

## PEMISAHAN DESA/KELURAHAN DI KABUPATEN SEMARANG MENURUT STATUS DAERAH MENGGUNAKAN ANALISIS DISKRIMINAN KUADRATIK KLASIK DAN DISKRIMINAN KUADRATIK *ROBUST*

Afianti Sonya Kurniasari<sup>1</sup>, Diah Safitri<sup>2\*</sup>, Sudarno<sup>3</sup>

<sup>1</sup>Mahasiswa Jurusan Statistika FSM UNDIP

<sup>2,3</sup>Staff Pengajar Jurusan Statistika FSM UNDIP

### ABSTRACT

Semarang Regency is one of 29 counties and 6 towns in Central Java province. In the district there are rural areas and urban areas. Discriminant analysis is a technique related to the separation of objects into different groups that have been set previously, thus, discriminant analysis can be used to separate village in Semarang Regency into urban or rural groups. Linear discriminant analysis assumes that the covariance matrix of the two groups are equal, If the assumption of equality covariance matrix is denied, function of quadratic discriminant can be used for classification. Classical estimation for the sample mean vector and sample covariance matrix is very sensitive to the presence of outliers in the observations and the functioning of the separation can be non-robust. Estimators that can be used to cope with data containing outliers are the Minimum Covariance Determinant. Robust discriminant analysis is obtained by replacing the mean and covariance matrix using the classic MCD estimator. After analysis is performed, obtained result the data of 2011 Village Potential contains outlier, so that the robust quadratic discriminant analysis more appropriate because it can provide precision the results of separation 89,79% while classical quadratic discriminant analysis give exactness of 87,23%.

**Keywords** : Rural, Urban, Quadratic Discriminant, Classical, Robust

## 1. PENDAHULUAN

### 1.1. Latar Belakang

Kabupaten Semarang merupakan salah satu kabupaten dari duapuluh sembilan kabupaten dan enam kota yang ada di Provinsi Jawa Tengah. Secara administratif Kabupaten Semarang terbagi menjadi 19 kecamatan, 27 kelurahan, dan 208 desa. Pemerintahan Kabupaten Semarang berpusat di daerah Ungaran yaitu berada di sekitar Kecamatan Ungaran Barat dan Ungaran Timur (Badan Pusat Statistik, 2011).

Tidak semua wilayah kabupaten merupakan kawasan perdesaan, tetapi mempunyai bagian-bagian yang merupakan kawasan perkotaan. Oleh karena itu, sifat hubungan kota-desa juga berlangsung di dalam suatu wilayah kabupaten (Sadyohutomo, 2009). Dalam pelaksanaannya penentuan apakah suatu desa/kelurahan termasuk daerah perkotaan atau perdesaan dilakukan oleh BPS Pusat menggunakan indikator yang didasarkan pada variabel kepadatan penduduk per km<sup>2</sup>, persentase rumah tangga pertanian dan akses ke fasilitas umum (Badan Pusat Statistik, 2010). Setelah mengetahui status dari suatu desa/kelurahan termasuk dalam daerah perkotaan atau perdesaan diharapkan pemerintah mampu membangun daerah desa (Daldjoeni, 1998).

Analisis diskriminan merupakan teknik multivariat yang berkaitan dengan pemisahan objek dalam kelompok yang berbeda dan mengalokasikan objek ke dalam suatu kelompok yang telah ditetapkan sebelumnya (Johnson and Wichern, 2007). Sehingga, analisis diskriminan dapat digunakan untuk memisahkan desa/kelurahan di Kabupaten Semarang ke dalam kelompok perkotaan atau perdesaan. Seperti yang diketahui bahwa estimasi klasik untuk vektor rata-rata sampel dan matriks kovarian

sampel sangat sensitif terhadap adanya pencilan pada pengamatan sehingga menyebabkan fungsi klasifikasi menjadi tidak *robust*. *Estimator* yang dapat digunakan untuk mengatasi data yang mengandung pencilan adalah *Minimum Covariance Determinant* karena estimator ini dapat dihitung untuk sekumpulan data yang besar dalam waktu yang sangat singkat (Hubert and Driessen, 2004).

