

PENGKLASIFIKASIAN KEJADIAN DIABETES
DENGAN ANALISIS DISKRIMINAN KUADRATIK
(Studi Kasus: Puskesmas Kecamatan Pesanggrahan Kota Jakarta Selatan)

Skripsi
Disusun untuk melengkapi syarat-syarat
guna memperoleh gelar Sarjana Sains



BETY SURYANI PUTRI

3125121975

PROGRAM STUDI MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI JAKARTA

2017

ABSTRACT

BETY SURYANI PUTRI, 3125121975. Classification Of Diabetes Case Using Quadratic Discriminant Analysis (Case Study: Puskesmas Kecamatan Pesanggrahan Kota Jakarta Selatan). Thesis. Faculty of Mathematics and Natural Science, Jakarta State University. 2017.

Classifying diabetes case accurately and stably is an important thing because it can be used as an early prevention before the occurrence of complications. One of the statistic methods which can analyze classification of diabetes case is using linear discriminant analysis. However, classification of diabetes case using linear discriminant analysis can give unstable classification because the homogeneity assumption could not be fulfilled. Quadratic discriminant analysis can be the best solution for that problem because the heterogeneity matrix would not be ignored. In this thesis, data that will be used is primer data of diabetes patient in Puskesmas Kecamatan Pesanggrahan Kota Jakarta Selatan. Based on linear discriminant function, it gets that someone who have more weight, more triglycerides level, and more systolic blood pressure level will make s/he classified into diabetes patient. Then, based on APER test, it gets that the percentage of misclassification on testing data using linear discriminant analysis is 35.13% and using quadratic discriminant analysis is 27.02%. Furthermore, based on accuracy and stability test using quadratic discriminant analysis is evaluated to be more accurate and more stable than using linear discriminant analysis. Therefore, classification of diabetes case using quadratic discriminant analysis is more optimal than using linear discriminant analysis.

Keywords : diabetes, heterogeneity, linear discriminant, quadratic discriminant, classification accuracy.

ABSTRAK

BETY SURYANI PUTRI, 3125121975. Pengklasifikasian Kejadian Diabetes Dengan Analisis Diskriminan Kuadratik (Studi Kasus: Puskesmas Kecamatan Pesanggrahan Kota Jakarta Selatan). Skripsi. Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta. 2017.

Pengklasifikasian kejadian diabetes secara akurat dan stabil merupakan hal penting karena dapat digunakan sebagai pencegahan dini sebelum terjadinya komplikasi. Salah satu metode statistika untuk menganalisis pengklasifikasian kejadian diabetes adalah dengan metode analisis diskriminan linear. Namun pengklasifikasian kejadian diabetes dengan analisis diskriminan linear dapat memberikan hasil yang tak stabil karena asumsi homogenitas yang tak terpenuhi. Untuk mengatasi masalah tersebut, maka dapat digunakan analisis diskriminan kuadratik karena matriks heterogenitas yang tak akan diabaikan. Pada penelitian ini digunakan data primer kejadian diabetes pada Puskesmas Kecamatan Pesanggrahan Kota Jakarta Selatan. Berdasarkan pada fungsi diskriminan linear, diperoleh seseorang dengan berat badan, trigliserida, dan tekanan darah sistolik yang tinggi akan cenderung diklasifikasikan ke dalam kelompok diabetes. Kemudian berdasarkan pada uji APER, diperoleh persentase misklasifikasi pada data *testing* dengan analisis diskriminan linear adalah sebesar 35.13% dan dengan analisis diskriminan kuadratik adalah sebesar 27.02%. Selain itu, berdasarkan pada uji keakuratan dan uji kestabilan, diperoleh bahwa dengan analisis diskriminan kuadratik lebih akurat dan lebih stabil dibandingkan dengan analisis diskriminan linear. Oleh karena itu, pengklasifikasian kejadian diabetes dengan analisis diskriminan kuadratik dinilai lebih optimal dibandingkan dengan analisis diskriminan linear.

Kata kunci : diabetes, heterogenitas, diskriminan linear, diskriminan kuadratik, ketepatan klasifikasi.

PERSEMBAHANKU...

”If you never try, you will never know”

”Just because you don’t see the road, it doesn’t mean the road doesn’t exist”

-Misaeng

Skripsi ini kupersembahkan untuk Bapak, Mama, Ardi, dan Bayu.

”Terima kasih atas dukungan, do’ā, serta kasih sayang kalian.”

KATA PENGANTAR

Puji syukur kepada Allah SWT atas pengetahuan dan kemampuan sehingga penulis dapat menyelesaikan skripsi yang berjudul ”Pengklasifikasian Kejadian Diabetes Dengan Analisis Diskriminan Kuadratik (Studi Kasus: Puskesmas Kecamatan Pesanggrahan Kota Jakarta Selatan)” yang merupakan salah satu syarat dalam memperoleh gelar Sarjana Program Studi Matematika Universitas Negeri Jakarta.

Skripsi ini berhasil diselesaikan tidak terlepas dari adanya bantuan dari berbagai pihak. Oleh karena itu, dalam kesempatan ini penulis ingin menyampaikan terima kasih terutama kepada

1. Allah SWT, yang alhamdulillah selalu mendengar doa penulis dan memberikan pelajaran berharga.
2. Bapak dan Mama yang selalu mendukung, memberi motivasi, dan setia membantu penulis dengan penuh cinta dan kasih sayang yang tulus.
3. Saudara laki-laki penulis, Ardi dan Bayu yang terus memberi semangat, mendoakan penulis, dan selalu menghibur ketika penulis mengalami kesulitan dalam penulisan skripsi ini.
4. Ibu Dra. Widyanti Rahayu, M. Si., selaku Dosen Pembimbing I dan Ibu Vera Maya Santi, M. Si., selaku Dosen Pembimbing II, yang telah me luangkan waktunya dalam memberikan bimbingan, saran, nasehat serta arahan sehingga skripsi ini dapat menjadi lebih baik dan terarah.
5. Ibu Dr. Lukita Ambarwati, S. Pd, M. Si., selaku Ketua Prodi Matematika FMIPA UNJ yang telah banyak membantu penulis.

6. Ibu Ratna Widyati, S. Si, M. Kom., selaku Pembimbing Akademik atas segala bimbingan dan kerja sama Ibu selama perkuliahan, dan seluruh Bapak/Ibu dosen atas pengajarannya yang telah diberikan, serta karyawan/karyawati FMIPA UNJ yang telah memberikan informasi yang penulis butuhkan dalam menyelesaikan skripsi.
7. ABDR (Ani, Apriliani, Arum, Dwi, Ratna, dan Rita), selaku teman masa SMA yang selalu memberikan *quote* akan kehidupan.
8. Cantik (Aan, Atika, Almira, Alphien, Dwi, Dewanti, Fatmah, Yohana, dan Zuhairini), selaku teman dekat masa perkuliahan yang selalu memberikan semangat dan seluruh mahasiswa matematika angkatan 2012 dan 2011, selaku teman dan kakak tingkat penulis yang selalu memberikan masukan ketika penulis menyusun skripsi ini.
9. Para dokter di Puskesmas Kecamatan Pesanggrahan (dr. Sonny, dr. Ellis, dsb), selaku dokter di Puskesmas Kecamatan Pesanggrahan yang telah memberikan kesempatan kepada penulis untuk melaksanakan penelitian.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Masukan dan kritikan akan sangat berarti. Semoga skripsi ini dapat bermanfaat bagi pembaca sekalian.

Jakarta, Juni 2017

Bety Suryani Putri

DAFTAR ISI

ABSTRACT	i
ABSTRAK	ii
KATA PENGANTAR	iv
DAFTAR ISI	vi
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Perumusan Masalah	3
1.3 Pembatasan Masalah	4
1.4 Tujuan Penulisan	4
1.5 Manfaat Penulisan	4
1.6 Metode Penelitian	5
II LANDASAN TEORI	6
2.1 Diabetes	6
2.2 Asumsi	11
2.3 Uji Perbedaan Signifikan Antar Kelompok	14
2.4 Analisis Diskriminan Linear	16
2.4.1 Aturan Pengklasifikasian Dengan Distribusi Normal Multivariat	17
2.4.2 Aturan Pengklasifikasian Dengan Metode ECM Minimum	19

2.4.3	Metode Pengklasifikasian Observasi Baru Dengan Nilai Tengah(<i>Cut-Off</i>)	23
2.5	Evaluasi Hasil Klasifikasi Dan Validasi	24
2.5.1	APER (<i>Apparent Error Rate</i>)	24
2.5.2	Uji Keakuratan	25
2.5.3	Uji Kestabilan	26
III PEMBAHASAN		28
3.1	Analisis Diskriminan Kuadratik	28
3.1.1	Aturan Pengklasifikasian Dengan Distribusi Normal Mul- tivariat	29
3.1.2	Aturan Pengklasifikasian Dengan Metode ECM Minimum	31
3.1.3	Metode Pengklasifikasian Observasi Baru	35
3.2	Metodologi Penelitian	35
3.2.1	Sumber Data	35
3.2.2	Variabel Penelitian	35
3.2.3	Diagram Alir Penelitian	37
3.3	Hasil Penelitian	40
3.3.1	Analisis Statistika Deskriptif	40
3.3.2	Asumsi	41
3.3.3	Uji Perbedaan Signifikan Antar Kelompok	43
3.3.4	Analisis Diskriminan Linear	44
3.3.5	Analisis Diskriminan Kuadratik	48
3.3.6	APER (<i>Apparent Error Rate</i>)	50
3.3.7	Uji Keakuratan	54
3.3.8	Uji Kestabilan	56

IV PENUTUP	58
4.1 Kesimpulan	58
4.2 Saran	60
DAFTAR PUSTAKA	62
LAMPIRAN-LAMPIRAN	64

DAFTAR TABEL

2.1	Biaya Akibat Kesalahan Klasifikasi	19
2.2	Tabel Klasifikasi	24
3.1	Biaya Akibat Kesalahan Klasifikasi	31
3.2	Statistika Deskriptif	40
3.3	Nilai VIF	43
3.4	Nilai F-Hitung Dan P-Value	44
3.5	Tabel Observasi Yang Diprediksi Salah Diklasifikasikan	47
3.6	Tabel Observasi Yang Diprediksi Salah Diklasifikasikan	50
3.7	Tabel Klasifikasi Data <i>Training</i> Dengan Menggunakan Analisis Diskriminan Linear	51
3.8	Tabel Klasifikasi Data <i>Training</i> Dengan Menggunakan Analisis Diskriminan Kuadratik	51
3.9	Tabel Klasifikasi Data <i>Testing</i> Dengan Menggunakan Analisis Diskriminan Linear	52
3.10	Tabel Klasifikasi Data <i>Testing</i> Dengan Menggunakan Analisis Diskriminan Kuadratik	53

DAFTAR GAMBAR

2.1	Grafik Penderita Diabetes Di Dunia	8
2.2	Grafik Normalitas Multivariat	12
2.3	Daerah Klasifikasi Untuk Dua Kelompok	18
2.4	Peluang Bersyarat	18
3.1	Daerah Klasifikasi Untuk Dua Kelompok	29
3.2	Peluang Bersyarat	30
3.3	Diagram Alir Penelitian	39
3.4	Grafik Normalitas Multivariat	41

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

IDF (*International Diabetes Federation*) menyebutkan bahwa diabetes merupakan salah satu penyakit yang paling sering dialami oleh masyarakat dunia pada era globalisasi. Pada tahun 2015, Indonesia menempati urutan ke-4 terbesar dalam jumlah penderita diabetes di dunia setelah India, China, dan Amerika Serikat. IDF (*International Diabetes Federation*) juga menyebutkan bahwa jumlah penderita diabetes di Indonesia adalah sekitar 9.1 juta penduduk dan diperkirakan jumlah tersebut akan meningkat menjadi 21.3 juta penduduk pada tahun 2030. Kementerian Kesehatan (Kemenkes) juga menyebutkan bahwa terdapat 14.7 % kejadian diabetes di perkotaan, 7.2% di pedesaan dan DKI Jakarta merupakan yang tertinggi se-Indonesia dengan 2.6 % diatas angka nasional sebesar 1.1 % (Sindonews, 2015).

Diabetes merupakan suatu penyakit kronik di mana kadar glukosa (gula sederhana) di dalam darah menjadi tinggi karena tubuh tidak dapat memproduksi atau mempergunakan insulin secara semestinya. Diabetes dapat disebabkan oleh banyak hal, seperti faktor keturunan, pola hidup yang tak sehat, jarang berolahraga, kelebihan berat badan (obesitas), dsb. Meskipun faktor keturunan memberikan pengaruh yang besar, namun ada faktor-faktor lain yang mempengaruhi seseorang mengidap penyakit ini. Diabetes juga dapat menimbulkan komplikasi penyakit, seperti kebutaan, gagal ginjal, penyakit jantung, dsb.

Penderita diabetes memang didominasi oleh masyarakat pada usia lanjut. Masyarakat pada usia lanjut diberikan penyuluhan yang bertujuan agar penyakit diabetes tidak menyebabkan komplikasi terhadap organ tubuh lainnya, selain kadar gula darah yang semakin meningkat. Seseorang jika memiliki kadar gula darah dimana sedikit melebihi batas normal belum dapat dikatakan sebagai penderita diabetes. Keadaan seperti itu masih dapat dikatakan sebagai penderita pradiabetes. Penyebab kadar gula darah yang semakin meningkat dapat disebabkan oleh banyak hal, seperti makanan tinggi lemak, pemanis buatan, dsb.

Klasifikasi antara penderita pradiabetes dan diabetes menjadi sangat dibutuhkan karena dapat digunakan sebagai pencegahan dini sebelum terjadinya komplikasi dan dapat diperoleh pula variabel-variabel signifikan yang menjadi pengaruh perubahan kadar gula darah seseorang. Ada banyak metode statistika yang mana dapat mengkaji klasifikasi kejadian diabetes, seperti analisis komponen utama, analisis regresi berganda, dan analisis diskriminan. Pengklasifikasian dapat dilakukan dengan metode diskriminan dengan alasan kejadian diabetes yang diamati terdiri dari penderita pradiabetes dan diabetes. Selain itu, didukung pula dengan alasan bahwa variabel terikat (y) berupa variabel berskala kualitatif/nonmetrik dan variabel bebas (x) berskala kuantitatif/metrik.

Analisis diskriminan merupakan salah satu metode pada analisis multivariat yang bertujuan untuk mengevaluasi ketepatan klasifikasi dan juga dapat menentukan variabel bebas yang memberikan sumbangannya terbesar terhadap terbentuknya pengklasifikasian. Analisis diskriminan linear merupakan salah satu jenis analisis diskriminan yang berfokus pada pemisahan observasi (pengamatan) dimana setiap populasi berdistribusi normal multivariat dan mempunyai kesamaan matriks varian kovarian (homogenitas). Namun pada

kasus-kasus tertentu seringkali ditemukan matriks varian kovarian yang tidak sama (heterogenitas). Untuk mengatasi permasalahan tersebut maka dapat digunakan pendekatan dengan analisis diskriminan kuadratik disebabkan jika masih menggunakan analisis diskriminan linear, maka hasil klasifikasi yang terbentuk menjadi tidak optimal.

Beberapa penelitian sebelumnya yang berkaitan dengan diabetes dan analisis diskriminan kuadratik, seperti Aswa (2015) yang telah melakukan penelitian mengenai analisis diskriminan kernel dengan metode *cross validation* (studi kasus: faktor-faktor yang berhubungan dengan kejadian hipertensi pada Puskesmas Usuku Wakatobi Sulawesi Tenggara pada tahun 2013), Gradianta (2014) yang telah melakukan penelitian mengenai klasifikasi pasien penderita diabetes mellitus tipe II menggunakan metode *hybrid genetic algorithm - discriminant analysis*, dan juga Vita (2013) yang telah melakukan penelitian mengenai aplikasi analisis regresi komponen utama terhadap faktor-faktor yang mempengaruhi penyakit diabetes mellitus (studi kasus di Puskesmas Tempeh Kab. Lumajang). Berdasarkan uraian di atas, akan dikaji suatu klasifikasi kejadian diabetes dengan metode analisis diskriminan kuadratik.

1.2 Perumusan Masalah

Perumusan masalah yang akan dikaji dalam skripsi ini adalah sebagai berikut.

1. Bagaimana kontribusi setiap variabel bebas dalam membedakan klasifikasi kejadian diabetes?
2. Bagaimana ketepatan klasifikasi pada kasus kejadian diabetes dengan metode analisis diskriminan kuadratik ?

1.3 Pembatasan Masalah

Pembatasan masalah yang akan digunakan dalam skripsi ini adalah sebagai berikut.

1. Klasifikasi kejadian diabetes yang akan dikaji dalam skripsi ini adalah penderita pradiabetes dan penderita diabetes.
2. Variabel bebas yang digunakan dalam skripsi ini, diantaranya usia, berat badan, kolesterol, trigliserida, tekanan darah sistolik, tekanan darah diastolik, dan asam urat.
3. Asumsi lain yang digunakan adalah nilai peluang dan biaya akibat kesalahan klasifikasi tidak diketahui.

1.4 Tujuan Penulisan

Tujuan yang ingin dicapai dalam skripsi ini adalah sebagai berikut.

1. Mengkaji metode statistika pada pengklasifikasian kejadian diabetes.
2. Mengidentifikasi kontribusi setiap variabel bebas dalam membedakan klasifikasi kejadian diabetes dengan analisis diskriminan linear.
3. Mengidentifikasi ketepatan klasifikasi dengan metode analisis diskriminan kuadratik.

1.5 Manfaat Penulisan

Manfaat yang akan diperoleh dari skripsi ini adalah sebagai berikut.

1. Untuk mahasiswa, sebagai referensi tambahan untuk mengetahui ketepatan klasifikasi metode analisis diskriminan kuadratik.

2. Untuk universitas, sebagai tambahan referensi/karya tulis demi melengkapi studi kepustakaan pada universitas ini.
3. Untuk lembaga sumber data yang diamati, sebagai tambahan informasi mengenai metode yang dapat digunakan pada studi kasus ini dan studi kasus lainnya.

1.6 Metode Penelitian

Skripsi ini merupakan aplikasi dari analisis diskriminan kuadratik yang diterapkan pada studi kasus yang terdapat di lingkungan kehidupan sehari-hari sekitar penulis, yaitu Puskesmas Kecamatan Pesanggrahan Kota Jakarta Selatan. Referensi utama yang digunakan adalah Johnson dan Wichern (2008).

BAB II

LANDASAN TEORI

Pada bab ini akan dibahas mengenai definisi diabetes, asumsi yang akan digunakan pada penelitian ini, uji perbedaan signifikan antar kelompok pada klasifikasi kejadian diabetes, definisi analisis diskriminan linear, dan apa saja evaluasi dan validasi yang akan mengkaji hasil klasifikasi.

2.1 Diabetes

Diabetes atau penyakit kencing manis adalah suatu penyakit kronis yang terjadi ketika pankreas tidak dapat menghasilkan insulin atau mempergunakan insulin secara semestinya. Hormon insulin berguna untuk menyeimbangkan kadar gula dalam darah. Jika jumlah insulin tidak mencukupi, maka terjadi penimbunan gula dalam darah sehingga menyebabkan diabetes.

Klasifikasi pada penderita diabetes terdiri atas tiga tipe, diantaranya

1. Diabetes Tipe I

Pada tipe ini berkaitan dengan ketidaksanggupan pankreas dalam menghasilkan insulin. Penderita diabetes tipe I sebagian besar terjadi pada orang di bawah umur 30 tahun.

2. Diabetes Tipe II

Berbeda dengan diabetes tipe I, pada tipe II masalahnya terdapat pada insulin yang dihasilkan tidak mencukupi. Diabetes tipe II merupakan jenis diabetes yang sebagian besar diderita oleh sebagian besar penduduk. Sekitar 90% hingga 95% penderita diabetes menderita diabetes tipe II.

Jenis diabetes ini paling sering diderita oleh orang dewasa yang berusia lebih dari 30 tahun dan cenderung semakin parah secara bertahap.

3. Diabetes Gestasional

Diabetes Mellitus yang muncul pada masa kehamilan umumnya bersifat sementara, tetapi merupakan faktor risiko untuk diabetes tipe II. Sekitar 4-5% wanita hamil diketahui menderita diabetes gestasional dan umumnya terdeteksi pada atau setelah trimester kedua (Ditjen Bina Farmasi dan ALKES, 2005).

IDF (*International Diabetes Federation*) menyebutkan bahwa jumlah penderita diabetes di Indonesia menempati urutan ke-7 terbesar di dunia dengan jumlah mencapai 8.5 juta penduduk pada tahun 2013. Penjelasan lebih lengkap akan ditunjukkan pada Gambar 2.1. Selanjutnya, Perkumpulan Endokrinologi (Perkeni) menyatakan bahwa jumlah penderita diabetes di Indonesia telah meningkat menjadi 9.1 juta penduduk pada tahun 2015 dan membuat Indonesia menempati urutan ke-4 terbesar di dunia. WHO (*World Health Organization*) juga memperkirakan bahwa jumlah penderita diabetes di Indonesia akan meningkat menjadi 21.3 juta penduduk pada tahun 2030.

