

# ANALISIS DISKRIMINAN PADA KLASIFIKASI DESA DI KABUPATEN TABANAN MENGGUNAKAN METODE *K-FOLD* *CROSS VALIDATION*

Ida Ayu Made Supartini<sup>1§</sup>, I Komang Gde Sukarsa<sup>2</sup>, I Gusti Ayu Made Srinadi<sup>3</sup>

<sup>1</sup>Jurusan Matematika, Fakultas MIPA – Universitas Udayana [Email: dayu.supartini@gmail.com]

<sup>2</sup>Jurusan Matematika, Fakultas MIPA – Universitas Udayana [Email: sukarsakomang@yahoo.com]

<sup>3</sup>Jurusan Matematika, Fakultas MIPA – Universitas Udayana [Email: srinadi@unud.ac.id]

<sup>§</sup>Corresponding Author

## ABSTRACT

*Tabanan Regency is one of the eight regencies and one municipality in Bali Province. Administratively, it is divided into 10 districts and 133 villages. There are rural areas and urban areas in the regions. Discriminant analysis is a technique related to the separation of objects into different groups that have been set previously. The purpose of this research is to classify 133 villages in Tabanan Regency into urban or rural groups with discriminant analysis. Linear discriminant analysis assumes that the covariance matrix of the two groups are equals, if the assumption of equality of covariance matrix is violated, quadratic discriminant analysis can be used for classification. This research uses k-fold cross validation method for calculating the accuracy of quadratic discriminant function where  $k = 2,3,4$ . Quadratic discriminant function is obtained by  $k = 4$  with the smallest APER value (0.09). All of classification results are stable and consistence.*

*Keywords: quadratic discriminant analysis, K-fold cross validation, clasiffication, rural, urban*

## 1. PENDAHULUAN

Seiring dengan perkembangan teknologi komputer dan pengaplikasian perangkat lunak statistika dalam analisis data, ilmu statistika mengalami perkembangan yang pesat dalam hal pengkajian data. Salah satu kajian dalam analisis statistika adalah kajian yang membahas tentang pengelompokkan suatu individu baru ke dalam kelompok yang sudah ada berdasarkan karakteristik data. Analisis peubah ganda yang digunakan untuk tujuan tersebut adalah analisis diskriminan.

Pada analisis diskriminan, estimasi parameter bisa menggunakan metode *Maximum Likelihood Estimation* (MLE). Metode MLE digunakan sebagai penduga parameter apabila asumsi sebaran normal ganda terpenuhi. Apabila asumsi sebaran normal ganda tidak terpenuhi, salah satu metode yang dapat digunakan untuk mengestimasi parameter adalah metode *bootstrap*. Fungsi yang terbentuk dalam analisis diskriminan disebut fungsi diskriminan. Selain estimasi parameter, pada analisis diskriminan bisa dilakukan suatu validasi keakuratan model

fungsi diskriminan. Validasi keakuratan model bisa menggunakan metode *cross validation*. Prinsip dasar metode *cross validation* adalah membagi keseluruhan data menjadi data *training* dan data *testing* (Davidson & Hinkley, 1997). Penelitian ini menggunakan metode *k-fold cross validation*.

Analisis diskriminan bisa diterapkan pada bidang pendidikan, industri, kesehatan, sosial, ekonomi, dan lain-lain. Penerapan analisis diskriminan pada penelitian ini yaitu pada klasifikasi wilayah desa di Kabupaten Tabanan. Menurut Badan Pusat Statistik (2010), wilayah Indonesia dibagi ke dalam beberapa tingkat wilayah administratif, yaitu provinsi, kabupaten/kota, kecamatan, dan desa atau disebut dengan nama lain yang merupakan wilayah administratif terkecil. Secara administrative Kabupaten Tabanan dibagi menjadi 10 kecamatan dan 133 desa.

Wilayah kabupaten mempunyai bagian-bagian yang merupakan kawasan perdesaan dan kawasan perkotaan. Setiap desa mempunyai

karakteristik sosial, ekonomi, kondisi dan akses lingkungan yang berbeda-beda dan akan terus berubah seiring dengan kemajuan tingkat pembangunan di suatu desa. Badan Pusat Statistik menggunakan kondisi yang berbeda dan terus mengalami perubahan tersebut sebagai indikator untuk menggolongkan suatu desa ke dalam desa perkotaan atau desa perdesaan. Pada pelaksanaannya, penentuan apakah suatu desa/kelurahan termasuk daerah perkotaan atau perdesaan dilakukan oleh Badan Pusat Statistik berdasarkan Peraturan Kepala Badan Pusat Statistik Nomor 37 Tahun 2010 tentang Klasifikasi Perkotaan dan Perdesaan di Indonesia (Badan Pusat Statistik, 2010).

Data klasifikasi desa perkotaan dan desa perdesaan digunakan sebagai dasar untuk merencanakan kegiatan sensus atau survei (Badan Pusat Statistik, 2010). Selain digunakan sebagai dasar untuk merencanakan kegiatan sensus atau survei, data klasifikasi desa perkotaan dan desa perdesaan juga dapat digunakan untuk perencanaan pembangunan dalam hal pemekaran wilayah. Berdasarkan Peraturan Kepala Badan Pusat Statistik Nomor 37 Tahun 2010 Pasal 4 menyebutkan bahwa apabila ada pembentukan desa/kelurahan/UPT baru, di mana desa/kelurahan baru tidak memiliki desa/kelurahan induk, maka status perkotaan/perdesaan dari desa/kelurahan baru tersebut harus ditentukan dengan mengimplementasikan kriteria wilayah perkotaan yang sama (Badan Pusat Statistik, 2010).

