengamatan dalam sampel, maka n vektor pengamatan dinotasikan oleh $\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_n}$. Secara umum dapat dituliskan sebagai:

$$\boldsymbol{x_i} = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix} \tag{2.7}$$

Vektor rataan sampel $\boldsymbol{x}$ diperoleh dari rata-rata n vektor pengamatan atau dengan perhitungan rata-rata dari p peubah lainnya secara terpisah (Rencher, 2002).

$$\overline{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n} x_i = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix} \tag{2.8}$$

dengan $\bar{x}_1$ merupakan rata-rata dari n pengamatan pada peubah pertama, $\bar{x}_2$ rata-rata dari peubah kedua, dan seterusnya.

Kesuluruhan rata-rata dari $\boldsymbol{x}$ dalam populasi disebut vektor rataan populasi atau nilai harapan dari $\boldsymbol{x}$. Hal ini didefinisikan sebagai suatu vektor nilai harapan dari setiap peubah.

$$E(\boldsymbol{X}) = E\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = \begin{bmatrix} E(x_1) \\ E(x_2) \\ \vdots \\ E(x_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \boldsymbol{\mu} \tag{2.9}$$

dimana μ_p adalah rata-rata populasi dari p peubah.

Hal ini bisa memperlihatkan bahwa nilai harapan dari $\bar{x}_p$ di $\bar{\boldsymbol{x}}$ adalah μ_p sehingga $E(\bar{x}_p) = \mu_p$. Dengan demikian, nilai harapan $\bar{\boldsymbol{x}}$ adalah:

$$E(\boldsymbol{X}) = E\begin{bmatrix}\bar{x}_1\\ \bar{x}_2\\ \vdots\\ \bar{x}_p\end{bmatrix} = \begin{bmatrix}E(\bar{x}_1)\\ E(\bar{x}_2)\\ \vdots\\ E(\bar{x}_p)\end{bmatrix} = \begin{bmatrix}\mu_1\\ \mu_2\\ \vdots\\ \mu_p\end{bmatrix} = \boldsymbol{\mu} \qquad (2.10)$$

Oleh karena itu, $\bar{\boldsymbol{x}}$ adalah penduga tak bias bagi $\boldsymbol{\mu}$.

2.4.2 Matriks Varian Kovarian

Menurut Raykov dan Marcoulides (2008), matriks varian kovarian merupakan suatu matriks simetris yang berisi varian pada diagonal utamanya dan kovarian pada elemen selainnya. Koefesien varian menggambarkan sebuah indeks dari hubungan linear antara dua peubah penjelas.

Menurut Everitt (2005), varian populasi dari dua peubah, x_i dan x_j didefinisikan oleh:

$$Cov(x_i, x_j) = E[(x_i - \mu_i)(x_j - \mu_j)] \qquad (2.11)$$

Kovarian dari x_i dan x_j biasanya dinotasikan oleh σ_{ij}. Jadi, varian dari peubah x_i sering dinotasikan oleh σ_{ii} dari pada σ_i^2.

Dengan p peubah, x_1, x_2, ..., x_p, ada p varian dan $\frac{p(p-1)}{2}$ kovarian. Secara umum, perhitungan ini dihasilkan dari suatu $p \times p$ matriks simetris $\sum$, yaitu:

$$\sum = E[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})']$$

$$= E\left(\begin{bmatrix} X_1-\mu_1 \\ X_2-\mu_2 \\ \vdots \\ X_p-\mu_p \end{bmatrix}\begin{bmatrix} X_1-\mu_1 & X_2-\mu_2 & \cdots & X_p-\mu_p \end{bmatrix}\right)$$

$$= \begin{bmatrix} E(X_1-\mu_1)^2 & E(X_1-\mu_1)(X_2-\mu_2) & \cdots & E(X_1-\mu_1)(X_p-\mu_p) \\ E(X_2-\mu_2)(X_1-\mu_1) & E(X_2-\mu_2)^2 & \cdots & E(X_2-\mu_2)(X_p-\mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_p-\mu_p)(X_1-\mu_1) & E(X_p-\mu_p)(X_2-\mu_2) & \cdots & E(X_p-\mu_p)^2 \end{bmatrix}$$

$$= \begin{bmatrix} var(x_1) & cov(x_1,x_2) & \cdots & cov(x_1,x_p) \\ cov(x_2,x_1) & var(x_2) & \cdots & cov(x_2,x_p) \\ \vdots & \vdots & \ddots & \vdots \\ cov(x_p,x_1) & cov(x_p,x_2) & \cdots & var(x_p) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix} \qquad (2.12)$$

dengan $\sigma_{ij} = \sigma_{ji}$. Matriks ini biasanya disebut matriks varian kovarian. Matriks $\sum$ diduga oleh matriks **S**.