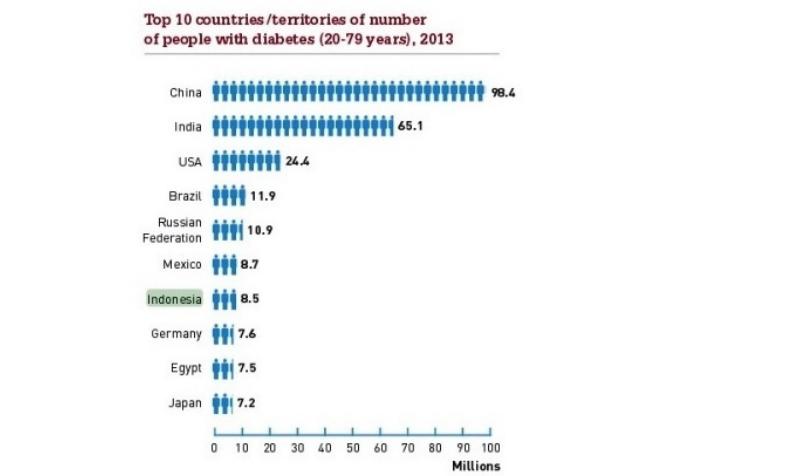
Terdapat beberapa faktor penyebab terjadinya diabetes, diantaranya

- Faktor Keturunan

Bagi orang tua yang memiliki riwayat penderita diabetes, kemungkinan besar terdapat resiko kejadian diabetes pada keturunannya. Tidak hanya dilihat dari sudut pandang orang tua, saudara kandung yang memiliki riwayat menderita diabetes juga patut diwaspada karena kemungkinan besar saudara lainnya juga terdapat resiko kejadian diabetes.

- Berat Badan

Mengkonsumsi makanan secara berlebih dengan tinggi lemak dan gula



Sumber: <http://sehat.link/data-prevalensi-penderita-diabetes-di-indonesia.info>

Gambar 2.1: Grafik Penderita Diabetes Di Dunia

merupakan salah satu gaya hidup yang tidak sehat yang mana dapat menyebabkan seseorang menderita berat badan yang berlebih (kegemukan). Salah satu penentu seseorang menderita kegemukan adalah berdasarkan nilai Indeks Massa Tubuh (IMT), dengan nilai IMT > 25. Kegemukan dapat menjadi salah satu faktor penyebab terjadinya diabetes pada seseorang karena hubungan antara lemak dalam tubuh dan gula dalam darah.

- Usia

Diabetes tidak hanya dialami oleh orang dewasa dan manula, namun juga remaja dan bahkan anak-anak. Prof. Dr. Achmad Rudijanto menyatakan bahwa usia penderita diabetes kini semakin muda. Pada tahun 2015, 1 dari 5 penderita diabetes masih berusia di bawah 40 tahun dengan jumlah sebanyak 1.6 juta penduduk. Sedangkan, pada rentang usia 40-59 tahun terdapat penderita diabetes dengan jumlah sebanyak 4.6 juta penduduk. Kelompok terakhir terdiri dari penderita diabetes berusia

60-79 tahun dengan jumlah sebanyak 2 juta penduduk. Kelompok usia dewasa tua (> 45 tahun) perlu waspada karena pada rentang usia tersebut banyak organ-organ vital melemah dan tubuh mulai tidak peka terhadap insulin.

- Kurangnya Aktivitas Fisik

Salah satu dampak dari kurangnya aktifitas fisik pada seseorang adalah dapat mengalami kegemukan dan dapat melemahkan kerja organ-organ vital seperti jantung, liver, ginjal, dan juga pankreas.

- Tingkat Kolesterol

Mengkonsumsi makanan dengan tingkat kolesterol yang tinggi dapat meningkatkan kadar kolesterol dalam tubuh. Kadar kolesterol yang tinggi dalam tubuh dapat memberikan kontribusi yang cukup besar terhadap mudahnya seseorang menderita diabetes.

- Tekanan darah

Konsumsi garam yang berlebih dapat meningkatkan tekanan darah pada seseorang yang pada akhirnya berperan dalam meningkatkan resiko seseorang menderita diabetes.

- Asam Urat

Salah satu penelitian dari Harvard Medical School menyebutkan bahwa asam urat dapat dikaitkan dengan 70% peningkatan risiko diabetes tipe II. Bertambahnya kadar asam urat setiap 1 mg / dL dapat dikaitkan dengan 18% peningkatan risiko diabetes yang signifikan. Hubungan tersebut tetap signifikan bahkan setelah disesuaikan untuk kadar glukosa dan insulin puasa.

- Kadar Gula Dalam Darah

Kadar gula darah pada seseorang dapat dikatakan normal jika kadar Gula Darah Puasa (GDP), Gula Darah 2 jam *Post Prandial* (G2PP), dan Gula Darah Sewaktu (GDS) berada pada rentang kadar semestinya. Dalam ilmu medis, kadar gula darah yang tinggi disebut hiperglikemia. Salah satu faktor penyebab meningkatnya kadar gula darah pada seseorang adalah mengkonsumsi makanan/minuman tinggi gula secara berlebihan dan terdapat gangguan pada hormon insulin.

1. Gula Darah Puasa (GDP)

Kadar glukosa darah dalam keadaan puasa pagi hari pada seseorang dapat dikatakan normal jika berada pada rentang 70-100 mg/dL. Seseorang dapat dikatakan terkena Glukosa Puasa Terganggu jika kadar GDP berada pada rentang 100-126 mg/dL dan seseorang dapat didiagnosa menderita diabetes jika kadar GDP > 126 mg/dL.

2. Gula Darah 2 jam *Post Prandial* (G2PP)

Kadar glukosa darah 2 jam sesudah makan pada seseorang dapat dikatakan normal jika memiliki kadar < 140 mg/dL. Seseorang dapat dikatakan terkena Toleransi Glukosa Terganggu jika kadar G2PP berada pada rentang 140-200 mg/dL dan seseorang dapat didiagnosa menderita diabetes jika kadar G2PP > 200 mg/dL.

3. Glukosa Darah Sewaktu (GDS)

Kadar glukosa darah sewaktu sesaat diperiksa pada seseorang dapat dikatakan normal jika memiliki kadar ≤ 180 mg/dL dan seseorang dapat didiagnosa menderita diabetes jika kadar GDS > 180 mg/dL.

Berdasarkan pada penjelasan sebelumnya, seseorang dapat diklasifikasikan ke dalam kelompok pradiabetes atau diabetes jika memenuhi aturan

berikut ini, diantaranya

1. Jika seseorang memiliki kadar GDP sekitar 100-126 mg/dL dan G2PP sekitar 140-200 mg/dL, maka seseorang tersebut dapat diklasifikasikan ke dalam kelompok pradiabetes.
2. Jika seseorang memiliki kadar GDP sekitar > 126 mg/dL dan G2PP sekitar > 200 mg/dL, maka seseorang tersebut dapat diklasifikasi ke dalam kelompok diabetes.

2.2 Asumsi

Terdapat tiga asumsi yang harus dipenuhi pada analisis diskriminan, diantaranya

- Uji Normalitas

Salah satu metode dalam uji normalitas adalah dengan menghitung jarak kuadrat/jarak *Mahalanobis* untuk setiap observasi/ pengamatan (Johnson & Wichern, 2008). Hipotesis yang akan digunakan adalah sebagai berikut.

$$H_0 : \text{Data berdistribusi normal multivariat}$$

$$H_a : \text{Data tidak berdistribusi normal multivariat}$$

Uji normalitas ini dilakukan dengan menggunakan plot jarak *Mahalanobis* (D_i^2) dan distribusi khi-kuadrat ($\chi_h^2(\frac{i-1}{n})$). Langkah-langkah pada pengujian adalah sebagai berikut.

1. Menghitung jarak kuadrat/jarak *Mahalanobis* dengan formula

$$D_i^2 = (x_i - \bar{x})' \mathbf{S}^{-1} (x_i - \bar{x})$$

dengan $i = 1, 2, \dots, n$; dimana

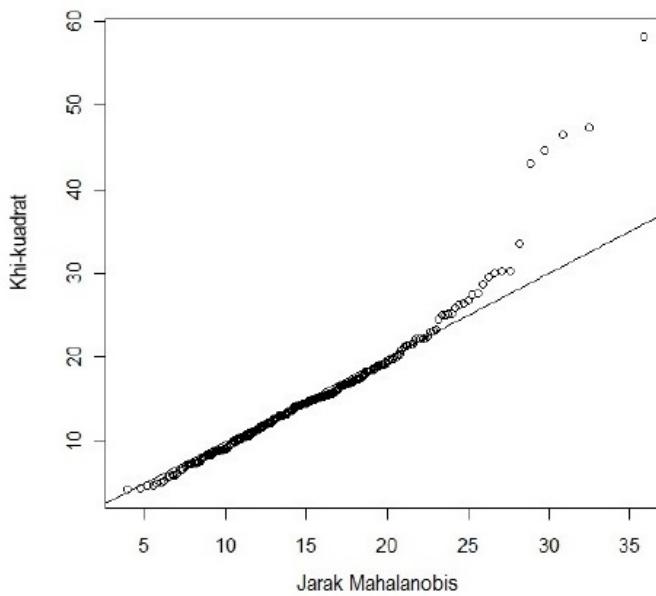
x_i = Observasi/pengamatan ke-i

\mathbf{S}^{-1} = Invers matriks varian kovarian \mathbf{S}

Masing-masing D_i^2 akan mengikuti sebaran khi-kuadrat $\chi_h^2(\frac{i-1}{n})$.

$\chi_h^2(\frac{i-1}{n})$ adalah distribusi khi-kuadrat dengan h adalah banyaknya variabel bebas.

2. Mengurutkan nilai D_i^2 dari nilai D_i^2 terkecil sampai nilai D_i^2 terbesar atau $D_{(1)}^2 \leq D_{(2)}^2 \leq \dots \leq D_{(n)}^2$. Plot khi-kuadrat akan memeriksa apakah D_i^2 mengikuti sebaran khi-kuadrat atau tidak dengan mengurutkan nilai D_i^2 .
3. Membuat plot dengan titik koordinat $(D_i^2, \chi_h^2(\frac{i-1}{n}))$. Hasil dari plot tersebut dapat diilustrasikan pada Gambar 2.2.



Gambar 2.2: Grafik Normalitas Multivariat

Berdasarkan pada Gambar 2.2, titik-titik menyebar membentuk garis lurus dan akan terima H_0 jika lebih dari 50% nilai $D_i^2 \leq \chi_h^2(\frac{i-1}{n})$.

- Uji Kesamaan Matriks Varian Kovarian (Homogenitas)

Pengujian kesamaan matriks varian kovarian antar kelompok akan menggunakan statistik uji *Box's M*. Apabila terdapat dua kelompok, maka hipotesis yang akan digunakan adalah sebagai berikut.

$$H_0 : \Sigma_1 = \Sigma_2$$

$$H_a : \Sigma_1 \neq \Sigma_2$$

Jika,

$$u = [\sum_l \frac{1}{(n_l - 1)} - \frac{1}{\sum_l (n_l - 1)}] \left[\frac{2h^2 + 3h - 1}{6(h+1)(g-1)} \right] \text{ dengan } l = 1, 2, \dots, g$$

$$M = (\sum_l (n_l - 1)) \ln |\mathbf{S}_{\text{gabungan}}| - \sum_l ((n_l - 1) \ln |\mathbf{S}_l|)$$

dimana h adalah jumlah variabel bebas dan g adalah jumlah kelompok, maka

$$c = (1 - u)M$$

$$= (1 - u)((\sum_l (n_l - 1)) \ln |\mathbf{S}_{\text{gabungan}}| - \sum_l ((n_l - 1) \ln |\mathbf{S}_l|))$$

memiliki distribusi χ^2 dengan derajat bebas adalah

$$v = g \frac{1}{2} h(h+1) - \frac{1}{2} h(h+1) = \frac{1}{2} h(h+1)(g-1)$$

Dengan tingkat signifikansi α , maka akan tolak H_0 jika nilai $c > \chi_v^2$.

- Uji Multikolinearitas

Pengujian ini memiliki tujuan untuk menemukan adanya korelasi antar variabel bebas atau tidak. Salah satu asumsi yang harus dipenuhi pada analisis diskriminan adalah tidak terjadi multikolinearitas yang mana berarti tidak ditemukan adanya korelasi di antara variabel bebas. Salah satu cara untuk menguji ada atau tidaknya multikolinearitas adalah

dengan mengitung nilai VIF (*Variance Inflation Factor*) dari variabel bebas terhadap variabel terikat. Hipotesis yang akan digunakan adalah sebagai berikut.

H_0 : Tidak terjadi multikolinearitas antar variabel bebas

H_a : Terjadi multikolinearitas antar variabel bebas

Jika nilai korelasi antar variabel bebas j dan g didefinisikan dengan

$$r_{jq} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{iq} - \bar{x}_q)}{\sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2} \sqrt{\sum_{i=1}^n (x_{iq} - \bar{x}_q)^2}}$$

dimana $j = 1, 2, \dots, h$ dan $q = 1, 2, \dots, h$; maka

$$VIF = \frac{1}{1 - r_{jq}}$$

Jika nilai VIF telah diperoleh, maka dapat diambil keputusan dengan rincian sebagai berikut.

1. Jika nilai $VIF \geq 10$, maka akan tolak H_0 .
2. Jika nilai $VIF < 10$, maka akan terima H_0 .

2.3 Uji Perbedaan Signifikan Antar Kelompok

Tahap ini dilakukan untuk melihat apakah terdapat perbedaan yang nyata antar kelompok yang terbentuk dari setiap variabel bebas. Jika nilai rata-rata antar kelompok tersebut berbeda maka analisis diskriminan dapat dilakukan, dengan syarat $\geq 50\%$ variabel bebas yang signifikan. Apabila hanya terdapat dua kelompok, maka hipotesis yang akan digunakan adalah sebagai berikut.

$$H_0 : \mu_1 = \mu_2$$

$$H_a : \mu_1 \neq \mu_2$$

dimana μ_1 dan μ_2 merupakan rata-rata populasi pada kelompok 1 dan kelompok 2. T^2 Hotelling dapat didefinisikan dengan sebagai berikut.

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{x}_1 - \bar{x}_2)' \mathbf{S}_{gabungan}^{-1} (\bar{x}_1 - \bar{x}_2)$$

dimana n_1 dan n_2 adalah jumlah observasi pada kelompok 1 dan 2. Dengan statistik uji,

$$F_{hitung} = \left(\frac{n_1 + n_2 - h - 1}{(n_1 + n_2 - 2)h} \right) T^2$$

dimana h adalah jumlah variabel bebas dan

$$\begin{aligned} \mathbf{S}_{gabungan} &= \left[\frac{(n_1 - 1)}{(n_1 - 1) + (n_2 - 1)} \right] \mathbf{S}_1 + \left[\frac{(n_2 - 1)}{(n_1 - 1) + (n_2 - 1)} \right] \mathbf{S}_2 \\ &= \frac{(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2}{n_1 + n_2 - 2} \end{aligned}$$

dengan

$$\begin{aligned} \mathbf{S}_1 &= \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)(X_{1i} - \bar{X}_1)' \\ \mathbf{S}_2 &= \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (X_{2i} - \bar{X}_2)(X_{2i} - \bar{X}_2)' \end{aligned}$$

Dengan demikian, akan tolak H_0 apabila nilai $F_{hitung} > F_{\alpha, (v_1, v_2)}$ dengan derajat bebas yang digunakan adalah

$$v_1 = n_1 + n_2 - h - 1$$

$$v_2 = h$$

dan dapat diperoleh keputusan bahwa pembeda kelompok telah signifikan secara statistik.

2.4 Analisis Diskriminan Linear

Analisis diskriminan adalah salah satu teknik multivariat yang bertujuan untuk memisahkan objek-objek ke dalam kelompok yang berbeda, mengidentifikasi kontribusi h variabel bebas terhadap pengklasifikasian / pembentukan / pemisahan kelompok, mengklasifikasikan objek baru ke dalam suatu kelompok, dan mengevaluasi ketepatan klasifikasi. Analisis diskriminan terdiri dari dua variabel, yaitu variabel terikat dan variabel bebas. Variabel terikat pada analisis diskriminan berupa data nonmetrik (nominal atau ordinal) dan variabel bebas pada analisis diskriminan berupa data metrik (interval atau rasio). Analisis diskriminan dapat digunakan untuk variabel terikat yang terdiri dari dua kelompok atau lebih.

Ada beberapa jenis analisis diskriminan, diantaranya

1. Analisis Diskriminan Linear

Analisis diskriminan linear digunakan jika data h buah variabel bebas berdistribusi normal multivariat dan antar kelompoknya memiliki matriks varian kovarian yang sama.

2. Analisis Diskriminan Kuadratik

Analisis diskriminan kuadratik digunakan jika data h buah variabel bebas berdistribusi normal multivariat tetapi matriks varian kovarian tidak sama antar kelompoknya.

3. Analisis Diskriminan Fisher

Analisis diskriminan fisher digunakan jika data h buah variabel bebas tidak berdistribusi normal multivariat tetapi antar kelompoknya memiliki matriks varian kovarian yang sama.

4. Analisis Diskriminan Nonparametrik

Analisis diskriminan nonparametrik digunakan jika data h buah variabel bebas tidak berdistribusi normal multivariat dan matriks varian kovarian tidak sama antar kelompoknya.

Khusus dalam penelitian ini yang akan dibahas adalah mengenai analisis diskriminan linear dan analisis diskriminan kuadratik.

Fungsi diskriminan linear adalah suatu kombinasi linear variabel-variabel bebas yang memberikan nilai sejauh mungkin antar kelompok dan nilai sedekat mungkin dalam satu kelompok. Berikut tahap-tahap yang harus dilalui dalam penggunaan analisis diskriminan linear untuk dua kelompok.

2.4.1 Aturan Pengklasifikasian Dengan Distribusi Normal Multivariat

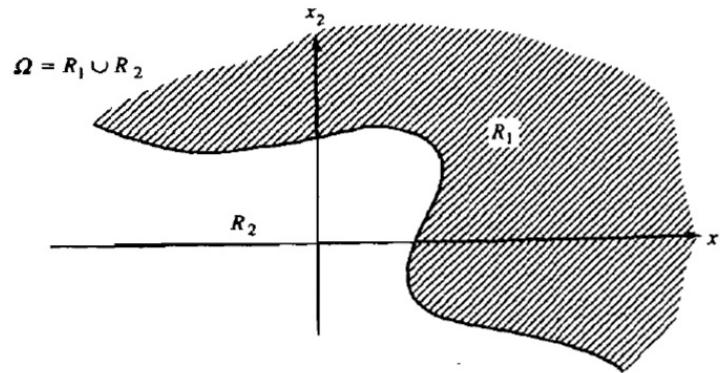
Misalkan $f_1(x)$ dan $f_2(x)$ adalah fungsi kepadatan peluang untuk populasi kelompok 1 (ω_1) dan kelompok 2 (ω_2) dimana

$$f_l(x) = \frac{1}{(2\pi)^{\frac{h}{2}} |\Sigma_l|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x - \mu_l)' \Sigma_l^{-1} (x - \mu_l)\right] \quad (2.1)$$

untuk $l = 1, 2$. Suatu observasi harus termasuk ke dalam salah satu di antara populasi ω_1 dan ω_2 . Misalkan Ω adalah ruang sampel jumlah observasi populasi ω_1 dan ω_2 . Jika R_1 merupakan daerah semua observasi pada kelompok 1, maka $R_2 = \Omega - R_1$ merupakan daerah semua observasi pada kelompok 2. Ilustrasi akan ditampilkan pada Gambar 2.3.

Peluang bersyarat $P(2|1)$ yang mana merupakan peluang suatu observasi yang berasal dari kelompok 1 namun diprediksi sebagai kelompok 2 dan dapat didefinisikan dengan

$$P(2|1) = P(\mathbf{X} \in R_2 | \omega_1) = \int_{R_2 = \Omega - R_1} f_1(x) dx$$

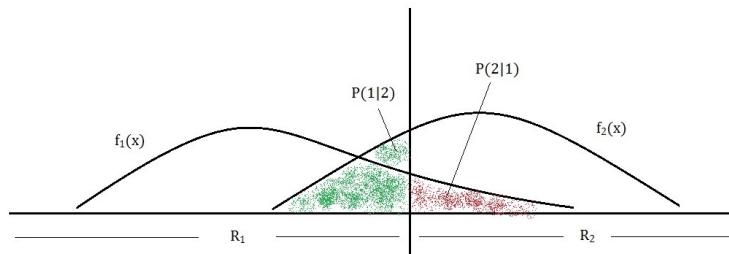


Gambar 2.3: Daerah Klasifikasi Untuk Dua Kelompok

dan peluang bersyarat $P(1|2)$ yang mana merupakan peluang suatu observasi yang berasal dari kelompok 2 namun diprediksi sebagai kelompok 1 dan dapat didefinisikan pula dengan

$$P(1|2) = P(\mathbf{X} \in R_1 | \omega_2) = \int_{R_1} f_2(x) dx$$

Peluang bersyarat $P(2|1)$ dan $P(1|2)$ dapat diilustrasikan pada Gambar 2.4.



Gambar 2.4: Peluang Bersyarat

Jika p_1 adalah peluang prior dari populasi ω_1 dan p_2 adalah peluang prior dari populasi ω_2 , maka peluang keseluruhan klasifikasi observasi secara tepat atau tidak tepat dapat diperoleh dengan perkalian peluang prior dengan peluang klasifikasi bersyarat.

- $P(\text{Observasi tepat di } \omega_1) = P(\mathbf{X} \in R_1|\omega_1)P(\omega_1) = P(1|1)p_1$
- $P(\text{Observasi tidak tepat di } \omega_1) = P(\mathbf{X} \in R_1|\omega_2)P(\omega_2) = P(1|2)p_2$
- $P(\text{Observasi tepat di } \omega_2) = P(\mathbf{X} \in R_2|\omega_2)P(\omega_2) = P(2|2)p_2$
- $P(\text{Observasi tidak tepat di } \omega_2) = P(\mathbf{X} \in R_2|\omega_1)P(\omega_1) = P(2|1)p_1$

2.4.2 Aturan Pengklasifikasian Dengan Metode ECM Minimum

Selain peluang kesalahan klasifikasi, ada hal lainnya yang sering diperhitungkan dalam mengklasifikasikan suatu observasi. Hal lain tersebut merupakan biaya akibat kesalahan klasifikasi. Biaya akibat kesalahan klasifikasi suatu observasi dapat didefinisikan pada tabel dibawah ini.

