

2.8	Evaluasi Hasil Klasifikasi	11
2.9	Defenisi Desa dan Kelurahan.....	12
BAB III METODE PENELITIAN.....		14
3.1	Sumber Data.....	14
3.2	Deskripsi Variabel	14
3.3	Tahap Analisis Data.....	15
BAB IV HASIL DAN PEMBAHASAN		18
4.1	Statistik Deskriptif	18
4.2	Pengujian Asumsi	19
4.2.1	Uji Kesamaan Vektor Rata-Rata.....	19
4.2.2	Uji Distribusi Normal Multivariat.....	19
4.2.3	Uji Kesamaan Matriks Varian Kovarian	20
4.3	Analisis Diskriminan Kuadratik Klasik	21
4.4	Pendeteksian Pencilan.....	23
4.5	Analisis Diskriminan Kuadratik <i>Robust</i>	24
4.5.1	Menduga Vektor Rata-Rata dan Matriks Varian Kovarian dengan Algoritma <i>fast-MCD</i>	25
4.5.2	Analisis Diskriminan Kuadratik <i>Robust</i>	28
4.6	Evaluasi Hasil Klasifikasi	30
BAB V PENUTUP.....		32
5.1	Kesimpulan	32
5.2	Saran	32
DAFTAR PUSTAKA		33

DAFTAR GAMBAR

Gambar 4. 1 QQ-Plot Hasil Pendataan Desa di Kab. Wajo Tahun 2019 20

DAFTAR TABEL

Tabel 2.1	Matriks Konfusi Hasil Klasifikasi Dua Kelompok	12
Tabel 3.1	Deskripsi Variabel Independen Hasil Pendataan Desa Kab. Wajo Tahun 2019	14
Tabel 4.1	Statistik Deskriptif Pendataan Desa di Kab. Wajo Tahun 2019 ..	19
Tabel 4.2	Hasil Klasifikasi Desa di Kab. Wajo Tahun 2019 Menggunakan Analisis Diskriminan Kuadratik Klasik	24
Tabel 4.3	Pengamatan Pencilan Hasil Pendataan Desa di Kab. Wajo Tahun 2019	25
Tabel 4.4	Contoh Data Multivariat Untuk Menduga Vektor Rata-Rata dan Matriks Varian Kovarian dengan Algoritma <i>fast-MCD</i>	26
Tabel 4.5	Jarak Mahalanobis contoh data multivariat terhadap \bar{x}_1 dan S_1 ...	27
Tabel 4.6	Hasil Pembobotan Pengamatan Contoh Data Multivariat	28
Tabel 4.7	Hasil Klasifikasi Desa di Kab. Wajo Tahun 2019 Menggunakan Analisis Diskriminan Kuadratik <i>Robust</i>	30
Tabel 4.8	Ketepatan Klasifikasi Desa di Kab. Wajo Tahun 2019 Menggunakan Analisis Diskriminan Kuadratik Klasik	31
Tabel 4.9	Ketepatan Klasifikasi Desa di Kab. Wajo Tahun 2019 Menggunakan Analisis Diskriminan Kuadratik <i>Robust</i>	32

DAFTAR LAMPIRAN

Lampiran 1.	Hasil Pendataan Desa di Kabupaten Wajo Tahun 2019	37
Lampiran 2.	<i>Output Uji Wilks Lambda</i>	38
Lampiran 3.	Nilai d_i^2 Setiap Data Desa di Kab. Wajo Tahun 2019	39
Lampiran 4.	<i>Output Uji Box's M</i>	40
Lampiran 5.	Hasil Klasifikasi Desa di Kab. Wajo Tahun 2019 Menggunakan Analisis Diskriminan Kuadratik Klasik	41
Lampiran 6.	Hasil Klasifikasi Desa di Kab. Wajo Tahun 2019 Menggunakan Analisis Diskriminan Kuadratik <i>Robust</i>	46

BAB I

PENDAHULUAN

1.1 Latar Belakang

Wilayah Indonesia dibagi ke dalam beberapa tingkat wilayah administratif, yaitu provinsi, kabupaten/kota, kecamatan, dan desa/kelurahan yang merupakan wilayah administratif terkecil. Setiap wilayah mempunyai karakteristik sosial ekonomi, akses ke fasilitas perkotaan, ciri dan tipologi lingkungan yang berbeda. Kondisi yang berbeda dan terus berubah tersebut dijadikan sebagai indikator untuk mengklasifikasi suatu wilayah menjadi desa atau kelurahan (Badan Pusat Statistik, 2010).