S adalah penduga matriks varian kovarian kelompok ke-*i* yang didefinisikan oleh:

$$\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})' \qquad (2.13)$$

dengan $\boldsymbol{x}_i' = [x_{i1}, x_{i2}, \dots, x_{ip}]$ adalah vektor pengamatan untuk *i* pengamatan. Diagonal utama dari matriks **S** berisi varian dari peubah lainnya.

2.5 Metode Kemungkinan Maksimum Likelihood

Menurut Rencher (2002), ketika suatu distribusi seperti normal multivariat diasumsikan untuk semua populasi, nilai dugaan bagi parameter sering diperoleh dengan metode kemungkinan maksimum likelihood (*maximum likelihood estimation*). Vektor pengamatan $\boldsymbol{x_1}$, $\boldsymbol{x_2}$, ..., $\boldsymbol{x_n}$ dianggap diketahui dan nilai $\boldsymbol{\mu}$ dan $\sum$ dicari dengan memaksimumkan densitas bersamanya yang disebut fungsi likelihood, yaitu:

$$L\,(\mathbf{X};\,\boldsymbol{\mu},\textstyle\sum) = \prod_{i=1}^{n} f\,(\boldsymbol{x};\ \boldsymbol{\mu},\textstyle\sum)$$

$$= \frac{1}{(2\pi)^{np/2}\,|\sum|^{n/2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x_i}-\ \boldsymbol{\mu})'{\textstyle\sum}^{-1}\,(\boldsymbol{x_i}-\ \boldsymbol{\mu})\right]$$

$$= |\,2\pi\,|^{-np/2}\,|{\textstyle\sum}|^{-n/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x_i}-\ \boldsymbol{\mu})'{\textstyle\sum}^{-1}\,(\boldsymbol{x_i}-\boldsymbol{\mu})\right] \quad (2.14)$$

Untuk normal multivariat, penduganya adalah:

$$\hat{\boldsymbol{\mu}} = \overline{\boldsymbol{x}}$$

$$\hat{\textstyle\sum} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x_i}-\overline{\boldsymbol{x}})(\boldsymbol{x_i}-\overline{\boldsymbol{x}})'$$

$$= \frac{1}{n}\boldsymbol{W}$$

$$= \frac{n-1}{n}\boldsymbol{S} \quad (2.15)$$

dengan $\boldsymbol{W} = \sum_{i=1}^{n}(\boldsymbol{x_i}-\overline{\boldsymbol{x}})(\boldsymbol{x_i}-\overline{\boldsymbol{x}})'$ dan **S** adalah matriks varian kovarian sampel yang didefiniskan:

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \ldots & s_{pp} \end{bmatrix} \quad (2.16)$$

2.6 Analisis Diskriminan

Menurut Giri (2004), ide dasar analisis diskriminan yaitu dari pengelompokan suatu individu ke salah satu dari beberapa populasi berbeda yang ada berdasarkan pengamatan pada beberapa karakter individu. Misalkan diberikan k populasi berbeda Y_1, ..., Y_k, akan diklasifikasikan suatu individu dengan pengamatan $\boldsymbol{x}' = (x_1, \dots, x_p)$ ke salah satu dari populasi Y_1, ..., Y_k.

Analisis diskriminan merupakan suatu fungsi yang terdiri dari kombinasi linear dari dua atau lebih peubah bebas yang paling baik dalam membedakan antara dua kelompok atau lebih (Sartono, 2003). Jika **X** merupakan peubah acak berdimensi p-variat dan b_k merupakan koefesien diskriminan yang akan diduga, maka fungsi diskriminan dapat dituliskan:

$$\begin{aligned} Y_k &= b_{k1}X_1 + b_{k2}X_2 + \cdots + b_{kp}X_p \\ &= \boldsymbol{b}_k'\boldsymbol{X} \end{aligned} \tag{2.17}$$

Y_k : nilai diskriminan ke-k, dengan k = 1, 2, ..., s dan s ≤ min (m-1, p)

p : jumlah variabel bebas

m : jumlah populasi

b : koefesien diskriminan

X : variabel bebas

Dalam permasalahan pengklasifikasian pada dua kelompok, membedakan kedua kelompok yaitu kelompok Y_1 dan kelompok Y_2 menjadi tujuan utama dalam suatu pengelompokan data.

$$P(X \in Y_i) = p_i \tag{2.18}$$

Persamaan (2.18) merupakan peluang prior dari suatu pengamatan $X = x$ yang mengalokasikannya ke dalam kelompok Y_1 atau kelompok Y_2. Andaikan juga bahwa peluang kejadian X yang mengalokasikan ke dalam kelompok Y_i adalah:

$$P(X \in Y_i | X \in Y_i) = f_i(x) \tag{2.19}$$

Berdasarkan persamaan (2.18) dan (2.19), teorema Bayes menghasilkan peluang posterior, yaitu:

$$P(Y_i|x) = P(X \in Y_i|X = x) = \frac{f_i(x)p_i}{f_1(x)p_1 + f_2(x)p_2} \qquad (2.20)$$

maka pengamatan $\boldsymbol{x}$ akan dialokasikan ke dalam kelompok Y_i, $i = 1,2$.