## 1.2. Tujuan

1. Memisahkan desa/kelurahan di Kabupaten Semarang ke dalam kelompok desa perkotaan atau desa perdesaan menggunakan analisis diskriminan kuadrat klasik dan analisis diskriminan kuadrat *robust*.
2. Mengetahui persentase ketepatan hasil pemisahan desa/kelurahan menggunakan analisis diskriminan kuadrat klasik dan analisis diskriminan kuadrat *robust*.

## 2. TINJAUAN PUSTAKA

### 2.1. Penentuan Status Daerah

Untuk menentukan apakah suatu desa atau kelurahan tertentu termasuk dalam daerah perkotaan atau perdesaan, BPS menggunakan indikator yang didasarkan pada kepadatan penduduk per km<sup>2</sup>, persentase rumah tangga pertanian dan akses ke fasilitas umum. Dalam pelaksanaannya penentuan apakah suatu desa/kelurahan termasuk daerah perkotaan atau perdesaan dilakukan oleh BPS Pusat dengan menggunakan hasil pendataan Potensi Desa (PODES) (Badan Pusat Statistik, 2010).

Daerah perkotaan adalah suatu wilayah administratif setingkat desa/kelurahan yang memenuhi persyaratan tertentu dalam hal kepadatan penduduk per km<sup>2</sup>, persentase rumah tangga pertanian dan sejumlah fasilitas perkotaan seperti sarana pendidikan formal, sarana kesehatan umum dan sebagainya (Badan Pusat Statistik, 2010). Maka dari itu daerah perkotaan harus didukung oleh berbagai prasarana dan sarana yang cukup untuk jangka waktu yang lama (Daldjoeni, 1998).

Daerah perdesaan adalah suatu wilayah administratif setingkat desa/kelurahan yang belum memenuhi persyaratan tertentu dalam hal kepadatan penduduk per km<sup>2</sup>, persentase rumah tangga pertanian dan sejumlah fasilitas perkotaan seperti sarana pendidikan formal, sarana kesehatan umum dan sebagainya (Badan Pusat Statistik, 2010). Penduduk di daerah perdesaan umumnya bekerja pada sektor agraris. Lahan pada wilayah perdesaan relatif lebih luas daripada jumlah penduduknya (Daldjoeni, 1998).

### 2.2. Analisis Diskriminan

Analisis diskriminan merupakan teknik multivariat yang berkaitan dengan pemisahan objek dalam kelompok yang berbeda dan mengalokasikan objek tersebut ke dalam suatu kelompok yang telah ditetapkan sebelumnya. Tujuan dari analisis diskriminan adalah untuk mengetahui sejauh mana perbedaan dari beberapa populasi atau kelompok (Johnson and Wichern, 2007). Nilai-nilai yang dihasilkan dari fungsi diskriminan dikenal dengan skor diskriminan (Sharma, 1996).

### 2.3. Distribusi Normal Multivariat

Asumsi normal multivariat diperlukan untuk pengujian signifikansi dari variabel diskriminan dan fungsi diskriminan. Hasil klasifikasi juga akan terpengaruh jika data tidak berasal dari distribusi normal multivariat (Sharma, 1996).

Menurut Johnson and Wichern (2007), pada kasus multivariat, vektor peubah acak  $\mathbf{X}' = [X_1, X_2, \dots, X_p]$  mengikuti fungsi densitas probabilitas

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

dimana  $-\infty < x_k < \infty$ ,  $k = 1, 2, \dots, p$  yang diberi notasi  $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ .

Menurut Johnson and Wichern (2007), sebuah metode untuk menilai normalitas dari sekumpulan data didasarkan pada jarak kuadrat tergeneralisasi

$$d_{ij}^2 = (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)' \mathbf{S}_i^{-1} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i) \quad i = 1, \dots, l; j = 1, 2, \dots, n_i$$

dimana  $n_i$  adalah jumlah objek pada populasi ke- $i$ . Prosedur ini tidak terbatas pada kasus bivariat, tetapi dapat digunakan untuk semua  $p \geq 2$ . Langkah-langkah untuk membuat plot chi-kuadrat:

1. Mengurutkan  $d_{ij}^2$  dari yang terkecil sampai dengan yang terbesar seperti  $d_{i(1)}^2 \leq d_{i(2)}^2 \leq d_{i(3)}^2 \dots \leq d_{i(n_i)}^2$ .
2. Membuat gambar pasangan  $(q_{c,p} \left( (j - \frac{1}{2}) / n_i \right), d_{ij}^2)$  dimana  $q_{c,p} \left( (j - \frac{1}{2}) / n_i \right)$  adalah kuantil  $100(j - \frac{1}{2}) / n_i$  untuk distribusi chi-kuadrat dengan derajat bebas  $p$ .

Kuantil  $q_{c,p} \left( (j - \frac{1}{2}) / n_i \right)$  berkaitan dengan persentil atas dari distribusi chi-kuadrat.

Secara khusus  $q_{c,p} \left( (j - \frac{1}{2}) / n_i \right) = \chi_p^2 \left( (j - \frac{1}{2}) / n_i \right)$ . Plot harus menyerupai garis lurus. Pola yang melengkung menunjukkan penyimpangan normalitas.

#### 2.4. Kesamaan Matriks Varian Kovarian

Analisis diskriminan linier mengasumsikan bahwa matriks varian kovarian dari dua kelompok adalah sama. Pelanggaran pada asumsi ini akan mempengaruhi pengujian signifikansi dan hasil klasifikasi. Jika asumsi kesamaan matriks varian kovarian ditolak, dapat digunakan fungsi diskriminan kuadratik untuk fungsi klasifikasi (Sharma, 1996). Menurut Johnson and Wichern (2007), uji yang digunakan untuk mengetahui kesamaan matriks varian kovarian adalah uji *Box's M* sebagai berikut:

Hipotesis:

$$H_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$$

$$H_1 : \boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$$

Statistik uji :

$$C = (1 - u)M = (1 - u) \sum_{i=1}^l (n_i - 1) \ln |\mathbf{S}_{pooled}| - \sum_{i=1}^l (n_i - 1) \ln |\mathbf{S}_i|$$

dengan:

$$\mathbf{S}_{pooled} = \frac{1}{\sum_{i=1}^l (n_i - 1)} \sum_{i=1}^l (n_i - 1) \mathbf{S}_i$$

$$u = \left[ \sum_{i=1}^l \frac{1}{(n_i - 1)} - \frac{1}{\sum_{i=1}^l (n_i - 1)} \right] \left[ \frac{2p^2 + 3p - 1}{6(p + 1)(l - 1)} \right]$$

dimana  $p$  adalah jumlah variabel dan  $l$  adalah jumlah kelompok

Kriteria penolakan :

$$H_0 \text{ ditolak jika nilai } C > \chi_{p(p+1)(g-1)/2}^2(\alpha) \text{ atau nilai sig} < \alpha$$

## 2.5. Pemilihan Variabel Pembeda

Pemilihan variabel terbaik yang secara signifikan dapat membedakan antara dua kelompok adalah tujuan pertama analisis diskriminan. Pengujian statistik untuk mengetahui perbedaan rata-rata dari dua kelompok adalah menggunakan uji-t sampel independen (Sharma, 1996). Ketika ukuran sampel yang akan diuji berukuran besar dapat digunakan uji  $-z$  (Dajan, 1986).

