

Comparison of Linear Discriminant Analysis with Robust Linear Discriminant Analysis

Fitri Hayati, Dodi Vionanda*, Yenni Kurniawati, Tessy Octavia Mukhti

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

*Corresponding author: dodi_vionanda@fmipa.unp.ac.id

Submitted : 02 Agustus 2024

Revised : 12 Agustus 2024

Accepted : 13 Agustus 2024

ABSTRACT

Discriminant analysis is a multivariate method for dividing things into discrete groups and assigning new objects to existing categories. A discriminant function, which is a linear combination of independent variables used to categorize things into two or more groups or categories, is the result of discriminant analysis. The independent variables in a linear discriminant analysis must be multivariate normally distributed, and the covariance matrices for each group must be equal. In linear discriminant analysis, it is also essential to identify outliers because their existence in the data set can undermine the assumptions made by the method and lead to incorrect classification results. Therefore, in discriminant analysis, handling outliers with robust approaches is required. One such robust method in discriminant analysis is the Minimum Covariance Determinant (MCD), which is highly effective in dealing with outliers and relatively easier to apply compared to other robust methods. The aim of this study is to compare the classification results of linear discriminant analysis with robust linear discriminant analysis on the dataset of diabetes patients at RSUD Padangsidempuan in 2023. The results obtained from this dataset indicate that linear discriminant analysis achieved an accuracy of 85,71%, while robust linear discriminant analysis achieved an accuracy of 80,95%. These findings suggest that the use of linear discriminant analysis and robust linear discriminant analysis can yield different results depending on the characteristics of the data and the number of outliers in the dataset.

Keywords: Accuracy, Discriminant analysis, Minimum Covariance Determinant (MCD), Outliers



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Analisis multivariat adalah analisis statistika yang berkaitan dengan analisis simultan yang bertujuan untuk memahami hubungan antara dua atau lebih variabel. Analisis multivariat sangat populer karena mampu dalam pengambilan keputusan di berbagai bidang, seperti kesehatan, keuangan, teknologi, dll. Salah satu teknik analisis multivariat yang umum digunakan dalam pengambilan keputusan adalah analisis diskriminan (Hair et al., 2010).

Analisis diskriminan diperkenalkan oleh ahli statistik bernama Sir Ronald Aylmer Fisher pada tahun 1936. Menurut Johnson dan Wichern (2002), analisis diskriminan merupakan teknik multivariat yang digunakan untuk memisahkan kumpulan objek yang berbeda dan mengalokasikan objek baru pada kelompok yang telah ditentukan sebelumnya. Analisis diskriminan melibatkan suatu variabel diskriminan, dimana variabel dependen berupa kategori dan variabel independen berupa non-kategori. Analisis diskriminan akan menghasilkan suatu fungsi diskriminan. Dimana fungsi diskriminan didefinisikan sebagai kombinasi linier dari variabel-variabel independen yang digunakan untuk mengklasifikasikan objek-objek ke dalam dua atau lebih kelompok atau kategori (Denis, 2016).

Berdasarkan fungsi diskriminan, analisis diskriminan terdiri dari dua jenis, yaitu analisis diskriminan non-linier dan analisis diskriminan linier. Pada penelitian ini analisis yang digunakan adalah analisis diskriminan linier yang bertujuan untuk membangun model yang dapat memisahkan kumpulan objek ke dalam dua atau lebih kelompok berdasarkan kombinasi linier dari variabel-variabel independen. Menurut Hair et al. (2010), Asumsi yang harus dipenuhi dalam analisis diskriminan linier meliputi variabel independen berdistribusi normal multivariat serta homogenitas matriks varians-kovarians di setiap kelompok. Dalam analisis diskriminan linier juga perlu dilakukan identifikasi *outlier*. *Outlier* atau pencilan merupakan data yang berbeda secara signifikan dari pola umum dalam suatu kumpulan data. Dengan adanya *outlier* pada kumpulan data dapat mengakibatkan asumsi analisis diskriminan linier tidak terpenuhi dan hasil klasifikasi analisis diskriminan linier tidak akurat. Oleh karena itu, perlu dilakukan penanganan terhadap

outlier menggunakan pendekatan metode *robust* dalam analisis diskriminan, yang dikenal sebagai analisis diskriminan linier *robust* (Huber dan Ronchetti 2008).