Menurut Giri (2004), misalkan seluruh ruang berdimensi p dari X dilambangkan E^p dan R adalah ruang atau daerah dari semua observasi $\boldsymbol{x}$, akan ditentukan aturan untuk membagi E^p ke dalam k daerah yaitu R_1, ..., R_k, sehingga jika $\boldsymbol{x}$ jatuh di R_i, maka $\boldsymbol{x}$ akan diklasifikasikan sebagai anggota populasi Y_i. Namun ada kemungkinan bahwa $\boldsymbol{x}$ yang semestinya merupakan anggota populasi Y_i diklasifikasikan menjadi anggota populasi Y_j. Biaya kesalahan klasifikassi suatu individu ke Y_j yang seharusnya ke Y_i dinotasikan dengan C(j|i).

Peluang kesalahan klasifikasi suatu individu dengan pengamatan $\boldsymbol{x}$ dari Y_i yang diklasifikasikan ke Y_j adalah

$$P(j|i, R) = \int_{R_j} f_i(x)\, dx \qquad (2.21)$$

Nilai kesalahan klasifikasi suatu pengamatan dari suatu populasi Y_i yaitu:

$$r_i(R) = \sum_{j=1, j\neq i}^{k} C(j|i)P(j|i), \qquad i = 1, \dots, k. \qquad (2.22)$$

Dalam menggambarkan suatu aturan klasifikasi terbaik, dibutuhkan pembanding vektor kesalahan $r(R) = \big(r_1(R), \dots, r_k(R)\big)$ untuk daerah R yang berbeda.

Misalkan p_i melambangkan proporsi dari kelompok Y_i, $i = 1, \dots, k$. Jika p_i diketahui, dapat ditetapkan rata-rata kesalahan klasifikasi suatu individu. Karena peluang yang menggambarkan suatu pengamatan dari Y_i adalah p_i dan secara tepat pengelompokan ke dalam Y_i diberikan oleh $p_i P(i|i,R), i = 1, \dots, k$. Begitu juga untuk peluang yang menggambarkan suatu pengamatan dari Y_i. Kemudian kesalahan klasifikasi ke Y_j $(i \neq j)$ adalah $p_i P(j|i,R)$.

$$\sum_{i=1}^{k} p_i \sum_{j=1, j\neq i}^{k} C(j|i)P(j|i,R) \quad (2.23)$$

adalah rata-rata kesalahan klasifikasi untuk aturan klasifikasi dengan mempertimbangkan peluang prior $p = (p_1, \dots, p_k)$.

2.6.1 Analisis Diskriminan Kuadratik

Pada kasus data yang variannya tidak homogen, aturan klasifikasi tergatung pada matriks varian-kovarian yaitu $\sum_1$, $\sum_2$, ..., $\sum_k$ (Sartono, 2003).

Suatu pengamatan $\boldsymbol{x}$ akan dimasukkan ke populasi ke-*t* jika

$$d_t^2(x) = \min_{j=1,\cdots,k} \left\{ d_j^2(x) \right\} \quad (2.24)$$

Dengan $d_t^2(x)$ adalah kuadrat jarak yang didefinisikan oleh:

$$d_j^2(x) = \left[\boldsymbol{x} - \overline{\boldsymbol{x}}_{\boldsymbol{j}}\right]' \boldsymbol{S}^{-1} \left[\boldsymbol{x} - \overline{\boldsymbol{x}}_{\boldsymbol{j}}\right] + ln\left|\boldsymbol{S}_{\boldsymbol{j}}\right| - 2ln(\pi_t) \text{ ; j=1, 2, ..., k} \quad (2.25)$$

Pengklasifikasian data ke dalam populasi juga dapat dilakukan dengan peluang posterior terbesar. Peluang tersebut besarnya diperoleh dari:

$$P(t|x) = \frac{e^{-\frac{1}{2}d_j^2(x)}}{e^{-\frac{1}{2}d_1^2(x)} + e^{-\frac{1}{2}d_2^2(x)}} \quad (2.26)$$

Johnson dan Wichern (2007), mempertimbangkan densitas normal multivariat yang mana $\sum_{\boldsymbol{i}}$, *i=1*,*2*, ..., *n* menggantikan $\sum$. Menurutnya, dengan begitu matriks varian kovarian maupun vektor nilai tengah menjadi tidak homogen untuk setiap populasi. Ketika densitas normal multivariat mempunyai matriks varian kovarian yang tidak homogen, hubungan dalam rasio densitas yang menyangkut $|\sum_i|^{\frac{1}{2}}$ tidak dihilangkan seperti yang dilakukan ketika $\sum_1 = \sum_2$. Selain itu, bentuk kuadratik dalam eksponen dari $f_1(x)$ dan $f_2(x)$ tidak digabungkan untuk memberikan hasil yang lebih sederhana.

