ESTIMASI: Journal of Statistics and Its Application
Vol. 7, No. 1, Januari, 2026, Hal. 97-115
DOI: 10.20956/ejsa.v7i1.44982

Perbandingan Performa *Quadratic Discriminant Analysis Klasik* dan *Robust* pada Data Hasil *Principal Component Analysis* untuk Klasifikasi Jenis Kaca

M. Fatta Arya Irwanda[1*], Syafriandi[2]
[12]Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang, Padang, 25171, Indonesia
*** Corresponding author, email: mfattaaryairwanda@gmail.com**

Abstract

This study aims to compare the performance of classical Quadratic Discriminant Analysis (QDA) and Robust Quadratic Discriminant Analysis (RQDA) after applying dimensionality reduction using Principal Component Analysis (PCA) on the Glass Identification Dataset. The dataset consists of eight chemical composition variables used to classify glass types based on their elemental characteristics. Prior to classification, discriminant analysis assumptions were examined, multivariate outliers were identified, and PCA was applied to address multicollinearity and enhance data stability. The PCA results indicate that the eight original variables can be reduced to five principal components, which collectively explain 93.20% of the total data variability. Classification was then performed using classical QDA and RQDA, where the latter incorporates the Minimum Covariance Determinant (MCD) estimator to obtain robust estimates of the mean vector and covariance matrix. Model performance was evaluated using a confusion matrix and the Apparent Error Rate (APER). The results show that both QDA and RQDA achieve the same classification accuracy of 63.7%, corresponding to an APER of 36.3%. These findings suggest that the application of PCA contributes to stabilizing the data structure and reducing the influence of outliers, thereby diminishing the advantage of robust estimation in this case. Nevertheless, RQDA remains a valuable alternative for classification tasks involving datasets with strong outliers or significant deviations from multivariate normality.

Keywords: *Quadratic Discriminant Analysis, Robust QDA, Principal Component Analysis, Multivariate Outliers, Glass Identification.*

Abstrak

Penelitian ini bertujuan untuk membandingkan kinerja metode *Quadratic Discriminant Analysis* (QDA) klasik dan *Robust Quadratic Discriminant Analysis* (RQDA) setelah dilakukan reduksi dimensi menggunakan *Principal Component Analysis* (PCA) pada *Glass Identification Dataset*. Dataset ini terdiri dari delapan variabel kimia yang digunakan untuk mengklasifikasikan jenis kaca berdasarkan komposisi unsur penyusunnya. Pada tahap awal, dilakukan pengujian asumsi analisis diskriminan, pendeteksian pencilan multivariat, serta reduksi dimensi menggunakan PCA guna mengatasi multikolinearitas dan meningkatkan kestabilan struktur data. Hasil PCA menunjukkan bahwa delapan variabel awal dapat direduksi menjadi lima komponen utama yang secara kumulatif mampu menjelaskan sebesar 93,20% variasi total data. Selanjutnya, klasifikasi dilakukan menggunakan QDA klasik dan RQDA yang mengadopsi estimator *Minimum Covariance Determinant* (MCD). Evaluasi kinerja model dilakukan menggunakan confusion matrix dan *Apparent Error Rate* (APER). Hasil penelitian menunjukkan bahwa kedua metode menghasilkan tingkat akurasi yang sama, yaitu 63,7%, dengan nilai APER sebesar 36,3%. Temuan ini mengindikasikan bahwa penerapan PCA berperan dalam mereduksi pengaruh pencilan sehingga penggunaan pendekatan *robust* tidak memberikan peningkatan kinerja yang signifikan dibandingkan QDA klasik. Meskipun demikian, RQDA tetap relevan sebagai alternatif klasifikasi pada kondisi data yang mengandung pencilan kuat atau penyimpangan dari asumsi normalitas multivariat.

Kata Kunci: Analisis Diskriminan Kuadratik, RQDA, PCA, Pencilan Multivariat, Identifikasi Kaca.

1.Pendahuluan

Kaca merupakan material penting yang banyak digunakan dalam berbagai sektor, seperti industri konstruksi, otomotif, elektronik, dan forensik. Setiap jenis kaca memiliki karakteristik fisik dan kimia yang berbeda, terutama pada komposisi unsur kimia penyusunnya. Perbedaan komposisi tersebut memengaruhi sifat mekanik, ketahanan termal, dan fungsi akhir dari produk kaca. Oleh karena itu, klasifikasi jenis kaca berdasarkan komposisi kimia menjadi langkah penting untuk mendukung pengendalian kualitas, identifikasi material, dan analisis forensik.

Dalam konteks analisis statistik, klasifikasi jenis kaca termasuk ke dalam permasalahan klasifikasi multivariat, karena melibatkan banyak variabel yang saling berkorelasi. Salah satu metode klasik yang banyak digunakan untuk tujuan ini adalah *Quadratic Discriminant Analysis* (QDA). Metode QDA efektif ketika data mengikuti asumsi normal multivariat dan setiap kelompok memiliki matriks kovarians yang berbeda [1, 2]. QDA memanfaatkan informasi rata-rata dan struktur kovarians masing-masing kelompok untuk membentuk fungsi diskriminan kuadratik yang mampu memisahkan kelompok secara optimal.

Namun demikian, dalam praktiknya, data empiris sering kali mengandung pencilan (*outlier*) yang dapat memengaruhi estimasi parameter, khususnya matriks kovarians. Keberadaan pencilan dapat menyebabkan fungsi diskriminan QDA klasik menjadi tidak stabil dan menurunkan akurasi klasifikasi [3]. Oleh karena itu, dikembangkan pendekatan *Robust Discriminant Analysis* yang dirancang agar tahan terhadap pengaruh pengamatan ekstrem.

Salah satu pendekatan *robust* yang banyak digunakan adalah *Robust Quadratic Discriminant Analysis* (RQDA) yang mengandalkan estimator kovarians *robust*, seperti *Minimum Covariance Determinant* (MCD). Estimator MCD bertujuan untuk memperoleh estimasi vektor rata-rata dan matriks kovarians yang stabil dengan meminimalkan determinan kovarians dari subset data yang paling representatif [4]. Penggunaan MCD dalam analisis diskriminan telah terbukti meningkatkan ketahanan model terhadap pencilan dan menghasilkan klasifikasi yang lebih andal pada data yang tidak sepenuhnya memenuhi asumsi klasik [5, 6].