Menurut Tarigan (2003), perencanaan pembangunan wilayah tersebut mencakup berbagai aspek yang tentunya mempertimbangkan peran keterkaitan antara desa dan kota. Sehingga status dari suatu desa/kelurahan apakah termasuk dalam daerah perdesaan atau perkotaan sangat penting untuk diketahui oleh pemerintah dalam hal perencanaan pembangunan di daerah desa. Berdasarkan kriteria desa perkotaan dan desa perdesaan yang tercantum dalam peraturan tersebut penulis bermaksud melakukan pengklasifikasian desa di Kabupaten Tabanan ke dalam kelompok daerah perkotaan atau perdesaan dengan teknik analisis diskriminan menggunakan *k-fold cross validation*.

Tujuan dari penelitian ini adalah untuk mengestimasi model klasifikasi desa di Kabupaten Tabanan dengan teknik analisis diskriminan menggunakan *k-fold cross validation*, untuk mengetahui hasil klasifikasi

ke fasilitas perkotaan, ciri dan tipologi desa di Kabupaten Tabanan dengan teknik analisis diskriminan menggunakan *k-fold cross validation*, dan untuk mengetahui ketepatan hasil klasifikasi desa di Kabupaten Tabanan dengan teknik analisis diskriminan menggunakan *Apparent Error Rate* (APER).

## 2. TINJAUAN PUSTAKA

### Analisis Diskriminan

Analisis Diskriminan adalah metode analisis peubah ganda yang bertujuan untuk memisahkan objek pengamatan yang berbeda dan mengalokasikan objek pengamatan baru ke dalam kelompok yang telah didefinisikan (Johnson & Wichern, 2007). Bentuk umum dari fungsi diskriminan yaitu:

$$Z_{jk} = a + W_1X_{1k} + W_2X_{2k} + \dots + W_iX_{ik} \quad (1)$$

Keterangan:

$Z_{jk}$  = skor diskriminan dari fungsi diskriminan ke- $j$  untuk objek ke- $k$ ;  $j = 1, 2, \dots, s$  dan  $s \leq \min(n-1, p)$ ,  $k = 1, 2, \dots, n$ ;

$a$  = intersep atau koefisien persamaan fungsi diskriminan;

$W_i$  = bobot diskriminan untuk peubah bebas ke- $i$ ;  $i = 1, 2, \dots, p$ ;

$X_{ik}$  = peubah bebas ke- $i$  untuk objek ke- $k$ ;  $i = 1, 2, \dots, p$  dan  $k = 1, 2, \dots, n$ .

Jika matriks ragam-peragam antar kelompok tidak homogen maka skor yang dibentuk adalah skor diskriminan kuadratik (Johnson & Wichern, 2007). Fungsi diskriminan kuadratik dirumuskan sebagai berikut:

$$d_k^Q(x) = -\frac{1}{2} \ln |\Sigma_k| - \frac{1}{2} (x - \mu_k)' \Sigma_k^{-1} (x - \mu_k) + \ln p_k \quad (2)$$

Keterangan:

$\Sigma_k$  = matriks ragam-peragam kelompok ke- $k$ ,  
 $\Sigma_k^{-1}$  = *inverse* matriks ragam-peragam kelompok ke- $k$ ,

$\mu_k$  = vektor rata-rata kelompok ke- $k$ ,

$p_k$  = *prior probability* kelompok ke- $k$ .

### Uji Distribusi Normal Ganda

Uji distribusi normal ganda dapat dilakukan menggunakan plot pasangan nilai jarak Mahalanobis dan nilai khi-kuadrat ( $d_i^2, \chi^2 p((i - \frac{1}{2})/n)$ ). Nilai jarak Mahalanobis dihitung dengan persamaan sebagai berikut:

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

Keterangan:

$x_i$  = sampel pengamatan,

$\bar{x}$  = vektor rata-rata,

$S^{-1}$  = inverse matriks ragam-peragam.

Selanjutnya nilai  $d_i^2$  diurutkan dari nilai terkecil sampai nilai terbesar dan dibuat plot  $d_i^2$  dan nilai  $\chi^2 p((i - \frac{1}{2})/n)$ . Jika plot yang dihasilkan dapat didekati dengan garis lurus atau berada di sekitar garis lurus, maka disimpulkan bahwa data berdistribusi normal ganda (Johnson & Wichern, 2007). Uji hipotesis untuk mengetahui apakah data telah berdistribusi normal ganda atau tidak yaitu:

$H_0$  : data berdistribusi normal ganda,

$H_1$  : data tidak berdistribusi normal ganda.

Statistik uji:

Jika pada  $\alpha = 0,05$ , lebih dari 50% nilai dari  $d_i^2 < \chi^2_{(0,05;p)}$  maka terima  $H_0$ , yang artinya bahwa data berdistribusi normal ganda dan sebaliknya.

### Uji Vektor Nilai Rata-Rata

Pengujian hipotesis terhadap vektor nilai rata-rata antar kelompok yaitu:

$H_0$  :  $\mu_1 = \mu_2 = \dots = \mu_k$ ,

$H_1$  : minimal ada dua kelompok yang berbeda dengan  $\mu_i \neq \mu_j$ ;  $i \neq j$  dengan  $i$  dan  $j = 1, 2, \dots, k$ .