Dalam skema klasifikasi sering kali dinilai dari peluang kesalahan tapi mengabaikan biaya kesalahan klasifikasi $c(i|j)$ sehingga dapat menimbulkan masalah. Oleh karena itu, rata-rata biaya kesalahan klasifikasi atau *Expexted Cost of Misclassification* (ECM) dan total peluang kesalahan klasifikasi atau *Total Probability of Misclassfication* (TPM) haruslah minimum. ECM dan TPM minimum bergantung pada rasio densitas $f_1(x)/f_2(x)$ atau sebanding dengan logaritma natural dari rasio densitas $ln[f_1(x)/f_2(x)]$. Kaidah pengklasifikasian objek ke dalam populasi yang meminimumkan rata-rata biaya kesalahan klasifikasi diberikan oleh:

$$R_1 = -\frac{1}{2}\boldsymbol{x}'(\Sigma_1^{-1} - \Sigma_2^{-1})\boldsymbol{x} + (\boldsymbol{\mu}_1'\Sigma_1^{-1} - \boldsymbol{\mu}_2'\Sigma_2^{-1})\boldsymbol{x} - k \geq ln\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right] \qquad (2.27)$$

dengan $k = \frac{1}{2} ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) + \frac{1}{2}\left(\boldsymbol{\mu}_1'\Sigma_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2'\Sigma_2^{-1}\boldsymbol{\mu}_2\right)$

$$p = \frac{n_i}{N}$$

$$c = \left|Y_{sebenarnya} - Y_{klasifikasi}\right|$$

Menurut Izenman (2008), persamaan (2.27) memiliki bentuk fungsi kuadratik dari $\boldsymbol{x}$ yang dapat dituliskan sebagai:

$$R = \beta_0 + \beta' x + x'\Omega x \qquad (2.28)$$

dengan,

$$\Omega = -\frac{1}{2}\left(\boldsymbol{\Sigma_1^{-1}} - \boldsymbol{\Sigma_2^{-1}}\right) \tag{2.29}$$

$$\beta = \boldsymbol{\Sigma_1^{-1}\mu_1} - \boldsymbol{\Sigma_2^{-1}\mu_2} \tag{2.30}$$

$$\beta_0 = -\frac{1}{2}\left\{ln\frac{|\Sigma_1|}{|\Sigma_2|} + \boldsymbol{\mu_1'\Sigma_1^{-1}\mu_1} - \boldsymbol{\mu_2'\Sigma_2^{-1}\mu_2}\right\} \tag{2.31}$$

2.7 Asumsi Analisis Diskriminan Kuadratik

2.7.1 Uji Distribusi Normal Multivariat

Secara umum, untuk pengujian data berdistribusi normal mutivariat, digunakan hipotesis:

$H_0 = X_1, X_2, \ldots, X_n$ berdistribusi multivariat normal

$H_1 = X_1, X_2, \ldots, X_n$ tidak berdistribusi multivariat normal

Pengujian asumsi yang digunakan adalah *Shapiro-Wilk's Test.* Uji Statistik Shapiro-Wilk didasarkan pada suatu sampel acak berukuran n, x_1, ..., x_n yang didefiniskan sebagai:

$$W_X = \frac{\tilde{\sigma}_X^2}{S_X^2} \tag{2.32}$$

dengan $S_X^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$ dan $\tilde{\sigma}_X^2 = \left[\sum_{i=1}^{n} a_i\, x_{(i)}\right]^2$ dimana $x_{(1)} < \cdots < x_{(n)}$ adalah statistik order, serta a_i adalah anggota ke-i dari vektor $\boldsymbol{a} = (a_1, \ldots, a_n)'$. Vektor $\boldsymbol{a}$ diberikan oleh:

$$\boldsymbol{a} = (a_1, \ldots, a_n)' = \frac{\boldsymbol{m'V^{-1}}}{\left(\boldsymbol{m'V^{-1}V^{-1}m}\right)^{\frac{1}{2}}} \tag{2.33}$$

dengan $\boldsymbol{m}' = E(\boldsymbol{Z})$ dan $\boldsymbol{V} = cov(\boldsymbol{Z})$ dimana **Z** merupakan vektor statistik order berukuran n.

Uji ini akan tolak H_0 dengan suatu ukuran taraf nyata α jika $W_X < k_\alpha$ dengan k_α merupakan persentil $100\alpha\%$ dari distribusi W_X (Alva dan Estrada, 2009).