Selain permasalahan pencilan, analisis klasifikasi multivariat juga sering dihadapkan pada isu dimensi data yang tinggi dan multikolinearitas antar variabel. Kondisi ini dapat menyebabkan estimasi matriks kovarians menjadi tidak stabil. Salah satu teknik yang umum digunakan untuk mengatasi permasalahan tersebut adalah *Principal Component Analysis* (PCA). PCA mereduksi dimensi data dengan membentuk sejumlah komponen utama yang saling tidak berkorelasi dan mampu merepresentasikan sebagian besar variasi data asli [1,7]. Dengan demikian, penerapan PCA sebelum analisis diskriminan diharapkan dapat meningkatkan kestabilan estimasi parameter dan menyederhanakan proses klasifikasi.

Beberapa penelitian sebelumnya telah membahas penerapan QDA dan RQDA pada berbagai jenis data dan menunjukkan bahwa pendekatan *robust* cenderung lebih unggul ketika data mengandung pencilan [8,9]. Selain itu, penelitian terkait deteksi pencilan multivariat juga menunjukkan bahwa jarak Mahalanobis sering digunakan sebagai alat identifikasi pengamatan ekstrem dalam data berdimensi tinggi [10,11]. Meskipun demikian, masih terbatas penelitian yang secara eksplisit membandingkan kinerja QDA klasik dan RQDA setelah penerapan PCA, khususnya pada kasus klasifikasi jenis kaca berdasarkan komposisi kimia.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk membandingkan kinerja *Quadratic Discriminant Analysis* (QDA) klasik dan *Robust Quadratic Discriminant Analysis* (RQDA) pada Glass Identification Dataset setelah dilakukan reduksi dimensi menggunakan *Principal Component Analysis* (PCA). Perbandingan kinerja dilakukan berdasarkan tingkat akurasi dan nilai *Apparent Error Rate* (APER). Hasil penelitian ini diharapkan dapat memberikan gambaran empiris mengenai sejauh mana penerapan pendekatan *robust* memberikan keunggulan dibandingkan metode klasik, khususnya ketika data telah ditransformasi melalui PCA.

2.Material dan Metode

2.1. Pengujian Asumsi Analisis Diskriminan

2.1.1. Uji Kesamaan Vektor Rata-Rata

Uji kesamaan vektor rata-rata dilakukan untuk mengetahui apakah terdapat perbedaan yang signifikan antara vektor rata-rata kelompok yang dibandingkan. Dalam konteks analisis diskriminan, pengujian ini bertujuan untuk memastikan bahwa terdapat perbedaan karakteristik multivariat antar kelompok sehingga pembentukan fungsi diskriminan layak dilakukan Johnson & Wichern [1,2].

Pengujian kesamaan vektor rata-rata antar kelompok dilakukan menggunakan pendekatan MANOVA dengan statistik uji *Wilks' Lambda* (Λ). Statistik *Wilks' Lambda* selanjutnya ditransformasikan menggunakan aproksimasi Bartlett sehingga mengikuti distribusi *chi-kuadrat* dengan derajat bebas tertentu [12].

Statistik *V-Bartlett* dihitung dengan rumus berikut:

$$V = -\left[(n-1) - \frac{(p+1)(k-1)}{2}\right] ln(\Lambda) \quad (1)$$

dengan:

n : jumlah pengamatan,

p : jumlah variabel,

k : jumlah kelompok,

Λ : *Wilk's Lamda*, dimana:

$$\Lambda = \frac{|W|}{|B + W|} \tag{2}$$

dengan:
W : matriks dalam kelompok (hasil penjumlahan kuadrat dan hasil kali data dalam kelompok),
B : matriks antar kelompok (hasil penjumlahan kuadrat dan hasil kali data antar kelompok).

Jika $V > \chi^2_{p(k-1)(1-\alpha)}$, artinya terdapat perbedaan nilai vector rata-rata antar kelompok [1, 12].

2.1.2. Uji Asumsi Normalitas Multivariat

Untuk pengujian data berdistribusi normal ganda dapat dilakukan dengan membuat plot jarak kuadrat Mahalanobis dengan nilai *chi-square*. Jarak kuadrat Mahalanobis untuk setiap objek dapat dihitung dengan menggunakan persamaan sebagai berikut:

$$d_i^2 = (x_i - \bar{x}_i)' S^{-1} (x_i - \bar{x}_i); i = 1,2, \dots, n \tag{3}$$

dengan:
d_i^2 : Jarak Mahalanobis pengamatan ke-*i*
x_i : Vektor observasi ke-*i*
$\bar{x}_i$: Vektor *mean* variabel independen
S^{-1} : Matriks kovarians invers

Setelah nilai d_i^2 didapatkan, kemudian nilai d_i^2 diurutkan dari yang terkecil ke yang terbesar, selanjutnya d_i^2 diplotkan dengan nilai *chi-square*. Apabila plot d_i^2 dengan nilai *chi-square* dapat didekati dengan garis lurus dan ada lebih dari 50% nilai $d_i^2 < \chi^2_{0,5,p}$ maka dapat disimpulkan bahwa data menyebar normal [1,12].

2.1.3 Uji Kesamaan Matriks Varian Kovarian

Uji kesamaan matriks varian–kovarian dilakukan untuk mengetahui apakah matriks kovarians antar kelompok adalah sama. Dalam analisis diskriminan, pengujian ini penting karena pelanggaran asumsi kesamaan matriks kovarians menjadi dasar pemilihan QDA dibandingkan *Linear Discriminant Analysis* (LDA) [1,2].