Tabel 2.1: Biaya Akibat Kesalahan Klasifikasi

	Kelompok Prediksi 1	Kelompok Prediksi 2
Kelompok Aktual 1	0	$c(2 1)$
Kelompok Aktual 2	$c(1 2)$	0

Keterangan:

$c(1|2)$ = Biaya kesalahan klasifikasi ketika sebuah observasi dari kelompok 2 namun diprediksi sebagai kelompok 1

$c(2|1)$ = Biaya kesalahan klasifikasi ketika sebuah observasi dari kelompok 1 namun diprediksi sebagai kelompok 2

Selain itu, dapat dihitung pula biaya dugaan akibat kesalahan klasifikasi atau *Expected Cost of Misclassification* (ECM) dengan formula

$$ECM = c(2|1)P(2|1)p_1 + c(1|2)P(1|2)p_2$$

Klasifikasi akan dikatakan tepat jika memiliki nilai ECM yang kecil atau minimum. ECM akan bernilai minimum apabila memenuhi aturan berikut.

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq \left(\frac{c(1|2)}{c(2|1)} \right) \left(\frac{p_2}{p_1} \right) \quad (2.2)$$

$$R_2 : \frac{f_1(x)}{f_2(x)} < \left(\frac{c(1|2)}{c(2|1)} \right) \left(\frac{p_2}{p_1} \right) \quad (2.3)$$

Selain itu, terdapat beberapa kasus khusus pada aturan menentukan daerah nilai ECM, diantaranya

- Jika $p_2/p_1 = 1$, maka

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq \left(\frac{c(1|2)}{c(2|1)} \right) \quad R_2 : \frac{f_1(x)}{f_2(x)} < \left(\frac{c(1|2)}{c(2|1)} \right) \quad (2.4)$$

- Jika $c(1|2)/c(2|1) = 1$, maka

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq \left(\frac{p_2}{p_1} \right) \quad R_2 : \frac{f_1(x)}{f_2(x)} < \left(\frac{p_2}{p_1} \right) \quad (2.5)$$

- Jika $p_2/p_1 = \frac{1}{c(1|2)/c(2|1)}$, maka

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq 1 \quad R_2 : \frac{f_1(x)}{f_2(x)} < 1 \quad (2.6)$$

Jika x_0 merupakan observasi baru yang ingin diklasifikasikan sedangkan nilai peluang dan biaya akibat kesalahan tidak diketahui maka dapat digunakan kasus khusus pada Persamaan (2.6). Selanjutnya, x_0 dapat diklasifikasikan ke dalam populasi ω_1 jika

$$\frac{f_1(x)}{f_2(x)} \geq 1 \text{ atau } f_1(x) \geq f_2(x)$$

atau x_0 dapat diklasifikasikan ke dalam populasi ω_2 jika

$$\frac{f_1(x)}{f_2(x)} < 1 \text{ atau } f_1(x) < f_2(x)$$

Selanjutnya terdapat dua kasus khusus pada pengklasifikasian dua kelompok, diantaranya

- **Nilai μ_1, μ_2 , dan $\Sigma_1 = \Sigma_2 = \Sigma$ diketahui**

Jika fungsi kepadatan peluang dari x adalah

$$f_l(x) = \frac{1}{(2\pi)^{\frac{h}{2}} |\Sigma_l|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x - \mu_l)' \Sigma_l^{-1} (x - \mu_l)\right] \quad (2.7)$$

untuk $l = 1, 2$; maka

$$\begin{aligned} \frac{f_1(x)}{f_2(x)} &= \frac{\frac{1}{(2\pi)^{\frac{h}{2}} |\Sigma_1|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x - \mu_1)' \Sigma_1^{-1} (x - \mu_1)\right]}{\frac{1}{(2\pi)^{\frac{h}{2}} |\Sigma_2|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x - \mu_2)' \Sigma_2^{-1} (x - \mu_2)\right]} \\ &= \exp\left[-\frac{1}{2}(x - \mu_1)' \Sigma_1^{-1} (x - \mu_1) + \frac{1}{2}(x - \mu_2)' \Sigma_2^{-1} (x - \mu_2)\right] \end{aligned} \quad (2.8)$$

Karena $\Sigma_1 = \Sigma_2 = \Sigma$, maka Persamaan (2.8) akan disederhanakan menjadi

$$\begin{aligned} &= \ln(\exp\left[-\frac{1}{2}(x - \mu_1)' \Sigma^{-1} (x - \mu_1) + \frac{1}{2}(x - \mu_2)' \Sigma^{-1} (x - \mu_2)\right]) \\ &= -\frac{1}{2}(x - \mu_1)' \Sigma^{-1} (x - \mu_1) + \frac{1}{2}(x - \mu_2)' \Sigma^{-1} (x - \mu_2) \\ &= -\frac{1}{2}[(x' \Sigma^{-1} - \mu_1' \Sigma^{-1})(x - \mu_1) - (x' \Sigma^{-1} - \mu_2' \Sigma^{-1})(x - \mu_2)] \\ &= -\frac{1}{2}[x' \Sigma^{-1} x - \mu_1' \Sigma^{-1} x - x' \Sigma^{-1} \mu_1 + \mu_1' \Sigma^{-1} \mu_1 - x' \Sigma^{-1} x + x' \Sigma^{-1} \mu_2 + \mu_2' \Sigma^{-1} x - \mu_2' \Sigma^{-1} \mu_2] \\ &= \frac{1}{2}[2\mu_1' \Sigma^{-1} x - 2x' \Sigma^{-1} \mu_2 - (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 + \mu_2)] \\ &= (\mu_1 - \mu_2)' \Sigma^{-1} x - \frac{1}{2}(\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 + \mu_2) \end{aligned}$$

Selanjutnya, daerah yang memenuhi aturan nilai ECM minimum berdasarkan pada Persamaan (2.2) dan (2.3) diantaranya

$$R_1 : (\mu_1 - \mu_2)' \Sigma^{-1} x - \frac{1}{2}(\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 + \mu_2) \geq \left(\frac{c(1|2)}{c(2|1)}\right) \left(\frac{p_2}{p_1}\right)$$

$$R_2 : (\mu_1 - \mu_2)' \Sigma^{-1} x - \frac{1}{2}(\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 + \mu_2) < \left(\frac{c(1|2)}{c(2|1)} \right) \left(\frac{p_2}{p_1} \right)$$

Sehingga, x_0 dapat diklasifikasikan ke dalam populasi ω_1 jika

$$(\mu_1 - \mu_2)' \Sigma^{-1} x - \frac{1}{2}(\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 + \mu_2) \geq \ln \left(\frac{c(1|2)}{c(2|1)} \right) \left(\frac{p_2}{p_1} \right) \quad (2.9)$$

atau x_0 dapat diklasifikasikan ke dalam populasi ω_2 jika

$$(\mu_1 - \mu_2)' \Sigma^{-1} x - \frac{1}{2}(\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 + \mu_2) < \ln \left(\frac{c(1|2)}{c(2|1)} \right) \left(\frac{p_2}{p_1} \right) \quad (2.10)$$

- **Nilai μ_1, μ_2 , dan $\Sigma_1 = \Sigma_2 = \Sigma$ tidak diketahui**

Jika pada kelompok 1 terdapat n_1 observasi dan pada kelompok 2 terdapat n_2 observasi dengan $n_1 + n_2 - 2 \geq h$, maka \bar{x}_1, \bar{x}_2 , dan $\mathbf{S}_{gabungan}$ dapat disubstitusi ke dalam Persamaan (2.9) dan (2.10). Dimana

$$\bar{x}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} x_{1i} \quad \bar{x}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} x_{2i}$$

dan

$$\begin{aligned} \mathbf{S}_1 &= \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)(X_{1i} - \bar{X}_1)' \\ \mathbf{S}_2 &= \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (X_{2i} - \bar{X}_2)(X_{2i} - \bar{X}_2)' \\ \mathbf{S}_{gabungan} &= \frac{(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2}{n_1 + n_2 - 2} \end{aligned}$$

Sehingga x_0 dapat diklasifikasikan ke dalam populasi ω_1 jika

$$(\bar{x}_1 - \bar{x}_2)' \mathbf{S}_{gabungan}^{-1} x_0 - \frac{1}{2}(\bar{x}_1 - \bar{x}_2)' \mathbf{S}_{gabungan}^{-1} (\bar{x}_1 + \bar{x}_2) \geq \ln \left(\frac{c(1|2)}{c(2|1)} \right) \left(\frac{p_2}{p_1} \right) \quad (2.11)$$

atau x_0 dapat diklasifikasikan ke dalam populasi ω_2 jika

$$(\bar{x}_1 - \bar{x}_2)' \mathbf{S}_{gabungan}^{-1} x_0 - \frac{1}{2}(\bar{x}_1 - \bar{x}_2)' \mathbf{S}_{gabungan}^{-1} (\bar{x}_1 + \bar{x}_2) < \ln \left(\frac{c(1|2)}{c(2|1)} \right) \left(\frac{p_2}{p_1} \right) \quad (2.12)$$

Berdasarkan pada Persamaan (2.6), jika $(\frac{c(1|2)}{c(2|1)})(\frac{p_2}{p_1}) = 1$ dan $\ln 1 = 0$ maka Persamaan (2.11) menjadi

$$(\bar{x}_1 - \bar{x}_2)' \mathbf{S}_{gabungan}^{-1} x_0 - \frac{1}{2} (\bar{x}_1 - \bar{x}_2)' \mathbf{S}_{gabungan}^{-1} (\bar{x}_1 + \bar{x}_2) \geq 0$$

$$(\bar{x}_1 - \bar{x}_2)' \mathbf{S}_{gabungan}^{-1} x_0 \geq \frac{1}{2} (\bar{x}_1 - \bar{x}_2)' \mathbf{S}_{gabungan}^{-1} (\bar{x}_1 + \bar{x}_2) \quad (2.13)$$

Berdasarkan pada Persamaan (2.13), maka dapat didefinisikan fungsi diskriminan linear untuk dua kelompok dengan sebagai berikut.

$$\hat{y} = (\bar{x}_1 - \bar{x}_2)' \mathbf{S}_{gabungan}^{-1} x \quad (2.14)$$

Sehingga, dapat diperoleh model dasar analisis diskriminan linear adalah dengan sebagai berikut.

$$y = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3 + \dots + \gamma_h x_h$$

dengan,

y = Fungsi diskriminan

γ_j = Koefisien diskriminan dengan $j = 1, 2, \dots, h$

x_j = Variabel bebas dengan $j = 1, 2, \dots, h$

2.4.3 Metode Pengklasifikasian Observasi Baru Dengan Nilai Tengah(*Cut-Off*)

- Jika $n_1 = n_2$, maka x_0 dievaluasi dengan nilai tengah antara nilai rata-rata diskriminan populasi ω_1 dan ω_2 . Nilai tengah dapat didefinisikan dengan sebagai berikut.

$$\begin{aligned} \hat{m} &= \frac{1}{2} (\bar{x}_1 - \bar{x}_2)' \mathbf{S}_{gabungan}^{-1} (\bar{x}_1 + \bar{x}_2) \\ &= \frac{1}{2} (\bar{y}_1 + \bar{y}_2) \end{aligned}$$

x_0 dapat diklasifikasikan ke dalam populasi ω_1 jika $\hat{y}_0 \geq \hat{m}$ atau x_0 dapat diklasifikasikan ke dalam populasi ω_2 jika $\hat{y}_0 < \hat{m}$.

- Jika $n_1 \neq n_2$, maka x_0 dievaluasi dengan nilai tengah antara nilai rata-rata diskriminan populasi ω_1 dan ω_2 . Nilai tengah dapat didefinisikan dengan sebagai berikut.

$$\hat{m} = \frac{n_1\bar{y}_1 + n_2\bar{y}_2}{n_1 + n_2}$$

x_0 dapat diklasifikasikan ke dalam populasi ω_1 jika $\hat{y}_0 \geq \hat{m}$ atau x_0 dapat diklasifikasikan ke dalam populasi ω_2 jika $\hat{y}_0 < \hat{m}$.

2.5 Evaluasi Hasil Klasifikasi Dan Validasi

2.5.1 APER (*Apparent Error Rate*)

APER merupakan salah satu evaluasi ketepatan klasifikasi yang bertujuan untuk menghitung peluang kesalahan klasifikasi, sehingga dapat diketahui persentase observasi pada populasi yang salah diklasifikasikan oleh fungsi diskriminan yang terbentuk. Selanjutnya, APER dapat dihitung dengan terlebih dahulu membuat tabel klasifikasi. Tabel klasifikasi untuk dua kelompok akan ditampilkan pada Tabel 2.2.

Tabel 2.2: Tabel Klasifikasi

KA	KP		JUMLAH OBSERVASI
	1	2	
1	n_{11}	n_{12}	n_1
2	n_{21}	n_{22}	n_2

Setelah dibuat tabel klasifikasi, maka sudah dapat ditentukan nilai APER-nya. Nilai APER dapat didefinisikan dengan sebagai berikut.

$$APER = \left(\frac{n_{12} + n_{21}}{n_1 + n_2} \right) \times 100\% \quad (2.15)$$

Jika KA = Kelompok Aktual dan KP = Kelompok Prediksi, maka

n_{11} = Jumlah observasi dari kelompok 1 tepat diklasifikasikan sebagai kelompok 1

n_{12} = Jumlah observasi dari kelompok 1 namun diprediksi sebagai kelompok 2

n_{21} = Jumlah observasi dari kelompok 2 namun diprediksi sebagai kelompok 1

n_{22} = Jumlah observasi dari kelompok 2 tepat diklasifikasikan sebagai kelompok 2

n_1, n_2 = Jumlah observasi masing-masing pada kelompok 1 dan 2

Selain APER terdapat pula AER (*Apparent Correct Classification Rate*).

AER merupakan salah satu evaluasi ketepatan klasifikasi yang bertujuan untuk menghitung peluang ketepatan klasifikasi sehingga dapat diketahui persentase observasi pada populasi yang tepat diklasifikasikan oleh fungsi diskriminan yang terbentuk. Nilai AER dapat diperoleh dengan sebagai berikut.

$$AER = 1 - APER$$

$$= \left(\frac{n_{11} + n_{22}}{n_1 + n_2} \right) \times 100\% \quad (2.16)$$

2.5.2 Uji Keakuratan

Pengujian ini dilakukan dengan tujuan untuk menentukan apakah pengelompokan yang dihasilkan memiliki akurasi yang tinggi, oleh karenanya maka digunakan *Proportional Change Criterion* (C_{pro}). C_{pro} dapat didefinisikan dengan sebagai berikut.

$$C_{pro} = r^2 + s^2 \quad (2.17)$$

Selain itu, terdapat *Maximum Change Criterion* (C_{max}) yang mana digunakan untuk menentukan persentase klasifikasi yang benar untuk kelompok

yang dipilih. C_{max} dapat didefinisikan dengan sebagai berikut.

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

dimana,

r = Proporsi jumlah sampel di kelompok 1

s = Proporsi jumlah sampel di kelompok 2

n_{max} = Jumlah sampel terbesar pada salah satu kelompok

N = Jumlah sampel secara keseluruhan

Hipotesis yang akan digunakan adalah sebagai berikut

$$H_0 : \text{Klasifikasi akurat}$$

$$H_a : \text{Klasifikasi tidak akurat}$$

Dengan demikian, akan terima H_0 apabila $Hit_{Ratio} > C_{max} > C_{pro}$.

2.5.3 Uji Kestabilan

Pengujian ini dilakukan dengan tujuan untuk mengkaji apakah pengalokasiannya dari tiap sampel dalam kelompok relatif stabil atau tidak. Hal tersebut dapat disebabkan akibat adanya perubahan perbedaan jumlah sampel yang diteliti. Uji ini dilakukan dengan cara menghitung nilai $press's Q$ yang dapat didefinisikan dengan sebagai berikut.

$$press's Q = \frac{(N - (n \times g))^2}{N(g - 1)} \quad (2.19)$$

dimana,

N = Total sampel

n = Jumlah pengamatan yang tepat diklasifikasikan

g = Jumlah kelompok

Hipotesis yang akan digunakan adalah sebagai berikut.

H_0 : Klasifikasi konsisten

H_a : Klasifikasi tidak konsisten

Dengan demikian, akan tolak H_0 apabila $\text{press}'sQ \leq \chi^2_{(db,\alpha)}$ atau akan terima H_0 apabila $\text{press}'sQ > \chi^2_{(db,\alpha)}$.

BAB III

PEMBAHASAN

Pada bab ini akan dibahas mengenai definisi analisis diskriminan kuadratik, metodologi penelitian, dan hasil penelitian penerapan analisis diskriminan kuadratik pada studi kasus.

3.1 Analisis Diskriminan Kuadratik

Pada subbab ini akan dijelaskan mengenai pengklasifikasian dengan analisis diskriminan kuadratik. Perbedaan analisis diskriminan linear dengan analisis diskriminan kuadratik adalah terdapat pada keputusan uji asumsinya. Jika pada analisis diskriminan linear, uji normalitas harus terpenuhi dan uji kesamaan varian kovarian harus terpenuhi. Namun pada analisis diskriminan kuadratik, uji normalitas terpenuhi tetapi uji kesamaan varian kovarian tidak terpenuhi. Aturan pengklasifikasian pada analisis diskriminan kuadratik tetap sama dengan aturan pengklasifikasian pada analisis diskriminan linear. Namun yang berbeda adalah ketika $\Sigma_1 \neq \Sigma_2$, dimana matriks varian kovarian pada kelompok 1 tidak sama dengan matriks varian kovarian pada kelompok

2. Diketahui bahwa

$$\Sigma_1 = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1h} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2h} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{h1} & \sigma_{h2} & \cdots & \sigma_{hh} \end{pmatrix}; \Sigma_2 = \begin{pmatrix} \varphi_{11} & \varphi_{12} & \cdots & \varphi_{1h} \\ \varphi_{21} & \varphi_{22} & \cdots & \varphi_{2h} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{h1} & \varphi_{h2} & \cdots & \varphi_{hh} \end{pmatrix}$$

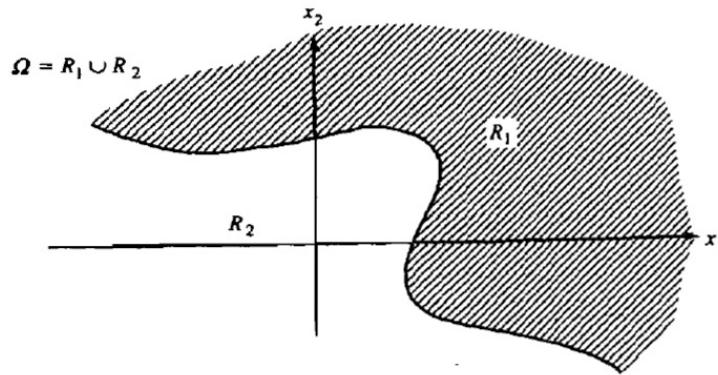
adalah matriks varian kovarian dengan h adalah jumlah variabel bebas. Berikut akan dijelaskan aturan pengklasifikasian dengan analisis diskriminan kuadratik untuk dua kelompok.

3.1.1 Aturan Pengklasifikasian Dengan Distribusi Normal Multivariat

Misalkan $f_1(x)$ dan $f_2(x)$ adalah fungsi kepadatan peluang untuk populasi populasi kelompok 1 (ω_1) dan kelompok 2 (ω_2) dimana

$$f_l(x) = \frac{1}{(2\pi)^{\frac{h}{2}} |\Sigma_l|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x - \mu_l)' \Sigma_l^{-1} (x - \mu_l)\right] \quad (3.1)$$

untuk $l = 1, 2$. Suatu observasi harus termasuk ke dalam salah satu antara populasi ω_1 dan ω_2 . Misalkan Ω adalah ruang sampel jumlah observasi populasi ω_1 dan ω_2 . Jika R_1 merupakan daerah semua observasi pada populasi ω_1 , maka $R_2 = \Omega - R_1$ merupakan daerah semua observasi pada populasi ω_2 . Ilustrasi akan ditampilkan pada Gambar 3.1.



Gambar 3.1: Daerah Klasifikasi Untuk Dua Kelompok

Peluang bersyarat $P(2|1)$ yang mana merupakan peluang suatu observasi yang berasal dari kelompok 1 namun diprediksi sebagai kelompok 2 dan dapat

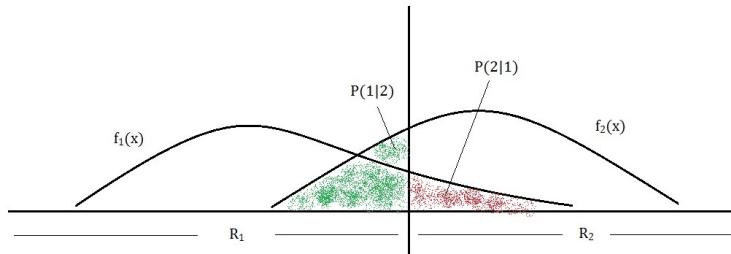
didefinisikan dengan

$$P(2|1) = P(\mathbf{X} \in R_2 | \omega_1) = \int_{R_2=\Omega-R_1} f_1(x) dx$$

dan peluang bersyarat $P(1|2)$ yang mana merupakan peluang suatu observasi yang berasal dari kelompok 2 namun diprediksi sebagai kelompok 1 dan dapat didefinisikan pula dengan

$$P(1|2) = P(\mathbf{X} \in R_1 | \omega_2) = \int_{R_1} f_2(x) dx$$

Peluang bersyarat $P(2|1)$ dan $P(1|2)$ dapat diilustrasikan pada Gambar 3.2.



Gambar 3.2: Peluang Bersyarat

Jika p_1 adalah peluang prior dari populasi ω_1 dan p_2 adalah peluang prior dari populasi ω_2 , maka peluang keseluruhan klasifikasi observasi secara tepat atau tidak tepat dapat diperoleh dengan perkalian peluang prior dengan peluang klasifikasi bersyarat.

