UNP JOURNAL OF STATISTICS AND DATA SCIENCE
Volume 2 Nomor 4 Tahun 2024 Halaman 469-474.
ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

Comparison of Quadratic Discrimination Analysis with Robust Quadratic Discrimination Analysis

Ully Martha, Dodi Vionanda*, Dony Permana, Zilrahmi
Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia
*Corresponding author: dodi_vionanda@fmipa.unp.ac.id

Submitted : 18 Oktober 2024
Revised : 10 November 2024
Accepted : 11 November 2024

ABSTRACT

This study compared the performance of quadratic discrimination analysis and robust quadratic discrimination analysis using the Iris dataset from Kaggle. The robust quadratic discriminant analysis, designed to handle outliers and non-normal distributions, shows better performance with an Average Percentage Error Rate (APER) of 2.5%. In contrast, the quadratic discriminant analysis, used for data with multivariate normal distribution and different variance-covariance matrices among groups, yields an APER of 3.03%. These results indicate that robust quadratic discriminant analysis is more accurate in classification on this dataset compared to quadratic discriminant analysis.

Keywords: *Apparent Error Rate, Quadratic Discrimination Analysis, Robust Quadratic Discrimination Analysis*

I. PENDAHULUAN

Analisis multivariat adalah metode untuk mengevaluasi data dari berbagai dimensi yang diobservasi secara simultan pada objek atau individu yang sama. Dalam analisis ini, dua pendekatan utama digunakan, pendekatan interdependensi dan pendekatan dependensi. Pendekatan pertama berusaha menjelaskan atau memprediksi variabel terikat dengan mempertimbangkan sejumlah variabel bebas yang mempengaruhinya. Johnson (2007) menyatakan bahwa analisis diskriminan digunakan untuk membedakan objek pengamatan yang berbeda dan memasukkan objek yang baru ke dalam kelompok yang sudah ditentukan. Setiap objek yang dikategorikan akan termasuk dalam salah satu kelompok tersebut. Ketika data pengamatan tersebar secara multivariat tetapi matriks varian-kovarians antar kelompok berbeda, analisis diskriminan kuadratik digunakan.

Ketika data pengamatan mengandung *outlier* digunakan analisis diskriminan kuadratik *robust*. *Outlier* adalah data yang sangat berbeda dari pola umum dataset. Karena matriks varian-kovarians dan rata-rata sampel sangat sensitif terhadap *outlier*, keberadaan *outlier* dapat mempengaruhi keakuratan hasil klasifikasi. Khiqmah et al. (2015) menjelaskan bahwa jika terdapat *outlier*, metode *robust* dapat digunakan untuk memastikan analisis diskriminan tetap optimal. Upadhayaya dan Singh (2012) serta Makkulau (2010) menekankan pentingnya penggunaan penaksir *robust* dalam situasi ini. Huber dan Ronchetti (2008) mengemukakan bahwa metode *robust* bertujuan untuk memaksimalkan hasil estimasi meskipun asumsi tidak sepenuhnya terpenuhi. *Minimum Covariance Determinant* (MCD) yang menurut Rousseeeuw (1999) sangat efektif dalam menangani *outlier* dan relatif mudah diterapkan dibandingkan dengan metode robust lainnya.

II. METODE PENELITIAN

A. Sumber Data dan Variabel Penelitian

Pada penelitian ini menggunakan data sekunder, yang diperoleh dari *website Kaggle.* Pada penelitian ini menggunakan gugus data yaitu *iris dataset.* Variabel yang digunakan yaitu *sepal.length* (X_1), *sepal.width* (X_2), *petal.length* (X_3), *petal.width* (X_4), *spesies* (Y).