2.7.2 Uji Homegenitas Matriks Varian Kovarian

Untuk menguji kehomogenan matriks varian kovarian ($\sum$) antar kelompok, digunakan hipotesis:

H_0 : $\sum_1 = \sum_2 = \ldots = \sum_k$

H_1 : $\sum_i \neq \sum_j$ (sedikitnya ada dua kelompok yang berbeda) $i \neq j = 1,2,\ldots,k$

Statistik uji yang digunakan adalah statistik Box's M, yaitu:

$$-2\, ln\, \lambda^* = (n-k)\, ln\, |\, \boldsymbol{W}/(n-k)\, | - \sum (ni-1)\, ln\, |\boldsymbol{S_j}| \qquad (2.34)$$

dengan:

$$\lambda^* = \frac{\prod |\boldsymbol{S_j}|^{(n_j-1)/2}}{|\boldsymbol{W}/(n-k)|^{(n-k)/2}}$$

k	= banyaknya kelompok
$\boldsymbol{W}/(n-k)$	= matriks varian kovarian dalam kelompok gabungan
$\boldsymbol{S_j}$	= matriks varian kovarian kelompok ke-j

Bila hipotesis nol benar, maka:

$(-2\, ln\, \lambda^*)/b$ akan mengikuti sebaran F dengan derajat bebas v_1 dan v_2 pada taraf nyata α, dimana:

$$v_1 = (1/2)(k-1)p(p+1)$$

$$v_2 = (v_1+2)/(a_2-a_1^2)$$

$$b = v_1/(1-a_1-v_1/v_2)$$

dengan,

$$a_1 = \frac{2p^2+3p-1}{6(k-1)(p+1)}\left[\sum \frac{1}{(n_j-1)} - \frac{1}{(n-k)}\right]$$

$$a_2 = \frac{(p-1)(p-2)}{6(k-1)}\left[\sum \frac{1}{(n_j-1)^2} - \frac{1}{(n-k)^2}\right]$$

p = jumlah variabel penjelas dalam fungsi diskriminan

Karena itu, apabila $(-2\ ln\ \lambda^*)/b > F_{v_1,v_2,\alpha}$ maka H_0 ditolak dan dapat disimpulkan bahwa terdapat kelompok yang memiliki matriks varian kovarian yang tidak homogen (Mattjik & Sumertajaya, 2011).

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung, pada semester ganjil Tahun Pelajaran 2015/2016.

3.2 Data Penelitian

Data penelitian yang digunakan adalah data yang dibangkitkan dengan simulasi menggunakan *software* R. Data yang dibangkitkan sebanyak 100 data dengan lima peubah penjelas (X_1 – X_5) untuk setiap kelompok I (Y_1) dan kelompok II (Y_2) dengan sebaran normal multivariat ($X \sim N_p(\mu, \Sigma)$) serta ragam tidak homogen ($\Sigma_\mathbf{1} \neq \Sigma_\mathbf{2}$). Bentuk data secara umum dapat dilihat pada Tabel 1 dan secara lengkap dapat dilihat pada Tabel 6 Lampiran 2.

Tabel 1. Struktur Data Pada Analisis Diskriminan

Populasi	Pengamatan	X_1	X_2	X_3	$\cdots$	X_k
Y_1	1 2 $\vdots$ j_1	x_{111} x_{121} $\vdots$ x_{1j_11}	x_{112} x_{122} $\vdots$ x_{1j_12}	x_{113} x_{123} $\vdots$ x_{1j_13}	$\cdots$ $\cdots$ $\cdots$ $\cdots$	x_{11k} x_{12k} $\vdots$ x_{1j_1k}
Y_2	1 2 $\vdots$ j_2 $\vdots$ j_i	x_{211} x_{221} $\vdots$ x_{2j_21} $\vdots$ x_{ij_i1}	x_{212} x_{222} $\vdots$ x_{2j_22} $\vdots$ x_{ij_i2}	x_{213} x_{223} $\vdots$ x_{2j_23} $\vdots$ x_{ij_i3}	$\cdots$ $\cdots$ $\cdots$ $\cdots$ $\cdots$ $\cdots$	x_{21k} x_{22k} $\vdots$ x_{2j_2k} $\vdots$ x_{ij_ik}

dengan:

x_{ijk} : Nilai pengamatan pada kelompok ke-i untuk pengulangan ke-j dan variabel ke-k

j_i : Pengamatan ke-j pada kelompok ke-i

3.3 Metode Penelitian

Metode yang digunakan dalam studi ini adalah studi pustaka dengan mempelajari buku-buku teks penunjang yang berhubungan dengan tugas akhir ini. Kemudian analisis pada data simulasi dalam penelitian ini akan menggunakan *software* R.