- $P(\text{Observasi tepat di } \omega_1) = P(\mathbf{X} \in R_1 | \omega_1)P(\omega_1) = P(1|1)p_1$
- $P(\text{Observasi tidak tepat di } \omega_1) = P(\mathbf{X} \in R_1 | \omega_2)P(\omega_2) = P(1|2)p_2$
- $P(\text{Observasi tepat di } \omega_2) = P(\mathbf{X} \in R_2 | \omega_2)P(\omega_2) = P(2|2)p_2$
- $P(\text{Observasi tidak tepat di } \omega_2) = P(\mathbf{X} \in R_2 | \omega_1)P(\omega_1) = P(2|1)p_1$

3.1.2 Aturan Pengklasifikasian Dengan Metode ECM Minimum

Selain peluang kesalahan klasifikasi, ada hal lainnya yang sering diperhitungkan dalam mengklasifikasikan suatu observasi. Hal lain tersebut merupakan biaya akibat kesalahan klasifikasi. Biaya akibat kesalahan klasifikasi suatu observasi dapat didefinisikan pada tabel di bawah ini.

Tabel 3.1: Biaya Akibat Kesalahan Klasifikasi

	Kelompok Prediksi 1	Kelompok Prediksi 2
Kelompok Aktual 1	0	$c(2 1)$
Kelompok Aktual 2	$c(1 2)$	0

Keterangan:

$c(1|2)$ = Biaya kesalahan klasifikasi ketika sebuah observasi dari kelompok 2 namun diprediksi sebagai kelompok 1

$c(2|1)$ = Biaya kesalahan klasifikasi ketika sebuah observasi dari kelompok 1 namun diprediksi sebagai kelompok 2

Selain itu, dapat dihitung pula biaya dugaan akibat kesalahan klasifikasi atau *Expected Cost of Misclassification* (ECM) dengan formula

$$ECM = c(2|1)P(2|1)p_1 + c(1|2)P(1|2)p_2$$

Klasifikasi dikatakan tepat jika memiliki nilai ECM yang kecil atau minimum.

ECM akan bernilai minimum apabila memenuhi aturan berikut.

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq \left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right) \quad (3.2)$$

$$R_2 : \frac{f_1(x)}{f_2(x)} < \left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right) \quad (3.3)$$

Selain itu, terdapat beberapa kasus khusus pada aturan menentukan daerah nilai ECM, diantaranya

- Jika $p_2/p_1 = 1$, maka

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq \left(\frac{c(1|2)}{c(2|1)} \right) \quad R_2 : \frac{f_1(x)}{f_2(x)} < \left(\frac{c(1|2)}{c(2|1)} \right) \quad (3.4)$$

- Jika $c(1|2)/c(2|1) = 1$, maka

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq \left(\frac{p_2}{p_1} \right) \quad R_2 : \frac{f_1(x)}{f_2(x)} < \left(\frac{p_2}{p_1} \right) \quad (3.5)$$

- Jika $P_2/P_1 = \frac{1}{c(1|2)/c(2|1)}$, maka

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq 1 \quad R_2 : \frac{f_1(x)}{f_2(x)} < 1 \quad (3.6)$$

Jika x_0 merupakan observasi baru yang ingin diklasifikasikan sedangkan nilai peluang dan biaya akibat kesalahan tidak diketahui maka dapat digunakan kasus khusus pada Persamaan (3.6). Selanjutnya, x_0 dapat diklasifikasikan ke dalam populasi ω_1 jika

$$\frac{f_1(x)}{f_2(x)} \geq 1 \text{ atau } f_1(x) \geq f_2(x)$$

atau x_0 dapat diklasifikasikan ke dalam populasi ω_2 jika

$$\frac{f_1(x)}{f_2(x)} < 1 \text{ atau } f_1(x) < f_2(x)$$

Selanjutnya terdapat dua kasus khusus pada pengklasifikasian dua kelompok, diantaranya

- **Nilai μ_1, μ_2 , dan $\Sigma_1 \neq \Sigma_2$ diketahui**

Jika diketahui fungsi kepadatan peluang dari x adalah

$$f_l(x) = \frac{1}{(2\pi)^{\frac{h}{2}} |\Sigma_l|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x - \mu_l)' \Sigma_l^{-1} (x - \mu_l)\right] \quad (3.7)$$

untuk $l = 1, 2$; maka

$$\frac{f_1(x)}{f_2(x)} = \frac{\frac{1}{(2\pi)^{\frac{h}{2}} |\Sigma_1|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x - \mu_1)' \Sigma_1^{-1} (x - \mu_1)\right]}{\frac{1}{(2\pi)^{\frac{h}{2}} |\Sigma_2|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x - \mu_2)' \Sigma_2^{-1} (x - \mu_2)\right]}$$

$$= \frac{|\Sigma_1|^{-\frac{1}{2}}}{|\Sigma_2|^{-\frac{1}{2}}} \exp\left[-\frac{1}{2}(x - \mu_1)' \Sigma_1^{-1} (x - \mu_1) + \frac{1}{2}(x - \mu_2)' \Sigma_2^{-1} (x - \mu_2)\right] \quad (3.8)$$

Persamaan (3.8) akan disederhanakan menjadi

$$\begin{aligned} &= -\frac{1}{2} \ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) - \frac{1}{2}(x - \mu_1)' \Sigma_1^{-1} (x - \mu_1) + \frac{1}{2}(x - \mu_2)' \Sigma_2^{-1} (x - \mu_2) \\ &= -\frac{1}{2} \ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) - \frac{1}{2}[(x' \Sigma_1^{-1} - \mu_1' \Sigma_1^{-1})(x - \mu_1) - (x' \Sigma_2^{-1} - \mu_2' \Sigma_2^{-1})(x - \mu_2)] \\ &= -\frac{1}{2} \ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) - \frac{1}{2}[(x' \Sigma_1^{-1} x - x' \Sigma_1^{-1} \mu_1 - \mu_1' \Sigma_1^{-1} x + \mu_1' \Sigma_1^{-1} \mu_1) - \\ &\quad (x' \Sigma_2^{-1} x - x' \Sigma_2^{-1} \mu_2 - \mu_2' \Sigma_2^{-1} x + \mu_2' \Sigma_2^{-1} \mu_2)] \\ &= -\frac{1}{2} \ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) - \frac{1}{2}[x' \Sigma_1^{-1} x - x' \Sigma_1^{-1} \mu_1 - \mu_1' \Sigma_1^{-1} x + \mu_1' \Sigma_1^{-1} \mu_1 - \\ &\quad x' \Sigma_2^{-1} x + x' \Sigma_2^{-1} \mu_2 + \mu_2' \Sigma_2^{-1} x - \mu_2' \Sigma_2^{-1} \mu_2] \\ &= -\frac{1}{2} \ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) - \frac{1}{2}[x' \Sigma_1^{-1} x - x' \Sigma_2^{-1} x + x' \Sigma_2^{-1} \mu_2 - x' \Sigma_1^{-1} \mu_1 + \\ &\quad \mu_2' \Sigma_2^{-1} x - \mu_1' \Sigma_1^{-1} x + \mu_1' \Sigma_1^{-1} \mu_1 - \mu_2' \Sigma_2^{-1} \mu_2] \\ &= -\frac{1}{2} \ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) - \frac{1}{2}(x' \Sigma_1^{-1} x - x' \Sigma_2^{-1} x) - \frac{1}{2}(x' \Sigma_2^{-1} \mu_2 - x' \Sigma_1^{-1} \mu_1) \\ &\quad - \frac{1}{2}(\mu_2' \Sigma_2^{-1} x - \mu_1' \Sigma_1^{-1} x) - \frac{1}{2}(\mu_1' \Sigma_1^{-1} \mu_1 - \mu_2' \Sigma_2^{-1} \mu_2) \\ &= -\frac{1}{2} x' (\Sigma_1^{-1} - \Sigma_2^{-1}) x - \frac{1}{2} x' (\Sigma_2^{-1} \mu_2 - \Sigma_1^{-1} \mu_1) - \frac{1}{2} (\mu_2' \Sigma_2^{-1} - \mu_1' \Sigma_1^{-1}) x \\ &\quad - \frac{1}{2} \ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) - \frac{1}{2} (\mu_1' \Sigma_1^{-1} \mu_1 - \mu_2' \Sigma_2^{-1} \mu_2) \end{aligned}$$

Selanjutnya, daerah yang memenuhi aturan nilai ECM minimum berdasarkan pada Persamaan (3.2) dan (3.3) diantaranya

$$\begin{aligned} R_1 : & -\frac{1}{2} x' (\Sigma_1^{-1} - \Sigma_2^{-1}) x - \frac{1}{2} x' (\Sigma_2^{-1} \mu_2 - \Sigma_1^{-1} \mu_1) - \frac{1}{2} (\mu_2' \Sigma_2^{-1} - \mu_1' \Sigma_1^{-1}) x - k \geq \delta \\ R_2 : & -\frac{1}{2} x' (\Sigma_1^{-1} - \Sigma_2^{-1}) x - \frac{1}{2} x' (\Sigma_2^{-1} \mu_2 - \Sigma_1^{-1} \mu_1) - \frac{1}{2} (\mu_2' \Sigma_2^{-1} - \mu_1' \Sigma_1^{-1}) x - k < \delta \end{aligned}$$

dengan $\delta = \ln\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)$, dimana

$$k = \frac{1}{2} \ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) + \frac{1}{2} (\mu_1' \Sigma_1^{-1} \mu_1 - \mu_2' \Sigma_2^{-1} \mu_2)$$

Sehingga, x_0 dapat diklasifikasikan ke dalam populasi ω_1 jika

$$-\frac{1}{2}x'(\Sigma_1^{-1} - \Sigma_2^{-1})x - \frac{1}{2}x'(\Sigma_2^{-1}\mu_2 - \Sigma_1^{-1}\mu_1) - \frac{1}{2}(\mu_2'\Sigma_2^{-1} - \mu_1'\Sigma_1^{-1})x - k \geq \delta \quad (3.9)$$

atau x_0 dapat diklasifikasikan ke dalam populasi ω_2 jika

$$-\frac{1}{2}x'(\Sigma_1^{-1} - \Sigma_2^{-1})x - \frac{1}{2}x'(\Sigma_2^{-1}\mu_2 - \Sigma_1^{-1}\mu_1) - \frac{1}{2}(\mu_2'\Sigma_2^{-1} - \mu_1'\Sigma_1^{-1})x - k < \delta \quad (3.10)$$

- **Nilai μ_1, μ_2 , dan $\Sigma_1 \neq \Sigma_2$ tidak diketahui**

Untuk selanjutnya $\bar{x}_1, \bar{x}_2, \mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_1^{-1}$, dan \mathbf{S}_2^{-1} dapat disubstitusi ke dalam Persamaan (3.9) dan (3.10). Kemudian persamaannya menjadi

$$-\frac{1}{2}x'(\mathbf{S}_1^{-1} - \mathbf{S}_2^{-1})x - \frac{1}{2}x'(\mathbf{S}_2^{-1}\bar{x}_2 - \mathbf{S}_1^{-1}\bar{x}_1) - \frac{1}{2}(\bar{x}_2'\mathbf{S}_2^{-1} - \bar{x}_1'\mathbf{S}_1^{-1})x - k \geq \delta \quad (3.11)$$

dan

$$-\frac{1}{2}x'(\mathbf{S}_1^{-1} - \mathbf{S}_2^{-1})x - \frac{1}{2}x'(\mathbf{S}_2^{-1}\bar{x}_2 - \mathbf{S}_1^{-1}\bar{x}_1) - \frac{1}{2}(\bar{x}_2'\mathbf{S}_2^{-1} - \bar{x}_1'\mathbf{S}_1^{-1})x - k < \delta \quad (3.12)$$

dimana

$$k = \frac{1}{2} \ln\left(\frac{|\mathbf{S}_1|}{|\mathbf{S}_2|}\right) + \frac{1}{2}(\bar{x}_1'\mathbf{S}_1^{-1}\bar{x}_1 - \bar{x}_2'\mathbf{S}_2^{-1}\bar{x}_2)$$

Berdasarkan pada Persamaan (3.6) jika $(\frac{c(1|2)}{c(2|1)})(\frac{p_2}{p_1}) = 1$ dan $\ln 1 = 0$, maka Persamaan (3.11) dan (3.12) menjadi

$$-\frac{1}{2}x'(\mathbf{S}_1^{-1} - \mathbf{S}_2^{-1})x - \frac{1}{2}x'(\mathbf{S}_2^{-1}\bar{x}_2 - \mathbf{S}_1^{-1}\bar{x}_1) - \frac{1}{2}(\bar{x}_2'\mathbf{S}_2^{-1} - \bar{x}_1'\mathbf{S}_1^{-1})x - k \geq 0 \quad (3.13)$$

dan

$$-\frac{1}{2}x'(\mathbf{S}_1^{-1} - \mathbf{S}_2^{-1})x - \frac{1}{2}x'(\mathbf{S}_2^{-1}\bar{x}_2 - \mathbf{S}_1^{-1}\bar{x}_1) - \frac{1}{2}(\bar{x}_2'\mathbf{S}_2^{-1} - \bar{x}_1'\mathbf{S}_1^{-1})x - k < 0 \quad (3.14)$$

3.1.3 Metode Pengklasifikasian Observasi Baru

Berdasarkan pada Persamaan (3.13) dan (3.14), maka fungsi diskriminan kuadratik untuk dua kelompok dapat didefinisikan dengan sebagai berikut.

$$\hat{z} = t - k \quad (3.15)$$

dimana,

$$t = -\frac{1}{2}x'(\mathbf{S}_1^{-1} - \mathbf{S}_2^{-1})x - \frac{1}{2}x'(\mathbf{S}_2^{-1}\bar{x}_2 - \mathbf{S}_1^{-1}\bar{x}_1) - \frac{1}{2}(\bar{x}_2'\mathbf{S}_2^{-1} - \bar{x}_1'\mathbf{S}_1^{-1})x$$

$$k = \frac{1}{2}\ln\left(\frac{|\mathbf{S}_1|}{|\mathbf{S}_2|}\right) + \frac{1}{2}(\bar{x}_1'\mathbf{S}_1^{-1}\bar{x}_1 - \bar{x}_2'\mathbf{S}_2^{-1}\bar{x}_2)$$

Jika x_0 merupakan observasi baru, maka x_0 akan diklasifikasikan ke dalam populasi ω_1 jika $\hat{z}_0 \geq 0$ atau x_0 akan diklasifikasikan ke dalam populasi ω_2 jika $\hat{z}_0 < 0$.

3.2 Metodologi Penelitian

3.2.1 Sumber Data

Sumber data yang digunakan dalam penelitian ini adalah data primer yang diperoleh dari Puskesmas Kecamatan Pesanggrahan Kota Jakarta Selatan Provinsi DKI Jakarta periode desember 2016 - januari 2017. Jumlah keseluruhan responden (pasien) pada penelitian ini adalah sebanyak 186 orang, dengan rincian yaitu 94 orang penderita pradiabetes dan 92 orang penderita diabetes.

3.2.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah

- Variabel terikat (y) yang terdiri atas kejadian diabetes dimana
 - Kategori 1 sebagai kelompok pradiabetes
 - Kategori 2 sebagai kelompok diabetes

- Variabel bebas (x) yang terdiri dari faktor-faktor yang mempengaruhi kejadian diabetes. Variabel bebas yang akan digunakan, diantaranya

1. Usia (x_1)

Menunjukkan usia dalam tahun pada setiap responden.

2. Berat Badan (x_2)

Menunjukkan berat badan dalam kg pada setiap responden.

3. Kolesterol (x_3)

Menunjukkan kadar kolesterol pada setiap responden. Seseorang dikatakan memiliki kadar kolesterol normal jika pada saat pengukuran menunjukkan nilai sekitar 150-250 mg/dL.

4. Trigliserida (x_4)

Menunjukkan kadar lemak yang dibawa dalam aliran darah. Seseorang dikatakan memiliki kadar trigliserida normal jika pada saat pengukuran menunjukkan nilai sekitar 65-155 mg/dL.

5. Tekanan Darah Sistolik (x_5)

Menunjukkan tekanan darah saat fase kontraksi dari siklus jantung pada setiap responden. Seseorang dikatakan memiliki tekanan darah sistolik normal jika pada saat pengukuran menunjukkan < 120 mmHg.

6. Tekanan Darah Diastolik (x_6)

Menunjukkan tekanan darah saat fase relaksasi dari siklus jantung pada setiap responden. Seseorang dikatakan memiliki tekanan darah diastolik normal jika pada saat pengukuran menunjukkan < 80 mmHg.

7. Asam Urat (x_7)

Menunjukkan kadar asam urat pada setiap responden. Seseorang

dikatakan memiliki kadar asam urat normal jika pada saat pengukuran menunjukkan nilai sekitar 3,5-6,0 mg/dL.

3.2.3 Diagram Alir Penelitian

Tahap awal analisis data pada penelitian ini adalah sebagai berikut.

1. Melakukan eksplorasi data yaitu dengan melakukan deskripsi variabel terikat mengenai kejadian diabetes dan faktor-faktor yang mempengaruhi yang mana dijadikan sebagai variabel bebas.
2. Melakukan analisis statistika deskriptif dengan menghitung rataan dan standar deviasi.
3. Melakukan uji asumsi, yaitu uji normalitas, uji kesamaan matriks varian kovarian dan uji multikolinearitas.
4. Melakukan uji perbedaan signifikan dengan uji T^2 Hotelling untuk melihat apakah terdapat perbedaan nyata antar kelompok yang terbentuk.
5. Memutuskan jenis analisis diskriminan yang diterapkan berdasarkan pengujian asumsi.
 - (a) Jika uji normalitas dan uji kesamaan matriks varian kovarian terpenuhi, maka dapat digunakan analisis diskriminan linear.
 - (b) Jika uji normalitas terpenuhi namun uji kesamaan matriks varian kovarian tidak terpenuhi, maka dapat digunakan analisis diskriminan kuadratik.
 - (c) Jika uji normalitas tidak terpenuhi namun uji kesamaan matriks varian kovarian terpenuhi, maka dapat digunakan analisis diskriminan fisher.

- (d) Jika uji normalitas dan uji kesamaan matriks varian kovarian tidak terpenuhi, maka dapat digunakan analisis diskriminan nonparametrik.

Tahap penerapan analisis diskriminan kuadratik dalam penelitian ini adalah sebagai berikut.

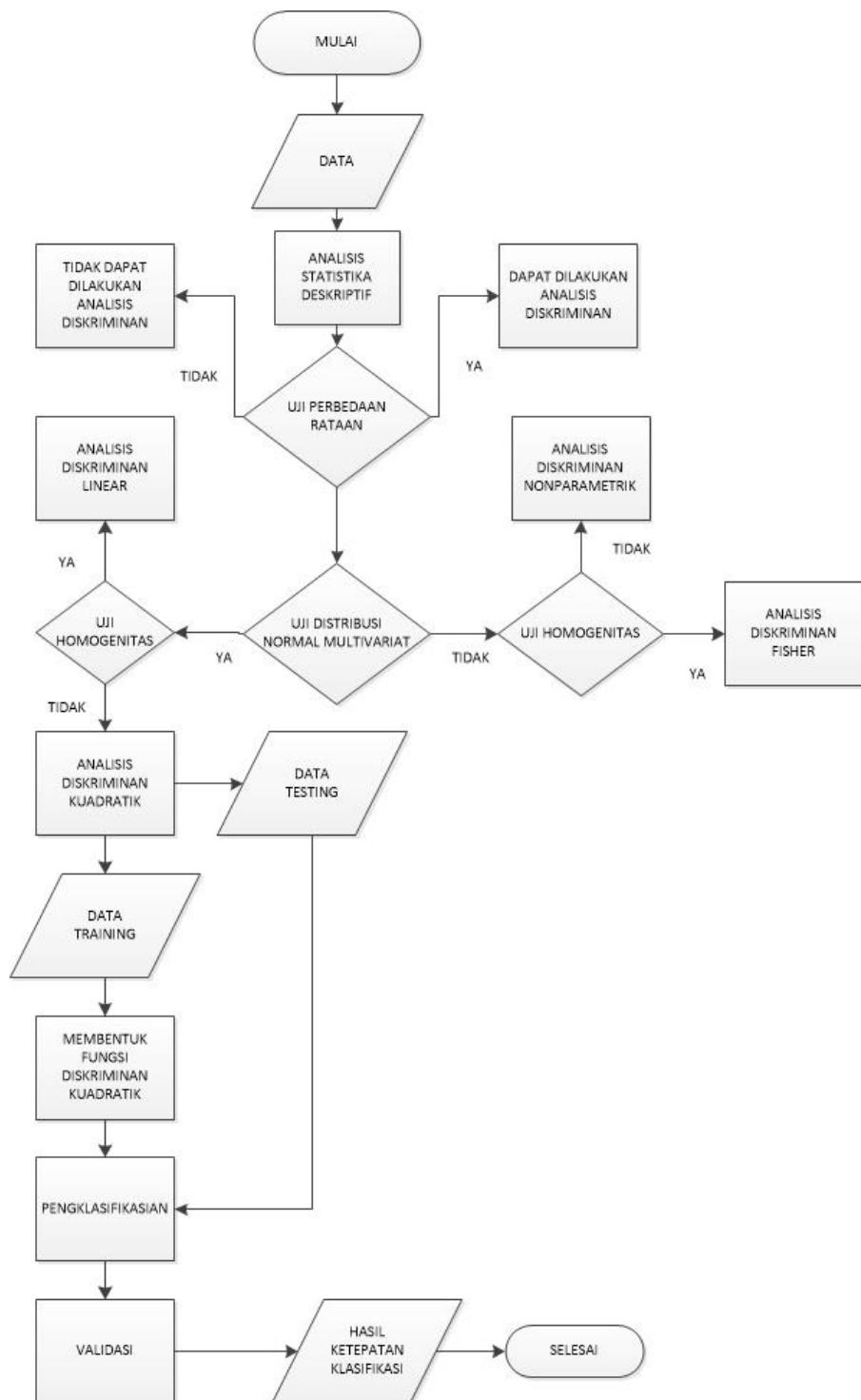
1. Membagi data menjadi dua bagian, yaitu data *training* dan data *testing*. Data *training* digunakan untuk membentuk fungsi diskriminan dan data *testing* digunakan untuk menguji ketepatan fungsi diskriminan yang telah terbentuk berdasarkan data *training*. Jumlah masing-masing observasi pada data *training* dan data *testing* adalah

$$N_{\text{training}} = \frac{80}{100} \times N$$

$$N_{\text{testing}} = \frac{20}{100} \times N$$

2. Membentuk fungsi diskriminan dengan data *training*.
3. Mengklasifikasikan data *testing* berdasarkan fungsi diskriminan yang telah terbentuk.
4. Melakukan evaluasi hasil klasifikasi dan validasi pada setiap data.
5. Hasil ketepatan klasifikasi akan diperoleh dan akan dibandingkan.