B. Teknik Analisis Data

Dengan menggunakan *software Rstudio,* maka langkah-langkah analisis data dengan menggunakan *robust* ialah sebagai berikut:

1. Persiapkan data yang digunakan.
2. Deskripsi Data

DOI: https://doi.org/10.24036/ujsds/vol2-iss4/315

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

3. Identifikasi *Outlier*

Identifikasi *outlier* perlu dilakukan karena keberadaan *outlier* dalam kumpulan data menyebabkan fungsi diskriminan yang dihasilkan kurang baik.

a. Menghitung *Minimum Covariance Determinant* (MCD):

Minimum Covariance Determinant (MCD) digunakan untuk mengestimasi rata-rata **μ** dan matriks varian-kovarians ∑ dari data multivariat dengan ketahanan terhadap *outlier*:

$$(\boldsymbol{x}_i - \boldsymbol{\mu}_{MCD})' \sum_{MCD}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_{MCD}) \geq \chi^2_{p,\alpha} \quad (1)$$

dimana, x_i adalah vektor amatan ke-i berukukan $p \times 1$, $\bar{\boldsymbol{x}}_{MCD}$ adalah vektor *mean* dari semua amatan dari penduga MCD berukuran $p \times 1$, $\sum_{MCD}^{-1}$ adalah nilai invers dari matriks varians-kovarians dari semua amatan yang berukuran $p \times p$. Untuk menentukan apakah suatu titik data dianggap sebagai *outlier*, digunakan nilai ambang batas berdasarkan distribusi *chi-square* dengan $\alpha = 0{,}05$. *Outlier* diidentifikasi dengan membuat scatter-plot jarak mahalanobis $\chi^2_{p,0,05}$, jika plot membentuk nilai garis lurus dari 50% dan jarak mahalanobis $\leq \chi^2_{p,0,05}$, maka variabel tersebut mengikuti sebaran normal multivariat.

b. Asumsi Analisis Diskriminan

Terdapat dua asumsi analisis diskriminan yang harus dipenuhi yaitu:

1) Distribusi Normal Multivariat

Tujuan dari asumsi distribusi normal multivariat adalah untuk mengetahui model dan residual berdistribusi normal. Pengujian distribusi normal multivariat ini dilakukan dengan menghitung nilai jarak mahalanobis.

$$d_j^2 = \left(\boldsymbol{x}_j - \bar{\boldsymbol{x}}\right)' \boldsymbol{S}^{-1} \left(\boldsymbol{x}_j - \bar{\boldsymbol{x}}\right), j = 1, 2, \ldots, n \quad (1)$$

Setelah didapatkan nilai jarak mahalanobis, urutkan nilai tersebut dari nilai terkecil hingga nilai terbesar, selanjutnya membuat plot *chi-square*. Jika plot *chi-square* mendekati garis lurus maka dapat disimpulkan bahwa data menyebar normal.

2) Kesamaan Matriks Varians Kovarians

Menurut Rencher (2002) pengujian kesamaan matriks varians kovarians dapat dilakukan dengan menggunakan χ^2 dan F terhadap distribusi M. untuk mengetahui apakah matriks varians kovarians dari beberapa kelompok adalah sama atau homogen, dilakukan perhitungan nilai u.

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

Nilai u digunakan untuk membandingkan antara nilai u dengan χ^2_α. Maka statistik ujinya tolak H_0 jika $u > \chi^2_\alpha$.

4. Analisis Diskriminan Kuadratik

Ketika kondisi data berdistribusi normal multivariat dan matriks varians kovarians antar kelompok tidak sama sehingga digunakan analisis diskriminan kuadratik.

$$\hat{d}_k^Q(x) = -\frac{1}{2} \ln \boldsymbol{S}_k - \frac{1}{2} (\boldsymbol{x} - \bar{\boldsymbol{x}}_k)' \boldsymbol{S}^{-1} \sum\nolimits_k^{-1} (\boldsymbol{x} - \bar{\boldsymbol{x}}_k) + \ln p_k \quad (3)$$

Untuk data $\boldsymbol{x}$ diklasifikasikan ke dalam kelompok dengan nilai $\hat{d}_k^Q(x)$ tertinggi yang menunjukkan kesesuaian terbesar antara data dan karakteristik kelompok.