Langkah-langkah yang dilakukan dalam penelitian ini adalah:

1. Menduga parameter distribusi normal multivariat dengan menggunakan metode *Maximum Likelihood Estimation*.
2. Menunjukkan model pengklasifikasi menggunakan persamaan fungsi diskriminan kuadratik. Aturan klasifikasi menggunakan:

$$R_1 = -\frac{1}{2}\boldsymbol{x}'(\Sigma_1^{-1} - \Sigma_2^{-1})\boldsymbol{x} + (\boldsymbol{\mu}_1'\Sigma_1^{-1} - \boldsymbol{\mu}_2'\Sigma_2^{-1})\boldsymbol{x} - k \geq ln\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right]$$

dengan:

$$k = \frac{1}{2} ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) + \frac{1}{2}(\boldsymbol{\mu}_1'\Sigma_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2'\Sigma_2^{-1}\boldsymbol{\mu}_2)$$

$$p = \frac{n_i}{N}$$

$$C(i|j) = |\prod_{j\ sebenarnya} - \prod_{j\ klasifikasi}|$$

3. Menerapkan pada data simulasi.

a. Membangkitkan data dengan menggunakan *software* R sebanyak 100 dengan lima peubah penjelas (X_1 – X_5) untuk setiap kelompok I (Y_1) dan kelompok II (Y_2) dengan sebaran normal peubah ganda ($X \sim N_p(\mu, \Sigma)$) serta ragam tidak homogen ($\Sigma_1 \neq \Sigma_2$) dengan parameter sebagai berikut:

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \\ 10 \end{bmatrix} \qquad \boldsymbol{\mu}_2 = \begin{bmatrix} 5 \\ 7 \\ 9 \\ 11 \\ 13 \end{bmatrix}$$

$$\Sigma_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 9 & 0 & 0 \\ 0 & 0 & 0 & 16 & 0 \\ 0 & 0 & 0 & 0 & 25 \end{bmatrix} \qquad \Sigma_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 & 0 \\ 0 & 0 & 27 & 0 & 0 \\ 0 & 0 & 0 & 64 & 0 \\ 0 & 0 & 0 & 0 & 125 \end{bmatrix}$$

b. Mencari nilai dugaan parameter dengan menghitung nilai rata-rata data $\bar{\boldsymbol{x}}_i$ dan matriks varian kovarian $\boldsymbol{S}_i$. Nilai rata-rata dan matriks varian kovarian diperoleh dengan menggunakan persamaan:

$$\bar{\boldsymbol{x}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij} \text{ dan } \hat{\Sigma}_i = \boldsymbol{S}_i = \frac{1}{n_i - 1}\left(\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)\right)$$

c. Membentuk model fungsi diskriminan kuadratik berdasarkan data simulasi dengan jumlah data tiap kelompok yaitu $n_1 = n_2$=100.

d. Mengklasifikasi data menggunakan aturan klasifikasi analisis diskriminan kuadratik.

e. Melakukan *resampling* pada data awal untuk memperoleh jumlah data kelompok $n_1 < n_2$ sehingga $p_1 < p_2$.

f. Membentuk model fungsi diskriminan kuadratik berdasarkan data simulasi dengan jumlah data tiap kelompok yaitu $n_1 < n_2$.

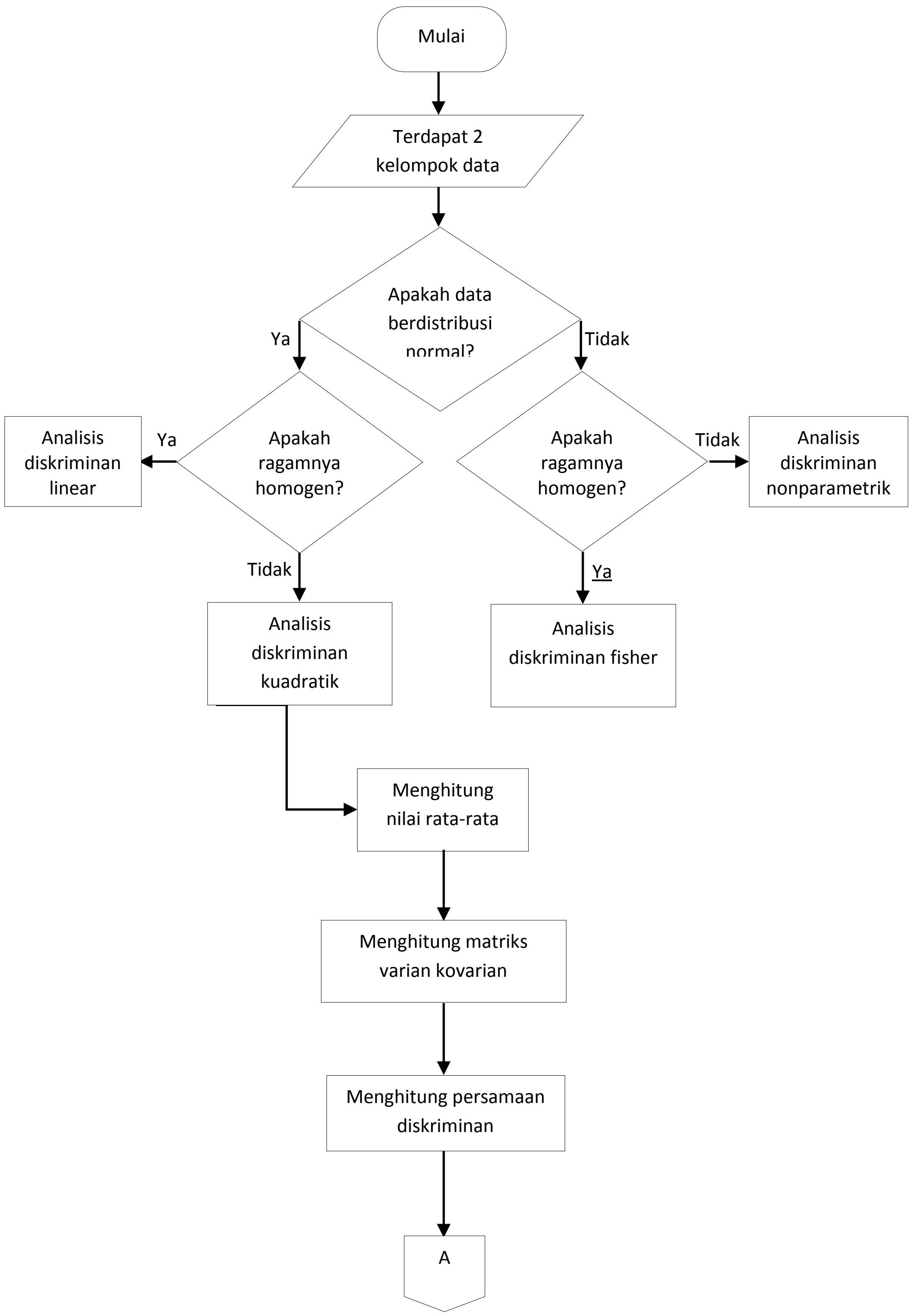
g. Mengklasifikasi data menggunakan aturan klasifikasi analisis diskriminan kuadratik.