Statistik uji yang digunakan untuk menguji kesamaan matriks peragam adalah statistik uji *Box's M.* Berikut adalah persamaan untuk statistik uji *Box's M.*

$$-2\ln\lambda^* = (\mathrm{n}-\mathrm{k})\ln\left|\frac{\mathrm{W}}{(\mathrm{n}-\mathrm{k})}\right| - \sum(\mathrm{n_i}-1)\ \ln\left|\mathrm{S_j}\right| \tag{4}$$

dimana:

$$\lambda^* = \frac{\prod\left|\mathrm{S_j}\right|^{\frac{\mathrm{n_j}-1}{2}}}{\left|\frac{\mathrm{W}}{(\mathrm{n}-\mathrm{k})}\right|^{\frac{\mathrm{n}-\mathrm{k}}{2}}} \tag{5}$$

dengan:

K : Banyaknya kelompok

$\left|\frac{\mathrm{W}}{(\mathrm{n}-\mathrm{k})}\right|$: Matriks kovarians dalam kelompok gabungan

S_j : Matriks kovarians kelompok ke-j

n_i : Banyaknya observasi pada kelompok ke-i

Nilai $\frac{(-2\ln\lambda^*)}{b}$ akan mengikuti distribusi F dengan derajat kebebasan $v1$ dan $v2$ pada tingkat signifikansi tertentu, dengan:

$$v1 = \frac{1}{2}(k-1)p(p+1) \tag{6}$$

$$v2 = \frac{v1+2}{a_2 - a_1^2} \tag{7}$$

dengan:

$$b = \frac{v1}{\left(1 - a_1 - \frac{v1}{v2}\right)} \tag{8}$$

Jika $\frac{(-2\ln\lambda^*)}{b} \le F_{v1,v2,\alpha}$ atau jika $p - value < 0{,}05$,maka seluruh matriks dianggap memiliki kovarians yang sama [1,12].

2.2. Pendeteksi Pencilan

Pencilan (*outlier*) merupakan pengamatan yang memiliki karakteristik berbeda secara signifikan dibandingkan dengan sebagian besar data lainnya dan dapat memberikan pengaruh yang besar terhadap hasil analisis statistik multivariat. Keberadaan pencilan dapat menyebabkan estimasi vektor rata-rata dan matriks kovarians menjadi bias, sehingga menurunkan kinerja metode klasifikasi berbasis kovarians, seperti analisis diskriminan [1, 10].

Salah satu metode yang umum digunakan untuk mendeteksi pencilan multivariat adalah dengan menghitung jarak Mahalanobis untuk setiap pengamatan. Jarak Mahalanobis mengukur jarak suatu pengamatan terhadap pusat data dengan mempertimbangkan struktur kovarians antar variabel. Suatu pengamatan ke-i dinyatakan sebagai pencilan apabila nilai jarak Mahalanobis kuadratnya melebihi nilai kritis distribusi chi-square pada derajat kebebasan sebesar jumlah variabel dan taraf signifikansi tertentu [1]. Secara matematis, kriteria pendeteksian pencilan dinyatakan sebagai berikut:.

$$d_i^2 > \chi^2_{p,\alpha} \tag{9}$$

dengan:

d_i^2 : nilai jarak Mahalanobis kuadrat pengamatan ke-i
p : jumlah variabel
α : taraf signifikansi yang digunakan

Pengamatan yang memenuhi kriteria tersebut dikategorikan sebagai pencilan multivariat dan perlu mendapatkan perhatian khusus dalam analisis lanjutan, terutama dalam proses estimasi matriks kovarians dan pembentukan fungsi diskriminan.

2.3. Analisis Komponen Utama (*Principal Component Analysis* - PCA)

Principal Component Analysis (*PCA*) merupakan suatu teknik statistik yang digunakan untuk mereduksi dimensi data multivariat dengan cara mengubah variabel asli yang saling berkorelasi menjadi sekumpulan variabel baru yang saling tidak berkorelasi yang disebut *komponen utama* (*principal components*). Tujuan utama dari *PCA* adalah untuk menyederhanakan struktur data tanpa kehilangan terlalu banyak informasi[1] .

Secara matematis, komponen utama merupakan kombinasi linear dari peubah yang diamati, di mana setiap komponen utama dibentuk sebagai gabungan dari seluruh variabel dengan bobot tertentu [1]. Jika $X_1, X_2, \dots, X_p$ menyatakan variabel acak yang diamati, maka komponen utama dapat dinyatakan dalam bentuk matriks sebagai berikut:

$$Y = AX \tag{10}$$

dengan:

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_p \end{bmatrix}, A = \begin{bmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{p1} & \cdots & a_{pp} \end{bmatrix}, X = \begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix} \tag{11}$$

Sehingga *komponen utama* dapat ditulis sebagai berikut:

$$Y_1 = a_1' = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \quad (12)$$

$$Y_2 = a_2' = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p \quad (13)$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$Y_p = a_p' = a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p \quad (14)$$

Dimana a_1 adalah vektor konstanta *p* yaitu $a_{11}, a_{12,} \dots, a_{1p}$ dan tanda $\square'$ menyatakan transpos [1].

Dengan menggunakan *PCA*, variabel-variabel asli yang mungkin berkorelasi diubah menjadi sejumlah *komponen utama* yang saling bebas (*uncorrelated*). Tujuan utamanya adalah untuk mereduksi dimensi data tanpa kehilangan informasi yang berarti, meminimalisir *multikolinearitas*, serta menyederhanakan proses klasifikasi pada analisis diskriminan (QDA dan RQDA).

Salah satu kaidah yang digunakan untuk menentukan jumlah komponen utama yang layak dipertahankan dalam PCA adalah *Kaiser's Rule*. Kaidah ini menyatakan bahwa hanya komponen utama dengan eigenvalue lebih besar dari 1 yang sebaiknya dipertahankan dalam analisis.

Alasannya, komponen dengan *eigenvalue* < 1 dianggap mengandung informasi yang lebih sedikit dibandingkan satu variabel asli, sehingga keberadaannya kurang efektif untuk dipertahankan. Oleh karena itu, hanya komponen dengan *eigenvalue* > 1 yang memberikan kontribusi informasi yang signifikan terhadap struktur data secara keseluruhan. Kriteria ini banyak digunakan karena bersifat sederhana dan intuitif dalam konteks reduksi dimensi data multivariat [1,7].

2.4. Pembentukan Fungsi Diskriminan Kuadratik Klasik

Dalam metode diskriminan, salah satu asumsi dasar yang digunakan adalah bahwa data berasal dari populasi yang terdistribusi normal multivariat, serta matriks kovarian dari populasi-populasi yang dibandingkan dianggap sama. Namun demikian, dalam beberapa situasi, asumsi kesamaan matriks kovarian tidak terpenuhi, sehingga diperlukan fungsi diskriminan kuadratik untuk proses klasifikasi [2].