5. Analisis Diskriminan Kuadratik *Robust*

Analisis diskrimininan dengan menggunakan metode *robust* yang tidak memenuhi kedua asumsi diskriminan disebut juga analisis diskriminan kuadratik *robust* .

$$\hat{d}_k^Q(x) = -\frac{1}{2} \ln \boldsymbol{S}_{kMCD} - \frac{1}{2} (\boldsymbol{x} - \bar{\boldsymbol{x}}_{kMCD})' \boldsymbol{S}^{-1} \sum\nolimits_k^{-1} (\boldsymbol{x} - \bar{\boldsymbol{x}}_{kMCD}) + \ln p_k \quad (4)$$

Untuk mengklasifikasikan suatu pengamatan $\boldsymbol{x}$ akan termasuk dalam kategori k, jika $\hat{d}_k^Q(x) = \max \left\{\hat{d}_k^Q(x); k = 1{,}2\right\}$.

6. Nilai *Apparent Error Rate* (APER)

Untuk mengetahui persentase kesalahan dakam analisis diskrimininan digunakan nilai APER.

$$APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2} \times 100\% \quad (5)$$

Jika suatu fungsi diskriminan mempunyai nilai APER kecil maka mpdel diskriminan cukup baik. Selanjutnya dapat dihitung ketepatan klasifikasi dengan:

$$ketepatan\ klasifikasi = 1 - APER \quad (6)$$

Apabila diperoleh ketepatan klasifikasi mencapai diatas 50%, maka fungsi yang didapat sudah valid.

DOI: https://doi.org/10.24036/ujsds/vol2-iss4/315

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

III. HASIL DAN PEMBAHASAN

A. Deskripsi Data

Deskripsi data ini digunakan untuk memberikan gambaran awal *dataset* yang meliputi karakteristik dan ringkasan umum dari masing-masing variabel yang akan digunakan. Statistika deskriptif dapat dilihat pada Tabel 1 yang terdiri dari jumlah objek, *mean*, nilai max dan nilai min dari setiap variabel.

Tabel 1. Statistika deskriptif *Iris Dataset*

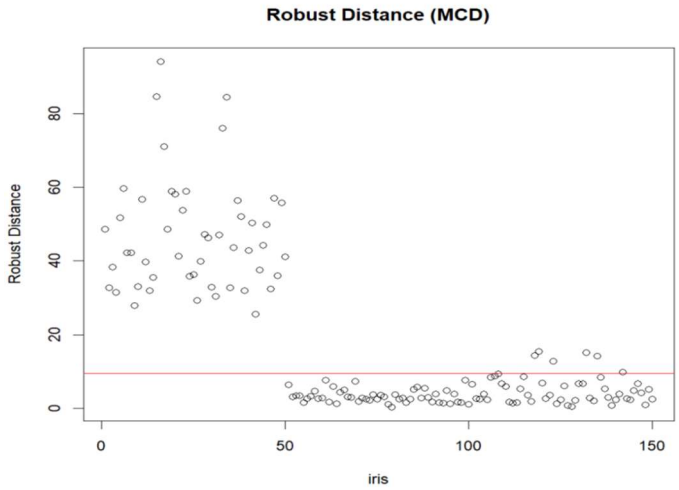
Variabel	Jumlah Objek	Mean	Max	Min
X_1	150	5,843	7,900	4,300
X_2	150	3,054	4,400	2,000
X_3	150	3,759	6,900	1,000
X_4	150	1,199	2,500	0,100

Pada Tabel 1 dapat dilihat bahwa *Iris Dataset* dengan jumlah objek 150, yang memperlihatkan variabel X_1 memiliki rata-rata 5,843 dan nilai maksimum 7,900 yang merupakan nilai tertinggi. Sementara itu, variabel X_4 memiliki rata-rata 1,199 dan nilai minimum 0,100 yang merupakan nilai terendah.

B. Analisis Diskriminan

1. Identifikasi *Outlier*

Fungsi diskriminan yang dihasilkan menjadi kurang baik ketika ada *outlier* dalam kumpulan data. Oleh karena itu, rata-rata dan matriks varian-kovarians dari data multivariat dengan ketahanan terhadap *outlier* dihitung menggunakan MCD yang dapat dilihat pada Gambar 1 berikut.