Menurut Dajan (1986), penggunaan uji- $z$  untuk mengetahui perbedaan rata-rata dari dua kelompok adalah sebagai berikut:

Hipotesis:

$$H_0 : \mu_{1k} = \mu_{2k}. \text{ (Rata-rata variabel ke-}k\text{ kelompok 1 dan 2 sama)}$$

$$H_1 : \mu_{1k} \neq \mu_{2k}. \text{ (Rata-rata variabel ke-}k\text{ kelompok 1 dan 2 tidak sama)}$$

Statistik Uji:

$$z = \frac{(\bar{X}_{1k} - \bar{X}_{2k})}{\sqrt{\frac{s_{1k}^2}{n_{1-1}} + \frac{s_{2k}^2}{n_{2-1}}}}, \quad k = 1, 2, \dots, p$$

Kriteria Penolakan:

$$H_0 \text{ ditolak jika nilai } z > \frac{z_{\alpha}}{2} \text{ atau } z < -\frac{z_{\alpha}}{2}$$

## 2.6. Analisis Diskriminan Kuadratik Klasik

Pada analisis diskriminan, ada asumsi yang harus dipenuhi yaitu data berdistribusi normal dan matriks varian kovarian dari populasi adalah sama. Namun terkadang ditemukan matriks varian kovarian tidak sama. Jika asumsi kesamaan matriks kovarian ditolak, dapat digunakan fungsi diskriminan kuadratik untuk fungsi klasifikasi (Sharma, 1996).

Menurut Johnson and Wichern (2007), untuk dua kelompok ( $i = 2$ ), himpunan dari kemungkinan semua hasil pada sampel dibagi menjadi dua wilayah,  $R_1$  dan  $R_2$ . Misalkan  $f_1(\mathbf{x})$  dan  $f_2(\mathbf{x})$  fungsi densitas peluang dari  $p \times 1$  vektor peubah acak  $\mathbf{X}$  untuk populasi  $\pi_1$  dan  $\pi_2$ .  $R_1$  adalah serangkaian nilai  $\mathbf{x}$  untuk objek yang kita klasifikasikan sebagai  $\pi_1$  sehingga  $R_2 = \Omega - R_1$ , dengan  $\Omega$  adalah ruang sampel yang berisi kumpulan dari semua observasi yang mungkin dari  $\mathbf{x}$ . Jika rasio harga kesalahan klasifikasi tidak ditentukan, maka rasio tersebut diambil bernilai satu. Sehingga daerah klasifikasinya menjadi:

$$R_1 : \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{p_2}{p_1}$$

$$R_2 : \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < \frac{p_2}{p_1}$$

Sehingga skor diskriminan kuadratik didefinisikan sebagai:

$$d_i^Q(\mathbf{x}) = \ln p_i - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)' \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) - \frac{1}{2} \ln(|\boldsymbol{\Sigma}_i|) \quad (2.1)$$

Menurut Jin and An (2011),  $\boldsymbol{\mu}_i$  dan  $\boldsymbol{\Sigma}_i$  tidak diketahui. Sehingga harus dicari estimasi dari sampelnya. Jadi, estimasi dari skor diskriminan kuadratik dapat menjadi:

$$\hat{d}_i^Q(\mathbf{x}) = \ln p_i - \frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}}_i)' \mathbf{S}_i^{-1}(\mathbf{x} - \bar{\mathbf{x}}_i) - \frac{1}{2} \ln(|\mathbf{S}_i|) \quad (2.2)$$

Maka, taksiran pengelompokkannya adalah alokasikan  $\mathbf{x}$  ke dalam  $\pi_2$  jika skor diskriminan kuadratik:

$$\hat{d}_1^Q(\mathbf{x}) < \hat{d}_2^Q(\mathbf{x}) \quad (2.3)$$

Untuk mengestimasi probabilitas anggota  $p_i$  pada (2.2), dapat digunakan dua pendekatan yang umum. Yang pertama  $p_i$  diasumsikan sama dari semua populasi, maka

$p_i = \frac{1}{l}$  untuk setiap  $i$ . Kedua,  $p_i$  diestimasi sebagai frekuensi relatif dari observasi pada setiap kelompok, sehingga  $p_i = \frac{n_i}{n}$  (Hubert and Driessen, 2004).