Statistik V-Bartlett didefinisikan sebagai berikut:

$$V = - \left[ (n - 1) - \frac{p+k}{2} \right] \ln (\Lambda) \quad (4)$$

$$\text{dengan Wilk Lambda } \Lambda = \frac{|W|}{|B+W|}. \quad (5)$$

Jika  $V > \chi^2_{p(k-1), (1-\alpha)}$  maka  $H_0$  ditolak. Hal ini menunjukkan bahwa terdapat perbedaan vektor nilai rata-rata antarkelompok (Mattjik & Sumertajaya, 2011).

Sedangkan jika menggunakan perangkat lunak statistika seperti SPSS, uji ini dilakukan secara univariat atau yang diuji bukan berupa vektor tetapi dengan bantuan tabel *Tests of Equality of Group Means*. Statistik uji: Jika pada  $\alpha = 0,05$ , nilai  $p\text{-value} > \alpha$  maka terima  $H_0$ . Hal ini berarti bahwa tidak ada perbedaan rata-rata dalam kelompok. Sebaliknya, jika pada  $\alpha = 0,05$ , nilai  $p\text{-value} < \alpha$  maka tolak  $H_0$ , yang artinya bahwa ada perbedaan rata-rata dalam kelompok.

### Uji Kehomogenan Ragam ( Uji M Box)

Statistik uji M Box dapat digunakan untuk menguji asumsi kehomogenan ragam. Hipotesis uji M Box yaitu:

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

$H_1$ : minimal ada dua kelompok yang berbeda dengan  $\Sigma_i \neq \Sigma_j$  untuk  $i \neq j$  dengan  $i$  dan  $j = 1, 2, \dots, k$ .

Statistik uji M Box yaitu:

$$-2 \ln \lambda^* = (n - k) \ln \left| \frac{W}{(n-k)} \right| - \sum_{i=1}^k (n_i - 1) \ln |S_i| \quad (6)$$

$$\text{dengan } \lambda^* = \frac{\prod_{i=1}^k |S_i|^{(n_i-1)/2}}{\left| \frac{W}{(n-k)} \right|^{(n-k)/2}}. \quad (7)$$

Keterangan:

$k$  = banyaknya kelompok,

$p$  = banyaknya peubah prediktor,

$n$  = banyaknya pengamatan,

$a$  = vektor pembobot,

$S_i$  = matriks ragam-peragam kelompok ke- $i$ ,

$W$  = matriks jumlah kuadrat dan hasil kali data dalam kelompok,

$v_1, v_2$  = derajat bebas.

Jika diperoleh  $\frac{-2 \ln \lambda^*}{b} \leq F_{v_1, v_2, \alpha}$  dan  $p - value > \alpha$  maka  $H_0$  diterima. Hal ini menunjukkan bahwa semua kelompok mempunyai matriks ragam-peragam yang homogen (Mattjik & Sumertajaya, 2011).

**Pencilan Peubah Ganda**

Pengamatan ke- $i$  didefinisikan sebagai data pencilan peubah ganda jika jarak Mahalanobisnya lebih besar dari nilai khi-kuadratnya pada  $p$  buah peubah (Johnson & Wichern, 2007).

$$d_{MD}^2 = (x_i - \bar{x})' S^{-1} (x_i - \bar{x}) > \chi^2_{p, (1-\alpha)} \quad (8)$$

Keterangan:

- $x_i$  = data pengamatan ke- $i$ ,
- $\bar{x}$  = vektor rata-rata kelompok,
- $S^{-1}$  = inverse matriks ragam-peragam kelompok.

**Metode K- Fold Cross Validation**

Cross Validation merupakan teknik untuk memvalidasi keakuratan sebuah model yang dibangun berdasarkan data set tertentu. Data yang digunakan dalam proses pembentukan model disebut sebagai data latih atau *training* dan data yang digunakan untuk memvalidasi model disebut sebagai data *testing* (Davidson & Hinkley, 1997).

Pada pendekatan metode *k-fold cross validation*, data set dibagi menjadi sejumlah  $k$  buah partisi secara acak. Selanjutnya, dilakukan sejumlah  $k$ -kali eksperimen dengan masing-masing eksperimen menggunakan data partisi ke- $k$  sebagai data *testing* dan menggunakan sisa partisi lainnya sebagai data *training*. Eksperimen yang akan dilakukan sesuai dengan jumlah partisi yang dilakukan.

**Apparent Error Rate (APER)**

Apparent Error Rate (APER) didefinisikan sebagai nilai dari besar kecilnya jumlah observasi yang salah diklasifikasikan oleh fungsi klasifikasi (Johnson & Wichern, 2007). Tingkat

kesalahan klasifikasi diperoleh menggunakan tabel kesalahan klasifikasi berikut.

Tabel 1. Kesalahan Klasifikasi

Hasil Observasi ( <i>actual class</i> )	Hasil Prediksi ( <i>predicted class</i> )	
	Kelompok 1	Kelompok 2
Kelompok 1	$n_{11}$	$n_{12}$
Kelompok 2	$n_{21}$	$n_{22}$

$$APER = \frac{\sum n_{ij}}{N}; i \neq j. \quad (9)$$

**Uji Keakuratan**

Analisis kriteria kemungkinan proporsional (*proportional chance criterion*) dapat digunakan untuk membandingkan proporsi pengamatan yang diklasifikasikan dengan benar dan sesuai dengan peluang proporsi yang diharapkan (Hair *et al.*, 2009) yang dirumuskan sebagai berikut:

$$C_{pro} = p^2 + (1 - p)^2. \quad (10)$$

Keterangan:

- $p$  = proporsi jumlah sampel di kelompok 1,
- $(1 - p)$  = proporsi jumlah sampel di kelompok 2.