Misalkan terdapat dua populasi $(i = 1,2)$ yang masing-masing memiliki himpunan pengamatan dengan sebaran normal multivariat. Fungsi diskriminan kuadratik dibangun untuk memisahkan kedua populasi ini berdasarkan vektor peubah acak $X \in \mathbb{R}^p$, di mana setiap populasi memiliki vektor rata-rata dan matriks kovarians yang berbeda. Dalam kondisi tersebut, fungsi diskriminan kuadratik dibangun untuk memisahkan kedua populasi secara optimal dengan mempertimbangkan perbedaan struktur kovarians antar kelompok [1]. Daerah klasifikasi untuk populasi ke-1 dan ke-2 dapat dituliskan sebagai berikut:

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq \left(\frac{c(2)}{c(1)}\right)\left(\frac{p_2}{p_1}\right) \tag{15}$$

$$R_2 : \frac{f_1(x)}{f_2(x)} < \left(\frac{c(2)}{c(1)}\right)\left(\frac{p_2}{p_1}\right) \tag{16}$$

Dimana p_i menyatakan probabilitas aprior untuk populasi ke-i $f_i(x)$ adalah fungsi densitas probabilitas untuk populasi ke-i.

Fungsi diskriminan kuadratik dirumuskan sebagai berikut:

$$d_i^Q(x) = \ln p_i - \frac{1}{2}(x - \mu_i)' \sum_1^{-1} (x - \mu_i) - \frac{1}{2}\ln(|\Sigma_i|) \tag{17}$$

Dengan μ_i adalah vektor rata-rata dan $\sum_i$ adalah matriks kovarian populasi ke-i. Dalam praktiknya, parameter μ_i dan $\sum_i$ biasanya tidak diketahui, sehingga perlu diestimasi dari data sampel. Estimasi fungsi diskriminan kuadratik menjadi:

$$\hat{d}_i^Q(x) = \ln p_i - \frac{1}{2}(x - \bar{x}_i)' S_1^{-1}(x - \bar{x}_i) - \frac{1}{2}\ln(|S_i|) \tag{18}$$

Dimana $\bar{x}_i$ adalah vektor rata-rata sampel S_i adalah matriks kovarian sampel untuk populasi ke-i.

Keputusan klasifikasi dilakukan dengan mengalokasikan observasi x ke populasi ke-2 apabila:

$$\hat{d}_1^Q(x) < \hat{d}_2^Q(x) \tag{19}$$

Untuk menentukan probabilitas apriori p_i , dapat digunakan dua pendekatan. Pendekatan pertama mengasumsikan bahwa kedua populasi memiliki peluang yang sama, yaitu:

$$p_i = \frac{1}{2} \text{ untuk setiap } i \tag{20}$$

Sedangkan pendekatan kedua menggunakan frekuensi relatif berdasarkan jumlah sampel pada masing-masing populasi, yaitu:

$$p_i = \frac{n_i}{n} \tag{20}$$

Dengan n_i adalah ukuran sampel dari populasi ke-*i,* dan *n* adalah total ukuran sampel [5].

2.5. Estimasi Parameter *Robust*

Estimasi parameter *robust* digunakan untuk menghitung nilai rata-rata dan matriks kovarians yang tahan terhadap pencilan. Salah satu metode yang sering digunakan adalah *Minimum Covariance Determinant* (MCD). Metode ini bertujuan untuk mencari subset data dengan determinan matriks kovarians terkecil sehingga nilai rata-rata dan matriks kovarians yang dihasilkan lebih stabil [4].

Langkah-langkah dalam metode MCD adalah sebagai berikut:

a. Menentukan banyaknya pengamatan yang diambil, yaitu:

$$h = \left\lceil \frac{n+p+1}{2} \right\rceil \tag{21}$$

dengan:
n = jumlah pengamatan
p = jumlah variabel.

b. Memilih secara acak *h* pengamatan untuk membentuk himpunan awal H_1.

c. Menghitung rata-rata ($\bar{x}_{H_1}$) dan matriks kovarians $\left(S_{H_1}\right)$dari H_1:

$$\bar{x}_{H_1} = \frac{1}{h} \sum_{i \in H_1} x_i \tag{23}$$

$$S_{H_1} = \frac{1}{h-1} \sum_{i \in H_1} (x_i - \bar{x}_{H_1})(x_i - \bar{x}_{H_1})' \tag{24}$$

[4].

d. Menghitung jarak Mahalanobis kuadrat setiap pengamatan terhadap rata-rata dan kovarians awal:

$$d_i^2 = \left(x_i - \bar{x}_{H_1}\right)' S_{H_1}^{-1}\left(x_i - \bar{x}_{H_1}\right) \tag{25}$$

e. Mengurutkan hasil perhitungan jarak Mahalanobis kuadrat secara ascending dan memilih *h* pengamatan dengan nilai terkecil untuk membentuk himpunan baru H_2 .

f. Mengulangi langkah c-e hingga determinan matriks kovarians stabil *(konvergen)* atau sampai ditemukan $\det(S_{m+1}) \leq \det(S_m)$.

g. Setelah iterasi stabil, digunakan bobot [4]:

h.

$$w_i = \begin{cases} 1\ jika\ d_i^2 \leq \chi^2_{p,\alpha} \\ 0\ jika\ d_i^2 > \chi^2_{p,\alpha} \end{cases}$$

i. Menghitung *mean robust*:

$$\bar{x}_{MCD} = \frac{\sum w_i x_i}{\sum w_i} \tag{26}$$

j. Menghitung matriks kovarians *robust*:

$$S_{MCD} = \frac{\sum w_i (x_i - \bar{x}_{MCD})(x_i - \bar{x}_{MCD})'}{\sum w_i - 1} \tag{27}$$

Dengan menggunakan *mean robust* dan matriks kovarians *robust* ini, fungsi diskriminan *robust* dapat dibentuk untuk klasifikasi yang lebih tahan terhadap pencilan [3,4].