## 2.7. Evaluasi Hasil Klasifikasi

Menurut Johnson and Wichern (2007), cara penting untuk menilai kinerja dari setiap prosedur klasifikasi adalah untuk menghitung tingkat kesalahan atau probabilitas kesalahan klasifikasi. Metode untuk menghitung probabilitas kesalahan klasifikasi adalah *Apparent Error Rate* (APER). Tingkat kesalahan dapat dengan mudah dihitung dengan menggunakan matriks *confusion*, yang menunjukkan jumlah keanggotaan secara aktual dengan jumlah keanggotaan secara prediksi. Untuk  $n_1$  pengamatan dari  $\pi_1$  dan  $n_2$  pengamatan dari  $\pi_2$ , matriks *confusion* memiliki bentuk:

		Keanggotaan yang diprediksi		
		$\pi_1$	$\pi_2$	
Keanggotaan yang aktual	$\pi_1$	$n_{1C}$	$n_{1M} = n_1 - n_{1C}$	$n_1$
	$\pi_2$	$n_{2M} = n_2 - n_{2C}$	$n_{2C}$	$n_2$

Kemudian *Apparent Error Rate* dihitung dengan:

$$APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2} \quad (2.4)$$

## 2.8. Pendeteksian *Outlier* dengan Jarak Mahalanobis

Menurut Giménez *et al.* (2012), pendekatan klasik untuk mendeteksi adanya *outlier* adalah dengan menghitung jarak mahalanobis dari masing-masing pengamatan. Jarak mahalanobis didefinisikan sebagai:

$$d_{ij}^2 = (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)' \mathbf{S}_i^{-1} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i) \quad (2.5)$$

Jika data mempunyai nilai  $d_{ij}^2 > \chi_{p,1-\alpha}^2$  maka data tersebut termasuk *outlier*.

## 2.9. Penaksir *Robust Minimum Covariance Determinant*

Estimasi klasik untuk  $\bar{\mathbf{x}}_i$  dan  $\mathbf{S}_i$  sangat sensitif terhadap adanya data *outlier*. Sehingga membuat aturan analisis diskriminan kuadratik secara klasik menjadi kurang sesuai pada data yang mengandung *outlier* tersebut. Salah satu metode yang dapat digunakan adalah penaksir *Minimum Covariance Determinant* atau MCD. *Minimum Covariance Determinant* digunakan karena estimator ini dapat dihitung dalam waktu yang relatif singkat untuk kumpulan data yang besar (Hubert and Driessen, 2004).

MCD menyediakan estimasi yang *robust* pada serangkaian pengamatan multivariat. Penaksir *Minimum Covariance Determinant* ini bertujuan untuk menemukan pengamatan dari jumlah  $h$  yang telah didefinisikan sebelumnya (tanpa sampel keseluruhan) yang mana memiliki matriks kovarian dengan determinan paling kecil. Jumlah  $h$  yang telah ditentukan adalah  $(n + p + 1)/2$  (Giménez *et al.*, 2012).

## 2.10. Analisis Diskriminan Kuadratik *Robust*

Menurut Hubert and Driessen (2004) analisis diskriminan *robust* diperoleh dengan mengganti rata-rata dan matriks kovarian klasik pada rumus (2.2) dengan menggunakan penaksir MCD. Skor diskriminan kuadratik didefinisikan sebagai:

$$\hat{d}_i^{RQ}(\mathbf{x}) = \ln p_i - \frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}}_{iMCD})' \mathbf{S}_{iMCD}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_{iMCD}) - \frac{1}{2} \ln(|\mathbf{S}_{iMCD}|) \quad (2.6)$$

