

PENERAPAN ANALISIS DISKRIMINAN KUADRATIK PADA PENDUGAAN KLASIFIKASI KELURAHAN DI KOTA DUMAI

SUCHI RACHMADANY, YUDIANTRI ASDI*, IZZATI RAHMI HG

*Program Studi S1 Matematika,
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas,
Kampus UNAND Limau Manis Padang, Indonesia.
email : rachmadanysuchi@gmail.com, yudiantriasdi@sci.unand.ac.id,
izzatirahmihg@sci.unand.ac.id*

Diterima 17 Februari 2020 Direvisi 7 Maret 2020 Dipublikasikan 29 April 2020

Abstrak. Perencanaan pembangunan di Indonesia memerlukan klasifikasi daerah perdesaan dan daerah perkotaan. Klasifikasi tersebut telah diatur berdasarkan Peraturan Kepala Badan Pusat Statistik Nomor 37 Tahun 2010 dengan 10 (sepuluh indikator). Namun, terkadang sulit untuk memperoleh data dengan keseluruhan indikator, sehingga diperlukan cara untuk menduga klasifikasi desa/kelurahan termasuk daerah perdesaan atau perkotaan dengan indikator yang tersedia dan menghasilkan pendugaan klasifikasi yang mendekati klasifikasi yang dihasilkan oleh indikator yang lengkap. Oleh karena itu, digunakan analisis diskriminan kuadrat. Penelitian ini bertujuan untuk menduga fungsi diskriminan kuadrat pada klasifikasi kelurahan di Kota Dumai, serta menentukan tingkat keakuratan dan kekonsistenan klasifikasi. Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik Kota Dumai yaitu Kota Dumai dalam Angka Tahun 2018. Metode yang digunakan pada penelitian ini adalah analisis diskriminan kuadrat dengan 6 (enam) peubah prediktor (X) dan 2 (dua) peubah respon (Y).

Pada penelitian ini diperoleh fungsi diskriminan kuadrat pada pendugaan klasifikasi kelurahan di Kota Dumai yaitu $\hat{d}_1^Q(x)$ adalah fungsi diskriminan kuadrat untuk daerah perdesaan dan $\hat{d}_2^Q(x)$ adalah fungsi diskriminan kuadrat untuk daerah perkotaan. Fungsi diskriminan kuadrat $\hat{d}_1^Q(x)$ dan $\hat{d}_2^Q(x)$ tersebut mempunyai peluang kesalahan klasifikasi sebesar 0,1818, mempunyai hasil klasifikasi yang akurat, dan mempunyai nilai $QPress > \chi_{(1,\alpha)}^2$ dengan $\alpha = 0,05$ yang menunjukkan klasifikasi konsisten.

Dengan menggunakan fungsi pendugaan klasifikasi ini daerah dapat menduga apakah termasuk dalam klasifikasi daerah perdesaan atau perkotaan dengan indikator yang tersedia dari daerah tersebut.

Kata Kunci: Analisis diskriminan, Fungsi diskriminan kuadrat, Klasifikasi

1. Pendahuluan

Wilayah Indonesia dibagi ke dalam beberapa tingkat wilayah administratif, yaitu propinsi, kabupaten/kota, kecamatan, dan desa/kelurahan sebagai wilayah ad-

*penulis korespondensi

ministrasi terendah dalam hierarki pembagian wilayah administrasi di bawah kecamatan. Wilayah kabupaten/kota mempunyai bagian-bagian yang merupakan kawasan perkotaan dan kawasan perdesaan. Pada pelaksanaannya, penentuan apakah suatu desa/kelurahan termasuk daerah perkotaan atau perdesaan dilakukan Badan Pusat Statistik berdasarkan [1]. Namun terkadang sulit untuk memperoleh data dengan keseluruhan indikator, sehingga diperlukan cara untuk menduga klasifikasi desa/kelurahan termasuk daerah perkotaan atau perdesaan dengan indikator yang tersedia dan menghasilkan pendugaan klasifikasi yang mendekati klasifikasi yang dihasilkan oleh indikator yang lengkap.

Data klasifikasi kelurahan perkotaan dan kelurahan perdesaan dapat digunakan sebagai perencanaan pembangunan dalam hal pemekaran wilayah. Dalam [6] perencanaan pembangunan wilayah 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 diketahui oleh pemerintah dalam hal perencanaan pembangunan di daerah desa. Berdasarkan kriteria desa perkotaan dan desa perdesaan yang tercantum dalam peraturan tersebut penulis bermaksud menentukan suatu fungsi pembeda secara statistik dalam menduga klasifikasi suatu daerah di Kota Dumai apakah termasuk ke dalam kelompok daerah perkotaan atau perdesaan dengan teknik analisis diskriminan kuadratik.