Salah satu wilayah yang terletak di Sulawesi tepatnya di Provinsi Sulawesi Selatan yaitu Kabupaten Wajo. Sebagaimana lazimnya yang terjadi disebagian besar wilayah tak luput dari perkembangan dalam wilayah kabupaten. Kabupaten Wajo kaya akan potensi sumber daya alam dan sumber daya manusia yang menjadi potensi yang cukup memadai untuk melaksanakan program pembangunan diberbagai aspek kehidupan. Potensi sumber-sumber ekonomi masih terus tingkatkan dan dikembangkan untuk kesejahteraan penduduk (Badan Pusat Statistik Kabupaten Wajo, 2019).

Analisis diskriminan adalah suatu teknik statistik yang dapat digunakan dalam pemisahan objek ke dalam kelompok yang berbeda dan mengalokasikan objek ke dalam suatu kelompok yang telah ditetapkan sebelumnya (Johnson, 2007). Oleh karena itu, analisis diskriminan dapat digunakan untuk mengklasifikasi wilayah ke dalam desa atau kelurahan. Analisis diskriminan mempunyai asumsi, vektor rata-rata antar kelompok berbeda, pengamatan berdistribusi normal multivariat dan matriks varian kovarian antar kelompok sama. Jika terdapat matriks varian kovarian dari dua atau beberapa kelompok berbeda, maka digunakan analisis diskriminan kuadratik untuk proses klasifikasi (Kurniasari, 2014).

Analisis diskriminan kuadratik mengklasifikasi dengan cara membentuk fungsi diskriminan setiap kelompok. Fungsi diskriminan diperoleh dari peluang *prior*, vektor rata-rata dan matriks varian kovarian setiap kelompok. Proses selanjutnya dengan menghitung skor diskriminan untuk setiap pengamatan dari masing-masing fungsi diskriminan, kemudian mengelompokkannya menggunakan skor diskriminan ke dalam kelompok tertentu yang telah ditetapkan sebelumnya.

Analisis diskriminan kuadratik tidak dapat bekerja dengan baik apabila pengamatan mengandung pencilan. Hal ini karena rata-rata sampel dan matriks varian kovarian sangat sensitif terhadap adanya pengamatan yang pencilan. Agar analisis diskriminan kuadratik tetap optimal dalam pengklasifikasian meskipun terdapat pengamatan yang pencilan diperlukan penduga yang *robust*. Salah satu penduga *robust* yang digunakan untuk mengatasi pengamatan pencilan adalah *fast-Minimum Covariance Determinant (fast-MCD)*. Penduga ini bertujuan untuk mencari pengamatan yang memiliki determinan matriks varian kovarian terkecil. Analisis diskriminan kuadratik yang menggunakan metode *robust* selanjutnya disebut sebagai analisis diskriminan kuadratik *robust* (Khiqmah, 2015).

Dalam penelitian (Kurniasari, 2014) dengan judul Perbandingan Pemisahan Desa/Kelurahan Di Kabupaten Semarang Menurut Status Daerah Menggunakan Analisis Diskriminan Kuadratik Klasik dan Diskriminan Kuadratik *Robust*. Penelitian tersebut menghasilkan analisis diskriminan kuadratik *robust* memberikan ketepatan hasil klasifikasi yang lebih besar daripada analisis diskriminan kuadratik klasik. Penelitian (Khiqmah,2015) dengan judul Perbandingan Diskriminan Kuadratik Klasik dan Diskriminan Kuadratik *Robust* pada Kasus Pengklasifikasian Peminatan Peserta Didik diperoleh analisis diskriminan kuadratik *robust* dengan penduga *fast-MCD* lebih tepat digunakan pada kasus data yang mengandung pencilan.