Metode *robust* adalah metode yang digunakan untuk memaksimalkan hasil estimasi terhadap *outlier* dan dapat menangani asumsi yang tidak terpenuhi. Terdapat beberapa jenis metode *robust*, yaitu metode *Minimum Covariance Determinant* (MCD), *Minimum Volume Ellipsoid* (MVE), *S-estimator*, dan *M-estimator*. Jenis metode *robust* dalam analisis diskriminan yang akan digunakan pada penelitian ini adalah *Minimum Covariance Determinant* (MCD). Metode *robust* (MCD) merupakan metode yang sangat populer dan sering digunakan dalam analisis diskriminan linier. Hal ini disebabkan karena metode *robust* (MCD) memiliki kemampuan yang kuat dalam menangani *outlier* dan penerapannya relatif lebih mudah dibandingkan metode *robust* lainnya (Rousseeuw dan Driessen, 1999).

Terdapat beberapa penelitian yang telah dilakukan terkait metode *robust* MCD dalam analisis diskriminan linier. Salah satunya adalah penelitian yang dilakukan oleh Aisyah Nur Amini (2020) dengan tujuan penelitian untuk melihat perbandingan analisis diskriminan linier dengan analisis diskriminan linier *robust* menggunakan metode *robust* MCD dan kesimpulan dari penelitian ini adalah analisis diskriminan linier *robust* MCD lebih efektif dibandingkan analisis diskriminan linier. Sehingga penelitian ini akan menggunakan metode *robust* MCD dalam perbandingan analisis diskriminan linier dan analisis diskriminan linier *robust*. Untuk mengukur persentase kesalahan dalam analisis ini digunakan *Apparent Error Rate* (APER). APER adalah persentase observasi dalam sampel pelatihan yang salah diklasifikasikan oleh fungsi klasifikasi sampel (Johnson dan Wichern, 2002).

Pada penelitian ini akan dilakukan perbandingan analisis diskriminan linier dengan analisis diskriminan linier *robust* pada gugus data pasien diabetes RSUD Padangsidempuan tahun 2023 sebagai studi kasus penelitian. Gugus data ini diklasifikasikan menjadi dua yaitu diabetes tipe gestasional dan diabetes tipe II. Diabetes tipe gestasional terjadi pada ibu hamil, sedangkan diabetes tipe II adalah jenis diabetes yang paling banyak diderita pasien karena disebabkan faktor gaya hidup dan faktor genetik. Adapun variabel independen yang akan digunakan pada gugus data ini meliputi usia, tekanan darah sistolik, tekanan darah diastolik, kadar glukosa darah, kolesterol, dan asam urat. Pada gugus data tersebut terdapat dua variabel independen yang memiliki *outlier* yaitu variabel tekanan darah sistolik dan kolesterol, sehingga gugus data ini cocok digunakan untuk penelitian ini. Variabel independen pada gugus data ini memiliki satuan yang berbeda-beda, sehingga perlu dilakukan standarisasi data. Penelitian ini bertujuan untuk membandingkan kinerja antara analisis diskriminan linier dengan analisis diskriminan linier *robust* pada gugus data pasien diabetes di RSUD Padangsidempuan tahun 2023.

II. METODE PENELITIAN

A. Jenis dan Sumber Data

Jenis data yang akan digunakan dalam penelitian ini adalah data sekunder yang bersumber dari rekam medis pasien diabetes di RSUD Padangsidempuan pada tahun 2023. Data diklasifikasikan menjadi dua tipe diabetes yaitu diabetes tipe gestasional dan diabetes tipe II. Tabel 1 menyajikan variabel yang terdapat pada penelitian ini.

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Skala
Tipe diabetes (Y)	0 : DM Gestasional 1 : DM Tipe II	Nominal
Usia pasien (X_1)	Satuan : Tahun	Rasio
Tekanan darah sistolik (X_2)	Satuan : mmHg	Rasio
Tekanan darah diastolik (X_3)	Satuan : mmHg	Rasio
Kadas glukosa dalam darah pasien (X_4)	Satuan : mg/dL	Rasio
Kolesterol (X_5)	Satuan : mg/dL	Rasio
Asam Urat (X_6)	Satuan : mg/dL	Rasio

B. Teknik Analisis Data

Analisis data pada penelitian ini dilakukan menggunakan *software RStudio* dengan tahapan sebagai berikut:

1. Membuat statistika deskriptif pada gugus data penelitian.
2. Melakukan standarisasi data pada gugus data penelitian.
3. Membagi data secara acak menjadi dua subset, yaitu 80% untuk data pelatihan (*training set*) dan 20% sisanya untuk data pengujian (*testing set*) pada gugus data penelitian.
4. Melakukan pengidentifikasian *outlier* pada gugus data penelitian menggunakan *robust distance* MCD.