Gambar 1. Identifikasi *outlier* dengan MCD

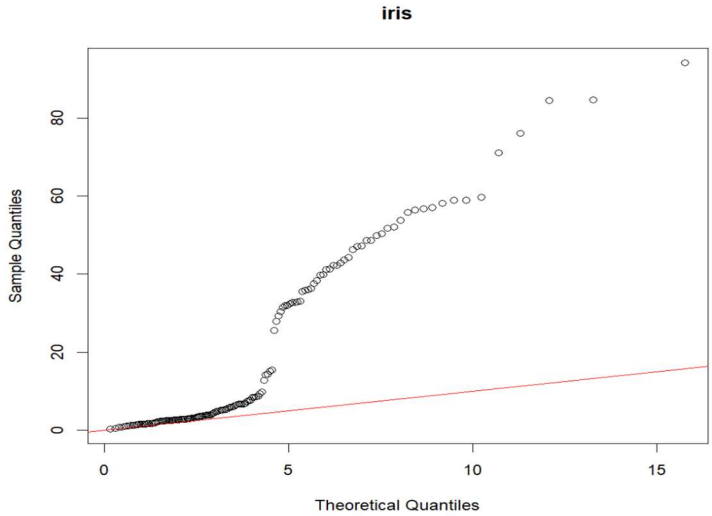
Pada Gambar 1 menyajikan plot identifikasi *outlier* dengan menggunakan MCD yang diperoleh dari total 150 objek dalam *dataset*, terdapat 56 objek di antaranya teridentifikasi sebagai *outlier* menggunakan metode MCD dengan persentase 37,33% dari total data.

2. Uji Distribusi Normal Multivariat

Tujuan asumsi distribusi normal multivariat adalah untuk mengetahui model dan residual berdistribusi normal yang dapat dilihat pada Gambar 2 berikut.

DOI: https://doi.org/10.24036/ujsds/vol2-iss4/315

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X



Gambar 2. Plot Distribusi Normal Multivariat

Uji distribusi normal multivariat pada *Iris Dataset* diperoleh nilai *p-value* < 0,05 terlihat bahwa *Iris Dataset* tidak terdistribusi normal secara multivariat. Oleh karena itu penanganan dilakukan menggunakan metode *robust*.

3. Uji Kesamaan Matriks Varians-Kovarians

Tujuan dari uji kesamaan matriks varians-kovarians multivariat adalah untuk mengetahui apakah matriks varians kovarians dari beberapa kelompok itu sama atau homogen. Diperoleh bahwa uji kesamaan matriks varians-kovarians *Iris dataset* diperoleh nilai *p-value* < 0,05 terlihat bahwa matriks varians-kovarians antar kelompok tidak homogen.

4. Fungsi Analisis Diskriminan Kuadratik

Pada analisis diskriminan kuadratik diperoleh fungsi kuadratik yang digunakan untuk mengkelompokkan data ke dalam kelompok berdasarkan variabel prediktor. Berikut merupakan fungsi diskriminan kuadratik:

$$\hat{d}_1^Q(x) = -\frac{1}{2}\ln\begin{vmatrix} 0{,}12728205 & 0{,}090410256 & 0{,}013128205 & 0{,}00825641 \\ 0{,}09041026 & 0{,}119076923 & 0{,}012769231 & 0{,}006358974 \\ 0{,}01312821 & 0{,}012769231 & 0{,}029512821 & 0{,}004512821 \\ 0{,}00825641 & 0{,}006358974 & 0{,}004512821 & 0{,}008615385 \end{vmatrix} - \frac{1}{2}(\boldsymbol{x} - \begin{pmatrix} 4{,}9800 \\ 3{,}3700 \\ 1{,}4650 \\ 0{,}2400 \end{pmatrix})' \begin{pmatrix} 0{,}12728205 & 0{,}090410256 & 0{,}013128205 & 0{,}00825641 \\ 0{,}09041026 & 0{,}119076923 & 0{,}012769231 & 0{,}006358974 \\ 0{,}01312821 & 0{,}012769231 & 0{,}029512821 & 0{,}004512821 \\ 0{,}00825641 & 0{,}006358974 & 0{,}004512821 & 0{,}008615385 \end{pmatrix}^{-1} (\boldsymbol{x} - \begin{pmatrix} 4{,}9800 \\ 3{,}3700 \\ 1{,}4650 \\ 0{,}2400 \end{pmatrix}) + \ln 0{,}3333$$