Software yang akan digunakan dalam penelitian ini adalah *Software R* dengan versi 3.3.2. Untuk selanjutnya, diagram alir penelitian akan diilustrasikan pada Gambar 3.3.



Gambar 3.3: Diagram Alir Penelitian

3.3 Hasil Penelitian

Pada penelitian ini, data yang digunakan akan dibagi menjadi dua bagian, yaitu data *training* dan data *testing*. Data *training* akan digunakan untuk membentuk fungsi diskriminan dengan 149 observasi sedangkan data *testing* akan digunakan untuk menguji ketepatan klasifikasi dengan fungsi diskriminan yang telah terbentuk berdasarkan data *training* dengan 37 observasi.

3.3.1 Analisis Statistika Deskriptif

Analisis statistika deskriptif bertujuan untuk mengetahui karakteristik data dan untuk digunakan pada keperluan selanjutnya. Statistika deskriptif yang akan dikalkulasi, terdiri dari rata-rata dan standar deviasi. Pada analisis statistika deskriptif, data yang digunakan adalah jumlah keseluruhan data yang digunakan.

Tabel 3.2: Statistika Deskriptif

Variabel	Diabetes		Pradiabetes	
	Rata-rata	Stdev	Rata-rata	Stdev
Usia	54.402174	9.202317	61.787234	7.277705
Berat Badan	64.934783	12.35892	60.223404	9.828808
Kolesterol	203.3	35.05777	193.276596	44.34818
Trigliserida	152.597826	72.708222	137.893617	69.41266
Tekanan Darah Sistolik	124.891304	13.05058	125.106383	10.85037
Tekanan Darah Diastolik	76.956522	6.912186	79.787234	6.388619
Asam Urat	5.678261	1.498118	6.824468	1.67965

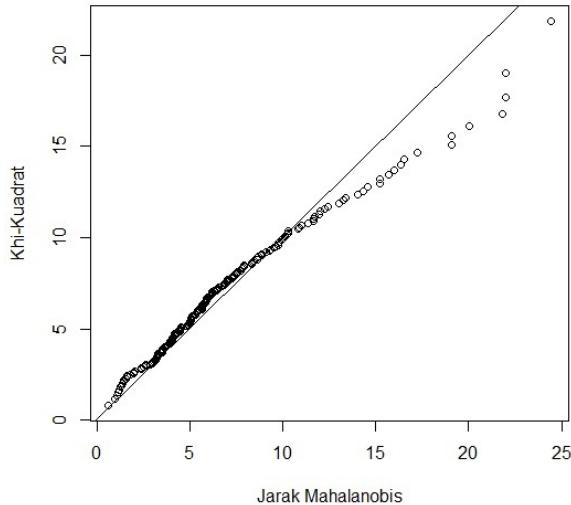
Pada Tabel 3.2, memperlihatkan bahwa pada kelompok diabetes menunjukkan variabel berat badan, kolesterol, dan trigliserida mempunyai nilai rata-rata yang lebih tinggi daripada kelompok pradiabetes.

3.3.2 Asumsi

Pada subbab ini akan dijelaskan mengenai uji asumsi yang harus dipenuhi pada analisis diskriminan dimana uji asumsi ini akan dilakukan pada seluruh data dan uji asumsi tersebut, diantaranya

- Uji Normalitas

Uji normalitas dilakukan dengan menghitung jarak Mahalanobis (D_i^2) untuk setiap observasi dan setelah itu akan diinterpretasikan ke dalam plot khi-kuadrat dengan titik koordinat $(D_i^2, \chi_h^2(\frac{i-\frac{1}{2}}{n}))$. Hasil pengujian normalitas multivariat akan diinterpretasikan ke dalam plot khi-kuadrat pada Gambar 3.4.



Gambar 3.4: Grafik Normalitas Multivariat

Berdasarkan pada Gambar 3.4, uji normalitas terpenuhi dengan titik-titik menyebar mendekati garis lurus dan terdapat 53.76 % jumlah nilai $D_i^2 \leq \chi_h^2(\frac{i-\frac{1}{2}}{n})$.

- Uji Kesamaan Matriks Varian Kovarian (Homogenitas)

Uji homogenitas dilakukan dengan menggunakan statistik uji *Box's M* dengan hipotesis sebagai berikut.

$$H_0 : \Sigma_1 = \Sigma_2$$

$$H_a : \Sigma_1 \neq \Sigma_2$$

dengan tingkat signifikansi $\alpha = 0.05$, maka akan tolak H_0 jika nilai $c > \chi^2_v$. Berdasarkan hasil perhitungan diperoleh nilai

$$\begin{aligned} u &= \left[\frac{1}{n_1 - 1} + \frac{1}{n_2 - 1} - \frac{1}{n_1 + n_2 - 2} \right] \left[\frac{2h^2 + 3h - 1}{6(h+1)(g-1)} \right] \\ &= \left[\frac{1}{94 - 1} + \frac{1}{92 - 1} - \frac{1}{94 + 92 - 2} \right] \left[\frac{(2 \times 7^2) + (3 \times 7) - 1}{6(7+1)(2-1)} \right] \\ &= 0.040087 \end{aligned}$$

$$M = [(n_1 + n_2 - 2) \times \ln|\mathbf{S}_{\text{gabungan}}|] [((n_1 - 1) \times \ln|\mathbf{S}_1|) + ((n_2 - 1) \times \ln|\mathbf{S}_2|)]$$

$$= [(184) \times 33.77071] [(93 \times 33.204) + (91 \times 33.7504)]$$

$$= 54.55323$$

sehingga,

$$c = (1 - u)M$$

$$= (1 - 0.040087) \times 54.55323$$

$$= 52.36631$$

$$\begin{aligned} v &= \frac{1}{2} \times h \times (h+1) \times (g-1) \\ &= \frac{1}{2} \times 7 \times (7+1) \times (2-1) \\ &= 28 \end{aligned}$$

$$\chi^2_{28} = 41.33714$$

Dengan demikian, disebabkan nilai $c > \chi^2_v$ maka diperoleh keputusan tolak H_0 yang berarti matriks varian kovarian tidak sama (heterogenitas). Oleh karena itu, disebabkan asumsi homogenitas tidak terpenuhi maka akan digunakan analisis diskriminan kuadratik.

- Uji Multikolinearitas

Pengujian ini digunakan untuk menemukan ada atau tidaknya korelasi di antara variabel bebas. Salah satu cara untuk mengetahuinya adalah dengan menghitung nilai VIF dari variabel bebas terhadap variabel terikat.

Hasil uji multikolinearitas akan ditampilkan pada Tabel 3.3.

Tabel 3.3: Nilai VIF

Variabel	VIF
Usia	1.231372
Berat Badan	1.121046
Kolesterol	1.210129
Trigliserida	1.157824
Tekanan Darah Sistolik	1.610958
Tekanan Darah Diastolik	1.629118
Asam Urat	1.151197

Berdasarkan Tabel 3.3, diperoleh pada setiap variabel nilai VIF-nya < 10 yang berarti tidak terjadi multikolinearitas.

3.3.3 Uji Perbedaan Signifikan Antar Kelompok

Pengujian ini dilakukan dengan tujuan untuk melihat apakah terdapat perbedaan yang nyata antar kelompok yang terbentuk dari setiap variabel bebas. Jika terdapat $\geq 50\%$ variabel bebas yang signifikan maka analisis diskriminan dapat dilakukan. Uji beda rataan ini akan dilakukan dengan uji T^2 *Hotelling*, dimana

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{x}_1 - \bar{x}_2)' \mathbf{S}_{gabungan}^{-1} (\bar{x}_1 - \bar{x}_2)$$

dengan statistik uji,

$$F_{hitung} = \left(\frac{n_1 + n_2 - h - 1}{(n_1 + n_2 - 2)h} \right) T^2$$

Perbedaan signifikan antar kelompok akan terjadi apabila nilai $F_{hitung} < F_{\alpha,(v_1,v_2)}$ dengan derajat bebas yang digunakan $v_1 = n_1 + n_2 - h - 1$ dan $v_2 = h$.

Hasil pengujian akan ditampilkan pada Tabel 3.4.

Tabel 3.4: Nilai F-Hitung Dan P-Value

Variabel	F-Hitung	P-Value
Usia	36.937*	6.87E-09
Berat Badan	8.2982*	0.00444
Kolesterol	2.916**	0.08939
Trigliserida	1.9907	0.16
Tekanan Darah Sistolik	0.015	0.9028
Tekanan Darah Diastolik	8.4178*	0.004169
Asam Urat	24.087*	2.02E-06

Keterangan:

(*) signifikan pada $\alpha = 0.05$.

(**) signifikan pada $\alpha = 0.1$.

Jika diketahui bahwa $F_{0.05,(178,7)} = 3.23$, maka variabel bebas yang signifikan adalah variabel usia, berat badan, tekanan darah diastolik, dan asam urat. Kesimpulan ini didukung dengan pembuktian jika sebuah variabel bebas memiliki nilai P-Value < 0.05 , maka variabel bebas tersebut sudah signifikan.

3.3.4 Analisis Diskriminan Linear

Analisis diskriminan linear merupakan salah satu metode statistika yang memiliki tujuan untuk mengklasifikasikan suatu observasi ke dalam salah satu kelompok berdasarkan variabel-variabel bebasnya dengan sebelumnya membentuk fungsi diskriminan. Di samping itu, analisis diskriminan juga dapat

digunakan untuk melihat variabel-variabel apa saja yang berpengaruh terhadap pembentukan klasifikasi tersebut. Pada penerapannya terdapat beberapa asumsi yang seharusnya dipenuhi, seperti asumsi normalitas dan asumsi homogenitas.

Pada subbab ini akan dijelaskan tahap-tahap pengklasifikasian kejadian diabetes dengan analisis diskriminan linear, diantaranya

- Pembentukan Fungsi Diskriminan Linear

Data *training* yang akan digunakan untuk membentuk fungsi diskriminan linear adalah berjumlah sebanyak 149 pasien dengan rincian 75 orang penderita pradiabetes dan 74 orang penderita diabetes. Data *training* tersebut akan diolah dengan menggunakan *Software R* dengan versi 3.3.2. Semua variabel bebas akan dimasukkan ke dalam proses dalam membentuk fungsi diskriminan linear sehingga fungsi diskriminan yang terbentuk merupakan kombinasi linear dari semua variabel bebas. Jika pada Persamaan (2.14) dimana

$$\hat{y} = (\bar{x}_1 - \bar{x}_2)' \mathbf{S}_{gabungan}^{-1} x$$

maka fungsi diskriminan linear yang terbentuk adalah sebagai berikut

$$\begin{aligned} \hat{y} = & 0.120896x_1 - 0.081616x_2 + 0.000899x_3 - 0.001785x_4 - 0.065966x_5 \\ & + 0.15656x_6 + 0.737623x_7 \end{aligned}$$

Koefisien-koefisien pada variabel bebas juga memberikan kontribusi terhadap terbentuknya klasifikasi tersebut, diantaranya

1. Koefisien pada variabel x_2 , x_4 , dan x_5 yang bertanda (-) akan memberikan pengurangan pada nilai skor diskriminan sehingga akan membuat suatu observasi cenderung diklasifikasikan ke dalam kelompok 2 (diabetes).

2. Koefisien pada variabel x_1, x_3, x_6 dan x_7 yang bertanda (+) akan memberikan penambahan pada nilai skor diskriminan sehingga akan membuat suatu observasi cenderung diklasifikasikan ke dalam kelompok 1 (pradiabetes).

Pada Lampiran 4, akan ditampilkan nilai skor diskriminan linear pada data *training* untuk setiap observasi.

- Nilai Tengah (*Cut-Off*) Untuk Masing-Masing Data

Nilai tengah merupakan nilai batas untuk menentukan suatu observasi dapat diklasifikasikan ke dalam kelompok 1 atau ke dalam kelompok 2. Disebabkan $n_1 \neq n_2$, maka nilai tengah untuk data *training* adalah

$$\begin{aligned}\hat{m} &= \frac{n_1\bar{y}_1 + n_2\bar{y}_2}{n_1 + n_2} \\ &= \frac{(75 \times 11.86669) + (74 \times 9.02712)}{149} \\ &= 10.45643\end{aligned}$$

dan nilai tengah untuk data *testing* adalah

$$\begin{aligned}\hat{m} &= \frac{n_1\bar{y}_1 + n_2\bar{y}_2}{n_1 + n_2} \\ &= \frac{(19 \times 1.31284) + (18 \times 9.837566)}{37} \\ &= 10.59514\end{aligned}$$

Berdasarkan pada fungsi diskriminan linear yang telah terbentuk, menjelaskan bahwa apabila $\hat{y}_0 < \hat{m}$ maka akan diklasifikasikan ke dalam kelompok 2 (diabetes). Namun apabila $\hat{y}_0 \geq \hat{m}$ maka akan diklasifikasikan ke dalam kelompok 1 (pradiabetes).

- Pengklasifikasian Observasi Baru

Salah satu cara untuk mengevaluasi fungsi diskriminan yang telah terbentuk adalah dengan pengklasifikasian observasi baru pada data *testing*. Data *testing* yang akan digunakan berjumlah sebanyak 37 pasien dengan rincian 19 orang penderita pradiabetes dan 18 orang penderita diabetes. Data *testing* tersebut akan diproses dengan fungsi diskriminan linear sebelumnya yang telah terbentuk. Setelah itu, akan diperoleh skor diskriminan dari setiap observasi yang mana selanjutnya akan digunakan untuk memperoleh nilai tengah. Pada Lampiran 5, akan ditampilkan skor diskriminan linear pada data *testing* untuk setiap observasi. Berdasarkan Lampiran 5, terdapat 13 observasi dimana diprediksi salah diklasifikasikan dengan rincian 4 observasi dari kelompok 1 namun diprediksi sebagai kelompok 2 (i_{12}) dan 9 observasi dari kelompok 2 namun diprediksi sebagai kelompok 1 (i_{21}). Rincian observasinya akan ditampilkan pada Tabel 3.5.

Tabel 3.5: Tabel Observasi Yang Diprediksi Salah Diklasifikasikan

i_{12}	i_{21}
Observasi 5	Observasi 20
Observasi 15	Observasi 21
Observasi 16	Observasi 22
Observasi 17	Observasi 23
	Observasi 24
	Observasi 25
	Observasi 28
	Observasi 30
	Observasi 33

3.3.5 Analisis Diskriminan Kuadratik

Analisis diskriminan kuadratik merupakan salah satu jenis analisis diskriminan selain analisis diskriminan linear. Namun, perbedaannya terdapat pada asumsi yang harus dipenuhi. Jika asumsi normalitas dan asumsi homogenitas terpenuhi, maka dapat diterapkan analisis diskriminan linear. Namun, jika asumsi normalitas terpenuhi dan asumsi homogenitas tidak terpenuhi, maka dapat diterapkan analisis diskriminan kuadratik.

Pada tahap awal penerapan analisis diskriminan kuadratik, tidak ada perbedaan dengan tahap awal penerapan analisis diskriminan linear pada subbab 3.3.4. Berikut tahap-tahap pengklasifikasian kejadian diabetes dengan analisis diskriminan kuadratik, diantaranya

- Pembentukan Fungsi Diskriminan Kuadratik

Data *training* akan digunakan kembali untuk membentuk fungsi diskriminan kuadratik dengan rincian 75 orang penderita pradiabetes dan 74 orang penderita diabetes. Dalam membentuk fungsi diskriminan kuadratik maka akan digunakan semua variabel bebas. Jika pada Persamaan (3.15) dimana

$$\hat{z} = t - k$$

dengan,

$$t = -\frac{1}{2}x'(\mathbf{S}_1^{-1} - \mathbf{S}_2^{-1})x - \frac{1}{2}x'(\mathbf{S}_2^{-1}\bar{x}_2 - \mathbf{S}_1^{-1}\bar{x}_1) - \frac{1}{2}(\bar{x}_2'\mathbf{S}_2^{-1} - \bar{x}_1'\mathbf{S}_1^{-1})x$$

$$k = \frac{1}{2}\ln\left(\frac{|\mathbf{S}_1|}{|\mathbf{S}_2|}\right) + \frac{1}{2}(\bar{x}_1'\mathbf{S}_1^{-1}\bar{x}_1 - \bar{x}_2'\mathbf{S}_2^{-1}\bar{x}_2)$$

maka fungsi diskriminan kuadratik yang terbentuk adalah sebagai berikut.

$$\hat{z} = -115.050885 - 0.0017030x_1^2 - 0.001275x_2^2 + 0.0001868x_3^2 - 0.000011x_4^2$$

$$\begin{aligned}
& -0.000258x_5^2 + 0.002705x_6^2 + 0.049251x_7^2 - 0.00393x_1x_2 \\
& -0.001264x_1x_3 + 0.000349x_1x_4 - 0.000463x_1x_5 - 0.006883x_1x_6 \\
& -0.029908x_1x_7 - 0.000529x_2x_3 - 0.00002x_2x_4 - 0.000018x_2x_5 \\
& -0.000817x_2x_6 + 0.002178x_2x_7 + 0.0036x_3x_4 - 0.000909x_3x_5 \\
& -0.000602x_3x_6 + 0.006293x_3x_7 + 0.000148x_4x_5 - 0.000394x_4x_6 \\
& -0.00092x_4x_7 - 0.002201x_5x_6 + 0.019293x_5x_7 + 0.000431x_6x_7 \\
& +1.56583x_1 + 0.474309x_2 + 0.148903x_3 - 0.006869x_4 \\
& +0.236946x_5 + 0.642924x_6 - 1.861631x_7
\end{aligned}$$

Setelah fungsi diskriminan kuadratik telah terbentuk, maka akan diperoleh nilai skor diskriminan untuk setiap observasi. Selanjutnya, jika nilai skor diskriminan telah diperoleh maka suatu observasi dapat diklasifikasikan ke dalam kelompok 1 atau kelompok 2. Jika $\hat{z} \geq 0$, maka observasi tersebut dapat diklasifikasikan ke dalam kelompok 1 atau jika $\hat{z} < 0$, maka observasi tersebut dapat diklasifikasikan ke dalam kelompok 2. Pada Lampiran 4, akan ditampilkan skor diskriminan kuadratik pada data *training* untuk setiap observasi.

- Pengklasifikasian Observasi Baru

Salah satu cara untuk mengevaluasi fungsi diskriminan kuadratik yang telah terbentuk adalah dengan pengklasifikasian observasi baru pada data *testing*. Selanjutnya akan diperoleh nilai skor diskriminan untuk setiap observasi. Suatu observasi baru akan diklasifikasikan ke dalam kelompok 1 jika $\hat{z}_0 \geq 0$ atau suatu observasi baru akan diklasifikasikan ke dalam kelompok 2 jika $\hat{z}_0 < 0$. Pada Lampiran 5, akan ditampilkan skor diskriminan kuadratik pada data *testing* untuk setiap observasi. Berdasarkan

Lampiran 5, terdapat 10 observasi dimana diprediksi salah diklasifikasikan dengan rincian 4 observasi dari kelompok 1 namun diprediksi sebagai kelompok 2 (i_{12}) dan 6 observasi dari kelompok 2 namun diprediksi sebagai kelompok 1 (i_{21}). Rincian observasinya akan ditampilkan pada Tabel 3.6.

Tabel 3.6: Tabel Observasi Yang Diprediksi Salah Diklasifikasikan

i_{12}	i_{21}
Observasi 15	Observasi 20
Observasi 17	Observasi 21
Observasi 18	Observasi 23
Observasi 19	Observasi 24
	Observasi 26
	Observasi 30

3.3.6 APER (*Apparent Error Rate*)

Untuk memperoleh nilai APER dan AER, maka terlebih dahulu akan dibuat tabel klasifikasi. Tabel klasifikasi akan dibuat pada setiap data dan pada setiap analisis diskriminan.

- Data *Training*

1. Analisis Diskriminan Linear

Pada tahap awal, data *training* akan diklasifikasikan dengan menggunakan fungsi diskriminan linear untuk dapat dilihat nilai APER dan AER-nya. Tabel klasifikasi akan ditampilkan pada Tabel 3.7.

Jika KA = Kelompok Aktual dan KP = Kelompok Prediksi, maka

Tabel 3.7: Tabel Klasifikasi Data *Training* Dengan Menggunakan Analisis Diskriminan Linear

KA	KP		Jumlah Observasi
	1	2	
1	63	12	75
2	12	62	74
Jumlah Observasi	75	74	149

Berdasarkan Tabel 3.7, dapat diperoleh nilai APER.

$$\begin{aligned} APER &= \frac{12 + 12}{149} \times 100\% \\ &= 16.11\% \end{aligned}$$

yang menunjukkan bahwa 16.11% dari data *training* diprediksi salah diklasifikasikan dan terdapat nilai

$$\begin{aligned} AER &= 1 - APER \\ &= 83.89\% \end{aligned}$$

yang menunjukkan bahwa 83.89% dari data *training* diprediksi tepat diklasifikasikan.