Berdasarkan latar belakang tersebut penulis bermaksud untuk melakukan pengklasifikasian wilayah yang ada di Kabupaten Wajo ke dalam kelompok desa atau kelompok kelurahan dengan teknik analisis diskriminan kuadratik klasik dan analisis diskriminan kuadratik *robust*.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang yang telah diuraikan, maka rumusan masalah yang didapatkan untuk tugas akhir ini yaitu:

1. Bagaimana analisis diskriminan kuadrat klasik dan analisis diskriminan kuadrat *robust* mengklasifikasi wilayah menjadi desa atau kelurahan?
2. Bagaimana akurasi ketepatan hasil klasifikasi desa atau kelurahan dengan teknik analisis diskriminan kuadrat klasik dan analisis diskriminan kuadrat *robust*?

1.3 Batasan Masalah

Penelitian ini dibatasi pada pengklasifikasian desa di Kabupaten Wajo Provinsi Sulawesi Selatan tahun 2019. Metode yang digunakan adalah analisis diskriminan kuadrat klasik dan analisis diskriminan kuadrat *robust* dengan penduga *fast-Minimum Covariance Determinant (fast-MCD)*.

1.4 Tujuan Penelitian

Berdasarkan masalah yang telah dirumuskan maka tujuan penulisan tugas akhir ini yaitu:

1. Untuk mengklasifikasi wilayah ke dalam kelompok desa atau kelompok kelurahan menggunakan analisis diskriminan kuadrat klasik dan analisis diskriminan kuadrat *robust*.
2. Untuk memperoleh akurasi hasil klasifikasi desa atau kelurahan dengan teknik analisis diskriminan kuadrat klasik dan analisis diskriminan *robust*.

1.5 Manfaat Penelitian

Penulisan tugas akhir ini dapat menambah wawasan pembaca mengenai teknik analisis diskriminan kuadrat klasik dan analisis diskriminan *robust*. Hasil pengklasifikasian wilayah termasuk desa atau kelurahan dapat menjadi informasi bagi pemerintah setempat dalam rangka pembangunan wilayah.

BAB II TINJAUAN PUSTAKA

2.1 Analisis Diskriminan

Analisis diskriminan adalah metode analisis yang bertujuan untuk memisahkan objek pengamatan yang berbeda dan mengalokasikan objek pengamatan baru ke dalam kelompok yang telah didefinisikan. Setiap objek yang diklasifikasikan akan menjadi anggota dari salah satu kelompok dan tidak ada objek yang menjadi anggota lebih dari satu kelompok (Johnson, 2007). Analisis diskriminan merupakan salah satu teknik analisis multivariat yang termasuk teknik dependensi, yaitu melihat pengaruh variabel dependen berdasarkan beberapa variabel independen. Dengan demikian, variabel dependen hasilnya tergantung dari data variabel independen. Analisis diskriminan digunakan pada kasus dimana variabel dependen berupa data kualitatif (nominal dan ordinal) dan variabel independen berupa data kuantitatif (interval dan rasio) (Annas & Irwan, 2015).

Menurut Rencher (1995), terdapat dua tujuan utama pemisahan kelompok dalam analisis diskriminan, yaitu:

- a. Aspek deskriptif, yaitu menggambarkan pemisahan kelompok, dengan fungsi diskriminan digunakan untuk mendeskripsikan atau menjelaskan perbedaan antara dua atau beberapa kelompok.
- b. Aspek prediksi, yaitu mengelompokkan pengamatan ke dalam kelompok, dengan fungsi diskriminan dari beberapa variabel digunakan untuk menentukan satu sampel individu atau objek ke dalam salah satu dari beberapa kelompok (Wati, 2013).

Model analisis diskriminan adalah sebuah persamaan yang menunjukkan suatu kombinasi linear dari berbagai variabel independen sebagai berikut:

$$D = b_0 + b_1X_1 + b_2X_2 + \dots + b_jX_j + \dots + b_pX_p, \quad (2.1)$$

dengan,

D = skor diskriminan

b_j = koefisien diskriminan variabel independen ke- j

X_j = variabel independen ke- j , dengan $j = 1, 2, \dots, p$

(Mega Selvia Tjahaya, 2022)

2.2 Uji Kesamaan Vektor Rata-Rata

Uji kesamaan vektor rata-rata dilakukan untuk menguji adanya perbedaan antar kelompok yang terbentuk dari setiap variabel independen. Pengujian ini dilakukan dengan hipotesis sebagai berikut:

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \dots = \boldsymbol{\mu}_g$$

$$H_1 : \text{Ada } \boldsymbol{\mu}_k \neq \boldsymbol{\mu}_{k'}, \text{ dengan } k \neq k', k = 1, 2, \dots, g$$

Statistik uji yang digunakan dalam pengujian ini adalah statistik *V-Barlett* yang menyebar mengikuti sebaran *chi-kuadrat* dengan derajat bebas $p(k - 1)$. Statistik *V-Barlett* diperoleh melalui:

$$V = - \left[(n - 1) - \frac{p+k}{2} \right] \ln(\Delta) \quad (2.2)$$

dengan,

$$\Delta = \frac{|W|}{|W+B|} = \text{wilks lambda} \quad (2.3)$$

$$W = \sum_{k=1}^g \sum_{i=1}^{n_k} (x_{ik} - \bar{x}_k)(x_{ik} - \bar{x}_k)' \quad (2.4)$$

$$B = \sum_{k=1}^g n_k (\bar{x}_k - \bar{x})(\bar{x}_k - \bar{x})' \quad (2.5)$$

x_{ik} = pengamatan ke- i kelompok ke- k

\bar{x}_k = vektor rata-rata kelompok ke- k

n_k = banyak pengamatan pada kelompok ke- k

\bar{x} = vektor rata-rata total

k = kelompok ke- k

g = banyaknya kelompok

Apabila $V > \chi_{p(k-1),\alpha}^2$ maka H_0 ditolak yang artinya terdapat perbedaan vektor rata-rata antar kelompok (Johnson, 2007).

2.3 Uji Distribusi Normal Multivariat

Menurut Johnson (2007), uji yang digunakan untuk mengetahui data berdistribusi normal multivariat atau tidak adalah menggunakan jarak Mahalanobis. Pengujian ini dilakukan dengan menggunakan QQ-Plot jarak mahalanobis (d_i^2) dan chi-kuadrat ($\chi_p^2(\frac{i-0,5}{n})$). Hipotesis untuk menguji data berdistribusi normal multivariat adalah sebagai berikut:

H_0 : data berdistribusi normal multivariat

H_1 : data tidak berdistribusi normal multivariat

Pengujian ini dilakukan dengan menggunakan jarak mahalanobis (d_i^2) dengan persamaan yaitu:

$$d_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}})'S^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}); i = 1, 2, \dots, n \quad (2.6)$$

dengan,

d_i^2 = Jarak mahalanobis pengamatan ke- i

\mathbf{x}_i = Vektor nilai pengamatan ke- i

$\bar{\mathbf{x}}$ = Vektor rata-rata variabel independen

S^{-1} = Invers matriks varian kovarian

Apabila titik-titik di QQ-Plot jarak mahalanobis dan chi-kuadrat cenderung membentuk garis lurus dan lebih dari 50% nilai dari d_i^2 lebih besar nilai $d_i^2 \leq \chi_p^2\left(\frac{i-0,5}{n}\right)$ maka H_0 diterima yang artinya data berdistribusi normal multivariat (Wati, 2013).

2.4 Uji Kesamaan Matriks Varian Kovarian

Menurut Johnson (2007), uji yang digunakan untuk mengetahui matriks varian kovarian antar kelompok sama atau tidak dapat dilakukan dengan menggunakan uji *Box's M*. Prosedur pada uji *Box's M* yaitu dengan menggunakan pendekatan distribusi chi-kuadrat dengan derajat bebas $\frac{1}{2}(g-1)p(p+1)$. Dengan hipotesis sebagai berikut:

$$H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_g$$

$$H_1 : \text{Ada } \Sigma_k \neq \Sigma_{k'}, \text{ dengan } k \neq k', k = 1, 2, \dots, g$$

Statistik ujinya dapat didefinisikan sebagai,

$$M = \sum_{k=1}^g (n_k - 1) \ln|S| - \sum_{k=1}^g (n_k - 1) \ln|S_k| \quad (2.7)$$

$$C^{-1} = 1 - \left(\frac{2p^2 + 3p - 1}{6(p+1)(g-1)} \right) \left(\sum_{k=1}^g \frac{1}{n_k - 1} - \frac{1}{\sum_{k=1}^g (n_k - 1)} \right) \quad (2.8)$$

dengan,

$$S = \frac{\sum_{k=1}^g (n_k - 1) S_k}{\sum_{k=1}^g n_k - 1} \quad (2.9)$$

S_k = matriks varian kovarian kelompok ke- k

$\bar{\mathbf{x}}_k$ = vektor rata-rata kelompok ke- k

n_k = banyaknya pengamatan kelompok ke- k

Jika nilai $MC^{-1} > \chi_{\alpha, \frac{1}{2}(k-1)p(p+1)}^2$, artinya tolak H_0 , maka matriks varian kovarian dari g kelompok berbeda (Johnson, 2007).