$$\hat{d}_2^Q(x) = -\frac{1}{2}\ln\begin{vmatrix} 0{,}4055092 & 0{,}1405340 & 0{,}2703109 & 0{,}0911865 \\ 0{,}1405340 & 0{,}1251085 & 0{,}1397109 & 0{,}0718078 \\ 0{,}2703109 & 0{,}1397109 & 0{,}2708914 & 0{,}1065702 \\ 0{,}0911865 & 0{,}0718078 & 0{,}1065702 & 0{,}0555503 \end{vmatrix} - \frac{1}{2}(\boldsymbol{x} - \begin{pmatrix} 5{,}9400 \\ 2{,}7700 \\ 4{,}2325 \\ 1{,}3275 \end{pmatrix})' \begin{pmatrix} 0{,}4055092 & 0{,}1405340 & 0{,}2703109 & 0{,}0911865 \\ 0{,}1405340 & 0{,}1251085 & 0{,}1397109 & 0{,}0718078 \\ 0{,}2703109 & 0{,}1397109 & 0{,}2708914 & 0{,}1065702 \\ 0{,}0911865 & 0{,}0718078 & 0{,}1065702 & 0{,}0555503 \end{pmatrix}^{-1} (\boldsymbol{x} - \begin{pmatrix} 5{,}9400 \\ 2{,}7700 \\ 4{,}2325 \\ 1{,}3275 \end{pmatrix}) + \ln 0{,}3333$$

$$\hat{d}_3^Q(x) = -\frac{1}{2}\ln\begin{vmatrix} 0{,}48291667 & 0{,}10054487 & 0{,}35169872 & 0{,}04320513 \\ 0{,}10054487 & 0{,}09753205 & 0{,}07125000 & 0{,}03884615 \\ 0{,}35169872 & 0{,}07125000 & 0{,}33358333 & 0{,}04146154 \\ 0{,}04320513 & 0{,}03884615 & 0{,}04146154 & 0{,}07087179 \end{vmatrix} - \frac{1}{2}(\boldsymbol{x} - \begin{pmatrix} 6{,}6375 \\ 3{,}0125 \\ 5{,}6225 \\ 2{,}0700 \end{pmatrix})' \begin{pmatrix} 0{,}48291667 & 0{,}10054487 & 0{,}35169872 & 0{,}04320513 \\ 0{,}10054487 & 0{,}09753205 & 0{,}07125000 & 0{,}03884615 \\ 0{,}35169872 & 0{,}07125000 & 0{,}33358333 & 0{,}04146154 \\ 0{,}04320513 & 0{,}03884615 & 0{,}04146154 & 0{,}07087179 \end{pmatrix}^{-1} (\boldsymbol{x} - \begin{pmatrix} 6{,}6375 \\ 3{,}0125 \\ 5{,}6225 \\ 2{,}0700 \end{pmatrix}) + \ln 0{,}3333$$

Dari data *Iris Dataset* diperoleh tiga fungsi dikriminan kuadratik yang digunakan untuk menentukan kelas asal dari sebuah pengamatan $\boldsymbol{x}$ dengan menghitung dan membandingkan nilai dari ketiga fungsi diskriminan, yang dimana pengamatan $\boldsymbol{x}$ akan diklasifikasikan ke dalam kelas dengan nilai fungsi diskriminan kuadratik tertinggi.