Pengidentifikasi *outlier* pada data multivariat menggunakan jarak mahalanobis. Namun, penggunaan jarak mahalanobis kurang efektif karena jarak mahalanobis tidak dapat mengatasi *masking effect*. Sehingga jarak mahalanobis dikembangkan menjadi *robust distance* menggunakan penaksir *robust* MCD untuk penduga rata-rata μ dan matriks varians kovarians Σ (Huber dan Debruyne, 2010). Berdasarkan *robust distance* MCD, maka suatu pengamatan dikatakan *outlier*, jika:

$$(\mathbf{x}_i - \boldsymbol{\mu}_{MCD})' \boldsymbol{\Sigma}_{MCD}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_{MCD}) > \chi_{p,\alpha}^2 \quad (1)$$

dimana, \mathbf{x}_i merupakan vektor amatan ke- i berukuran $p \times 1$, $\boldsymbol{\mu}_{MCD}$ merupakan vektor rata-rata dari seluruh amatan dari penduga MCD berukuran $p \times 1$, dan $\boldsymbol{\Sigma}_{MCD}^{-1}$ merupakan nilai invers dari matriks varians kovarians dari seluruh amatan berukuran $p \times p$.

5. Melakukan uji asumsi analisis diskriminan pada gugus data penelitian.

- a. Pengujian distribusi normal multivariat.

Tujuan pengujian distribusi normal multivariat adalah untuk mengetahui apakah distribusi sebuah data mengikuti atau mendekati distribusi normal. Pengujian ini dapat dilakukan dengan cara membuat *Q-Q plot*. Data dinyatakan berdistribusi normal multivariat jika *Q-Q plot* cenderung membentuk garis lurus lebih dari 50%.

- b. Pengujian kesamaan matriks varians-kovarians

Menurut Rencher (2002) pengujian kesamaan matriks varian-kovarians dapat dilakukan dengan pendekatan χ^2 dan F terhadap distribusi M. Pengujian ini berdasarkan salah satu pendekatan disebut *Box's M-Test*. Adapun hipotesis dari pengujian ini menggunakan *Box's M-Test*, sebagai berikut:

$$H_0 : \Sigma_1 = \Sigma_2 = \Sigma_3 = \dots = \Sigma_k$$

$$H_1 : \Sigma_1 \neq \Sigma_2 \neq \Sigma_3 \neq \dots \neq \Sigma_k$$

Statistik uji:

$$u = -2(1 - c_1) \ln M \quad (2)$$

Kriteria uji:

Tolak H_0 , jika $u > \chi_{\alpha}^2$ dengan derajat kebebasan $[1/2(k - 1)p(p + 1)]$ yang berarti matriks kovarians berbeda.

6. Melakukan perhitungan analisis diskriminan linier dan analisis diskriminan linier *robust*.

- a. Analisis diskriminan linier

Tujuan analisis diskriminan linier adalah menemukan kombinasi linier dari variabel independen yang memaksimalkan perbedaan antara kelompok. Analisis diskriminan linier akan menghasilkan suatu fungsi diskriminan, sebagai berikut:

$$\hat{y} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \boldsymbol{\Sigma}_{pooled}^{-1} \mathbf{x} = \hat{\mathbf{a}}' \mathbf{x} \quad (3)$$

dimana, \hat{y} merupakan fungsi diskriminan linier, $\hat{\mathbf{a}}$ adalah koefisien, dan \mathbf{x} adalah variabel independen.