2.6. Pembentukan Fungsi Diskriminan Kuadratik *Robust*

Setelah diperoleh nilai rata-rata robus*t* ($\bar{x}_{MCD}$) dan matriks kovarians *robust* (S_{MCD}) untuk setiap kelompok, fungsi diskriminan kuadratik *robust* dapat dibentuk untuk mengklasifikasikan data ke dalam kelompok yang sesuai. Pendekatan ini menggunakan estimator MCD untuk mengurangi pengaruh pencilan terhadap estimasi parameter, sehingga fungsi diskriminan yang dihasilkan menjadi lebih stabil dan andal [5]. Fungsi diskriminan kuadratik *robust* untuk kelompok ke-i dirumuskan sebagai berikut:

$$\hat{d}_i^{RQ}(x) = \ln p_i - \frac{1}{2}(x - \bar{x}_{iMCD})' S_{iMCD}^{-1}(x - \bar{x}_{iMCD}) - \frac{1}{2} ln|S_{iMCD}| \tag{28}$$

dengan:
$\hat{d}_i^{RQ}(x)$: skor diskriminan kuadratik *robust* kelompok ke-i
p_i : peluang *prior* ke-i
$\bar{x}_{iMCD}$: vektor rata-rata MCD kelompok ke-i
S_{iMCD}^{-1} : invers matriks varian kovarian MCD kelompok ke-i
$|S_{iMCD}|$: determinan matriks varian kovarian MCD kelompok ke-i [5].

Keputusan klasifikasi dilakukan dengan mengalokasikan suatu pengamatan x ke kelompok dengan nilai skor diskriminan kuadratik *robust* terbesar. Untuk kasus dua kelompok, pengamatan x dialokasikan ke kelompok ke-2 apabila memenuhi kondisi berikut:

$$\hat{d}_1^{RQ}(x) < \hat{d}_2^{RQ}(x) \tag{29}$$

2.7. Klasifikasi dan Evaluasi Model

Setelah fungsi diskriminan kuadratik *robust* terbentuk, langkah selanjutnya adalah menggunakannya untuk mengklasifikasikan data baru. Setiap pengamatan baru dihitung nilai skor diskriminannya untuk setiap kelompok menggunakan rumus persamaan (28).Setelah skor dihitung, pengamatan diklasifikasikan ke kelompok dengan nilai diskriminan tertinggi [3]. Tahap terakhir dalam analisis ini adalah evaluasi kinerja model diskriminan kuadratik klasik dan *robust* menggunakan nilai *Apparent Error Rate* (APER). APER merupakan ukuran proporsi kesalahan klasifikasi yang terjadi dalam data yang digunakan untuk membentuk model (data training). Semakin kecil nilai APER, maka semakin baik kinerja model dalam mengklasifikasikan data. Rumus untuk menghitung APER adalah sebagai berikut:

$$APER = \frac{Jumlah\ klasifikasi\ salah}{Jumlah\ total\ pengamatan} \times 100\% \tag{30}$$

$atau$

$$APER = 1 - Akurasi \tag{31}$$

Evaluasi ini penting untuk melihat apakah model klasifikasi yang dihasilkan sudah memadai dan dapat diandalkan dalam aplikasi nyata [13].

2.8. Metode Penelitian

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari *Glass Identification Dataset* yang diunduh melalui UCI Machine Learning Repository. Data ini terdiri dari 214 data sampel kaca yang memiliki 8 variabel independen yaitu sodium (X_1), magnesium (X_2), aluminium (X_3), silicon (X_4), potassium (X_5), calcium (X_6), barium (X_7), dan iron (X_8). Sedangkan variabel dependennya adalah tipe kaca yang dikelompokkan menjadi 7 kelas (yaitu tipe kaca 1 hingga tipe kaca 8).

Dalam penelitian ini hanya digunakan dua tipe kaca, yaitu tipe kaca 1 dan tipe kaca 2, dari keseluruhan 7 tipe kaca yang tersedia dalam dataset *Glass Identification*. Pemilihan dua tipe kaca ini didasarkan pada pertimbangan untuk menyederhanakan proses klasifikasi dan memfokuskan analisis terhadap dua kelompok yang secara komposisi kimia memiliki perbedaan yang lebih jelas serta jumlah sampel yang memadai. Selain itu, penggunaan dua kelompok mempermudah pengujian asumsi dan interpretasi hasil pada metode *Quadratic Discriminant Analysis* (QDA) dan *Robust Quadratic Discriminant Analysis* (RQDA), terutama dalam tahapan pengujian normalitas multivariat, kesamaan matriks kovarians, dan evaluasi fungsi diskriminan.

Berdasarkan pemilahan data tersebut, jumlah sampel yang digunakan dalam penelitian ini menjadi 146 data, yang terdiri dari 70 sampel tipe kaca 1 dan 76 sampel tipe kaca 2.

Langkah-langkah yang dilakukan dalam menganalisis data adalah sebagai berikut:

1. Pengujian asumsi
    a. Menguji kesamaan vektor rata-rata menggunakan statistik *V-Barlett.*
    b. Menguji asumsi normal multivariat menggunakan jarak *Mahalanobis*.
    c. Menguji kesamaan matriks varian kovarian menggunakan *Box's M.*
2. Mendeteksi pencilan.
3. Reduksi Dimensi Menggunakan *Principal Component Analysis* (PCA)
4. Estimasi parameter
    a. Estimasi parameter vektor rata-rata dan matriks varian-kovarian secara klasik untuk metode QDA klasik.
    b. Estimasi parameter *robust* menggunakan Fast-MCD untuk metode RQDA.
5. Pembentukan fungsi Diskriminan.
    a. Pembentukan fungsi diskriminan QDA klasik berdasarkan hasil estimasi parameter klasik
    b. Pembentukan fungsi diskriminan RQDA berdasarkan hasil estimasi parameter *robust* dari Fast-MCD.
6. Evaluasi Model Klasifikasi.

3.Hasil dan Diskusi

3.1. Pengujian Asumsi

3.1.1 Uji Kesamaan Vektor Rata-rata

Sebelum dilakukan analisis diskriminan kuadratik *Robust*, perlu dilakukan pengujian asumsi. Uji asumsi yang pertama adalah uji kesamaan vektor rata-rata antar kelompok berdasarkan 8 variabel independen (*Na, Mg, Al, Si, K, Ca, Ba,* dan *Fe)* menggunakan statistik *V-Bartlett*. Uji ini bertujuan untuk mengetahui apakah terdapat perbedaan signifikan pada vektor rata-rata antar dua kelompok tipe kaca.

Berdasarkan hasil perhitungan menggunakan perangkat lunak R, diperoleh nilai statistik $V = 51{,}43549$, sedangkan nilai kritis *chi-kuadrat* dengan derajat bebas sebesar $\chi^2_{8(2-1)(1-0{,}05)} = 15{,}50731$. Karena nilai statistik lebih besar dari nilai kritis $(51{,}43549 > 15{,}50731)$, yang berarti terdapat perbedaan vektor rata-rata yang signifikan antar kelompok tipe kaca. Dengan demikian, asumsi perbedaan rata-rata antar kelompok terpenuhi sehingga analisis diskriminan kuadratik *robust* dapat dilanjutkan.