Nilai  $C_{max}$  adalah ekspektasi klasifikasi yang benar untuk kelompok yang dipilih. Perhitungan  $C_{max}$  berdasarkan asumsi bahwa semua pengamatan dikategorikan berasal dari kelompok tersebut dan dirumuskan sebagai berikut:

$$C_{max} = \left( \frac{n_{max}}{N} \right) \times 100\%. \quad (11)$$

Keterangan:

- $n_{max}$  = jumlah sampel terbesar pada salah satu kelompok,
- $N$  = jumlah keseluruhan sampel.

Uji hipotesis keakuratan hasil klasifikasi yaitu:

- $H_0$  : Klasifikasi akurat,
- $H_1$  : Klasifikasi tidak akurat.

Statistik uji:

Jika  $Hit_{Ratio} > C_{max} > C_{pro}$  maka terima  $H_0$ , yang artinya bahwa hasil klasifikasi sudah akurat. Sebaliknya, jika  $C_{pro} < Hit_{Ratio} < C_{max}$  maka tolak  $H_0$ . Hal tersebut mengartikan bahwa hasil klasifikasi tidak akurat (Arisona, 2015).

### Uji Kestabilan

Nilai Q Press digunakan untuk menguji apakah pengalokasian dari setiap sampel dalam kelompok relatif stabil atau tidak sebagai akibat adanya perubahan perbedaan jumlah sampel yang diteliti. Nilai Q Press dirumuskan sebagai berikut:

$$Q \text{ Press} = \frac{[N-(nk)]^2}{N(k-1)}. \quad (12)$$

Uji hipotesis kestabilan hasil klasifikasi yaitu:

$H_0$  : Klasifikasi konsisten,

$H_1$  : Klasifikasi tidak konsisten.

Statistik uji: Jika nilai  $Q \text{ Press} > \chi^2_{(1,\alpha)}$  maka dapat disimpulkan bahwa keakuratan pengklasifikasian adalah konsisten dan sebaliknya (Hair *et al.*, 2009).

## 3. METODE PENELITIAN

### Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik Kabupaten Tabanan. Data tersebut merupakan data hasil pendataan setiap kecamatan yang ada di Kabupaten Tabanan pada tahun 2015.

### Identifikasi Peubah Penelitian

Peubah prediktor pada penelitian ini adalah kepadatan penduduk per  $km^2$  ( $X_1$ ), banyak pasar yang tersedia ( $X_2$ ), banyak Sekolah Taman Kanak-Kanak yang tersedia (TK) ( $X_3$ ), banyak Sekolah Menengah Pertama (SMP) yang tersedia ( $X_4$ ), banyak Sekolah Menengah Umum (SMU) yang tersedia ( $X_5$ ), banyak pertokoan yang tersedia ( $X_6$ ), dan banyak hotel yang

tersedia ( $X_7$ ). Status daerah desa di Kabupaten Tabanan yang digunakan sebagai peubah respon ( $Y$ ) dibedakan menjadi dua kelompok yaitu daerah perkotaan dan daerah perdesaan.

### Metode Analisis Data

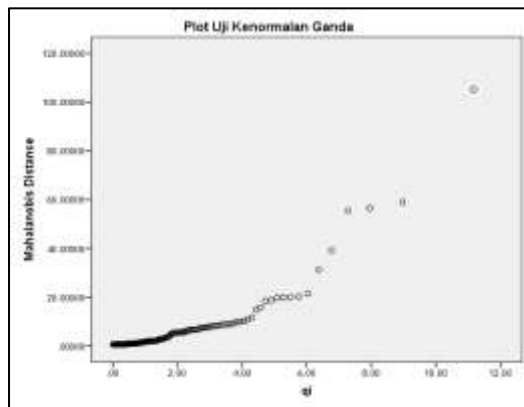
1. Mendeskripsikan data karakteristik klasifikasi desa di Kabupaten Tabanan terdiri dari tujuh peubah prediktor dan dua kelompok peubah respon.
2. Uji asumsi dasar analisis diskriminan yang terdiri dari:
  - a. Uji distribusi normal ganda pada data menggunakan plot antara nilai jarak Mahalanobis dan nilai khi-kuadrat.
  - b. Uji kesamaan vektor rata-rata menggunakan uji V-Bartlett atau tabel *Tests of Equality of Group Means*.
  - c. Uji kehomogenan matriks ragam-peragam pada data menggunakan uji M Box.
  - d. Menguji pencilan (*outlier*) pada data menggunakan jarak kuadrat Mahalanobis.
3. Mengevaluasi signifikansi peubah pembeda.
4. Melakukan uji *k-fold cross validation* pada data. Data dibagi menjadi data *training* dan data *testing* dengan persentase proporsi tertentu secara acak.
5. Mengestimasi fungsi diskriminan.
6. Melakukan klasifikasi menggunakan analisis diskriminan.
7. Uji ketepatan klasifikasi yang terdiri dari:
  - a. Uji tingkat kesalahan klasifikasi menggunakan *Apparent Error Rate* (APER).
  - b. Uji keakuratan klasifikasi menggunakan analisis kriteria kemungkinan proporsional (*proportional chance criterion*).
  - c. Uji kestabilan klasifikasi menggunakan nilai Q Press.