Dengan diperolehnya fungsi diskriminan linier, maka dapat dilakukan pengalokasian objek baru ke dalam suatu kelompok yang telah ditentukan dengan membandingkan dugaan fungsi diskriminan dengan median dari kelompok.

$$\hat{m} = \frac{1}{2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \boldsymbol{\Sigma}_{pooled}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) = \frac{1}{2} (\bar{y}_1 - \bar{y}_2) \quad (4)$$

dimana,

$$\begin{aligned} \hat{y}_1 &= (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \boldsymbol{\Sigma}_{pooled}^{-1} \bar{\mathbf{x}}_1 = \hat{\mathbf{a}}' \bar{\mathbf{x}}_1 \\ \hat{y}_2 &= (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \boldsymbol{\Sigma}_{pooled}^{-1} \bar{\mathbf{x}}_2 = \hat{\mathbf{a}}' \bar{\mathbf{x}}_2 \end{aligned} \quad (5)$$

dengan kriteria uji jika $\hat{y}_0 \geq \hat{m}$, maka objek pengamatan akan ditempatkan dalam kelompok 1 dan jika $\hat{y}_0 \leq \hat{m}$, maka objek pengamatan akan ditempatkan dalam kelompok 2.

- b. Analisis diskriminan linier *robust*

Analisis diskriminan linier *robust* menggunakan metode *robust* MCD dilakukan dengan mengganti rata-rata μ dan matriks varians kovarians Σ dalam rumus fungsi diskriminan dengan $\boldsymbol{\mu}_{MCD}$ dan $\boldsymbol{\Sigma}_{MCD}$ yang merupakan vektor rata-rata dan matriks kovarians dengan metode *robust* MCD. Rumus fungsi diskriminan linier *robust* untuk dua kelompok adalah

$$\begin{aligned} y_{1R}(x) &= \mathbf{x}' \boldsymbol{\Sigma}_{MCD}^{-1} \boldsymbol{\mu}_{1MCD} - \frac{1}{2} \boldsymbol{\mu}'_{1MCD} \boldsymbol{\Sigma}_{MCD}^{-1} \boldsymbol{\mu}_{1MCD} + \ln(p_1) \\ y_{2R}(x) &= \mathbf{x}' \boldsymbol{\Sigma}_{MCD}^{-1} \boldsymbol{\mu}'_{2MCD} - \frac{1}{2} \boldsymbol{\mu}'_{2MCD} \boldsymbol{\Sigma}_{MCD}^{-1} \boldsymbol{\mu}_{2MCD} + \ln(p_2) \end{aligned} \quad (6)$$

Hasil prediksi responden akan termasuk dalam kelompok p_k jika skor diskriminan linier *robust* sebagai berikut: $y_{kR}(x) = \max\{y_{kR}(x); k = 1,2\}$.

7. Melakukan perhitungan nilai APER pada analisis diskriminan linier dan analisis diskriminan linier *robust*.

Menurut Johnson dan Wichern (2002), *Apparent Error Rate* (APER) adalah persentase dari observasi dalam sampel data pelatihan yang diklasifikasikan secara tidak tepat oleh fungsi klasifikasi. APER dapat dihitung dengan menggunakan *confusion matrix*, yang menunjukkan perbandingan antara klasifikasi aktual dengan klasifikasi hasil prediksi. Untuk n_1 pengamatan dari π_1 dan n_2 pengamatan dari π_2 . Berikut Tabel 2 disajikan *confusion matrix*.

Tabel 2. *Confusion Matrix*

		<i>Predicted membership</i>		Total
		π_1	π_2	
<i>Actual membership</i>	π_1	n_{1C}	$n_{1M} = n_1 - n_{1C}$	n_1
	π_2	$n_{2M} = n_2 - n_{2C}$	n_{2C}	n_2

dimana,

n_{1C} = total pengamatan dari π_1 tepat diklasifikasikan sebagai π_1

n_{1M} = total pengamatan dari π_1 salah diklasifikasikan sebagai π_2

n_{2C} = total pengamatan dari π_2 tepat diklasifikasikan sebagai π_2

n_{2M} = total pengamatan dari π_2 salah diklasifikasikan sebagai π_1

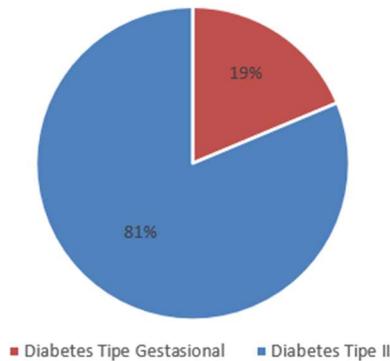
maka, nilai APER dapat dihitung dengan:

$$APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2} \tag{7}$$

III. HASIL DAN PEMBAHASAN

A. Statistika Deskriptif

Data pasien diabetes RSUD Padangsidempuan tahun 2023 terdiri dari 107 pasien, dimana diabetes tipe gestasional sebanyak 20 pasien dan diabetes tipe II sebanyak 87 pasien. Gambar 2 menyajikan diagram lingkaran variabel dependen data pasien diabetes RSUD Padangsidempuan tahun 2023.