3.1.2. Uji Asumsi Normalitas Multivariat

Asumsi selanjutnya adalah bahwa data harus berdistribusi normal multivariat. Uji ini dilakukan menggunakan jarak Mahalanobis, dengan membandingkan nilai d^2 terhadap nilai kritis $\chi^2_{(0{,}05;8)}$, karena terdapat 8 variabel . Berdasarkan hasil uji menggunakan perangkat lunak R, diperoleh jumlah nilai $d^2 < \chi^2_{(0{,}05;8)}$ sebanyak 133 dari total 146 pengamatan, atau sekitar 91,1%, maka dapat disimpulkan bahwa data berdistribusi normal multivariat.

3.1.3. Uji Kesamaan Matriks Varian Kovarian

Asumsi selanjutnya adalah kesamaan matriks varian kovarian antar kelompok yang diuji menggunakan *Box's M* test. Berdasarkan hasil uji menggunakan perangkat lunak R, diperoleh nilai Box's M sebesar $356{,}15$ dengan derajat bebas 180 dan $nilai\, p-value < 2.2 \times 10^{-16}$ Karena nilai p lebih kecil dari 0,05, maka dapat disimpulkan bahwa matriks varian kovarian antar kelompok berbeda.

3.2. Mendeteksi Pencilan

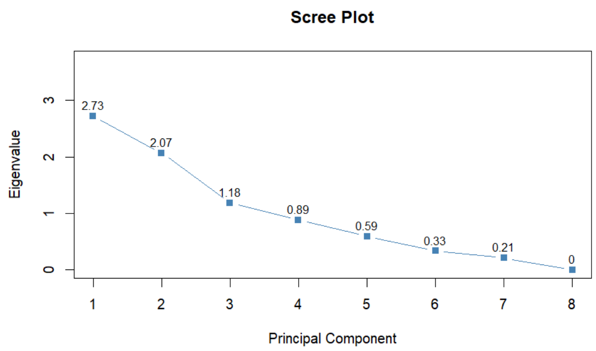
Dalam analisis diskriminan kuadratik *robust*, deteksi pencilan merupakan langkah penting untuk memastikan bahwa model tidak terpengaruh oleh pengamatan ekstrem yang dapat mendistorsi hasil klasifikasi. Pendeteksian pencilan dilakukan menggunakan jarak Mahalanobis, di mana data dengan nilai $d^2 > \chi^2_{(0{,}05;8)} = 15{,}50731$ dikategorikan sebagai pencilan.

Berdasarkan hasil analisis, diperoleh sebanyak 13 pengamatan yang teridentifikasi sebagai pencilan dari total 146 data. Keberadaan pencilan ini memperkuat alasan penggunaan pendekatan *robust* pada diskriminan kuadratik, karena metode RQDA memiliki ketahanan terhadap pengaruh data ekstrem, sehingga tetap mampu menghasilkan klasifikasi yang akurat dan stabil meskipun terdapat pencilan dalam data.

3.3. Analisis Komponen Utama (*Principal Component Analysis* - PCA)

Analisis Komponen Utama (PCA) digunakan untuk mereduksi dimensi data sehingga variabel-variabel awal dapat direpresentasikan dalam sejumlah komponen utama yang memuat sebagian besar informasi dari data tersebut. PCA bertujuan mengubah variabel-variabel asli yang saling berkorelasi menjadi serangkaian variabel baru yang tidak saling berkorelasi (*principal components*) berdasarkan *eigenvalue* tertinggi.

Hasil PCA pada dataset menunjukkan bahwa nilai eigenvalue untuk delapan komponen berturut-turut adalah 2.73, 2.07, 1.18, 0.89,0.59, 0.33, 0.21, dan 0.00. Nilai eigenvalue ini digambarkan pada Gambar 1 melalui scree plot berikut:



Gambar 1. Scree Plot Komponen Utama (PCA)

Berdasarkan Gambar 1, tampak bahwa terdapat penurunan tajam nilai eigenvalue pada komponen 1 hingga komponen 2, diikuti penurunan lebih landai hingga komponen 5. Setelah komponen 5, penurunan eigenvalue cenderung kecil, mendekati nol, yang mengindikasikan bahwa komponen setelah PC5 tidak lagi memberikan informasi signifikan.

Tabel 1 berikut menunjukkan hasil lengkap PCA menggunakan perangkat lunak R:

Tabel 1. Ringkasan Hasil Analisis Komponen Utama (PCA)

Komponen Utama	Standar Deviasi	Proporsi Variansi	Akumulasi Proporsi
PC1	1,653	0,341	0,341
PC2	1,438	0,2585	0,5994
PC3	1,0872	0,1477	0,7472
PC4	0,941	0,1107	0,8579
PC5	0,7703	0,0742	0,932
PC6	0,5765	0,0416	0,9763

PC7	0,4577	0,0262	0,9998
PC8	0,0443	0,0002	1,0000

Dari Tabel 1, terlihat bahwa komponen pertama (PC1) menjelaskan variansi sebesar 34.10%. Komponen kedua (PC2) menambah penjelasan variansi sebesar 25.85%, sehingga dua komponen pertama sudah mencakup 59.94% variansi data. Dengan menambahkan PC3, PC4, dan PC5, akumulasi proporsi variansi yang dapat dijelaskan meningkat menjadi 93.20%.

Pemilihan lima komponen utama (PC1 hingga PC5) dalam penelitian ini didasarkan pada beberapa pertimbangan statistik. Pertama, berdasarkan *scree plot* (Gambar 1), terjadi penurunan tajam *eigenvalue* pada PC1 dan PC2, diikuti dengan penurunan yang lebih landai hingga PC5. Setelah PC5, nilai *eigenvalue* menunjukkan penurunan yang sangat kecil dan cenderung mendekati nol, mengindikasikan bahwa informasi tambahan yang diberikan oleh komponen setelah PC5 tidak signifikan.

Kedua, sesuai dengan kaidah Kaiser's Rule, komponen utama yang memiliki eigenvalue > 1 dianggap layak untuk dipertahankan. Dalam kasus ini, hanya tiga komponen pertama yang memenuhi kriteria tersebut, yaitu PC1 hingga PC3. Namun demikian, dua komponen tambahan (PC4 dan PC5) tetap dipertahankan berdasarkan informasi tambahan dari scree plot dan nilai kumulatif variansi yang dijelaskan.