2.5 Analisis Diskriminan Kuadratik

Analisis diskriminan bertujuan untuk membentuk fungsi diskriminan yang mampu membedakan kelompok. Analisis ini dilakukan berdasarkan suatu perhitungan statistik terhadap objek-objek yang telah diketahui dengan jelas pengelompokannya. Pada analisis diskriminan, ada asumsi yang harus dipenuhi yaitu data berdistribusi normal multivariat dan matriks varian kovarian pada kelompoknya tidak beda. Namun terkadang ditemukan dari beberapa kasus terdapat matriks varian kovarian yang berbeda. Jika matriks varian kovarian berbeda, maka dapat digunakan fungsi diskriminan kuadratik untuk proses klasifikasi (Kurniasari, 2014) .

Pada kasus multivariat, vektor variabel acak $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p]'$ mengikuti fungsi kepadatan peluang $f(\mathbf{x})$. Apabila data sampel berdistribusi normal multivariat, maka fungsi kepadatan peluang $f(\mathbf{x})$ untuk dengan vektor rata-rata $\boldsymbol{\mu}$ dan matriks varian kovarian Σ dapat ditulis:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})}{2}\right) \quad (2.10)$$

Menurut Johnson (2007), Pengelompokan kelompok ke- k memiliki fungsi kepadatan normal multivariat dengan vektor rata-rata $\boldsymbol{\mu}_k$ dan matriks varian kovarian Σ_k

$$f_k(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma_k|^{\frac{1}{2}}} \exp\left(-\frac{(\mathbf{x} - \boldsymbol{\mu}_k)' \Sigma_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k)}{2}\right) \quad (2.11)$$

Perbedaan antar kelompok dapat dilihat dari fungsi kepadatannya $f_k(\mathbf{x})$ dengan peluang prior p_k dimana,

$$p_k f_k(\mathbf{x}) = p_k \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma_k|^{\frac{1}{2}}} \exp\left(-\frac{(\mathbf{x} - \boldsymbol{\mu}_k)' \Sigma_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k)}{2}\right) \quad (2.12)$$

Skor diskriminan kuadratik dapat dibentuk dari Persamaan (2.12)

$$d_k^Q(\mathbf{x}) = \ln p_k f_k(\mathbf{x}) = \ln \left\{ p_k \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma_k|^{\frac{1}{2}}} \exp\left(-\frac{(\mathbf{x} - \boldsymbol{\mu}_k)' \Sigma_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k)}{2}\right) \right\} \quad (2.13)$$

$$\ln p_k f_k(\mathbf{x}) = \ln p_k - (\ln 2\pi)^{\frac{p}{2}} - \ln(|\Sigma_k|)^{\frac{1}{2}} - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_k)' \Sigma_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k) \quad (2.14)$$

$$\ln p_k f_k(\mathbf{x}) = \ln p_k - \frac{p}{2} \ln(2\pi) - \frac{1}{2} \ln(|\Sigma_k|) - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_k)' \Sigma_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k) \quad (2.15)$$

Karena konstanta $\frac{p}{2} \ln(2\pi)$ bernilai sama untuk semua kelompok maka dapat diabaikan, sehingga Persamaan (2.15) menjadi:

$$\ln p_k f_k(\mathbf{x}) = \ln p_k - \frac{1}{2} \ln(|\Sigma_k|) - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_k)' \Sigma_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k) \quad (2.16)$$

Berdasarkan pengelompokan pada distribusi normal multivariat, maka skor diskriminan untuk kelompok- k didefinisikan sebagai:

$$d_k^Q(\mathbf{x}) = \ln p_k - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_k)' \Sigma_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k) - \frac{1}{2} \ln(|\Sigma_k|) \quad (2.17)$$