Gambar 2. Diagram Lingkaran Variabel Dependen Data Pasien Diabetes RSUD Padangsidempuan Tahun 2023

Berikut adalah deskriptif data dari variabel independen data pasien diabetes RSUD Padangsidempuan tahun 2023:

Tabel 3. Deskriptif Data untuk Diabetes Tipe Gestasional

Variabel Independen	n	Min	Median	Mean	Max
Usia Pasien	20	25,00	32,00	32,70	42,00
Tekanan Darah Sistolik	20	90,00	120,00	123,50	150,00
Tekanan Darah Diastolik	20	60,00	70,00	72,50	100,00
Kadar Glukosa	20	91,00	321,50	196,10	476,00
Kolesterol	20	90,00	183,00	183,70	378,00
Asam Urat	20	2,50	5,10	5,54	9,50

Tabel 4. Deskriptif Data untuk Diabetes Tipe II

Variabel Independen	n	Min	Median	Mean	Max
Usia Pasien	87	36,00	57,00	56,79	88,00
Tekanan Darah Sistolik	87	80,00	120,00	125,40	190,00
Tekanan Darah Diastolik	87	50,00	70,00	73,91	100,00
Kadar Glukosa	87	70,00	286,00	295,30	620,00
Kolesterol	87	90,00	170,00	182,60	451,00
Asam Urat	87	2,00	4,80	5,31	9,40

Berdasarkan Tabel 3 dan Tabel 4 diperoleh nilai-nilai minimum, median, rata-rata, maksimal, dan jumlah pengamatan masing-masing variabel independen berdasarkan tipe diabetes pasien pada data pasien diabetes RSUD Padangsidimpuan tahun 2023.

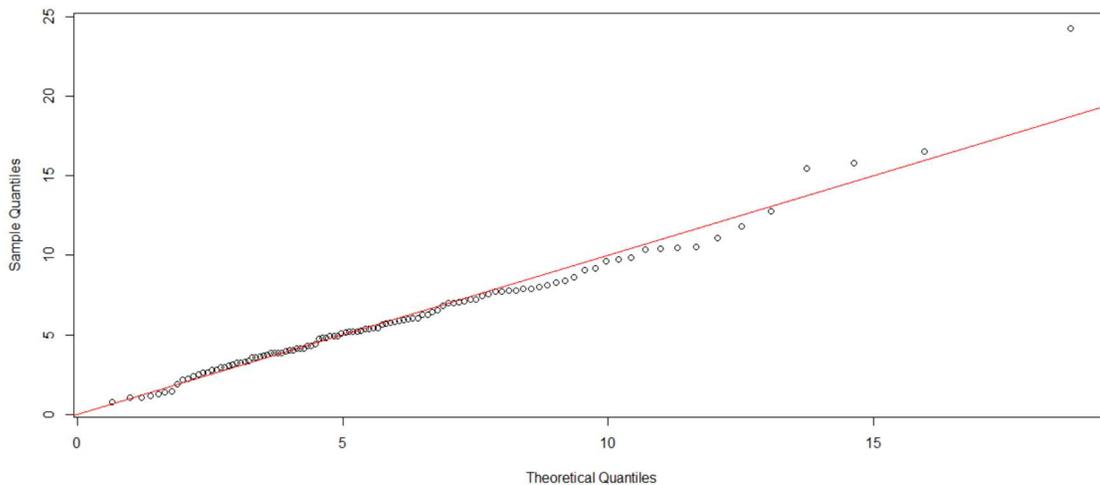
B. Identifikasi Pencilan/Outlier

Berdasarkan pengidentifikasian *outlier* menggunakan *robust distance* (MCD) dengan bantuan *software RStudio*, diperoleh hasil bahwa sebanyak 12 pengamatan dari 86 pengamatan dari data *training set* teridentifikasi sebagai *outlier*. *Outlier* tersebut antara lain berada pada amatan ke 3, 14, 18, 54, 56, 64, 68, 73, 77, 89, 95, dan 98. Dengan demikian, dapat disimpulkan bahwa persentase *outlier* sebanyak 13,95%.