Dalam [2] analisis diskriminan adalah salah satu kajian dalam statistika yang membahas tentang pengelompokan suatu individu baru ke dalam kelompok yang sudah ada berdasarkan karakteristik data. Fungsi yang terbentuk dalam analisis diskriminan disebut fungsi diskriminan. Fungsi diskriminan merupakan sebuah varian dari variabel bebas yang terpilih untuk kekuatan diskriminatif yang digunakan dalam memprediksi keanggotaan grup.

Penerapan analisis diskriminan pada penelitian ini yaitu analisis diskriminan kuadratik pada klasifikasi wilayah kelurahan di Kota Dumai. Secara administratif Kota Dumai dibagi menjadi 7 kecamatan dan 33 kelurahan pada tahun 2017.

Berdasarkan uraian latar belakang di atas, rumusan masalah pada penelitian ini adalah bagaimana teknik pengklasifikasian kelurahan di Kota Dumai menggunakan metode analisis diskriminan kuadratik.

Tujuan penelitian ini adalah menduga fungsi diskriminan kuadratik pada klasifikasi kelurahan di Kota Dumai, serta menentukan tingkat keakuratan dan kekonsistenan klasifikasi.

2. Landasan Teori

2.1. Analisis Diskriminan

Dalam [2] analisis diskriminan berkaitan dengan memperoleh sebuah variasi. Variasi diskriminan adalah kombinasi linier dari dua (atau lebih) variabel bebas yang akan membedakan antar objek ke dalam grup yang telah didefinisikan dengan bentuk umum sebagai berikut:

$$Z_{jk} = a + W_1X_{1k} + W_2X_{2k} + \dots + W_pX_{pk} \tag{2.1}$$

dimana

- Z_{jk} : skor Z diskriminan dari fungsi diskriminan j untuk kelompok ke- k ,
- $j : 1, 2, \dots, s, s \leq \min n - 1, p$;
- a : titik potong pada sumbu Z ,
- W_p : bobot diskriminan untuk peubah bebas ke- p ,
- X_{pk} : peubah bebas p untuk kelompok ke- k ,
- p : banyak peubah bebas,
- k : banyak kelompok.

Pada analisis diskriminan terdapat uji asumsi dasar yaitu uji distribusi normal ganda dan uji kehomogenan ragam.

2.2. Uji Distribusi Normal Ganda

Dalam [3] uji distribusi normal ganda digunakan untuk mengetahui apakah data telah berdistribusi normal ganda atau tidak. Uji ini dilakukan dengan membandingkan nilai jarak Mahalanobis dan nilai khi-kuadrat.

Hipotesis uji:

H_0 : data berdistribusi normal ganda;

H_1 : data tidak berdistribusi normal ganda.

Nilai jarak Mahalanobis dihitung dengan persamaan sebagai berikut:

$$d_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) \quad (2.2)$$

dimana

\mathbf{x}_i : vektor pengamatan ke- i ,

$\bar{\mathbf{x}}$: vektor rata-rata,

\mathbf{S}^{-1} : inverse matriks varian-kovarian.

Jika pada $\alpha = 0,5$, lebih dari 50% nilai dari $d_i < \chi_{(db, \alpha)}^2$ maka terima H_0 , yang artinya bahwa data berdistribusi normal ganda dan sebaliknya.

2.3. Uji Kehomogenan Ragam (Uji M Box)

Dalam [5] uji M Box digunakan untuk menguji kesamaan matriks varian-kovarian antar kelompok.

Hipotesis uji:

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

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

Statistik uji M Box yaitu:

$$-2 \ln \lambda^* = (n - k) \ln \left| \frac{\mathbf{W}}{(n - k)} \right| - \sum_{i=1}^k (n_i - 1) \ln |\mathbf{S}_i|, \quad (2.3)$$

dengan

$$\lambda^* = \frac{\prod_{i=1}^k |\mathbf{S}_i|^{(n_i-1)/2}}{\left| \frac{\mathbf{W}}{(n-k)} \right|^{(n-k)/2}}$$

dimana

- \mathbf{S}_i : matriks varian-kovarian kelompok ke- i ,
- \mathbf{W} : matriks jumlah kuadrat dan hasil kali data dalam kelompok
- k : banyak kelompok,
- p : banyak peubah prediktor,
- n : banyak pengamatan,
- c : vektor pembobot.