#### 4. HASIL DAN PEMBAHASAN

Berdasarkan data Kabupaten Tabanan dalam angka tahun 2015, sejumlah 105 desa (79%) yang berada di Kabupaten Tabanan merupakan daerah perdesaan, sedangkan sisanya yaitu 28 desa (21%) merupakan daerah perkotaan.

##### Uji Distribusi Normal Ganda

Plot pasangan nilai jarak Mahalanobis dan nilai khi-kuadrat ( $d_i^2, \chi^2 p((i - \frac{1}{2})/n)$ ) dapat dilihat pada Gambar 1 berikut:



Gambar 1. Plot Distribusi Normal Ganda Seluruh Peubah Prediktor

Plot yang dibentuk oleh nilai jarak Mahalanobis ( $d_i^2$ ) dan nilai khi-kuadrat ( $\chi^2 p(i - \frac{1}{2})/n$ ) pada Gambar 1 cenderung membentuk garis lurus. Selanjutnya dilakukan pengujian asumsi distribusi normal ganda menggunakan jarak Mahalanobis seperti pada persamaan (3). Uji hipotesis untuk mengetahui apakah data telah berdistribusi normal ganda atau tidak yaitu:

$H_0$  : data berdistribusi normal ganda,

$H_1$  : data tidak berdistribusi normal ganda.

Statistik uji: Jika pada  $\alpha = 0,05$ , lebih dari 50% nilai dari  $d_i^2 < \chi^2_{(0,05;p)}$  maka tolak  $H_0$  dan sebaliknya.

Hasil untuk nilai jarak Mahalanobis yaitu: untuk  $n = 1$ , nilai  $d_i^2 = 6,62576 < \chi^2_{(0,05;7)} = 14,067$ ; untuk  $n = 2$ , nilai  $d_i^2 = 18,68419 > \chi^2_{(0,05;7)} = 14,067$ ; untuk  $n = 3$ , nilai  $d_i^2 = 3,96284 < \chi^2_{(0,05;7)} = 14,067$  dan seterusnya sampai  $n = 133$ . Oleh

karena lebih dari 50% nilai dari  $d_i^2 < \chi^2_{(0,05;7)}$  yaitu sebesar (88,72%) maka terima  $H_0$ , yang artinya bahwa data kecamatan di Kabupaten Tabanan dalam angka tahun 2015 memenuhi asumsi distribusi normal ganda.

##### Uji Vektor Nilai Rata-Rata

Pada output SPSS, untuk peubah kepadatan penduduk, jumlah pasar, jumlah TK, jumlah SMP, jumlah SMU, dan jumlah pertokoan diperoleh masing-masing nilai  $p\text{-value} = 0,000 < \alpha = 0,05$  maka tolak  $H_0$ . Jadi dapat disimpulkan bahwa keenam peubah tersebut memberikan perbedaan rata-rata pada pengelompokan desa perdesaan dan desa perkotaan di Kabupaten Tabanan. Sedangkan untuk peubah jumlah hotel, diperoleh nilai  $p\text{-value} = 0,128 > \alpha = 0,05$  maka terima  $H_0$ . Hal ini mengartikan bahwa peubah jumlah hotel tidak memberikan perbedaan rata-rata pada pengelompokan desa di Kabupaten Tabanan.

##### Uji Kehomogenan Matriks Ragam Peragam

Pada output hasil Uji M Box, diperoleh nilai  $p\text{-value} = 0,000 < \alpha = 0,05$ . Dapat disimpulkan bahwa matriks ragam-peragam tidak bersifat homogen. Sehingga analisis diskriminan yang digunakan untuk langkah selanjutnya adalah analisis diskriminan kuadratik.

##### Uji Pencilan (Outlier)

Terdapat tiga desa yang terdeteksi sebagai pencilan yaitu Lalang Linggah, Jatiluwih, dan Pujungan untuk desa yang berstatus perdesaan. Sedangkan untuk desa yang berstatus perkotaan, terdapat 12 desa yang terdeteksi sebagai pencilan yaitu Bajera, Dauh Peken, Delod Peken, Dajan Peken, Beraban, Pejaten, Kediri, Abian Tuwung, Banjar Anyar, Kukuh, Baturiti, dan Candikuning.

##### Analisis Diskriminan Bertatar (Stepwise Discriminant Analysis)

Pada output hasil uji *stepwise* menggunakan SPSS, peubah prediktor yang akan digunakan dalam pembentukan fungsi diskriminan adalah

peubah kepadatan penduduk ( $X_1$ ) dan jumlah pasar ( $X_2$ ). Kedua peubah prediktor ini merupakan peubah yang mendominasi dalam pembentukan fungsi diskriminan karena mempunyai nilai  $F$  terbesar, nilai  $Wilk\ Lambda$  terkecil, nilai minimum jarak Mahalanobis yang signifikan, dan memiliki nilai signifikansi lebih kecil dari taraf signifikansi 5% yaitu  $p\text{-value} = 0,000 < \alpha = 0,05$ .

### Metode 2-Fold Cross Validation

Pada eksperimen pertama, data  $K_1$  sebagai data *testing* dan sisanya yaitu data  $K_2$  sebagai data *training*. Selanjutnya pada eksperimen kedua dilakukan penukaran fungsi yaitu data  $K_2$  sebagai data *testing* dan data  $K_1$  sebagai data *training*.