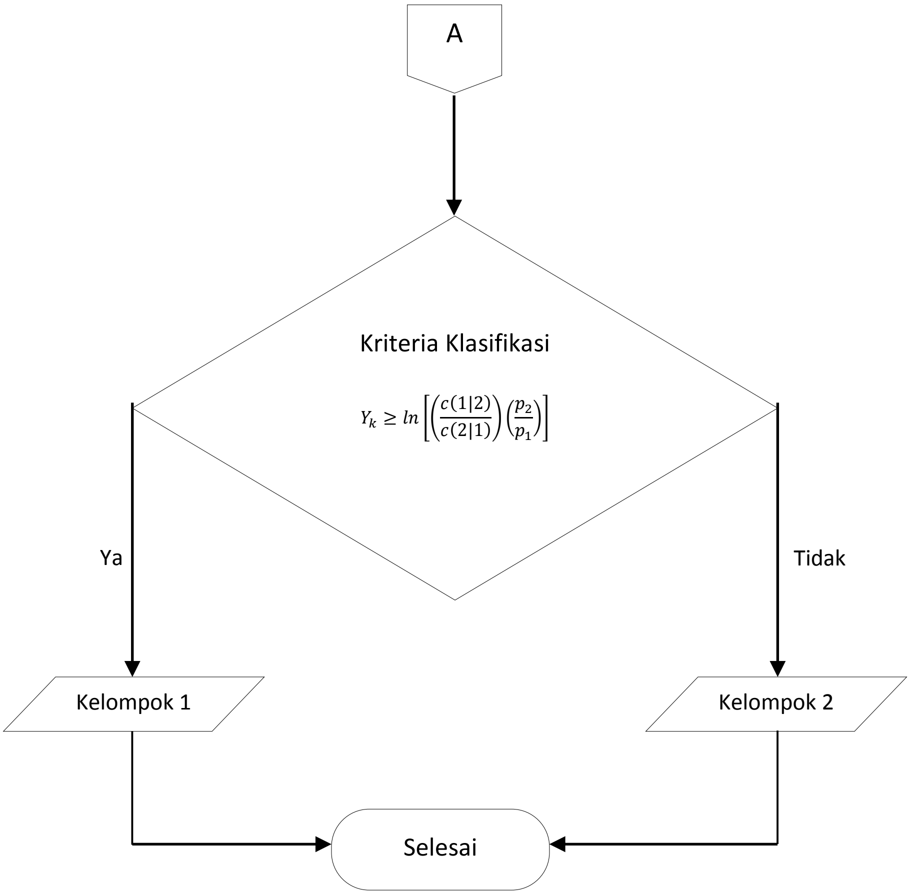
h. Melakukan *resampling* pada data awal untuk memperoleh jumlah data kelompok $n_1 > n_2$ sehingga $p_1 > p_2$.

i. Membentuk model fungsi diskriminan kuadratik berdasarkan data simulasi dengan jumlah data tiap kelompok yaitu $n_1 > n_2$.

j. Mengklasifikasi data menggunakan aturan klasifikasi analisis diskriminan kuadratik.