C. Uji Asumsi Analisis Diskriminan

1. Distribusi Normal Multivariat

Gambar 3 menyajikan hasil *Q-Q plot* data pasien diabetes RSUD Padangsidimpuan tahun 2023.



Gambar 3. *Q-Q Plot* Data Pasien Diabetes RSUD Padangsidimpuan tahun 2023

Berdasarkan Gambar 3, data pasien diabetes RSUD Padangsidimpuan tahun 2023 menunjukkan pola distribusi yang mendekati normal multivariat karena sebagian besar titik-titik menyebar membentuk garis lurus yang menunjukkan bahwa adanya kesesuaian pola jarak mahalanobis terhadap sebaran $\chi^2_{p;0,05}$. Namun, kehadiran beberapa *outlier* di bagian atas menunjukkan bahwa data tidak sepenuhnya berdistribusi normal.

2. Kesamaan Matriks Varians-Kovarians

Dengan menggunakan statistik uji pada persamaan (2), maka tabel 5 menyajikan hasil pengujian *box's M-test*.

Tabel 5. *Box's M-test for Homogeneity of Covariance Matrices*

<i>Chi-Sq (approx.)</i>	<i>df</i>	<i>p-value</i>
23,756	21	0,4643

Berdasarkan Tabel 5 diketahui bahwa *p-value* sebesar 0,4643 besar dari taraf signifikansi 5% yang berarti gagal tolak H_0 , sehingga dapat disimpulkan bahwa nilai kesamaan matriks varians kovarians berdasarkan tipe diabetes sama antar kategori atau kelompok.

D. Analisis Diskriminan Linier

Berdasarkan hasil analisis diskriminan linier menggunakan bantuan *Software RStudio* diperoleh nilai *prior probabilities* pengklasifikasian data pasien diabetes RSUD Padangsidempuan tahun 2023 untuk kategori 0 (Diabetes Tipe Gestasional) sebesar 19% dan kategori 1 (Diabetes Tipe II) sebesar 81%. Sehingga fungsi diskriminan dapat dibentuk sebagai berikut:

$$\hat{y} = 1,54Usia - 0,37TS + 0,66TD + 0,13Glukosa - 0,04Kolesterol + 0,06Asam.Urat$$

Dari persamaan tersebut dapat diketahui bahwa koefisien positif pada variabel usia, tekanan darah diastolik, kadar glukosa, dan asam urat menunjukkan bahwa variabel-variabel tersebut mampu meningkatkan skor diskriminan. Sebaliknya, koefisien negatif pada variabel tekanan darah sistolik dan kolesterol menunjukkan bahwa variabel-variabel tersebut mampu menurunkan skor diskriminan. Dengan demikian, persamaan tersebut menggabungkan seluruh variabel yang bertujuan untuk menghasilkan skor diskriminan yang dapat digunakan untuk mengklasifikasikan tipe diabetes pasien ke dalam kelompok atau kategori yang berbeda berdasarkan karakteristik variabelnya.

E. Analisis Diskriminan Linier Robust

Berdasarkan hasil analisis diskriminan linier *robust* menggunakan bantuan *Software RStudio* diperoleh nilai *prior probabilities* pengklasifikasian untuk kategori 0 (Diabetes Tipe Gestasional) sebesar 19% dan kategori 1 (Diabetes Tipe II) sebesar 81%. Sehingga fungsi diskriminan dapat dibentuk sebagai berikut:

$$\hat{y} = -4,93Usia + 1,437TS - 2,95TD - 0,68Glukosa + 0,12Kolesterol - 0,18Asam.Urat$$

Dari persamaan tersebut dapat diketahui bahwa koefisien positif pada variabel tekanan darah sistolik dan kolesterol menunjukkan bahwa variabel-variabel tersebut mampu meningkatkan skor diskriminan. Sebaliknya, koefisien negatif pada variabel usia, tekanan darah diastolik, kadar glukosa, dan asam urat menunjukkan bahwa variabel-variabel tersebut mampu menurunkan skor diskriminan. Dengan demikian, persamaan tersebut menggabungkan seluruh variabel yang bertujuan untuk menghasilkan skor diskriminan yang dapat digunakan untuk mengklasifikasikan tipe diabetes pasien ke dalam kelompok atau kategori yang berbeda berdasarkan karakteristik variabelnya.