1. Eksperimen 1 untuk analisis diskriminan kuadratik dengan metode *2-fold cross validation*

a. Berdasarkan persamaan (2), fungsi diskriminan kuadratik untuk desa dengan status perdesaan dilambangkan dengan  $\hat{d}_1^Q$ , sedangkan untuk status perkotaan dilambangkan dengan  $\hat{d}_2^Q$ . Sehingga fungsi diskriminan kuadratik yang terbentuk pada eksperimen 1 untuk data *training* ( $K_2$ ) yaitu:

$$\hat{d}_1^Q(x) = -\frac{1}{2}\ln|S_1| - \frac{1}{2}(x - \bar{x}_1)'S_1^{-1}(x - \bar{x}_1) + \ln\left(\frac{53}{67}\right) \quad (13)$$

$$\hat{d}_2^Q(x) = -\frac{1}{2}\ln|S_2| - \frac{1}{2}(x - \bar{x}_2)'S_2^{-1}(x - \bar{x}_2) + \ln\left(\frac{14}{67}\right). \quad (14)$$

Pada eksperimen 1 terdapat tujuh desa pada kelompok perdesaan yang salah klasifikasi yaitu pada nomor 10, 30, 32, 35, 38, 47 dan 49. Hal ini berarti bahwa tujuh desa yang diprediksikan masuk ke dalam kelompok perkotaan, pada data asli seharusnya masuk ke dalam kelompok perdesaan. Sedangkan terdapat lima desa pada kelompok perkotaan yang salah klasifikasi yaitu pada nomor 56, 61, 63, 64 dan 67, yang artinya bahwa lima desa yang diprediksikan masuk ke dalam kelompok perdesaan, pada data asli seharusnya masuk ke dalam kelompok perkotaan. Jadi jumlah kesalahan klasifikasi

pada eksperimen 1 untuk data *training* ( $K_2$ ) adalah 12 desa.

b. Aplikasi pada data *testing* ( $K_1$ )

Pada eksperimen 1 untuk data *testing* ( $K_1$ ) terdapat lima desa pada kelompok perdesaan yang salah klasifikasi yaitu pada nomor 4, 18, 27, 45, dan 46. Hal ini mengartikan bahwa lima desa yang diprediksikan masuk ke dalam kelompok perkotaan, pada data asli seharusnya masuk ke dalam kelompok perdesaan. Sedangkan terdapat satu desa pada kelompok perkotaan yang salah klasifikasi yaitu pada nomor 66 yang artinya bahwa satu desa yang diprediksikan masuk ke dalam kelompok perdesaan, pada data asli seharusnya masuk ke dalam kelompok perkotaan. Jadi jumlah kesalahan klasifikasi pada eksperimen 1 untuk data *testing* ( $K_1$ ) adalah enam desa.

2. Eksperimen 2 untuk analisis diskriminan kuadratik dengan metode *2-fold cross validation*

a. Fungsi diskriminan kuadratik yang terbentuk pada eksperimen 2 untuk data ( $K_1$ ) yaitu:

$$\hat{d}_1^Q(x) = -\frac{1}{2}\ln|S_1| - \frac{1}{2}(x - \bar{x}_1)'S_1^{-1}(x - \bar{x}_1) + \ln\left(\frac{52}{66}\right) \quad (15)$$

$$\hat{d}_2^Q(x) = -\frac{1}{2}\ln|S_2| - \frac{1}{2}(x - \bar{x}_2)'S_2^{-1}(x - \bar{x}_2) + \ln\left(\frac{14}{66}\right). \quad (16)$$

Pada eksperimen 2 terdapat lima desa pada kelompok perdesaan yang salah klasifikasi yaitu pada nomor 4, 18, 27, 45, dan 46. Sedangkan terdapat satu desa pada kelompok 2 yang salah klasifikasi yaitu pada nomor 66. Jadi jumlah kesalahan klasifikasi pada eksperimen 2 untuk data *training* ( $K_1$ ) adalah enam desa.

b. Aplikasi pada data *testing* ( $K_2$ )

Untuk memeriksa apakah model yang diperoleh pada persamaan (15) dan (16) dapat diterapkan pada data baru, maka dilakukan validasi terhadap data *testing*. Terdapat tujuh desa pada kelompok perdesaan yang salah klasifikasi yaitu pada nomor 10, 30, 32, 35, 38, 47 dan 49. Sedangkan terdapat lima desa pada kelompok 2 yang salah klasifikasi yaitu pada nomor 56, 61, 63, 64 dan 67. Jadi jumlah

kesalahan klasifikasi pada eksperimen 2 untuk data *testing* ( $K_2$ ) adalah 12 desa.

### Metode 3-Fold Cross Validation

Pada eksperimen pertama, data  $K_1$  sebagai data *testing*, dan sisanya yaitu data  $K_2$  dan  $K_3$  sebagai data *training*. Pada eksperimen kedua, data  $K_2$  menjadi data *testing* dan sisanya yaitu data  $K_1$  dan  $K_3$  sebagai data *training*. Selanjutnya pada eksperimen ketiga, data  $K_3$  sebagai data *testing* dan sisanya yaitu data  $K_1$  dan  $K_2$  sebagai data *training*.