5. Fungsi Analisis Diskriminan Kuadratik *Robust*

Diperoleh fungsi analisis diskriminan kuadratik *robust* untuk meningkatkan ketepatan klasifikasi ketika data yang digunakan mengandung *outlier* atau distribusi yang tidak normal, berikut fungsi diskriminan kuadratik *robust:*

DOI: https://doi.org/10.24036/ujsds/vol2-iss4/315

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

$$\hat{d}_1^Q(x) = -\frac{1}{2}\ln\begin{vmatrix}0{,}256242470 & 0{,}152205742 & 0{,}031914452 & 0{,}001826655\\ 0{,}152205742 & 0{,}135600917 & 0{,}007570930 & 0{,}004221618\\ 0{,}031914452 & 0{,}007570930 & 0{,}035535980 & -0{,}002825774\\ 0{,}001826655 & 0{,}004221618 & -0{,}002825774 & 0{,}002438190\end{vmatrix} - \frac{1}{2}(\boldsymbol{x} - \begin{pmatrix}4{,}925381\\ 3{,}327896\\ 1{,}456075\\ 0{,}2110514\end{pmatrix})' \begin{pmatrix}0{,}256242470 & 0{,}152205742 & 0{,}031914452 & 0{,}001826655\\ 0{,}152205742 & 0{,}135600917 & 0{,}007570930 & 0{,}004221618\\ 0{,}031914452 & 0{,}007570930 & 0{,}035535980 & -0{,}002825774\\ 0{,}001826655 & 0{,}004221618 & -0{,}002825774 & 0{,}002438190\end{pmatrix}^{-1} (\boldsymbol{x} - \begin{pmatrix}4{,}925381\\ 3{,}327896\\ 1{,}456075\\ 0{,}2110514\end{pmatrix}) + \ln 0{,}3333$$

$$\hat{d}_2^Q(x) = -\frac{1}{2}\ln\begin{vmatrix}0{,}39543779 & 0{,}15333975 & 0{,}26043563 & 0{,}09212109\\ 0{,}15333975 & 0{,}13396597 & 0{,}13971462 & 0{,}07312508\\ 0{,}26043563 & 0{,}13971462 & 0{,}24202296 & 0{,}09679503\\ 0{,}09212109 & 0{,}07312508 & 0{,}09679503 & 0{,}05295615\end{vmatrix} - \frac{1}{2}(\boldsymbol{x} - \begin{pmatrix}6{,}040221\\ 2{,}811821\\ 4{,}232017\\ 1{,}3155421\end{pmatrix})' \begin{pmatrix}0{,}39543779 & 0{,}15333975 & 0{,}26043563 & 0{,}09212109\\ 0{,}15333975 & 0{,}13396597 & 0{,}13971462 & 0{,}07312508\\ 0{,}26043563 & 0{,}13971462 & 0{,}24202296 & 0{,}09679503\\ 0{,}09212109 & 0{,}07312508 & 0{,}09679503 & 0{,}05295615\end{pmatrix}^{-1} (\boldsymbol{x} - \begin{pmatrix}6{,}040221\\ 2{,}811821\\ 4{,}232017\\ 1{,}3155421\end{pmatrix}) + \ln 0{,}3333$$

$$\hat{d}_3^Q(x) = -\frac{1}{2}\ln\begin{vmatrix}0{,}40118727 & 0{,}09144196 & 0{,}28335510 & 0{,}03266951\\ 0{,}09144196 & 0{,}06992111 & 0{,}06450348 & 0{,}03426770\\ 0{,}28335510 & 0{,}06450348 & 0{,}27234996 & 0{,}05313279\\ 0{,}03266951 & 0{,}03426770 & 0{,}05313279 & 0{,}09541048\end{vmatrix} - \frac{1}{2}(\boldsymbol{x} - \begin{pmatrix}6{,}462907\\ 2{,}983293\\ 5{,}449425\\ 2{,}0369872\end{pmatrix})' \begin{pmatrix}0{,}40118727 & 0{,}09144196 & 0{,}28335510 & 0{,}03266951\\ 0{,}09144196 & 0{,}06992111 & 0{,}06450348 & 0{,}03426770\\ 0{,}28335510 & 0{,}06450348 & 0{,}27234996 & 0{,}05313279\\ 0{,}03266951 & 0{,}03426770 & 0{,}05313279 & 0{,}09541048\end{pmatrix}^{-1} (\boldsymbol{x} - \begin{pmatrix}6{,}462907\\ 2{,}983293\\ 5{,}449425\\ 2{,}0369872\end{pmatrix}) + \ln 0{,}3333$$

Dari data *Iris Dataset* diperoleh tiga fungsi diskriminan kuadratik *robust* yang bertujuan untuk mengelompokkan pengamatan $\boldsymbol{x}$ ke dalam salah satu dari tiga kelas, dengan mempertimbangkan *outlier* dan distribusi data yang tidak normal. Untuk menentukan kelas yang paling tepat, nilai dari ketiga fungsi diskriminan kuadratik tersebut dibandingkan dan pengamatan $\boldsymbol{x}$ akan dikategorikan ke dalam kelas yang menghasilkan nilai fungsi diskriminan kuadratik *robust* tertinggi.