F. Perbandingan Hasil Estimasi Ketepatan Klasifikasi Analisis Diskriminan Linier dengan Analisis Diskriminan Linier Robust

Berdasarkan hasil perhitungan dari *confusion matrix*, diperoleh proporsi kesalahan pengklasifikasian dalam analisis diskriminan linier sebesar 14,29%, sedangkan proporsi kesalahan pengklasifikasian dalam analisis diskriminan linier *robust* sebesar 19,05%. Sehingga dapat disimpulkan bahwa analisis diskriminan linier lebih akurat pada gugus data pasien diabetes RSUD Padangsidempuan tahun 2023 dibandingkan analisis diskriminan linier *robust*. Hal ini dapat disebabkan karena jumlah outlier dalam gugus data relatif kecil dan sebagian besar data mengikuti distribusi normal multivariat.

IV. KESIMPULAN

Berdasarkan hasil perbandingan analisis diskriminan linier dengan analisis diskriminan linier *robust* pada penelitian ini, didapatkan $\hat{y} = 1,54Usia - 0,37TS + 0,66TD + 0,13Glukosa - 0,04Kolesterol + 0,06Asam.Urat$ sebagai fungsi diskriminan linier yang menghasilkan persentase kesalahan pengklasifikasian sebesar 14,29% dan $y = -4,93Usia + 1,437TS - 2,95TD - 0,68Glukosa + 0,12Kolesterol - 0,18Asam.Urat$ sebagai fungsi diskriminan linier *robust* yang menghasilkan persentase kesalahan pengklasifikasian sebesar 19,05%. Sehingga, penelitian ini dapat disimpulkan bahwa analisis diskriminan linier lebih akurat dibandingkan analisis diskriminan linier *robust* pada gugus data pasien diabetes RSUD Padangsidempuan tahun 2023.

DAFTAR PUSTAKA

- Johnson, R. A., and Wichern, D. W. 2002. *Applied Multivariate Statistical Analysis*. 5rd Edition. New Jersey: Prentice Hall.
- Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E. (2010). *Multivariate Data Analysis*. 7rd Edition. Upper Saddle River: Prentice Hall.

- Denis, D. J. 2016. *Applied Univariate, Bivariate, and Multivariate Statistics*. Hoboken: John Wiley & Sons, Inc.
- Singh, K., and Upadhyaya, S. (2012). Outlier detection: applications and techniques. *International Journal of Computer Science Issues (IJCSI)*, 9(1), 307.
- Huber, J. Peter, and Ronchetti, M. Elvesio. 2008. *Robust Statistics*. 2nd Edition. John Wiley & Sons, Inc. Canada.
- Maronna, A. Ricardo, Martin, R. Douglas, and Yohai, J. Victor. 2006. *Robust Statistics: Theory and Methods*. 1st Edition. John Wiley & Sons, Ltd. England.
- Rousseeuw, P. J., and Driessen, K. V. (1999). *A fast algorithm for the minimum covariance determinant estimator*. *Technometrics*, 41(3), 212-223.
- Hubert, M., Debruyne, M., and Rousseeuw, P. J. (2018). Minimum covariance determinant and extensions. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(3), e1421.
- Rencher, C. Alvin. 2002. *Methods of Multivariate Analysis*. 2nd Edition. John Wiley & Sons, Inc. Canada.
- Nur'en, Surni'a, dan Handayani, Lilies. (2019). Analisis Diskriminan Linear Robust pada Berat Bayi Lahir di RSUD Luwuk. *Statistika*. Vol 19. No 1, 19 - 27.
- Kartikawati, A., Mukid, M. A., & Ispriyanti, D. (2013). Perbandingan Analisis Diskriminan Linier Klasik dan Analisis Diskriminan Linier Robust Untuk Pengklasifikasian Kesejahteraan Masyarakat Kabupaten/Kota di Jawa Tengah. *Jurnal Gaussian*, 2(3), 157-166.