Secara garis besar langkah-langkah penelitian yang dilakukan dapat dilihat pada diagram alir dalam Gambar 1.


Mulai
Terdapat 2 kelompok data
Apakah data berdistribusi normal?
Ya
Tidak
Apakah ragamnya homogen?
Ya
Analisis diskriminan linear
Tidak
Analisis diskriminan kuadratik
Apakah ragamnya homogen?
Tidak
Analisis diskriminan nonparametrik
Ya
Analisis diskriminan fisher
Menghitung nilai rata-rata
Menghitung matriks varian kovarian
Menghitung persamaan diskriminan
A



Gambar 1. Diagram Alir Pengklasifikasian Data (2 kelompok) Menggunakan Analisis Diskriminan Kuadratik

V. KESIMPULAN

Adapun kesimpulan yang diperoleh dari hasil penelitian sebagai berikut:

1. Kaidah klasifikasi pada Analisis Diskriminan Kuadratik dengan *Expected Cost of Misclassification* (ECM) minimum bergantung pada rasio fungsi kepekatan peluang, rasio biaya kesalahan klasifikasi, dan rasio peluang prior sehingga diperoleh persamaan diskriminan kuadratik sebagai berikut:

$$R_1 = -\frac{1}{2}\boldsymbol{x}'\left(\boldsymbol{\Sigma_1^{-1}} - \boldsymbol{\Sigma_2^{-1}}\right)\boldsymbol{x} + \left(\boldsymbol{\mu}_1'\boldsymbol{\Sigma_1^{-1}} - \boldsymbol{\mu}_2'\boldsymbol{\Sigma_2^{-1}}\right)\boldsymbol{x} - k \geq ln\left[\left(\frac{C(1|2)}{C(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right]$$

2. ECM minimum jika R_1 yang memuat $x_1, \ldots, x_p$ sedemikian sehingga fungsi dari integralnya bernilai negatif, yaitu $C(1|2)p_2f_2(x) - C(1|2)p_2f_2(x) < 0$

3. Berdasarkan data simulasi, diperoleh aturan klasifikasi sebagai berikut:

$$R_1 = -\frac{1}{2}\boldsymbol{x}'\begin{bmatrix} -0{,}232494 & -0{,}050246 & -0{,}008870 & -0{,}009702 & -0{,}000804 \\ -0{,}050246 & 0{,}080959 & -0{,}014447 & 0{,}008266 & -0{,}016860 \\ -0{,}008870 & -0{,}014447 & 0{,}065173 & 0{,}010834 & 0{,}015990 \\ -0{,}009702 & 0{,}008266 & 0{,}010834 & 0{,}051537 & 0{,}003986 \\ -0{,}000804 & -0{,}016860 & 0{,}015990 & 0{,}003986 & 0{,}038256 \end{bmatrix}\boldsymbol{x}$$

$$+ [-4{,}108325 \quad -0{,}299451 \quad 0{,}401608 \quad 0{,}537062 \quad 0{,}3659638]\boldsymbol{x}$$

$$-\frac{1}{2}\ln(0{,}01412723) - \frac{1}{2}(-19{,}9735) \geq ln\left[\left(\frac{C(1|2)}{C(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right]$$

DAFTAR PUSTAKA

Alva, J. A. V. & Estrada, E. G. 2009. *A Generalization of Shapiro–Wilk's Test for Multivariate Normality*. Communications in Statistics - Theory and Methods. Vol. 38. No. 11. Page 1870-1883.

Anton, H. & Rorres, C. 2010. *Elementary Linear Algebra.* Tenth Edition. New York: John Wiley & Son, Inc.

Everitt, B.S. 2005. *An R and S-PLUS Companion to Multivariate Analysis.* London: Springer.

Giri, N.C. 2004. *Multivariate Statistical Analysis.* Second Edition. New York: Marcel Dekker, Inc.

Izenman, A.J. 2008. *Modern Multivariate Statistical Techniques*. Philadelphia: Springer.

Johnson, R. A. & Wichern, D. W. 2007. *Applied Multivariate Statistical Analysis*. Sixth Edition. New York: Prentice-Hall, Inc.

Mattjik, A. & Sumertajaya, I.M. 2011. *Sidik Peubah Ganda Menggunakan SAS*. Bogor: IPB Press.

Raykov, T. & Marcoulides, G.A. *An Introduction to Applied Multivariat Analysis*. New York: Taylor and Fracis Group.

Rencher, A.C. 2002. *Methods of Multivariate Analysis*. Second Edition. New York: John Wiley & Son, Inc.

Sartono, B. dkk. 2003. *Analisis Peubah Ganda*. Bogor: Institut Pertanian Bogor.