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 varian-kovarian yang homogen.

2.4. Fungsi Diskriminan Kuadratik

Dalam [3] analisis diskriminan kuadratik digunakan ketika data berdistribusi normal ganda dengan ragam tidak homogen. Fungsi diskriminan kuadratik yaitu:

$$\hat{d}_k^Q = -\frac{1}{2} \ln |\mathbf{S}_k| - \frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}}_k)' \mathbf{S}_k^{-1} (\mathbf{x} - \bar{\mathbf{x}}_k) + \ln pr_k; \quad k = 1, 2, \dots, g, \quad (2.4)$$

dimana

- \mathbf{S}_k : matriks varian-kovarian kelompok ke- k ,
- \mathbf{S}_k^{-1} : *inverse* matriks varian-kovarian kelompok ke- k ,
- $\bar{\mathbf{x}}_k$: vektor rata-rata kelompok ke- k ,
- pr_k : *prior probability* kelompok ke- k .

2.5. Apparent Error Rate (APER)

Dalam [3] *Apparent Error Rate* (APER) adalah nilai yang menunjukkan perbandingan keanggotaan grup aktual dengan keanggotaan grup prediksi yang ditunjukkan pada tabel kesalahan klasifikasi sebagai berikut:

Tabel 1. Kesalahan Klasifikasi

Hasil Pengamatan	Hasil Prediksi	
	Kelompok 1	Kelompok 2
Kelompok 1	n_{11}	n_{12}
Kelompok 2	n_{21}	n_{22}

dan dihitung menggunakan persamaan berikut:

$$APER = \frac{n_{12} + n_{21}}{n_1 + n_2}. \quad (2.5)$$

2.6. Uji Keakuratan

Uji keakuratan dilakukan dengan membandingkan nilai Hit_{Rasio} , C_{pro} , dan C_{max} . Dalam [2] ketepatan fungsi diskriminan dalam memprediksi dapat diukur dengan Hit_{Rasio} sebagai berikut:

$$Hit_{Rasio} = \left(\frac{n_{11} + n_{22}}{n_1 + n_2} \right) \times 100\%. \quad (2.6)$$

Analisis kriteria kemungkinan proporsional (*proportional chance criterion*) dapat digunakan untuk mengidentifikasi keanggotaan grup/kelompok yang didefinisikan dengan tepat dari semua grup/kelompok, tidak hanya grup/kelompok yang terbesar, ketika ukuran grup/kelompok berbeda, yang dirumuskan sebagai berikut:

$$C_{pro} = p^2 + (1 - p)^2. \quad (2.7)$$

C_{max} adalah nilai ekspektasi klasifikasi yang benar untuk kelompok yang dipilih dan dirumuskan sebagai berikut:

$$C_{max} = \left(\frac{n_{max}}{N} \right) \times 100\%. \quad (2.8)$$

Hipotesis uji keakuratan hasil klasifikasi yaitu:

H_0 : Klasifikasi akurat;

H_1 : Klasifikasi tidak akurat.

Kriteria uji:

Dalam [4] jika $Hit_{Ratio} > C_{max} > C_{pro}$ maka terima H_0 , yang artinya bahwa hasil klasifikasi sudah akurat.

2.7. Uji Kestabilan

Dalam [2] *Q Press* adalah uji statistik untuk menguji kekuatan diskriminatif dari klasifikasi bila dibandingkan dengan model kemungkinan. Nilai yang dihitung kemudian dibandingkan dengan nilai kritis (nilai khi-kuadrat untuk derajat bebas 1 pada tingkat kepercayaan yang diinginkan). Nilai *Q Press* dirumuskan sebagai berikut:

$$Q\ Press = \frac{[N - (nk)]^2}{N(k - 1)} \quad (2.9)$$

dimana

N : ukuran sampel total,

n : banyak pengamatan yang diklasifikasikan dengan tepat

k : banyak kelompok.

Hipotesis uji kestabilan pengklasifikasian yaitu:

H_0 : Keakuratan pengklasifikasian konsisten;

H_1 : Keakuratan pengklasifikasian tidak konsisten.

Kriteria uji:

Jika nilai $Q\ Press > \chi^2_{(1,\alpha)}$ maka dapat disimpulkan bahwa keakuratan pengklasifikasian adalah konsisten, dan sebaliknya.