1. Eksperimen 1 untuk analisis diskriminan kuadratik dengan metode *3-fold cross validation*  
a. Fungsi diskriminan kuadratik yang terbentuk pada eksperimen 1 untuk data *training* ( $K_2$  dan  $K_3$ ) yaitu:

$$\hat{d}_1^Q(x) = -\frac{1}{2}\ln|S_1| - \frac{1}{2}(x - \bar{x}_1)'S_1^{-1}(x - \bar{x}_1) + \ln\left(\frac{70}{89}\right) \quad (17)$$

$$\hat{d}_2^Q(x) = -\frac{1}{2}\ln|S_2| - \frac{1}{2}(x - \bar{x}_2)'S_2^{-1}(x - \bar{x}_2) + \ln\left(\frac{19}{89}\right). \quad (18)$$

Pada eksperimen 1 terdapat lima desa pada kelompok perdesaan yang salah klasifikasi yaitu pada nomor 9, 16, 48, 51, dan 60. Sedangkan terdapat empat desa pada kelompok perkotaan yang salah klasifikasi yaitu pada nomor 41, 82, 87, dan 88. Jadi jumlah kesalahan klasifikasi pada eksperimen 1 untuk data *training* ( $K_2$  dan  $K_3$ ) adalah sembilan desa.

b. Aplikasi pada data *testing* ( $K_1$ )

Pada eksperimen 1 untuk data *testing* ( $K_1$ ) terdapat empat desa pada kelompok perdesaan yang salah klasifikasi yaitu pada nomor 5, 21, 27, dan 35. Sedangkan terdapat tiga desa pada kelompok perkotaan yang salah klasifikasi yaitu pada nomor 36, 37 dan 39. Jadi jumlah kesalahan klasifikasi pada eksperimen 1 untuk data *testing* ( $K_1$ ) adalah tujuh desa.

Langkah yang sama dilakukan pada eksperimen 2 dan 3 untuk analisis diskriminan kuadratik dengan metode *3-fold cross validation*.

### Metode 4-Fold Cross Validation

Pada eksperimen pertama, data  $K_1$  sebagai data *testing*, dan sisanya yaitu data  $K_2$ ,  $K_3$ , dan  $K_4$  sebagai data *training*. Pada eksperimen kedua, data  $K_2$  menjadi data *testing* dan sisanya yaitu data  $K_1$ ,  $K_3$ , dan  $K_4$  sebagai data *training*. Pada eksperimen ketiga, data  $K_3$  sebagai data *testing* dan sisanya yaitu data  $K_1$ ,  $K_2$ , dan  $K_4$  sebagai data *training*. Selanjutnya pada eksperimen keempat, data  $K_4$  sebagai data *testing* dan sisanya yaitu data  $K_1$ ,  $K_2$ , dan  $K_3$  sebagai data *training*.

1. Eksperimen 1 untuk analisis diskriminan kuadratik dengan metode *4-fold cross validation*

a. Fungsi diskriminan kuadratik yang terbentuk pada eksperimen 1 untuk data *training* ( $K_2$ ,  $K_3$  dan  $K_4$ ) yaitu:

$$\hat{d}_1^Q(x) = -\frac{1}{2}\ln|S_1| - \frac{1}{2}(x - \bar{x}_1)'S_1^{-1}(x - \bar{x}_1) + \ln\left(\frac{79}{100}\right) \quad (19)$$

$$\hat{d}_2^Q(x) = -\frac{1}{2}\ln|S_2| - \frac{1}{2}(x - \bar{x}_2)'S_2^{-1}(x - \bar{x}_2) + \ln\left(\frac{21}{100}\right). \quad (20)$$

Pada eksperimen 1 terdapat 10 desa pada kelompok perdesaan yang salah klasifikasi yaitu pada nomor 9, 11, 26, 35, 38, 55, 56, 71, 80, dan 84. Sedangkan terdapat 4 desa pada kelompok perkotaan yang salah klasifikasi yaitu pada nomor 27, 28, 33, dan 65. Jadi jumlah kesalahan klasifikasi pada eksperimen 1 untuk data *training* ( $K_2$ ,  $K_3$ , dan  $K_4$ ) adalah 14 desa.

b. Aplikasi pada data *testing* ( $K_1$ )

Pada eksperimen 1 untuk data *testing* ( $K_1$ ) terdapat satu desa pada kelompok perdesaan yang salah klasifikasi yaitu pada nomor 26. Sedangkan terdapat dua desa pada kelompok perkotaan yang salah klasifikasi yaitu pada nomor 31 dan 32. Jadi jumlah kesalahan klasifikasi pada eksperimen 1 untuk data *testing* ( $K_1$ ) adalah tiga desa.

Langkah yang sama dilakukan pada eksperimen 2, 3, dan 4 untuk analisis diskriminan kuadratik dengan metode *4-fold cross validation*.



**Peluang Kesalahan Klasifikasi (APER)**

1. Metode *2-fold cross validation*  
Untuk eksperimen 1

Tabel . Hasil klasifikasi data training (K<sub>2</sub>)

Kelompok Aktual	Kelompok Prediksi		Jumlah Observasi
	1	2	
1	46	5	51
2	7	9	16

$$APER = \frac{5+7}{51+16} = 0,18.$$

Tabel 3. Hasil klasifikasi data testing (K1)

Kelompok Aktual	Kelompok Prediksi		Jumlah Observasi
	1	2	
1	47	1	48
2	5	13	18

$$APER = \frac{1+5}{48+18} = 0,09.$$

Pada tabel 2 dan 3 total peluang kesalahan klasifikasi pada eksperimen 1 =  $\frac{0,18+0,09}{2} = 0,135$ . Perhitungan yang sama dilakukan pada masing-masing eksperimen untuk  $k = 2,3,4$ .