2. Analisis Diskriminan Kuadratik

Untuk selanjutnya, data *training* akan diklasifikasikan dengan menggunakan fungsi diskriminan kuadratik. Tabel klasifikasi akan ditampilkan pada Tabel 3.8.

Tabel 3.8: Tabel Klasifikasi Data *Training* Dengan Menggunakan Analisis Diskriminan Kuadratik

KA	KP		Jumlah Observasi
	1	2	
1	66	9	75
2	11	63	74
Jumlah Observasi	77	72	149

Berdasarkan Tabel 3.8, dapat diperoleh nilai APER.

$$\begin{aligned} APER &= \frac{9 + 11}{149} \times 100\% \\ &= 13.42\% \end{aligned}$$

yang menunjukkan bahwa 13.42% dari data *training* diprediksi salah diklasifikasikan dan terdapat nilai

$$\begin{aligned} AER &= 1 - APER \\ &= 86.58\% \end{aligned}$$

yang menunjukkan bahwa 86.58% dari data *training* diprediksi tepat diklasifikasikan.

- Data *Testing*

1. Analisis Diskriminan Linear

Selain data *training*, data *testing* juga akan diklasifikasikan dengan menggunakan fungsi diskriminan linear terlebih dahulu. Ini bertujuan untuk melihat ketepatan klasifikasi berdasarkan fungsi diskriminan yang telah terbentuk. Tabel klasifikasi akan ditampilkan pada Tabel 3.9.

Tabel 3.9: Tabel Klasifikasi Data *Testing* Dengan Menggunakan Analisis Diskriminan Linear

KA	KP		Jumlah Observasi
	1	2	
1	15	4	19
2	9	9	18
Jumlah Observasi	24	13	37

Berdasarkan Tabel 3.9, dapat diperoleh nilai APER.

$$APER = \frac{9 + 4}{37} \times 100\%$$

$$= 35.14\%$$

yang menunjukkan bahwa 35.14 % dari data *testing* diprediksi salah diklasifikasikan dan terdapat nilai

$$AER = 1 - APER$$

$$= 64.86\%$$

yang menunjukkan bahwa terdapat 64.86 % dari data *testing* diprediksi tepat diklasifikasikan.

2. Analisis Diskriminan Kuadratik

Untuk selanjutnya, data *testing* akan diklasifikasikan dengan menggunakan fungsi diskriminan kuadratik. Tabel klasifikasi akan ditampilkan pada Tabel 3.10.

Tabel 3.10: Tabel Klasifikasi Data *Testing* Dengan Menggunakan Analisis Diskriminan Kuadratik

KA	KP		Jumlah Observasi
	1	2	
1	15	4	19
2	6	12	18
Jumlah Observasi	21	16	37

Berdasarkan Tabel 3.10, dapat diperoleh nilai APER.

$$APER = \frac{4 + 6}{37} \times 100\%$$

$$= 27.03\%$$

yang menunjukkan bahwa 27.03% dari data *testing* diprediksi salah diklasifikasikan dan terdapat nilai

$$AER = 1 - APER$$

$$= 72.97\%$$

yang menunjukkan bahwa 72.97% dari data *testing* diprediksi tepat diklasifikasikan.

3.3.7 Uji Keakuratan

Uji ini digunakan untuk menentukan apakah pengelompokan yang dihasilkan memiliki akurasi yang tinggi dan untuk menentukan persentase klasifikasi yang benar untuk kelompok yang dipilih. Uji ini akan diterapkan berdasarkan hasil klasifikasi pada setiap data di setiap analisis diskriminan, tepatnya pada subbab 3.3.6. Sebelumnya telah dibuat hipotesis dengan sebagai berikut.

$$H_0 : \text{Klasifikasi akurat}$$

$$H_a : \text{Klasifikasi tidak akurat}$$

Hasil evaluasi dengan uji keakuratan akan dijelaskan dengan sebagai berikut.

- Data *Training*

1. Analisis Diskriminan Linear

Berdasarkan pada Persamaan (2.17) dan (2.18), telah diperoleh

$$Hit_{Ratio} = \frac{125}{149} \times 100\% = 83.89\%$$

$$C_{pro} = (0.503355^2 + 0.496644^2) \times 100\% = 50\%$$

$$C_{max} = \frac{75}{149} \times 100\% = 50.33\%$$

Berdasarkan pada penghitungan sebelumnya, dapat diperoleh keputusan terima H_0 disebabkan nilai $Hit_{Ratio} > C_{max} > C_{pro}$ dan ini mengartikan bahwa klasifikasi telah akurat.

2. Analisis Diskriminan Kuadratik

Berdasarkan pada Persamaan (2.17) dan (2.18), telah diperoleh

$$Hit_{Ratio} = \frac{129}{149} \times 100\% = 86.57\%$$

$$C_{pro} = (0.503355^2 + 0.496644^2) \times 100\% = 50\%$$

$$C_{max} = \frac{75}{149} \times 100\% = 50.33\%$$

Berdasarkan pada penghitungan sebelumnya, dapat diperoleh keputusan terima H_0 disebabkan nilai $Hit_{Ratio} > C_{max} > C_{pro}$ dan ini mengartikan bahwa klasifikasi telah akurat.

- Data *Testing*

1. Analisis Diskriminan Linear

Berdasarkan pada Persamaan (2.17) dan (2.18), telah diperoleh

$$Hit_{Ratio} = \frac{24}{37} \times 100\% = 64.86\%$$

$$C_{pro} = (0.513514^2 + 0.486486^2) \times 100\% = 50.02\%$$

$$C_{max} = \frac{19}{37} \times 100\% = 51.35\%$$

Berdasarkan pada penghitungan sebelumnya, dapat diperoleh keputusan terima H_0 disebabkan nilai $Hit_{Ratio} > C_{max} > C_{pro}$ dan ini mengartikan bahwa klasifikasi telah akurat.

2. Analisis Diskriminan Kuadratik

Berdasarkan pada Persamaan (2.17) dan (2.18), telah diperoleh

$$Hit_{Ratio} = \frac{27}{37} \times 100\% = 72.9\%$$

$$C_{pro} = (0.513514^2 + 0.486486^2) \times 100\% = 50.02\%$$

$$C_{max} = \frac{19}{37} \times 100\% = 51.35\%$$

Berdasarkan pada penghitungan sebelumnya, dapat diperoleh keputusan terima H_0 disebabkan nilai $Hit_{Ratio} > C_{max} > C_{pro}$ dan ini mengartikan bahwa klasifikasi telah akurat.

Setelah dievaluasi dengan uji keakuratan, maka diperoleh keputusan bahwa pada setiap analisis diskriminan dan pada setiap data menghasilkan klasifikasi yang akurat. Oleh sebab itu, selanjutnya akan dievaluasi kembali dengan uji kestabilan untuk memperoleh analisis diskriminan yang tepat untuk studi kasus ini.

3.3.8 Uji Kestabilan

Uji ini digunakan untuk mengkaji apakah pengalokasian dari tiap sampel dalam kelompok relatif stabil atau tidak. Uji ini akan diterapkan berdasarkan hasil klasifikasi pada data *testing* di setiap analisis diskriminan, tepatnya pada Tabel (3.9) dan (3.10). Sebelumnya telah dibuat hipotesis dengan sebagai berikut.

$$H_0 : \text{Klasifikasi konsisten}$$

$$H_a : \text{Klasifikasi tidak konsisten}$$

dan selanjutnya akan digunakan sebagai pengambilan keputusan. Jika diketahui bahwa $\chi^2_{(db,\alpha)} = \chi^2_{(1,0.05)} = 3.841$, maka

- Analisis Diskriminan Linear

Berdasarkan pada Persamaan (2.19), telah diperoleh

$$\begin{aligned} press's Q &= \frac{(N - (n \times g))^2}{N(g - 1)} \\ &= \frac{(37 - (24 \times 2))^2}{37} \end{aligned}$$

$$= 3.270$$

Berdasarkan pada penghitungan sebelumnya, dapat diambil kesimpulan tolak H_0 disebabkan nilai $_{press's}Q \leq \chi^2_{(1,0.05)}$ dan ini mengartikan bahwa klasifikasi tidak konsisten.

- Analisis Diskriminan Kuadratik

Berdasarkan pada Persamaan (2.19), telah diperoleh

$$\begin{aligned} _{press's}Q &= \frac{(N - (n \times g))^2}{N(g - 1)} \\ &= \frac{(37 - (27 \times 2))^2}{37} \\ &= 7.810 \end{aligned}$$

Berdasarkan pada penghitungan sebelumnya, dapat diambil kesimpulan terima H_0 disebabkan nilai $_{press's}Q > \chi^2_{(1,0.05)}$ dan ini mengartikan bahwa klasifikasi telah konsisten.

Berdasarkan hasil evaluasi pada uji kestabilan diatas, maka dapat diperoleh keputusan bahwa klasifikasi dengan analisis diskriminan kuadratik dinilai telah konsisten dibandingkan dengan analisis diskriminan linear.

BAB IV

PENUTUP

4.1 Kesimpulan

Berdasarkan hasil penelitian pada Bab 3, maka dapat diperoleh kesimpulan dengan sebagai berikut.

1. Model umum analisis diskriminan kuadratik untuk dua kelompok adalah dengan sebagai berikut.

$$\hat{z} = t - k$$

dimana,

$$t = -\frac{1}{2}x'(\mathbf{S}_1^{-1} - \mathbf{S}_2^{-1})x - \frac{1}{2}x'(\mathbf{S}_2^{-1}\bar{x}_2 - \mathbf{S}_1^{-1}\bar{x}_1) - \frac{1}{2}(\bar{x}_2'\mathbf{S}_2^{-1} - \bar{x}_1'\mathbf{S}_1^{-1})x$$
$$k = \frac{1}{2}\ln\left(\frac{|\mathbf{S}_1|}{|\mathbf{S}_2|}\right) + \frac{1}{2}(\bar{x}_1'\mathbf{S}_1^{-1}\bar{x}_1 - \bar{x}_2'\mathbf{S}_2^{-1}\bar{x}_2)$$

2. Jika diketahui bahwa variabel bebas x_1 = Usia, x_2 = Berat Badan, x_3 = Kolesterol, x_4 = Triglicerida, x_5 = Tekanan Darah Sistolik, x_6 = Tekanan Darah Diastolik, dan x_7 = Asam Urat, maka dapat diperoleh fungsi diskriminan dengan sebagai berikut.

- (a) Fungsi diskriminan linear untuk dua kelompok berdasarkan data *training* pada pengklasifikasian kejadian diabetes adalah dengan sebagai berikut.

$$\hat{y} = 0.120896x_1 - 0.081616x_2 + 0.000899x_3 - 0.001785x_4 - 0.065966x_5 + 0.15656x_6 + 0.737623x_7$$

Berdasarkan model tersebut dapat diinterpretasikan bahwa

- Semakin tinggi berat badan, trigliserida, dan tekanan darah sistolik akan membuat suatu observasi cenderung diklasifikasikan ke dalam kelompok 2 (diabetes).
 - Semakin tinggi usia, kolesterol, tekanan darah diastolik, dan asam urat akan membuat suatu observasi cenderung diklasifikasikan ke dalam kelompok 1 (pradiabetes).
- (b) Fungsi diskriminan kuadratik untuk dua kelompok berdasarkan data *training* pada pengklasifikasian kejadian diabetes adalah dengan sebagai berikut.

$$\begin{aligned}\hat{z} = & -115.050885 - 0.0017030x_1^2 - 0.001275x_2^2 + 0.0001868x_3^2 \\& - 0.000011x_4^2 - 0.000258x_5^2 + 0.002705x_6^2 + 0.049251x_7^2 \\& - 0.00393x_1x_2 - 0.001264x_1x_3 + 0.000349x_1x_4 - 0.000463x_1x_5 \\& - 0.006883x_1x_6 - 0.029908x_1x_7 - 0.000529x_2x_3 - 0.00002x_2x_4 \\& - 0.000018x_2x_5 - 0.000817x_2x_6 + 0.002178x_2x_7 + 0.0036x_3x_4 \\& - 0.000909x_3x_5 - 0.000602x_3x_6 + 0.006293x_3x_7 + 0.000148x_4x_5 \\& - 0.000394x_4x_6 - 0.00092x_4x_7 - 0.002201x_5x_6 + 0.019293x_5x_7 \\& + 0.000431x_6x_7 + 1.56583x_1 + 0.474309x_2 + 0.148903x_3 \\& - 0.006869x_4 + 0.236946x_5 + 0.642924x_6 - 1.861631x_7\end{aligned}$$

3. Pada evaluasi hasil klasifikasi akan digunakan 3 evaluasi, yaitu APER (*Apparent Error Rate*), Uji Keakuratan, dan Uji Kestabilan.

(a) APER (*Apparent Error Rate*)

- Persentase misklasifikasi dengan analisis diskriminan linear pada data *training* adalah sebesar 16.10% dan pada data *testing* adalah sebesar 35.13 %.

- Persentase misklasifikasi dengan analisis diskriminan kuadratik pada data *training* adalah sebesar 13.42% dan pada data *testing* adalah sebesar 27.02 %.

Dengan demikian, persentase misklasifikasi dengan analisis diskriminan kuadratik pada data *training* ataupun pada data *testing* lebih kecil dibandingkan dengan analisis diskriminan linear.

(b) Uji Keakuratan

Pengklasifikasian kejadian diabetes dengan analisis diskriminan linear ataupun dengan analisis diskriminan kuadratik pada data *training* ataupun pada data *testing* telah akurat karena nilai $Hit_{Ratio} > C_{max} > C_{pro}$.

(c) Uji Kestabilan

Jika diketahui nilai $\chi^2_{(1,0.05)} = 3.841$, maka

- Pengklasifikasian kejadian diabetes dengan analisis diskriminan linear tidak stabil/tidak konsisten karena nilai $press'sQ \leq \chi^2_{(1,0.05)}$ dengan nilai $press'sQ = 3.270$.
- Pengklasifikasian kejadian diabetes dengan analisis diskriminan kuadratik telah stabil/telah konsisten disebabkan nilai $press'sQ > \chi^2_{(1,0.05)}$ dengan nilai $press'sQ = 7.810$.

4.2 Saran

Saran dibutuhkan untuk membuat sebuah penelitian menjadi lebih baik dari penelitian sebelumnya. Saran dari penulis untuk mengembangkan sebuah penelitian ini menjadi lebih baik adalah dengan sebagai berikut.

1. Menerapkan analisis diskriminan kuadratik pada studi kasus kejadian

diabetes namun dengan kasus ECM yang berbeda, seperti

- (a) Saat $p_2/p_1 = 1$
 - (b) Saat $c(1|2)/c(2|1) = 1$
2. Menerapkan analisis diskriminan kuadratik pada studi kasus yang berbeda, seperti pada kasus penyakit kanker serviks.
 3. Menerapkan metode lain dalam meneliti kejadian diabetes atau kejadian lainnya, seperti analisis diskriminan kuadratik robust, analisis diskriminan kanonik, analisis diskriminan logistik, atau analisis diskriminan kernel.

DAFTAR PUSTAKA

- Aswa, Rasmiati. 2015. *Analisis Diskriminan Kernel Dengan Metode Cross Validation (Studi Kasus: Faktor-Faktor Yang Berhubungan Dengan Kejadian Hipertensi Pada Puskesmas Usuku Wakatobi Sulawesi Tenggara Pada Tahun 2013)*. Makassar: Program Sarjana, Universitas Hasanudin.
- Dipiro, Joseph T., dkk. 2005. *Pharmacotherapy A Pathophysiologie Approach*. Edisi ke-6. USA: Mc. Graw-Hill.
- Ghozali, Imam. 2011. *Applikasi Analisis Multivariat Dengan Program IBM SPSS 21*. Semarang: BP Universitas Diponegoro.
- Gradianta, Ramadhana Dio. 2014. *Klasifikasi Pasien Penderita Diabetes Mellitus Tipe II Menggunakan Metode Hybrid Genetic Algorithm - Discriminant Analysis*. Surabaya: Program Sarjana, Institut Tekhnologi Sepuluh November.
- Huberty, Carl J dan Stephen Olejnik. 2006. *Applied MANOVA And Discriminant Analysis*. Edisi ke-2. New York: John Wiley & Sons, Inc.
- Johnson, Richard A dan Dean W. Wichern. 2008. *Applied Multivariate Statistical Analysis*. Edisi ke-6. London: Pearson Education, Inc.
- Rencher, Alvin C. 2002. *Methods Of Multivariate Analysis*. Edisi ke-2. New York: John Wiley & Sons, Inc.
- Purnomo, Hepy. 2003. *Metode Klasifikasi Menggunakan Fungsi Diskriminan*. Bogor: Program Sarjana, Institut Pertanian Bogor.
- Sharma, Subhash. 1996. *Applied Multivariate Techniques*. New York: John Wiley & Sons, Inc.
- Supranto, J. 2004. *Analisis Multivariat: Arti & Interpretasi*. Jakarta: Rineka Cipta.

- Syamiyah, Najah. 2014. *Faktor Resiko Kejadian Diabetes Mellitus Tipe 2 Pada Wanita Di Puskesmas Kecamatan Pesanggrahan Jakarta Selatan Tahun 2014*. Jakarta: Program Sarjana, Universitas Islam Negeri Syarif Hidayatullah.
- Vita, Livia Arif. 2013. *Aplikasi Analisis Regresi Komponen Utama Terhadap Faktor-Faktor Yang Mempengaruhi Penyakit Diabetes Mellitus (Studi Kasus Di Puskesmas Tempeh Kab. Lumajang)*. Jakarta: Program Sarjana, Universitas Negeri Malang.
- Wahyuningtyas, Yosiana Fitria. 2011. *Evaluasi Ketepatan Klasifikasi Kelulusan Tes Keterampilan SNMPTN Bidang Olahraga Menggunakan Analisis Diskriminan Kernel*. Surabaya: Program Sarjana, Institut Teknologi Sepuluh Nopember.
- Yulinah, Elin., dkk. 2008. *ISO Farmakoterapi*. Jakarta: PT ISFI Penerbitan.

LAMPIRAN-LAMPIRAN

LAMPIRAN 1

Data *Training*

$$|\mathbf{S}_1| = 3.757262 \times 10^4 \quad |\mathbf{S}_2| = 3.763395 \times 10^4$$

$$|\mathbf{S}_{gabungan}| = 5.190922 \times 10^4$$

$$\bar{x}_1 = \begin{pmatrix} 62.16 \\ 61.08 \\ 194.6667 \\ 140.64 \\ 126 \\ 80.13333 \\ 7.021333 \end{pmatrix}; \bar{x}_2 = \begin{pmatrix} 54.09459 \\ 64.94595 \\ 206.6027 \\ 158.3243 \\ 126.0811 \\ 77.02703 \\ 5.616216 \end{pmatrix}$$

$$\mathbf{S}_1 = \begin{pmatrix} 55.244324 & -15.175135 & -128.635135 & -90.941622 & 10.783784 & 0.518918 & -1.122378 \\ -15.175135 & 110.966486 & -44.878378 & 15.272432 & 1 & 11.205405 & 5.632054 \\ -128.635135 & -44.878378 & 2227.576577 & 1452.270270 & -91.891892 & -65.9009 & 4.926126 \\ -90.941621 & 15.272432 & 1452.270270 & 5793.233514 & -27.270270 & -34.140540 & 9.636162 \\ 10.783783 & 1 & -91.891892 & -27.270270 & 127.027027 & 41.081081 & 1.951351 \\ 0.518918 & 11.205405 & -65.900901 & -34.140541 & 41.081081 & 41.873873 & 0.091711 \\ -1.122378 & 5.632054 & 4.926126 & 9.636162 & 1.951351 & 0.091711 & 2.539538 \end{pmatrix}$$

$$\mathbf{S}_2 = \begin{pmatrix} 65.648464 & 14.964087 & -8.208478 & -87.044798 & 28.594965 & 15.353573 & 3.071048 \\ 14.964087 & 141.613476 & -20.145057 & 102.004073 & -5.146242 & 13.398741 & 6.661163 \\ -8.208478 & -20.145057 & 1209.755061 & 607.681303 & 154.257312 & 57.925953 & -13.967716 \\ -87.044798 & 102.004073 & 607.681303 & 6081.564606 & 39.370603 & 64.402073 & 5.461792 \\ 28.594965 & -5.146242 & 154.257312 & 39.370603 & 183.061829 & 64.901888 & -3.908182 \\ 15.353573 & 13.398741 & 57.925953 & 64.402073 & 64.901888 & 48.574602 & -1.184006 \\ 3.071048 & 6.661163 & -13.967716 & 5.461792 & -3.908182 & -1.184006 & 2.420281 \end{pmatrix}$$

$$\mathbf{S}_1^{-1} = \begin{pmatrix} 0.022577 & 0.003301 & 0.001368 & 0.000008 & -0.001893 & 0.002851 & 0.001322 \\ 0.003301 & 0.011133 & 0.000471 & -0.000069 & 0.001613 & -0.003916 & -0.025427 \\ 0.001368 & 0.000471 & 0.000654 & -0.000137 & 0.000135 & 0.000644 & -0.001316 \\ 0.000008 & -0.000069 & -0.000137 & 0.000207 & -0.00006 & 0.000031 & -0.000319 \\ -0.001893 & 0.001613 & 0.000135 & -0.00006 & 0.012276 & -0.012258 & -0.013441 \\ 0.002851 & -0.003916 & 0.000644 & 0.000031 & -0.012258 & 0.037924 & 0.016625 \\ 0.001322 & -0.025428 & -0.001316 & -0.000319 & -0.013441 & 0.016625 & 0.464244 \end{pmatrix}$$

$$\mathbf{S}_2^{-1} = \begin{pmatrix} 0.01917 & -0.000628 & 0.000103 & 0.000358 & -0.002357 & -0.004031 & -0.028586 \\ -0.000628 & 0.008783 & -0.000058 & -0.000089 & 0.001595 & -0.004734 & -0.023249 \\ 0.000103 & -0.000058 & 0.00103 & -0.0001 & -0.000773 & 0.000042 & 0.004976 \\ 0.000358 & -0.000089 & -0.0001 & 0.001854 & 0.000088 & -0.000363 & -0.00124 \\ -0.002357 & 0.001595 & -0.000773 & 0.000088 & 0.01176 & -0.01446 & 0.005851 \\ -0.004031 & -0.004734 & 0.000042 & -0.000363 & -0.01446 & 0.043335 & 0.017056 \\ -0.028586 & -0.023249 & 0.004976 & -0.00124 & 0.005851 & 0.017056 & 0.562748 \end{pmatrix}$$

LAMPIRAN 2

Fungsi Distribusi Normal Multivariat

Jika diketahui bahwa fungsi distribusi normal univariat adalah sebagai berikut.