6. Nilai APER

Berikut merupakan persentase kesalahan klasifikasi dengan menggunakan fungsi klasifikasi:

Tabel 2. Nilai APER *Iris Dataset*

Data	**APER**
Analisis Diskriminan Kuadratik	3,03%
Analisis Diskriminan Kuadratik *Robust*	2,5%

Pada Tabel 2 dapat dilihat bahwa Analisis Diskriminan Kuadratik *Robust* memiliki nilai APER yang lebih rendah sebesar 2,5% dibandingkan dengan Analisis Diskriminan Kuadratik yang memiliki nilai APER sebesar 3,03%, yang menunjukkan bahwa Analisis Diskriminan Kuadratik *Robust* memiliki tingkat akurasi klasifikasi yang lebih tinggi daripada Analisis Diskriminan Kuadratik berdasarkan nilai APER yang lebih rendah .

IV. KESIMPULAN

Pada analisis *Iris Dataset* menunjukkan bahwa metode analisis diskriminan kuadratik *robust* memiliki hasil yang lebih baik daripada analisis diskriminan kuadratik. Hal ini dibuktikan dengan nilai APER yang paling rendah pada analisis diskriminan kuadratik *robust* yang sebesar 2,5% yang menunjukkan tingkat kesalahan klasifikasi yang lebih kecil.

DAFTAR PUSTAKA

An, J., & Jin, J. (2011). Robust discriminant analysis and its application to identify protein coding region of rice genes. *Mathematical Biosciences, 232*, 96-100. https://doi.org/10.1016/j.mbs.2011.07.004

Annas, S., & Irwan, I. (2015). Penerapan analisis diskriminan dalam pengelompokan desa miskin di Kabupaten Wajo. *Jurnal Scientific Pinisi, 1*(1), 34-43. https://doi.org/10.26858/jsp.v1i1.546

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

DOI: https://doi.org/10.24036/ujsds/vol2-iss4/315

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

Hubert, M., Debruyne, M., & Rousseeuw, P. J. (2018). Minimum covariance determinant and extensions. *Wiley Interdisciplinary Reviews: Computational Statistics, 10*(3), e1421. https://doi.org/10.1002/wics.1421

Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

Katikawati, A., Mukid, M. A., & Ispriyanti, D. (2013). Perbandingan analisis diskriminan linear klasik dan analisis diskriminan linear robust untuk pengklasifikasian kesejahteraan masyarakat Kabupaten/Kota di Jawa Tengah. *Jurnal Gaussian, 2*(3), 157-166.

Makkulau, S. L. (2010). Pendeteksian outlier dan penentuan faktor-faktor yang mempengaruhi produksi gula dan tetes tebu dengan metode likelihood displacement statistic-Lagrange. *Jurnal Teknik Industri, 12*(2), 95.

Morrison, D. F. (1984). *Multivariate statistical methods* (2nd ed.). Toronto, Canada: McGraw-Hill.

Rencher, C. A. (2002). *Methods of multivariate analysis* (2nd ed.). Toronto, Canada: John Wiley & Sons.

Singh, K., & Upadhyaya, S. (2012). Outlier detection: Applications and techniques. *International Journal of Computer Science Issues (IJCSI), 9*(1), 307.

Supandi, E. D. (2020). *Statistika dan terapannya*. Refika Aditama.

Theodoridis, S., & Koutroumbas, K. (2006). *Pattern recognition*. Elsevier.

DOI: https://doi.org/10.24036/ujsds/vol2-iss4/315