$\boldsymbol{\mu}_k$ dan Σ_k merupakan parameter yang tidak diketahui sehingga harus dicari estimasi dari sampelnya. Oleh karena itu, estimasi dari skor diskriminan kuadratik menjadi:

$$\hat{d}_k^Q(\mathbf{x}) = \ln p_k - \frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}}_k)' S_k^{-1} (\mathbf{x} - \bar{\mathbf{x}}_k) - \frac{1}{2} \ln(|S_k|) \quad (2.18)$$

dengan,

$\hat{d}_k^Q(\mathbf{x})$ = skor diskriminan kuadratik klasik kelompok ke- k

p_k = peluang *prior* kelompok ke- k

$\bar{\mathbf{x}}_k$ = vektor rata-rata kelompok ke- k

S_k^{-1} = invers matriks varian kovarian kelompok ke- k

$|S_k|$ = determinan matriks varian kovarian kelompok ke- k

Fungsi diskriminan kuadratik klasik (2.18) dapat dibentuk aturan pengelompokannya dengan menempatkan x_i ke dalam kelompok ke- k jika skor diskriminan kuadratik klasik:

$$d_r^Q(x) = \max \{ \hat{d}_1^Q(x), \hat{d}_2^Q(x), \dots, \hat{d}_g^Q(x) \} \quad (2.19)$$

Pendekatan yang digunakan untuk mengestimasi peluang prior (p_k) yaitu $p_k = \frac{n_k}{n}$ untuk setiap kelompok dengan p_k sebagai frekuensi relatif dari observasi pada setiap kelompok.

(Khiqmah, 2015)

2.6 Pendeteksian Pencilan

Menurut Gimenez *et.al.* (2012), pendekatan untuk mendeteksi adanya pencilan adalah dengan menghitung jarak mahalanobis dari masing-masing pengamatan. Jarak mahalanobis didefinisikan seperti pada Persamaan (2.6). Jika data mempunyai nilai $d_i^2 > \chi_{p,\alpha}^2$ maka data tersebut termasuk pencilan (Kurniasari, 2014).

2.7 Analisis Diskriminan Kuadratik *Robust*

Agar analisis diskriminan kuadratik tetap optimal dalam pengklasifikasian meskipun dalam kondisi data yang mengandung pencilan maka diperlukan suatu penduga yang *robust* terhadap data yang mengandung pencilan. Analisis diskriminan kuadratik yang menggunakan penduga yang *robust* selanjutnya akan disebut sebagai analisis diskriminan kuadratik *robust*. Konsep dasar dari analisis diskriminan ini adalah mengganti vektor rata-rata dan matriks varian kovarian yang *robust* ke dalam fungsi diskriminan kuadratik (Budyandra, 2016). Beberapa penduga yang dapat digunakan pada analisis diskriminan *robust* diantaranya adalah metode *the minimum volume ellipsoid* (MVE), *fast minimum covariance determinant* (fast-MCD), *minimum weighted covariance determinant* (MWCD), dan *M-estimator*.

2.7.1 Penduga *Minimum Covariance Determinant*

MCD merupakan salah satu penduga *robust* untuk analisis diskriminan kuadratik yang cukup populer. Penduga ini merupakan penduga *robust* untuk membentuk rata-rata dan matriks varian kovarian yang bertujuan untuk mencari h pengamatan dari total pengamatan n yang memiliki determinan matriks varian kovarian terkecil, dimana h adalah bilangan bulat terkecil dari $(n+p+1)/2$. Misalkan terdapat vektor acak $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p]$ dari sejumlah n pengamatan yang terdiri dari p dan $n \geq p + 1$. Penduga MCD merupakan pasangan rata-rata sub sampel $\bar{\mathbf{x}}$ dan matriks varian kovarian S dari suatu sub sampel berukuran h pengamatan dimana $\frac{n+p+1}{2} \leq h < n$ dengan

$$\bar{\mathbf{x}} = \frac{1}{h} \sum_{i=1}^h x_{ih} \quad (2.20)$$

$$S = \frac{1}{h-1} \sum_{i=1}^h (x_i - \bar{x})(x_i - \bar{x})' \quad (2.21)$$

Algoritma yang terkenal dalam menaksir penduga MCD adalah algoritma *fast-MCD* yang diperkenalkan oleh Rousseeuw dan Van Driessen yang mampu bekerja lebih cepat dan mampu menangani himpunan data yang sangat besar. Untuk menghitung MCD dengan algoritma *fast-MCD* dilakukan dengan langkah-langkah sebagai berikut:

- a. Menentukan pengamatan yang diambil yaitu $h = (n + p + 1)/2$ dengan syarat $\frac{n+p+1}{2} \leq h < n$, dengan demikian akan terbentuk himpunan pengamatan baru sebanyak $\binom{n}{h}$
- b. Memilih secara acak himpunan pertama (H_1). Berdasarkan himpunan H_1 menghitung vektor rata-rata dan matriks varian kovarian (\bar{x}_1, S_1) . Kemudian selanjutnya menghitung $\det(S_1)$.
- c. Kemudian menghitung jarak mahalanobis dari setiap pengamatan terhadap rata-rata \bar{x}_1 dan kovarian S_1 menggunakan rumus:

$$d(i) = \sqrt{(x_i - \bar{x}_1)' S_1^{-1} (x_i - \bar{x}_1)} \quad (2.22)$$

- d. Mengurutkan pengamatan tersebut berdasarkan jarak mahalanobis dari yang terkecil ke yang terbesar.
- e. Mengambil elemen dari h pengamatan dengan jarak terkecil untuk menjadi elemen himpunan H_2 .
- f. Menghitung vektor rata-rata dan matriks varian kovarian (\bar{x}_2, S_2) dari himpunan H_2 .
- g. Membandingkan $\det(S_2)$ dengan $\det(S_1)$. Bila $\det(S_2) > \det(S_1)$ ulangi langkah poin b sampai d, sampai ditemukan $\det(S_{m+1}) \leq \det(S_m)$.
- h. Menetapkan anggota himpunan H_m sebagai himpunan dengan determinan matriks varian kovarian terkecil
- i. Berdasarkan himpunan H_m data selanjutnya diberi bobot

$$w_i = \begin{cases} 1 & \text{jika } (x_i - \bar{x}_m)' S_m^{-1} (x_i - \bar{x}_m) \leq \chi_{p,\alpha}^2 \\ 0 & \text{jika lainnya} \end{cases} \quad (2.23)$$

j. Berdasarkan pembobot pada (2.22), maka penduga *fast*-MCD dihitung:

$$\bar{x}_{MCD} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad (2.24)$$

$$S_{MCD} = \frac{\sum_{i=1}^n w_i (x_i - \bar{x}_{MCD})(x_i - \bar{x}_{MCD})'}{\sum_{i=1}^n w_i - 1} \quad (2.25)$$

(Suryana, 2008).

2.7.2 Analisis Diskriminan Kuadratik *Robust*

Menurut Hubert dan Driessen (2002), skor diskriminan kuadratik dengan menggunakan penduga *robust* MCD diperoleh dengan mengganti penduga vektor rata-rata dan matriks varian kovarian dengan penduga MCD, sebagai berikut:

$$\hat{d}_k^{QMCD}(x) = \ln p_k - \frac{1}{2} (x - \bar{x}_{MCD(k)})' S_{MCD(k)}^{-1} (x - \bar{x}_{MCD(k)}) - \frac{1}{2} \ln |S_{MCD(k)}| \quad (2.26)$$

dengan,

$\hat{d}_k^{QMCD}(x)$ = skor diskriminan kuadratik *robust* kelompok ke- k

p_k = peluang prior kelompok ke- k

$\bar{x}_{(MCD)k}$ = vektor rata-rata MCD kelompok ke- k

$S_{(MCD)k}^{-1}$ = invers matriks varian kovarian MCD kelompok ke- k

$|S_{(MCD)k}|$ = determinan matriks varian kovarian MCD kelompok ke- k

Pengamatan x_i akan masuk ke dalam kelompok ke- k jika skor diskriminan kuadratik

$$\hat{d}_k^{QMCD}(x) = maks \{ \hat{d}_1^{QMCD}(x), \hat{d}_2^{QMCD}(x), \dots, \hat{d}_g^{QMCD}(x) \} \quad (2.27)$$

(Khiqmah, 2015)

2.8 Evaluasi Hasil Klasifikasi

Setelah proses klasifikasi, selanjutnya dapat diketahui seberapa tepat aturan klasifikasi yang telah ditentukan dengan menghitung nilai *Error Rate* yaitu proporsi kesalahan klasifikasi. Metode untuk menghitung probabilitas kesalahan klasifikasi adalah *Apparent Error Rate* (APER).