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

dan jika

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

maka

$$= (x-\mu)'(\Sigma)^{-1}(x-\mu)$$

disebabkan varian $\sigma^2 = \Sigma$.

Jika diketahui konstanta pada univariat adalah sebagai berikut.

$$(2\pi)^{-\frac{1}{2}}(\sigma)^{-\frac{1}{2}}$$

maka konstanta pada multivariat adalah sebagai berikut.

$$(2\pi)^{-\frac{h}{2}}|\Sigma|^{-\frac{1}{2}}$$

disebabkan terdapat vektor acak $\mathbf{X}' = [X_1, X_2, \dots, X_h]$ dengan h adalah banyaknya variabel bebas.

Untuk selanjutnya, diperoleh fungsi distribusi normal multivariat dengan sebagai berikut.

$$f_l(x) = \frac{1}{(2\pi)^{\frac{h}{2}}|\Sigma_l|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x-\mu_l)' \Sigma_l^{-1} (x-\mu_l)\right] \quad -\infty < x_j < \infty$$

dengan l menunjukkan setiap kelompok dan $j = 1, 2, \dots, h$.

LAMPIRAN 3

Nilai Jarak *Mahalanobis* (D_i^2) Dan Khi-Kuadrat ($\chi_h^2(\frac{i-\frac{1}{2}}{n})$) Pada Seluruh Data

i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	D_i^2	$\chi_h^2(\frac{i-\frac{1}{2}}{n})$
1	67	56	167	46	140	80	6.8	4.862187	0.812732
2	59	49	203	183	120	70	7.4	4.272100	1.154574
3	68	64	171	103	120	80	8.6	3.170591	1.367051
4	61	91	198	270	140	90	8.8	11.701947	1.531952
5	72	52	158	96	120	80	7.2	4.073969	1.670674
6	62	56	170	149	130	80	8.8	3.926225	1.792450
7	74	51	144	261	130	70	5.5	15.727706	1.902212
8	71	58	174	168	130	80	5.3	3.555375	2.002940
9	61	51	205	139	110	70	5.8	3.279994	2.096593
10	58	57	184	126	110	80	6.9	3.635040	2.184532
11	64	60	171	112	120	80	7.2	1.397278	2.267747
12	63	80	204	117	130	80	11.1	10.299146	2.346984
13	56	50	209	147	130	90	4.2	7.057927	2.422821
14	58	65	198	99	110	70	8.6	4.556459	2.495715
15	66	59	173	112	130	80	6.4	1.166462	2.566035
16	62	62	173	83	130	80	10.9	9.543387	2.634083
17	55	61	223	165	110	80	6.3	3.929871	2.700113
18	71	59	264	153	130	90	9	11.670836	2.764335
19	71	56	142	107	130	80	4.8	4.986896	2.826930
20	64	57	221	83	130	80	9.5	6.756976	2.888054

21	66	64	220	126	130	80	5	2.625292	2.947842
22	67	40	245	110	130	70	6.5	9.929170	3.006410
23	71	53	164	91	130	70	7.8	7.858259	3.063862
24	67	57	183	63	130	90	6.5	5.417501	3.120289
25	55	41	209	128	110	80	5.6	7.927995	3.175773
26	63	75	143	243	140	90	8.3	9.787083	3.230386
27	63	64	165	119	130	80	6.9	1.114860	3.284194
28	67	65	295	302	110	80	7.8	15.991527	3.337256
29	58	53	189	166	130	90	5.4	5.417827	3.389625
30	69	65	207	139	110	70	5.5	6.080801	3.441351
31	74	55	197	144	130	80	6.6	3.518724	3.492478
32	71	58	147	54	110	80	8.3	7.241715	3.543047
33	65	51	185	177	130	90	6	5.827357	3.593095
34	72	66	204	65	120	80	6.3	5.302459	3.642657
35	65	70	206	114	130	80	5.9	1.990014	3.691766
36	52	48	190	121	130	90	5.8	7.727792	3.740451
37	62	62	173	83	160	90	10.9	20.03966	3.788741
38	59	63	198	99	120	80	7.9	1.982399	3.836660
39	53	71	266	145	150	90	7.9	10.139596	3.884233
40	64	59	171	112	120	80	7.2	1.459415	3.931484

41	60	69	145	89	130	80	7.8	3.288510	3.978433
42	61	51	205	139	110	70	5.8	3.279994	4.025101
43	63	73	143	243	140	90	8.3	9.638323	4.071506
44	55	41	209	128	110	80	5.6	7.927995	4.117668
45	57	56	267	183	140	80	8.2	7.340043	4.163601
46	71	70	176	114	130	80	6.8	3.004178	4.209324
47	69	59	188	386	130	80	7.6	16.364028	4.254852
48	47	60	299	91	120	70	5.6	10.844403	4.300198
49	67	52	183	63	130	90	6.5	6.132646	4.345378
50	43	78	197	264	120	80	10.2	13.293469	4.390405
51	55	80	200	108	110	80	6.9	5.673130	4.435291
52	47	62	358	333	110	70	7.7	22.015403	4.480049
53	57	92	143	45	130	90	5.4	14.352774	4.524692
54	65	51	185	177	130	90	6	5.827357	4.569229
55	74	76	114	81	120	80	5.2	10.883398	4.613674
56	56	51	118	74	130	80	5.2	6.513437	4.658036
57	53	65	235	175	150	90	3.7	7.815569	4.702326
58	47	62	358	333	110	70	7.7	22.015403	4.746554
59	71	60	155	200	110	70	4.4	10.303494	4.790729
60	54	64	127	68	130	80	4.9	5.494662	4.834862

61	55	80	200	108	110	80	6.9	5.673130	4.878962
62	43	79	168	102	120	80	10.2	13.011550	4.923038
63	66	58	173	112	130	80	6.1	1.251598	4.967098
64	52	70	149	56	120	70	7.9	6.866724	5.011152
65	59	54	235	354	130	70	7.8	14.543358	5.055208
66	64	41	199	71	150	80	6.2	10.213301	5.099275
67	73	49	176	113	130	80	7.3	4.105	5.143361
68	59	61	224	42	130	80	6	3.702183	5.187473
69	68	67	190	81	140	80	8.7	5.904138	5.231621
70	66	67	228	115	110	70	7.6	5.874339	5.27581
71	63	61	288	168	130	80	6.7	6.430396	5.320055
72	62	54	171	260	140	90	6.6	8.714099	5.364356
73	62	64	197	115	120	80	5.6	1.342770	5.408724
74	71	65	156	67	120	80	8.9	5.022751	5.453166
75	61	65	132	170	130	80	7.4	4.308823	5.497691
76	59	60	200	144	130	80	9.2	4.016175	5.542305
77	64	53	154	110	110	70	9.1	6.894447	5.587017
78	63	60	209	155	130	80	7.8	1.440334	5.631835
79	65	54	143	99	130	80	4.6	3.964033	5.676765
80	69	50	200	111	120	80	2	11.338096	5.721816

81	60	63	185	133	120	80	6.3	0.633584	5.766995
82	66	45	130	70	130	80	4	7.309076	5.812311
83	71	55	212	160	110	80	7.7	7.039189	5.857770
84	60	61	169	150	130	90	5	5.034064	5.903382
85	57	60	201	85	120	80	7	1.811691	5.94915
86	58	52	239	128	130	90	5.1	6.924529	5.995091
87	55	60	221	166	110	80	6.6	3.967599	6.041206
88	58	55	148	73	110	80	7	5.380424	6.087504
89	70	59	182	126	130	80	6	2.013797	6.133995
90	57	69	246	150	130	70	4.2	7.042816	6.180687
91	57	58	190	146	120	80	5.2	1.238171	6.227587
92	43	57	177	140	120	70	4.5	5.363312	6.274706
93	61	54	186	103	120	70	6.3	2.519865	6.322050
94	53	55	176	165	110	70	7.3	3.768381	6.369631
95	55	63	229	183	120	80	5.5	1.609664	6.417456
96	65	58	216	96	140	80	4.6	4.191239	6.465535
97	55	53	226	175	120	70	5.6	2.632761	6.513878
98	53	61	190	43	140	80	4.4	5.669416	6.562493
99	54	70	220	140	110	70	7	3.144760	6.611392
100	63	61	173	78	120	70	5.9	3.268556	6.660584

101	61	81	170	102	130	80	4.9	5.683350	6.710079
102	50	75	227	125	140	80	7.8	6.227801	6.759889
103	53	45	210	191	120	70	4.9	4.548824	6.810023
104	65	58	216	96	140	80	4.6	4.191239	6.860494
105	51	46	220	107	120	70	4.7	4.452171	6.911313
106	50	72	241	186	120	80	5.4	3.071970	6.962491
107	58	79	190	141	130	80	5.6	2.973527	7.014041
108	69	70	173	171	140	80	5	6.061858	7.065975
109	50	75	227	125	140	80	7.8	6.227801	7.118307
110	53	64	190	43	130	70	4.4	7.114846	7.171049
111	61	91	198	270	140	90	7.1	10.991569	7.224216
112	39	52	192	58	110	70	4.9	8.432096	7.277822
113	69	68	173	171	140	80	5	5.695667	7.331881
114	49	66	200	178	150	90	4.1	8.332929	7.386409
115	58	66	189	222	120	70	7.9	4.413172	7.441422
116	62	60	177	75	130	70	6.3	5.028018	7.496935
117	41	59	226	178	120	80	5.6	4.971323	7.552967
118	44	50	175	308	110	70	6.1	13.388443	7.609534
119	41	57	196	114	120	80	4.4	5.865582	7.666655
120	54	60	218	197	120	80	5.9	1.539936	7.724349

121	48	75	185	301	140	90	7	11.682074	7.782637
122	55	78	187	99	110	70	7.6	5.073557	7.841538
123	55	69	216	109	140	80	7.2	3.809696	7.901075
124	47	56	262	394	160	90	3.5	24.434538	7.961269
125	58	74	191	83	110	70	6.2	4.503886	8.022145
126	52	64	280	158	140	90	6.7	8.394151	8.083728
127	38	69	203	206	110	70	3.3	8.887584	8.146042
128	42	75	129	146	100	60	5.7	15.210922	8.209115
129	65	71	238	84	150	90	4.7	9.280339	8.272976
130	62	61	147	109	120	80	8.5	3.501628	8.337653
131	52	78	163	235	120	70	12	19.072673	8.403179
132	54	37	208	131	130	80	5.5	6.178856	8.469586
133	36	44	171	74	110	70	2.5	14.041215	8.536909
134	60	76	204	154	110	80	5	6.325050	8.605184
135	54	65	184	97	140	80	4.5	4.076480	8.674450
136	51	84	170	143	110	80	5.7	7.523580	8.744747
137	59	72	231	320	110	70	6	9.738344	8.816118
138	55	94	259	204	120	80	5.3	12.006884	8.888609
139	51	46	220	107	120	70	4.7	4.452171	8.962269
140	38	69	203	206	110	70	3.3	8.887584	9.037147

141	64	48	209	113	120	80	6.7	3.241545	9.113299
142	73	51	209	121	110	80	4.6	10.010892	9.190783
143	48	75	295	65	130	70	3.9	16.533035	9.269661
144	44	69	187	82	110	70	6.1	5.156300	9.349997
145	62	59	191	228	130	80	5.8	2.409738	9.431864
146	50	75	188	195	120	80	6.1	2.871773	9.515337
147	51	84	170	143	110	80	5.7	7.523580	9.600496
148	54	68	239	133	130	70	5.7	5.137662	9.687429
149	57	45	248	110	150	90	2.7	12.238004	9.776230
150	61	78	186	217	130	80	9.4	5.74019	9.867000
151	52	59	220	205	140	70	4.8	8.647455	9.959849
152	53	76	195	124	140	90	4.5	6.508135	10.054895
153	64	56	237	137	120	70	6.7	4.153173	10.152268
154	52	60	280	158	130	80	6.7	5.144298	10.252108
155	61	67	132	170	120	80	7.4	4.083242	10.354569
156	39	62	227	291	120	70	4.5	9.125755	10.459818
157	64	74	208	116	120	70	7.6	5.063056	10.568040
158	55	78	187	99	110	70	7.6	5.073557	10.679439
159	60	72	231	320	120	70	6	9.362012	10.794238
160	47	56	262	394	140	80	3.5	17.242804	10.912688

161	54	59	218	197	120	80	5.9	1.623950	11.035065
162	58	56	138	68	130	80	5.9	3.530965	11.161680
163	53	47	179.4	162	120	70	4.1	4.906995	11.292881
164	65	72	238	84	170	90	4.3	19.103003	11.429061
165	52	62	283	109	130	80	5	6.535047	11.570666
166	60	75	204	154	120	80	5	3.257300	11.718204
167	56	66	257.2	301	130	80	4	7.556982	11.872259
168	39	40	157	87	130	70	5.1	15.243793	12.033506
169	70	56	122	114	120	80	6	5.322912	12.202729
170	80	55	150	142	130	80	6	7.399162	12.380851
171	61	61	239	115	140	80	6.7	3.672975	12.568965
172	79	68	167	203	120	70	5.9	12.410951	12.768377
173	58	60	208	176	120	80	5.9	0.985538	12.980674
174	59	72	170	125	120	70	6.6	3.513533	13.207802
175	59	60	162	134	120	80	8	2.388952	13.452187
176	45	70	215	167	120	70	4.8	4.102017	13.716911
177	41	42	190	59	120	80	6.4	11.962588	14.005970
178	48	71	203	145	110	80	6.5	4.525552	14.324678
179	62	45	175	49	100	70	7.2	8.625573	14.680327
180	50	110	183	154	120	90	7.9	21.813174	15.083306

181	43	55	173	95	110	70	5.2	5.666567	15.549176
182	67	75	243	150	120	80	6	5.914231	16.102803
183	38	70	201	146	120	70	3.3	8.366133	16.787761
184	36	64	159	100	110	70	4.2	9.826852	17.691765
185	56	69	246	123	120	70	4.2	5.952826	19.039494
186	50	65	209	126	140	90	6	5.094619	21.857468

$$\begin{aligned}
 \text{Uji Normalitas} &= \frac{\sum D_i^2 \leq \chi_h^2(\frac{i-\frac{1}{2}}{n})}{\sum D_i^2 \leq \chi_h^2(\frac{i-\frac{1}{2}}{n}) + \sum D_i^2 > \chi_h^2(\frac{i-\frac{1}{2}}{n})} \times 100\% \\
 &= \frac{100}{186} \times 100\% \\
 &= 53.763\%
 \end{aligned}$$

LAMPIRAN 4

Skor Diskriminan Analisis Diskriminan Linear Dan Analisis Diskriminan
Kuadratik Pada Data *Training*

i	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	SD ADL (\hat{y})	SD ADK (\hat{z})
1	1	67	56	167	46	140	80	6.8	11.902902	1.633326
2	1	59	49	203	183	120	70	7.4	11.491058	0.683331
3	1	68	64	171	103	120	80	8.6	13.919722	2.617144
4	1	61	91	198	270	140	90	8.8	10.98964	-0.930051
5	1	72	52	158	96	120	80	7.2	14.350837	3.837061
6	1	62	56	170	149	130	80	8.8	13.252083	2.766292
7	1	74	51	144	261	130	70	5.5	10.887747	4.378519
8	1	71	58	174	168	130	80	5.3	11.564897	2.143476
9	1	61	51	205	139	110	70	5.8	11.129466	0.700568
10	1	58	57	184	126	110	80	6.9	12.658399	1.709863
11	1	64	60	171	112	120	80	7.2	12.713857	2.072395
12	1	63	80	204	117	130	80	11.1	13.198452	3.124061
13	1	56	50	209	147	130	90	4.2	11.227588	0.720197
14	1	58	65	198	99	110	70	8.6	11.754637	0.584507
15	1	66	59	173	112	130	80	6.4	11.789299	1.590844
16	1	62	62	173	83	130	80	10.9	14.431964	4.570556
17	1	55	61	223	165	110	80	6.3	11.492097	1.362691
18	1	71	59	264	153	130	90	9	15.885826	4.468951
19	1	71	56	142	107	130	80	4.8	11.439478	3.463868
20	1	64	57	221	83	130	80	9.5	14.092336	3.993181

Keterangan:

SD ADL (\hat{y}) = Skor Diskriminan Analisis Diskriminan Linear

SD ADK (\hat{z}) = Skor Diskriminan Analisis Diskriminan Kuadratik

21	1	66	64	220	126	130	80	5	10.365814	-0.550227
22	1	67	40	245	110	130	70	6.5	12.037393	2.158577
23	1	71	53	164	91	130	70	7.8	12.379957	2.409617
24	1	67	57	183	63	130	90	6.5	13.809297	3.393438
25	1	55	41	209	128	110	80	5.6	12.661574	1.616
26	1	63	75	143	243	140	90	8.3	12.167233	2.456338
27	1	63	64	165	119	130	80	6.9	11.367645	1.243786
28	1	67	65	295	302	110	80	7.8	13.542904	4.926927
29	1	58	53	189	166	130	90	5.4	12.057758	1.811265
30	1	69	65	207	139	110	70	5.5	10.734518	1.055547
31	1	74	55	197	144	130	80	6.6	13.194893	2.507627
32	1	71	58	147	54	110	80	8.3	15.276411	3.57909
33	1	65	51	185	177	130	90	6	13.486594	3.155582
34	1	72	66	204	65	120	80	6.3	12.641086	1.278384
35	1	65	70	206	114	130	80	5.9	10.427921	-0.601147
36	1	52	48	190	121	130	90	5.8	12.116779	1.718689
37	1	62	62	173	83	160	90	10.9	14.018568	6.618288
38	1	59	63	198	99	120	80	7.9	12.428366	1.923086
39	1	53	71	266	145	150	90	7.9	10.615671	0.993056
40	1	64	59	171	112	120	80	7.2	12.795473	2.146129
41	1	60	69	145	89	130	80	7.8	11.296326	1.609024
42	1	61	51	205	139	110	70	5.8	11.129466	0.700568
43	1	63	73	143	243	140	90	8.3	12.330466	2.657943
44	1	55	41	209	128	110	80	5.6	12.661574	1.616
45	1	57	56	267	183	140	80	8.2	11.571884	2.664221

46	1	71	70	176	114	130	80	6.8	11.790176	0.943885
47	1	69	59	188	386	130	80	7.6	12.561282	1.71846
48	1	47	60	299	91	120	70	5.6	8.065454	0.012371
49	1	67	52	183	63	130	90	6.5	14.217378	3.833449
50	1	43	78	197	264	120	80	10.2	10.670743	1.088698
51	1	55	80	200	108	110	80	6.9	10.465074	-0.078845
52	1	47	62	358	333	110	70	7.7	9.731768	9.364792
53	1	57	92	143	45	130	90	5.4	8.928553	0.460712
54	1	65	51	185	177	130	90	6	13.486594	3.155582
55	1	74	76	114	81	120	80	5.2	11.145808	3.751453
56	1	56	51	118	74	130	80	5.2	10.366519	1.034161
57	1	53	65	235	175	150	90	3.7	7.925891	-5.059282
58	1	47	62	358	333	110	70	7.7	9.731768	9.364792
59	1	71	60	155	200	110	70	4.4	10.417295	2.968997
60	1	54	64	127	68	130	80	4.9	8.861239	-0.063378
61	1	55	80	200	108	110	80	6.9	10.465074	-0.078845
62	1	43	79	168	102	120	80	10.2	10.852364	2.28376
63	1	66	58	173	112	130	80	6.1	11.649628	1.542465
64	1	52	70	149	56	120	70	7.9	9.477902	-1.625142
65	1	59	54	235	354	130	70	7.8	10.441747	1.298116
66	1	64	41	199	71	150	80	6.2	11.64635	0.063130
67	1	73	49	176	113	130	80	7.3	14.116507	3.558595
68	1	59	61	224	42	130	80	6	10.65563	-0.245110
69	1	68	67	190	81	140	80	8.7	12.485682	1.679257
70	1	66	67	228	115	110	70	7.6	11.819357	1.336404

71	1	63	61	288	168	130	80	6.7	11.488086	1.788384
72	1	62	54	171	260	140	90	6.6	12.501141	2.063225
73	1	62	64	197	115	120	80	5.6	10.98343	0.630731
74	1	71	65	156	67	120	80	8.9	14.472883	2.934514
75	1	61	65	132	170	130	80	7.4	11.292286	1.764305
76	2	55	63	229	183	120	80	5.5	10.05235	-0.228523
77	2	65	58	216	96	140	80	4.6	9.829881	-1.598233
78	2	55	53	226	175	120	70	5.6	9.388261	-1.082438
79	2	53	61	190	43	140	80	4.4	8.058025	-3.313604
80	2	54	70	220	140	110	70	7	9.629339	-0.821048
81	2	63	61	173	78	120	70	5.9	10.049352	-0.255521
82	2	61	81	170	102	130	80	4.9	8.297988	-1.852743
83	2	50	75	227	125	140	80	7.8	8.947461	-0.419943
84	2	53	45	210	191	120	70	4.9	9.240097	-2.293265
85	2	65	58	216	96	140	80	4.6	9.829881	-1.598233
86	2	51	46	220	107	120	70	4.7	8.928176	-2.886499
87	2	50	72	241	186	120	80	5.4	8.644996	-1.49564
88	2	58	79	190	141	130	80	5.6	8.563206	-2.105714
89	2	69	70	173	171	140	80	5	9.4565	-0.334865
90	2	50	75	227	125	140	80	7.8	8.947461	-0.419943
91	2	53	64	190	43	130	70	4.4	6.90724	-4.469231
92	2	61	91	198	270	140	90	7.1	9.735681	-2.685602
93	2	39	52	192	58	110	70	4.9	7.857245	-6.32314
94	2	69	68	173	171	140	80	5	9.619732	-0.084779
95	2	49	66	200	178	150	90	4.1	7.618903	-4.118955