Ketiga, pemilihan lima komponen utama juga mempertimbangkan ambang batas kumulatif proporsi variansi yang dijelaskan, di mana nilai kumulatif sebesar 93,20% dipandang sudah cukup tinggi untuk merepresentasikan informasi dari delapan variabel awal secara efektif. Dengan mempertahankan lima komponen, penelitian ini dapat mereduksi dimensi tanpa kehilangan informasi yang berarti, sekaligus menyederhanakan model klasifikasi yang akan dibentuk.

3.4. Analisis Diskriminan Kuadratik Klasik

Analisis diskriminan kuadratik klasik diterapkan untuk memisahkan data pada *Glass Identification Dataset* menjadi dua kelompok, yaitu *Type 1* dan *Type 2*. Pada Langkah sebelumnya telah dilakukan reduksi dimensi dengan Principal Component Analysis (PCA) terhadap variabel kimia seperti *Na, Mg, Al, Si, K, Ca, Fe, dan Ba*. Dari hasil PCA, dipilih lima komponen utama pertama (PC1 hingga PC5) yang mewakili sebagian besar variasi data. Langkah awal dalam analisis diskriminan kuadratik ini adalah menghitung vektor rata-rata dan matriks varian-kovarian dari masing-masing kelompok. Berdasarkan hasil tersebut, dibentuk fungsi diskriminan kuadratik klasik sebagai berikut:

$$\hat{d}_1^Q(x) = -\frac{1}{2}(x-\bar{x}_1)'S_1^{-1}(x-\bar{x}_1) - \frac{1}{2}\ln(|S_1|) + \ln(\frac{70}{146})$$

$$\hat{d}_2^Q(x) = -\frac{1}{2}(x-\bar{x}_2)'S_2^{-1}(x-\bar{x}_2) - \frac{1}{2}\ln(|S_2|) + \ln(\frac{76}{146})$$

Langkah berikutnya adalah menghitung nilai $\hat{d}_1^Q$ dan $\hat{d}_2^Q$ untuk setiap pengamatan pada dataset. Setiap data selanjutnya diklasifikasikan ke dalam *Type 1* atau *Type 2* berdasarkan nilai diskriminan yang lebih kecil. Apabila suatu pengamatan memiliki nilai $\hat{d}_1^Q < \hat{d}_2^Q$, maka data tersebut dikategorikan sebagai *Type 1*, sebaliknya akan diklasifikasikan sebagai *Type 2* jika $\hat{d}_2^Q < \hat{d}_1^Q$. Penerapan fungsi diskriminan kuadratik klasik pada dataset *Glass Identification* menghasilkan klasifikasi sebanyak 117 data termasuk ke dalam *Type 1* dan 29 data termasuk ke dalam *Type 2*. Hasil ini menunjukkan distribusi klasifikasi dari model QDA klasik terhadap data berdasarkan komposisi kimianya.

3.5. Analisis Diskriminan Kuadratik *Robust*

Salah satu tujuan dari penelitian ini adalah membandingkan kinerja antara model diskriminan kuadratik klasik dan *robust.* Penerapan analisis diskriminan kuadratik *robust* dilakukan pada *Glass Identification Dataset*, yang sebelumnya telah direduksi dimensinya menggunakan PCA. Dalam hal ini, digunakan lima komponen utama pertama (PC1 hingga PC5) sebagai representasi dari data asli. Dalam metode *robust*, digunakan estimator MCD untuk menghitung vektor rata-rata dan matriks kovarian yang tahan terhadap pengaruh pencilan, sehingga model diskriminan menjadi lebih stabil dalam kondisi data yang mengandung *outlier*.

Langkah-langkah yang dilakukan dalam analisis diskriminan kuadratik *robust* serupa dengan diskriminan kuadratik klasik. Pertama, dihitung vektor rata-rata dan matriks kovarian *robust* dari masing-masing kelompok *Type 1* dan *Type 2* menggunakan estimator MCD. Berdasarkan hasil tersebut, dibentuk fungsi diskriminan kuadratik *robust* sebagai berikut:

$$\hat{d}_1^{RQ}(x) = -\frac{1}{2}(x - \bar{x}_{1MCD})' S_{1MCD}^{-1}(x - \bar{x}_{1MCD}) - \frac{1}{2}\ln(|S_{1MCD}|) + \ln(\frac{70}{146})$$
$$\hat{d}_2^{RQ}(x) = -\frac{1}{2}(x - \bar{x}_{2MCD})' S_{2MCD}^{-1}(x - \bar{x}_{2MCD}) - \frac{1}{2}\ln(|S_{2MCD}|) + \ln(\frac{76}{146})$$

Langkah selanjutnya adalah menghitung nilai $\hat{d}_1^{RQ}(x)$ dan $\hat{d}_2^{RQ}(x)$ untuk setiap data. Data akan diklasifikasikan ke dalam *Type 1* $\hat{d}_1^{RQ}(x) < \hat{d}_2^{RQ}(x)$, dan ke dalam *Type 2* apabila 1 $\hat{d}_2^{RQ}(x) < \hat{d}_1^{RQ}(x)$. Penerapan fungsi diskriminan kuadratik *robust* pada dataset *Glass Identification* menghasilkan klasifikasi sebanyak 61 data termasuk ke dalam *Type 1* dan 85 data termasuk ke dalam *Type 2*. Hasil ini menunjukkan distribusi klasifikasi dari model RQDA terhadap data berdasarkan komposisi kimianya.

3.6. Evaluasi Model

Evaluasi kinerja model dilakukan dengan menggunakan *apparent error rate* (APER), yaitu proporsi kesalahan klasifikasi yang terjadi pada data pelatihan. Semakin kecil nilai APER, maka semakin baik kemampuan model dalam melakukan klasifikasi**.**

3.6.1. Evaluasi Model Diskriminan Kuadratik Klasik

Tabel kontingensi untuk model QDA klasik ditunjukkan pada Tabel berikut:

Tabel 2. Hasil *Confusion Matrix* Model QDA pada dataset *Glass Identification*

		Data Prediksi (Hasil Klasifikasi)	
		Type 1	***Type 2***
Data Asli	***Type 1***	67	3
(Aktual)	***Type 2***	50	26

Sumber : Data diolah, 2025

Berdasarkan Tabel 2, jumlah pengamatan yang salah diklasifikasikan adalah sebanyak 53 observasi, yang terdiri dari 50 pengamatan *Type 2* yang salah diklasifikasikan sebagai *Type 1* dan 3 pengamatan *Type 1* yang salah diklasifikasikan sebagai *Type 2*. Dengan demikian, nilai APER untuk model QDA adalah:

$$APPER_{QDA} = \frac{50+3}{146} = 36{,}3\%$$

Hasil tersebut menunjukkan bahwa tingkat ketepatan klasifikasi (accuracy) model QDA klasik pada dataset ini adalah sebesar 63,7%.