**Uji Keakuratan**

Uji keakuratan hasil klasifikasi untuk masing-masing eksperimen pada metode *k-fold cross validation* dengan  $k = 2,3$ , dan 4 yaitu:

1. Metode *2-fold cross validation*

Uji keakuratan untuk eksperimen 1 data *training* (K<sub>2</sub>):

$$Hit_{Ratio} = \frac{46 + 9}{67} \times 100\% = 82\%$$

$$C_{pro} = ((0,76)^2 + (0,24)^2) \times 100\% = 64\%$$

$$C_{max} = \frac{51}{67} \times 100\% = 76\%.$$

Diperoleh nilai  $Hit_{Ratio} > C_{max} > C_{pro}$  maka  $H_0$  diterima, yang artinya bahwa hasil klasifikasi untuk eksperimen 1 data *training* (K<sub>2</sub>) adalah akurat.

Perhitungan yang sama dilakukan pada masing-masing eksperimen sehingga diperoleh hasil klasifikasi untuk  $k = 2,3$ , dan 4 pada masing-masing eksperimen adalah akurat.

**Uji Kestabilan**

Nilai uji kestabilan hasil klasifikasi untuk masing-masing eksperimen pada metode *k-fold cross validation* dengan  $k = 2,3$ , dan 4 dapat dilihat pada tabel 4 berikut:

Tabel 4. Uji Kestabilan Hasil Klasifikasi

k	Eksperimen	Q Press	Keterangan
2	1	70,74	Konsisten
	2	70,74	Konsisten
3	1	76,69	Konsisten
	2	70,74	Konsisten
	3	70,74	Konsisten
4	1	73,69	Konsisten
	2	67,86	Konsisten
	3	67,86	Konsisten
	4	76,69	Konsisten

**5. KESIMPULAN DAN SARAN**

Berdasarkan hasil dan pembahasan yang diperoleh maka dapat diambil kesimpulan yaitu: Pertama, model pengklasifikasian desa di Kabupaten Tabanan ke dalam kelompok daerah perdesaan atau perkotaan dengan teknik analisis diskriminan kuadratik diperoleh pada saat  $k = 4$  untuk eksperimen 4 adalah

$$\hat{d}_1^Q(x) = -\frac{1}{2} \ln|S_1| - \frac{1}{2} (x - \bar{x}_1)' S_1^{-1} (x - \bar{x}_1) + \ln\left(\frac{78}{99}\right) \text{ dan}$$

$$\hat{d}_2^Q(x) = -\frac{1}{2} \ln|S_2| - \frac{1}{2} (x - \bar{x}_2)' S_2^{-1} (x - \bar{x}_2) + \ln\left(\frac{21}{99}\right).$$

Kedua, model diskriminan kuadratik  $\hat{d}_1^Q(x)$  dan  $\hat{d}_2^Q(x)$  tersebut mempunyai total peluang kesalahan klasifikasi yang terkecil diantara masing-masing eksperimen untuk  $k = 2,3$ , dan 4 yaitu sebesar 0,09 dan mempunyai hasil klasifikasi yang akurat dan konsisten. Ketiga, hasil pengklasifikasian desa di Kabupaten Tabanan dengan teknik analisis diskriminan

kuadratik menggunakan *k-fold cross validation* untuk  $k = 2, 3$ , dan  $4$  pada masing-masing eksperimen adalah akurat dan konsisten.

Adapun saran yang dapat diberikan pada penelitian selanjutnya yaitu untuk menggunakan teknik analisis diskriminan kuadratik dengan jumlah kelompok yang lebih besar dengan metode *k-fold cross validation* untuk  $k > 4$ . Kedua, disarankan untuk mengatasi pencilan (*outlier*) pada data dan menggunakan data dengan jumlah individu pada setiap kelompok yang jumlahnya seimbang. Ketiga, berdasarkan hasil uji pencilan (*outlier*) terutama untuk desa yang berstatus perkotaan, terdapat 12 desa yang terdeteksi sebagai pencilan dari 28 desa yang berstatus perkotaan. Hal ini memungkinkan terjadinya pengelompokan baru pada pencilan tersebut (desa transisi). Sehingga untuk penelitian selanjutnya disarankan untuk melakukan pemekaran wilayah desa atau pengelompokan lebih dari dua kelompok untuk status daerah desa tersebut.

## DAFTAR PUSTAKA

- Arisona, D. C. (2015). Analisis Diskriminan Linier Pada Klasifikasi Nasabah Menunggak Dan Tidak Menunggak Dengan Metode *Cross Validation*. *Skripsi*. FMIPA, Jurusan Matematika, Program Studi Statistika, Universitas Hasanuddin Makassar.
- Badan Pusat Statistik. (2010). *Klasifikasi Perkotaan Dan Perdesaan Di Indonesia*. Jakarta: Badan Pusat Statistik.
- Davidson, A., & Hinkley, D. (1997). *Bootstrap Methods and their Application*. New York: Cambridge University Press.
- Hair JR., J., Black, W., Babin, B., & Anderson, R. (2009). *Multivariate Data Analysis* (Seventh ed.). USA: Prentice-Hall.
- Johnson, R. A., & Wichern, D. (2007). *Applied Multivariate Statistical Analysis* (Sixth ed.). New Jersey: Pearson Education, Inc.
- Mattjik, A. A., & Sumertajaya, I. (2011). *Sidik Peubah Ganda Dengan Menggunakan SAS*. Bogor: IPB PRESS.
- Tarigan, A. (2003). *Rural-Urban Economic Linkages*. Tersedia pada: <http://www.bappenas.go.id/>. [diakses pada 1 Agustus 2016].