96	2	58	66	189	222	120	70	7.9	10.269255	-0.466738
97	2	62	60	177	75	130	70	6.3	9.654411	-0.672253
98	2	41	59	226	178	120	80	5.6	8.766265	-2.297774
99	2	44	50	175	308	110	70	6.1	9.048334	-6.177523
100	2	41	57	196	114	120	80	4.4	8.131666	-3.848833
101	2	54	60	218	197	120	80	5.9	10.436455	-0.082256
102	2	48	75	185	301	140	90	7	9.329074	-1.160503
103	2	55	78	187	99	110	70	7.6	9.583421	-1.66825
104	2	55	69	216	109	140	80	7.2	9.617746	-0.672517
105	2	47	56	262	394	160	90	3.5	6.761036	-8.466266
106	2	58	74	191	83	110	70	6.2	9.272074	-1.631602
107	2	52	64	280	158	140	90	6.7	10.829982	1.258816
108	2	38	69	203	206	110	70	3.3	4.914252	-8.201752
109	2	42	75	129	146	100	60	5.7	5.813097	-9.343828
110	2	65	71	238	84	150	90	4.7	9.789786	-4.424225
111	2	62	61	147	109	120	80	8.5	13.333131	2.492928
112	2	52	78	163	235	120	70	12	11.542138	0.879107
113	2	54	37	208	131	130	80	5.5	11.467791	-0.493874
114	2	36	44	171	74	110	70	2.5	6.329729	-10.34917
115	2	60	76	204	154	110	80	5	9.91598	-0.355195

116	2	54	65	184	97	140	80	4.5	7.824382	-3.049499
117	2	51	84	170	143	110	80	5.7	8.680387	-1.838997
118	2	59	72	231	320	110	70	6	9.021389	-1.174822
119	2	55	94	259	204	120	80	5.3	7.364199	-4.139157
120	2	51	46	220	107	120	70	4.7	8.928176	-2.886499
121	2	38	69	203	206	110	70	3.3	4.914252	-8.201752
122	2	64	48	209	113	120	80	6.7	13.356832	2.735126
123	2	73	51	209	121	110	80	4.6	13.296417	4.250441
124	2	48	75	295	65	130	70	3.9	5.091317	-5.931789
125	2	44	69	187	82	110	70	6.1	7.912038	-4.059186
126	2	62	59	191	228	130	80	5.8	10.672163	0.318011
127	2	50	75	188	195	120	80	6.1	8.852742	-1.881063
128	2	51	84	170	143	110	80	5.7	8.680387	-1.838997
129	2	54	68	239	133	130	70	5.7	7.543919	-2.876454
130	2	57	45	248	110	150	90	2.7	9.431954	-4.540046
131	2	61	78	186	217	130	80	9.4	11.671151	1.036245
132	2	52	59	220	205	140	70	4.8	6.56747	-4.60832
133	2	53	76	195	124	140	90	4.5	8.332983	-2.322685
134	2	64	56	237	137	120	70	6.7	11.120621	1.410801
135	2	52	60	280	158	130	80	6.7	10.25051	1.430075

136	2	61	67	132	170	120	80	7.4	11.788719	1.499336
137	2	39	62	227	291	120	70	4.5	5.701725	-7.24156
138	2	64	74	208	116	120	70	7.6	10.326811	-0.410431
139	2	55	78	187	99	110	70	7.6	9.583421	-1.66825
140	2	60	72	231	320	120	70	6	8.482619	-1.274066
141	2	47	56	262	394	140	80	3.5	6.514766	-6.12646
142	2	54	59	218	197	120	80	5.9	10.518071	-0.018379
143	2	58	56	138	68	130	80	5.9	10.74527	1.108502
144	2	53	47	179.4	162	120	70	4.1	8.511036	-3.150657
145	2	65	72	238	84	170	90	4.3	8.093788	-8.966609
146	2	52	62	283	109	130	80	5	8.923527	-1.170747
147	2	60	75	204	154	120	80	5	9.33793	-1.172809
148	2	56	66	257.2	301	130	80	4	7.976919	-2.819535
149	2	39	40	157	87	130	70	5.1	7.581561	-8.801306

LAMPIRAN 5

Skor Diskriminan Analisis Diskriminan Linear Dan Analisis Diskriminan

Kuadratik Pada Data *Testing*

i	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	SD ADL (\hat{y})	SD ADK (\hat{z})
1	1	59	60	200	144	130	80	9.2	12.893891	2.909977
2	1	64	53	154	110	110	70	9.1	13.769001	1.210932
3	1	63	60	209	155	130	80	7.8	12.333252	1.80574
4	1	65	54	143	99	130	80	4.6	10.744997	2.197508
5	1	69	50	200	111	120	80	2	10.326727	2.586647
6	1	60	63	185	133	120	80	6.3	11.296651	0.828883
7	1	66	45	130	70	130	80	4	11.197965	3.493533
8	1	71	55	212	160	110	80	7.7	14.947839	3.814416
9	1	60	61	169	150	130	90	5	11.362158	1.873949
10	1	57	60	201	85	120	80	7	11.795263	1.341182
11	1	58	52	239	128	130	90	5.1	12.030923	1.51233
12	1	55	60	221	166	110	80	6.6	11.791416	1.567568
13	1	58	55	148	73	110	80	7	12.957669	1.837619
14	1	70	59	182	126	130	80	6	11.960925	1.727052
15	1	57	69	246	150	130	70	4.2	6.69449	-3.932428
16	1	57	58	190	146	120	80	5.2	10.511938	0.0906822
17	1	43	57	177	140	120	70	4.5	6.818095	-6.445247
18	1	61	54	186	103	120	70	6.3	10.640968	-0.0634981
19	1	53	55	176	165	110	70	7.3	10.869751	-1.5721
20	2	70	56	122	114	120	80	6	12.832907	4.345046

Keterangan:

SD ADL (\hat{y}) = Skor Diskriminan Analisis Diskriminan Linear

SD ADK (\hat{z}) = Skor Diskriminan Analisis Diskriminan Kuadratik

21	2	80	55	150	142	130	80	6	13.438994	4.753021
22	2	61	61	239	115	140	80	6.7	10.637211	-0.330811
23	2	79	68	167	203	120	70	5.9	11.183737	2.721348
24	2	58	60	208	176	120	80	5.9	10.94855	0.479730
25	2	59	72	170	125	120	70	6.6	9.097689	-1.449448
26	2	59	60	162	134	120	80	8	12.652091	1.785645
27	2	45	70	215	167	120	70	4.8	6.20612	-5.251571
28	2	41	42	190	59	120	80	6.4	10.923986	-1.932754
29	2	48	71	203	145	110	80	6.5	9.994918	-0.405312
30	2	62	45	175	49	100	70	7.2	13.56615	1.028267
31	2	50	110	183	154	120	90	7.9	8.958224	-1.488384
32	2	43	55	173	95	110	70	5.2	8.234099	-5.428189
33	2	67	75	243	150	120	80	6	10.964045	-0.391037
34	2	38	70	201	146	120	70	3.3	4.278326	-8.846805
35	2	36	64	159	100	110	70	4.2	5.894136	-7.861294
36	2	56	69	246	123	120	70	4.2	7.281481	-2.913146
37	2	50	65	209	126	140	90	6	9.983529	-0.162672

LAMPIRAN 6

Kelompok Prediksi Analisis Diskriminan Linear Dan Analisis Diskriminan
Kuadratik Pada Data *Training*

Jika KA = Kelompok Aktual dan KP = Kelompok Prediksi, maka

i	KA	KP DTr ADL	KP DTr ADK
1	1	1	1
2	1	1	1
3	1	1	1
4	1	1	2
5	1	1	1
6	1	1	1
7	1	1	1
8	1	1	1
9	1	1	1
10	1	1	1
11	1	1	1
12	1	1	1
13	1	1	1
14	1	1	1
15	1	1	1

Keterangan:

KP DTr ADL = Kelompok Prediksi Data *Training* Analisis Diskriminan Linear

KP DTr ADK = Kelompok Prediksi Data *Training* Analisis Diskriminan Kuadratik

i	KA	KP DTr ADL	KP DTr ADK
16	1	1	1
17	1	1	1
18	1	1	1
19	1	1	1
20	1	1	1
21	1	2	2
22	1	1	1
23	1	1	1
24	1	1	1
25	1	1	1
26	1	1	1
27	1	1	1
28	1	1	1
29	1	1	1
30	1	1	1
31	1	1	1
32	1	1	1
33	1	1	1
34	1	1	1
35	1	2	2
36	1	1	1

i	KA	KP DTr ADL	KP DTr ADK
37	1	1	1
38	1	1	1
39	1	1	1
40	1	1	1
41	1	1	1
42	1	1	1
43	1	1	1
44	1	1	1
45	1	1	1
46	1	1	1
47	1	1	1
48	1	2	1
49	1	1	1
50	1	1	1
51	1	1	2
52	1	2	1
53	1	2	1
54	1	1	1
55	1	1	1
56	1	2	1
57	1	2	2

i	KA	KP DTr ADL	KP DTr ADK
58	1	2	1
59	1	2	1
60	1	2	2
61	1	1	2
62	1	1	1
63	1	1	1
64	1	2	2
65	1	2	1
66	1	1	1
67	1	1	1
68	1	1	2
69	1	1	1
70	1	1	1
71	1	1	1
72	1	1	1
73	1	1	1
74	1	1	1
75	1	1	1
76	2	2	2
77	2	2	2
78	2	2	2

i	KA	KP DTr ADL	KP DTr ADK
79	2	2	2
80	2	2	2
81	2	2	2
82	2	2	2
83	2	2	2
84	2	2	2
85	2	2	2
86	2	2	2
87	2	2	2
88	2	2	2
89	2	2	2
90	2	2	2
91	2	2	2
92	2	2	2
93	2	2	2
94	2	2	2
95	2	2	2
96	2	2	2
97	2	2	2
98	2	2	2
99	2	2	2

i	KA	KP DTr ADL	KP DTr ADK
100	2	2	2
101	2	2	2
102	2	2	2
103	2	2	2
104	2	2	2
105	2	2	2
106	2	2	2
107	2	1	1
108	2	2	2
109	2	2	2
110	2	2	2
111	2	1	1
112	2	1	1
113	2	1	2
114	2	2	2
115	2	2	2
116	2	2	2
117	2	2	2
118	2	2	2
119	2	2	2
120	2	2	2

i	KA	KP DTr ADL	KP DTr ADK
121	2	2	2
122	2	1	1
123	2	1	1
124	2	2	2
125	2	2	2
126	2	1	1
127	2	2	2
128	2	2	2
129	2	2	2
130	2	2	2
131	2	1	1
132	2	2	2
133	2	2	2
134	2	1	1
135	2	2	1
136	2	1	1
137	2	2	2
138	2	2	2
139	2	2	2
140	2	2	2

i	KA	KP DTr ADL	KP DTr ADK
141	2	2	2
142	2	1	2
143	2	1	1
144	2	2	2
145	2	2	2
146	2	2	2
147	2	2	2
148	2	2	2
149	2	2	2

LAMPIRAN 7

Kelompok Prediksi Analisis Diskriminan Linear Dan Analisis Diskriminan
Kuadratik Pada Data *Testing*

i	KA	KP DTs ADL	KP DTs ADK
1	1	1	1
2	1	1	1
3	1	1	1
4	1	1	1
5	1	2	1
6	1	1	1
7	1	1	1
8	1	1	1
9	1	1	1
10	1	1	1
11	1	1	1
12	1	1	1
13	1	1	1
14	1	1	1
15	1	2	2

Keterangan:

KP DTs ADL = Kelompok Prediksi Data *Testing* Analisis Diskriminan Linear

KP DTs ADK = Kelompok Prediksi Data *Testing* Analisis Diskriminan Kuadratik

i	KA	KP DTs ADL	KP DTs ADK
16	1	2	1
17	1	2	2
18	1	1	2
19	1	1	2
20	2	1	1
21	2	1	1
22	2	1	2
23	2	1	1
24	2	1	1
25	2	2	2
26	2	1	1
27	2	2	2
28	2	1	2
29	2	2	2
30	2	1	1
31	2	2	2
32	2	2	2
33	2	1	2
34	2	2	2
35	2	2	2
36	2	2	2
37	2	2	2

LAMPIRAN 8

Skrip Program Analisis Diskriminan Dengan *Software R* Versi 3.3.2

```
library(MASS)
library(car)

#Seluruh Data
all<-read.table("D:DT.txt",header=T)
xx1<-all$x1
xx2<-all$x2
xx3<-all$x3
xx4<-all$x4
xx5<-all$x5
xx6<-all$x6
xx7<-all$x7
yy<-all$y
all3<-cbind(xx1,xx2,xx3,xx4,xx5,xx6,xx7)
all3x<-as.matrix(all3)

nn<-nrow(all3)
pp<-ncol(all3)
g<-2
nn1<-94
nn2<-92
mean<-colMeans(all3)
sx<-cov(all3)
```

#Menghitung Jarak Mahalanobis

```

D<-mahalanobis(all3,mean,sx)
chisq<-qchisq(ppoints(nn),df=pp)
qqplot(D,chisq,main="chisquare-plot",ylab="Khi-Kuadrat",xlab="Jarak Ma-
halanobis")
abline(a=0,b=1)

NN1<-rbind(1:nn1)
all4<-all3[NN1,]
nkel1<-nrow(all4)
skel1<-cov(all4)
detskel1<-det(skel1)
trans_skel1<-t(skel1)
meankel1<-colMeans(all4)

NN2<-rbind((nn1+1):nn)
all5<-all3[NN2,]
nkel2<-nrow(all5)
skel2<-cov(all5)
trans_skel2<-t(skel2)
detskel2<-det(skel2)
meankel2<-colMeans(all5)

sgabung<-((nn1-1)*sdt+(nn2-1)*sdtt)/(nn1+nn2-2)
detsgabung<-det(sgabung)

```

#Uji Homogenitas

```
t<-((2*(pp^2))+(3*pp)-1)/(6*(pp+1)*(g-1))
r<-(1/(nn1-1))+(1/(nn2-1))-(1/(nn1+nn2-2))
u<-r*t
t<-1-u
M<-((nn1+nn2-2)*log(detsgabung))-(((nn1-1)*log(detskel1)+((nn2-1)-
log(detskel2)))
c<-t*M
v<-0.5*pp*(pp+1)*(g-1)
chisq<-qchisq(c(0.05),df=v,lower.tail=FALSE)
```

#Uji Multikolinearitas

```
reg<-lm(yy ~ xx1+xx2+xx3+xx4+xx5+xx6+xx7)
vif(reg)
```

#Uji Perbedaan Signifikan Antar Kelompok

```
ubv<-manova(cbind(xx1,xx2,xx3,xx4,xx5,xx6,xx7)~yy)
summary.aov(ubv,test="Hotelling-Lawley")
```

#Data *Training*

```
data<-read.table("D:Percobaan8.txt",header=T)
x1<-as.matrix(data$USIA)
x2<-as.matrix(data$BB)
x3<-as.matrix(data$KOLEST)
x4<-as.matrix(data$TRIGLI)
x5<-as.matrix(data$TKS)
x6<-as.matrix(data$TKD)
```

```

x7<-as.matrix(data$AU)
y<-as.matrix(data$PASIEN)
data1<-cbind(x1,x2,x3,x4,x5,x6,x7,y)
data2<-as.matrix(data1)
data3<-cbind(x1,x2,x3,x4,x5,x6,x7)
X<-as.matrix(data3)

```

n<-nrow(X)

p<-ncol(X)

g<-2

n1<-75

n2<-74

N1<-rbind(1:n1)

data4<-data3[N1,]

ndt<-nrow(data4)

sdt<-cov(data4)

sdt_inv<-solve(sdt)

mean1<-colMeans(data4)

trans_mean1<-t(mean1)

detsdt<-det(sdt)

N2<-rbind((n1+1):n)

data5<-data3[N2,]

ndtt<-nrow(data5)

sdtt<-cov(data5)

sdtt_inv<-solve(sdtt)

```

mean2<-colMeans(data5)
trans_mean2<-t(mean2)
detsdtt<-det(sdtt)

sgab<-((n1-1)*sd+ (n2-1)*sdtt)/(n1+n2-2)
detsgab<-det(sgab)
sgab_inv<-solve(sgab)

```

#Data Testing

```

data7<-read.table("D:Percobaan12.txt",header=T)
X1<-as.matrix(data7$USIA)
X2<-as.matrix(data7$BB)
X3<-as.matrix(data7$KOLEST)
X4<-as.matrix(data7$TRIGLI)
X5<-as.matrix(data7$TKS)
X6<-as.matrix(data7$TKD)
X7<-as.matrix(data7$AU)
Y<-as.matrix(data7$PASIEN)
data8<-cbind(X1,X2,X3,X4,X5,X6,X7,Y)
data9<-as.matrix(data8)
data10<-cbind(X1,X2,X3,X4,X5,X6,X7)
x<-as.matrix(data10)

```

#Analisis Diskriminan Linear

```

I<-t(mean1-mean2)%%sgab_inv%%t(X)
trans_I<-t(I)
III<-t(mean1-mean2)%%sgab_inv%%t(x)

```

```
trans_III<-t(III)
```

#Menentukan Nilai *Cut-Off* Untuk Data *Training*

```
w1<-rbind(1:75)
w2<-rbind(76:149)
W1<-trans_I[w1,]
W2<-trans_I[w2,]
meanW1<-sum(W1)/75
meanW2<-sum(W2)/74
m<-((75*meanW1)+(74*meanW2))/(149)
```

#Menentukan Nilai *Cut-Off* Untuk Data *Testing*

```
ww1<-rbind(1:19)
ww2<-rbind(20:37)
WW1<-trans_III[ww1,]
WW2<-trans_III[ww2,]
meanWW1<-sum(WW1)/19
meanWW2<-sum(WW2)/18
mm<-((19*meanWW1)+(18*meanWW2))/(37)
```

```
x0<-matrix(c(x1,x2,x3,x4,x5,x6,x7),ncol=1,nrow=7)
x0<-matrix(c(X1,X2,X3,X4,X5,X6,X7),ncol=1,nrow=7)
trans_x0<-t(x0)
```

#Analisis Diskriminan Kuadratik

```
A<-(-(1/2))*log(detsdt/detsdtt)
B<-(-(1/2))*((trans_mean1%*%sdt_inv%*%mean1)-(trans_mean2%*%
```

```
sdtt_inv%*%mean2))  
C<-(-(1/2))*trans_x0%*%(sdt_inv-sdtt_inv)%*%x0)  
D<-(-(1/2))*trans_x0%*%((sdtt_inv%*%mean2)-(sdt_inv%*%mean1)))  
E<-(-(1/2))*(((trans_mean2%*%sdtt_inv)-(trans_mean1%*%sdt_inv))%*%  
x0)  
QDA<-A+B+C+D+E
```

SURAT PERNYATAAN KEASLIAN SKRIPSI

Dengan ini saya yang bertanda tangan di bawah ini, mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta:

Nama : Bety Suryani Putri

No. Registrasi : 3125121975

Program Studi : Matematika

Menyatakan bahwa skripsi ini yang saya buat dengan judul "**Pengklasifikasian Kejadian Diabetes Dengan Analisis Diskriminan Kuadratik (Studi Kasus: Puskesmas Kecamatan Pesanggrahan Kota Jakarta Selatan)**" adalah :

1. Dibuat dan diselesaikan oleh saya sendiri.
2. Bukan merupakan duplikat skripsi yang pernah dibuat oleh orang lain atau jiplakan karya tulis orang lain.

Pernyataan ini dibuat dengan sesungguhnya dan saya bersedia menanggung segala akibat yang timbul jika pernyataan saya tidak benar.

Jakarta, Juni 2017

Yang membuat pernyataan

Bety Suryani Putri

DAFTAR RIWAYAT HIDUP



BETY SURYANI PUTRI. Lahir di Jakarta, 08 Oktober 1994. Anak kedua dari pasangan Bapak Giran dan Ibu Giyuk Ngadiyem. Saat ini bertempat tinggal di Jalan Masjid Darul Falah No 19 RT 09/ RW 03 Petukangan Utara, Pesanggrahan, Jakarta Selatan 12260.

No. Ponsel : 083872317456

Email : betysp.99@gmail.com

Riwayat Pendidikan : Penulis mengawali pendidikan di TK Islam Tunas Harapan selama 2 tahun, dan kemudian melanjutkan pendidikan di SDI Darunnajah Petukangan pada tahun 2000 - 2006. Setelah itu, penulis melanjutkan ke SMP Negeri 177 Jakarta hingga tahun 2009. Kemudian kembali melanjutkan ke SMA Negeri 47 Jakarta dan lulus tahun 2012. Di tahun yang sama, penulis melanjutkan ke Universitas Negeri Jakarta (UNJ), program studi Matematika, melalui jalur SNMPTN Tulis. Di pertengahan tahun 2017, penulis telah memperoleh gelar Sarjana Sains untuk Program Studi Matematika, FMIPA, UNJ.

Riwayat Organisasi : Selama di bangku perkuliahan, penulis aktif mengikuti suatu organisasi kemahasiswaan. Dalam dua tahun pertama, penulis mendapat kepercayaan sebagai salah satu staff Departemen Profeil BEMJ Matematika.

Riwayat Pekerjaan : Pada pertengahan tahun 2015, penulis memiliki pengalaman magang selama satu bulan di Jasa Raharja.