3. Pembahasan

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik Kota Dumai yaitu Kota Dumai dalam Angka 2018. Peubah prediktor pada penelitian ini adalah kepadatan penduduk per km^2 (X_1), banyak Sekolah Taman Kanak-Kanak (TK) (X_2), banyak Sekolah Menengah Pertama/ sederajat (SMP/ sederajat) (X_3), banyak Sekolah Menengah Umum/ sederajat (SMU/ sederajat) (X_4), banyak pasar (X_5), dan banyak hotel yang tersedia (X_6). Status daerah kelurahan di Kota Dumai yang digunakan sebagai peubah respon (Y) dibedakan menjadi dua kelompok yaitu daerah perkotaan dan daerah pedesaan. Berdasarkan data Peraturan Kepala Badan Pusat Statistik Nomor 37 Tahun 2010, sejumlah 14 kelurahan (42,4%) yang berada di Kota Dumai merupakan daerah pedesaan, sedangkan sisanya yaitu 19 kelurahan (57,6%) merupakan daerah perkotaan.

3.1. Uji Asumsi Dasar Analisis Diskriminan

Uji asumsi dasar analisis diskriminan terdiri dari uji distribusi normal ganda dan uji kehomogenan ragam. Uji distribusi normal ganda dilakukan menggunakan perbandingan nilai jarak Mahalanobis dan nilai khi kuadrat, diperoleh hasil lebih dari 50% nilai dari $d_i^2 < \chi_{(6;0,5)}^2$ yaitu sebesar 60,61% maka terima H_0 , yang artinya bahwa data berdistribusi normal ganda.

Selanjutnya akan dilakukan uji kehomogenan ragam. Uji ini dilakukan menggunakan aplikasi SPSS. Dari output hasil uji M Box, diperoleh nilai $p\text{-value} = 0.000 < \alpha = 0.05$, maka tolak H_0 dan dapat disimpulkan bahwa matriks varian-kovarian tidak homogen. Sehingga untuk langkah selanjutnya fungsi diskriminan yang digunakan adalah fungsi diskriminan kuadratik yang diperoleh menggunakan analisis diskriminan bertatar.

3.2. Analisis Diskriminan Bertatar (*Stepwise Discriminant Analysis*)

Pada output hasil analisis diskriminan bertatar (*stepwise discriminant analysis*) menggunakan SPSS yang ditunjukkan oleh Tabel 2, peubah prediktor yang akan digunakan dalam pembentukan fungsi diskriminan kuadratik adalah peubah kepadatan penduduk (X_1) dan banyak hotel (X_6).

Tabel 2. Hasil Analisis Diskriminan Bertatar

Langkah	Variabel	Wilk's Lambda	Nilai F	$p\text{-value}$
1	Kepadatan Penduduk	0,661	15,923	0,000
2	Hotel	0,557	11,933	0,000

3.3. Analisis Diskriminan Kuadratik

Fungsi diskriminan kuadratik diperoleh menggunakan analisis diskriminan bertatar (*stepwise discriminant analysis*) pada SPSS. Fungsi diskriminan kuadratik diperoleh

sebagai berikut:

$$\hat{d}_1^Q(\mathbf{x}) = -\frac{1}{2} \ln |\mathbf{S}_1| - \frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}}_1)' \mathbf{S}_1^{-1} (\mathbf{x} - \bar{\mathbf{x}}_1) + \ln \left(\frac{14}{33} \right) \quad (3.1)$$

$$\hat{d}_2^Q(\mathbf{x}) = -\frac{1}{2} \ln |\mathbf{S}_2| - \frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}}_2)' \mathbf{S}_2^{-1} (\mathbf{x} - \bar{\mathbf{x}}_2) + \ln \left(\frac{19}{33} \right) \quad (3.2)$$

Berdasarkan fungsi analisis diskriminan kuadratik pada (3.1) dan (3.2), diperoleh hasil terdapat enam kelurahan yang salah klasifikasi, yaitu kelurahan Pangkalan Se-sai, kelurahan Bagan Keladi, kelurahan Mekar Sari, kelurahan Bukit Timah, kelurahan Bukit Datuk, dan kelurahan Tanjung Palas.

3.4. Peluang Kesalahan Klasifikasi (APER)

Peluang kesalahan klasifikasi dihitung menggunakan persamaan (2.5) dan diperoleh hasil sebagai berikut sebagai berikut:

Tabel 3. Kesalahan Klasifikasi

Kelompok Aktual	Kelompok Prediksi		Jumlah Pengamatan
	1	2	
1	14	0	14
2	6	13	19

$$APER = \frac{(0 + 6)}{(14 + 19)} = 0,1818$$

Hal ini menunjukkan bahwa terdapat peluang kesalahan klasifikasi sebesar 0,1818.