3.6.2. Evaluasi Model Diskriminan Kuadratik *Robust*

Evaluasi kinerja model RQDA dilakukan menggunakan confusion matrix yang ditunjukkan pada Tabel 3.

Tabel 3. Hasil *Confusion Matrix* RQDA pada Dataset *Glass Identification*

		Data Prediksi (Hasil Klasifikasi)	
		Type 1	***Type 2***
Data Asli	***Type 1***	39	31
(Aktual)	***Type 2***	22	54

Sumber : Data diolah, 2025

Berdasarkan Tabel 3, jumlah pengamatan yang salah diklasifikasikan juga sebanyak 53 observasi, yang terdiri dari 31 pengamatan *Type 1* yang salah diklasifikasikan sebagai *Type 2* dan 22 pengamatan *Type 2* yang salah diklasifikasikan sebagai *Type 1*. Nilai APER untuk model RQDA diperoleh sebagai berikut:

$$APPER_{RQDA} = \frac{22 + 31}{146} = 36{,}3\%$$

Dengan demikian, tingkat klasifikasi model RQDA pada dataset ini juga sebesar 63,7%.

4.Kesimpulan

Penelitian ini membandingkan kinerja metode Quadratic Discriminant Analysis (QDA) klasik dan *Robust* Quadratic Discriminant Analysis (RQDA) setelah dilakukan reduksi dimensi menggunakan Principal Component Analysis (PCA) pada Glass Identification Dataset. Hasil PCA menunjukkan bahwa delapan variabel kimia dapat direduksi menjadi lima komponen utama yang secara kumulatif mampu menjelaskan 93,20% variasi total data, sehingga sebagian besar informasi penting tetap terwakili dalam ruang berdimensi lebih rendah.

Hasil klasifikasi menunjukkan bahwa QDA klasik dan RQDA menghasilkan tingkat akurasi yang sama, yaitu 63,7%, dengan nilai Apparent Error Rate (APER) sebesar 36,3%. Meskipun pada tahap awal teridentifikasi adanya pencilan dalam data, penggunaan estimator *robust* pada RQDA tidak memberikan peningkatan kinerja klasifikasi yang signifikan dibandingkan QDA klasik. Kondisi ini mengindikasikan bahwa penerapan PCA berperan dalam menstabilkan struktur data, sehingga pengaruh pencilan terhadap proses klasifikasi menjadi berkurang.

Temuan ini menunjukkan bahwa QDA klasik, yang lebih sederhana secara komputasi, masih dapat digunakan secara efektif sebagai metode klasifikasi awal pada permasalahan identifikasi kaca berdasarkan komposisi kimia, khususnya ketika data telah melalui proses reduksi dimensi yang memadai. Di sisi lain, RQDA tetap memiliki peran penting sebagai alternatif yang lebih andal ketika data yang dianalisis mengandung pencilan yang kuat atau menunjukkan penyimpangan yang signifikan dari asumsi normal multivariat.

Sebagai pengembangan penelitian selanjutnya, disarankan untuk menerapkan metode yang digunakan pada dataset empiris dari proses manufaktur kaca atau kasus forensik nyata, serta menggunakan teknik validasi silang (cross-validation) untuk mengevaluasi kemampuan generalisasi model secara lebih objektif. Pendekatan tersebut diharapkan dapat memberikan gambaran yang lebih komprehensif mengenai kinerja QDA dan RQDA pada kondisi data yang lebih beragam dan kompleks.

Daftar Pustaka

[1] Johnson, R. A., & Wichern, D. W. *Applied Multivariate Statistical Analysis*. Upper Saddle River, NJ: Pearson Prentice Hall, 2007.

[2] Sharma, S. *Applied Multivariate Techniques*. New York: John Wiley & Sons, 1996.

[3] Hubert, M., & Debruyne, M. Minimum Covariance Determinant. *WIREs Computational Statistics*, 2(1), 36–43, 2010.

[4] Rousseeuw, P. J., & Van Driessen, K. A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics*, 41(3), 212–223, 1999.

[5] Hubert, M., & Van Driessen, K. Fast and Robust Discriminant Analysis. *Computational Statistics & Data Analysis*, 45(2), 301–320, 2004.

[6] Croux, C., & Dehon, C. Robust Linear Discriminant Analysis Using S-Estimators. *Canadian Journal of Statistics*, 29(3), 473–493, 2001.

[7] Rencher, A. C. *Methods of Multivariate Analysis*. New York: John Wiley & Sons, 2002.

[8] Asnidar, Nirwan, & Raupong. Penerapan Analisis Diskriminan Kuadratik Robust pada Klasifikasi Desa. *ESTIMASI: Journal of Statistics and Its Application*, 4(1), 197–205, 2024.

[9] Martha, U. M., Vionanda, D., Permana, D., & Zilrahmi. Perbandingan Analisis Diskriminan Kuadratik dengan Analisis Diskriminan Kuadratik Robust. *UNP Journal of Statistics and Data Science*, 2(4), 469–474, 2024. https://doi.org/10.24036/ujsds/vol2-iss4/315

[10] Hadi, A. S. *Identifying Multiple Outliers in Multivariate Data*. New York: Wiley, 1992.

[11] Sari, P. P., Herdiani, E. T., & Sunusi, N. Outlier Detection Using Minimum Vector Variance Algorithm with Depth Function and Mahalanobis Distance. *Jurnal Matematika, Statistika & Komputasi*, 17(3), 418–427, 2021. https://doi.org/10.20956/j.v17i3.12629

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

[13] James, G., Witten, D., Hastie, T., & Tibshirani, R. *An Introduction to Statistical Learning* (2nd ed.). Springer, 2021.