Tingkat kesalahan dapat dengan mudah dihitung dengan menggunakan matriks konfusi, yang menunjukkan jumlah keanggotaan kelompok asli dengan jumlah keanggotaan kelompok hasil klasifikasi dengan bentuk:

Tabel 2. 1 Matriks Konfusi Hasil Klasifikasi Dua Kelompok

		Kelompok Hasil Klasifikasi	
		k_1	k_2
Kel. Asli	k_1	n_{11}	n_{12}
	k_2	n_{21}	n_{22}

Statistik uji nilai *APER* dirumuskan:

$$APER = \left(\frac{n_{12} + n_{21}}{n_1 + n_2} \right) \times 100\% \tag{2.28}$$

dengan,

n_{11} = jumlah dari k_1 yang tepat diklasifikasikan sebagai k_1

n_{12} = jumlah dari k_1 yang salah diklasifikasikan sebagai k_1

n_{21} = jumlah dari k_2 yang salah diklasifikasikan sebagai k_2

n_{22} = jumlah dari k_2 yang tepat diklasifikasikan sebagai k_2

n_1 = jumlah pengamatan dalam kelompok k_1

n_2 = jumlah pengamatan dalam kelompok k_2

(Khiqmah, 2015).

2.9 Defenisi Desa dan Kelurahan

Menurut Undang-Undang Nomor 6 Tahun 2014 tentang Desa, Desa adalah desa dan desa adat atau yang disebut dengan nama lain, selanjutnya disebut Desa adalah kesatuan masyarakat hukum yang memiliki batas wilayah yang berwenang untuk mengatur dan mengurus urusan pemerintahan, kepentingan masyarakat setempat berdasarkan Prakarsa masyarakat, hak asal usul, dan/atau hak tradisional yang diakui dan dihormati dalam system Pemerintah NKRI.

Menurut Undang-Undang No. 23 Tahun 2014 Tentang Pemerintah Daerah Kelurahan adalah pembagian wilayah administrasi di Indonesia di bawah kecamatan. Kelurahan dipimpin oleh seorang kepala kelurahan yang disebut Lurah selaku perangkat kecamatan dan bertanggung jawab kepada Camat. Lurah diangkat oleh Bupati/Walikota atas usul Sekretaris Daerah dari Pegawai Negeri Sipil yang memenuhi persyaratan sesuai dengan ketentuan peraturan perundang-undangan (Badan Pusat Statistik Kabupaten Wajo, 2019).

BAB III METODE PENELITIAN

3.1 Sumber Data

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari publikasi Badan Pusat Statistik Kabupaten Wajo yang diakses pada website <https://wajokab.bps.go.id> tanggal 25 Januari 2021. Data tersebut adalah hasil pendataan 176 Desa/Kelurahan di 14 Kecamatan yang ada di Kabupaten Wajo Tahun 2019 yang tertuang dalam buku Kecamatan dalam Angka 2020. Terdapat 128 desa dan 48 kelurahan terlampir pada Lampiran 1.

3.2 Deskripsi Variabel

Adapun variabel yang digunakan dalam penelitian ini yaitu:

1. Variabel Dependen

Variabel dependen yang digunakan dalam penelitian ini adalah dua kelompok yaitu kelompok desa dan kelompok kelurahan dengan kode sebagai berikut:

1 = desa

2 = kelurahan

2. Variabel Independen

Variabel independen yang digunakan dalam penelitian ini adalah 7 variabel seperti pada Tabel 3.1.

Tabel 3. 1 Deskripsi Variabel Independen Hasil Pendataan Desa Kab. Wajo Tahun 2019

Variabel	Nama Variabel	Skala Pengukuran
X_1	Kepadatan Penduduk	Rasio
X_2	TK	Rasio
X_3	SMP	Rasio
X_4	SMA	Rasio
X_5	Pasar	Rasio
X_6	Toko	Rasio
X_7	Fasilitas Kesehatan	Rasio

Sumber : Data diolah, 2022