3.5. Uji Keakuratan

Keakuratan hasil klasifikasi menggunakan fungsi diskriminan kuadratik diperoleh sebagai berikut:

$$\begin{aligned} Hit_{Ratio} &= \frac{(14 + 13)}{33} \times 100\% = 81,82\% \\ C_{pro} &= \frac{((0,42)^2 + (0,58)^2)}{33} \times 100\% = 51,28\% \\ C_{max} &= \frac{19}{33} \times 100\% = 57,58\% \end{aligned}$$

Diperoleh nilai $Hit_{ratio} > C_{max} > C_{pro}$, maka tidak tolak H_0 . Hal ini berarti bahwa hasil klasifikasi menggunakan fungsi diskriminan kuadratik akurat.

3.6. Uji Kestabilan

Nilai $Q Press$ dihitung menggunakan persamaan (2.9) dan diperoleh hasil sebagai berikut:

$$Q Press = \frac{[33 - (27)(2)]^2}{33(2 - 1)} = 13,36$$

Diketahui nilai $\chi_{(1,\alpha)}^2 = 3,841$ pada $\alpha = 0,05$, sehingga diperoleh $Q Press > \chi_{(1,\alpha)}^2$. Maka dapat disimpulkan bahwa keakuratan pengklasifikasian konsisten.

4. Kesimpulan

Pada hasil dan pembahasan diperoleh data berdistribusi normal ganda dengan ragam tidak homogen, sehingga metode yang digunakan adalah analisis diskriminan kuadratik dan diperoleh fungsi diskriminan kuadratik. Fungsi diskriminan kuadratik pada pendugaan klasifikasi kelurahan di Kota Dumai dalam kelompok daerah perdesaan atau perkotaan yaitu:

$$\begin{aligned} d_1^Q(\mathbf{x}) &= -\frac{1}{2} \ln |\mathbf{S}_1| - \frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}}_1)' \mathbf{S}_1^{-1} (\mathbf{x} - \bar{\mathbf{x}}_1) + \ln \left(\frac{14}{33} \right) \\ d_2^Q(\mathbf{x}) &= -\frac{1}{2} \ln |\mathbf{S}_2| - \frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}}_2)' \mathbf{S}_2^{-1} (\mathbf{x} - \bar{\mathbf{x}}_2) + \ln \left(\frac{19}{33} \right) \end{aligned}$$

Fungsi diskriminan kuadratik $\hat{d}_1^Q(x)$ dan $\hat{d}_2^Q(x)$ tersebut mempunyai peluang kesalahan klasifikasi sebesar 0,1818, mempunyai hasil klasifikasi yang akurat, dan mempunyai nilai $Q Press > \chi_{(1,\alpha)}^2$ pada $\alpha = 0,05$ yang menunjukkan keakuratan pengklasifikasian konsisten.

5. Ucapan Terima kasih

Terima kasih kepada ibu Hazmira Yozza, ibu Maiyastri, dan ibu Susila Bahri yang telah memberikan kritik dan saran untuk penulisan artikel ini.

Daftar Pustaka

- [1] Badan Pusat Statistik. 2010. *Klasifikasi Perkotaan dan Perdesaan di Indonesia*. Badan Pusat Statistik. Jakarta.
- [2] Hair, Joseph F., William C. Black, Barry J. Babin, Rolph E. Anderson. 2009. *Multivariate Data Analysis*. 7th Edition. Prentice Hall. USA.
- [3] Johnson, Richard A., Dean D. Wichern. 2007. *Applied Multivariate Statistical Analysis*. Pearson Education, Inc. New Jersey.
- [4] Made Supartini, Ida A., I Komang Gde Sukarsa, I Gusti Ayu Made Srinadi. (2017). Analisis Diskriminan pada Klasifikasi Desa di Kabupaten Tabanan Menggunakan Metode *K-Fold Cross Validation*. *E-Jurnal Matematika* 6(2): 106 – 115.
- [5] Mattjik, Ahmad A., I Made Sumertajaya. 2011. *Sidik Peubah Ganda dengan Menggunakan SAS*. Bogor: IPB PRESS.
- [6] Tarigan, A. 2003. *Rural-Urban Economic Linkages*. Tersedia pada: <http://www.bappenas.go.id/>. [diakses pada 23 Oktober 2019].