Alokasikan  $x$  ke dalam  $\pi_2$  jika skor diskriminan kudratik

$$\hat{d}_1^{RQ}(x) < \hat{d}_2^{RQ}(x) \quad (2.7)$$

### 3. METODOLOGI PENELITIAN

#### 3.1. Sumber Data

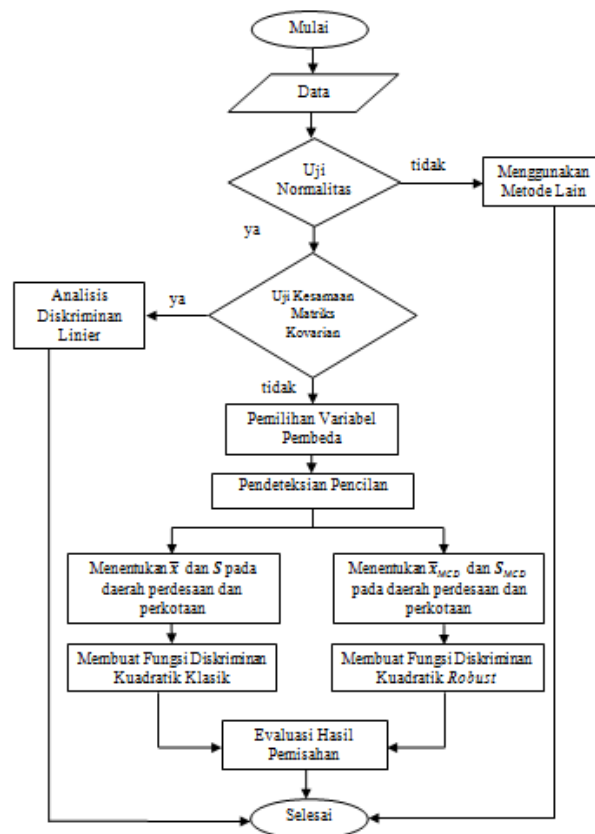
Data yang digunakan pada tugas akhir ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik. Data tersebut merupakan data hasil Pendataan Potensi Desa (PODES) Kabupaten Semarang Tahun 2011.

#### 3.2. Variabel

Pada tugas akhir ini ada delapan variabel pembeda yang digunakan yaitu kepadatan penduduk per km<sup>2</sup>, persentase rumah tangga pertanian, jarak ke pasar terdekat, jumlah Sekolah Dasar, jarak ke Sekolah Menengah Pertama terdekat, jarak ke Sekolah Menengah Atas terdekat, jarak ke rumah sakit terdekat, dan jumlah praktek bidan. Status daerah sebagai kelompok dibedakan menjadi dua golongan yaitu daerah perdesaan dan daerah perkotaan.

#### 3.3. Tahapan Analisis

Untuk lebih jelas mengenai tahapan analisis data yang dilakukan, dapat dilihat pada diagram alir Gambar 3.1.



Gambar 3.1 Diagram Alir Pengolahan Data

## **4. HASIL DAN PEMBAHASAN**

### **4.1. Statistik Deskriptif**

Berdasarkan data hasil Pendataan Potensi Desa (PODES) Kabupaten Semarang Tahun 2011, 26% desa/kelurahan yang berada di Kabupaten Semarang merupakan daerah perkotaan, sedangkan sisanya adalah daerah perdesaan. Di Kabupaten Semarang terdapat kecamatan yang seluruh desa/kelurahannya masih berstatus perdesaan yaitu Kecamatan Bancak dan Kecamatan Getasan.

### **4.2. Uji Normalitas**

Langkah pertama dalam penelitian ini adalah melakukan pengujian normalitas multivariat terhadap data yang digunakan. Pengujian normalitas multivariat dilakukan untuk setiap kelompok yaitu perdesaan dan perkotaan.

Pada pengujian normalitas multivariat diperoleh hasil bahwa nilai p-value  $(0,5692) > \alpha$ , artinya jarak mahalnobis berdistribusi chi square sehingga data perdesaan berdistribusi normal multivariat. Secara visual dapat dikatakan bahwa data perdesaan berdistribusi normal multivariat karena pola yang terbentuk cenderung mengikuti garis lurus.

Pada pengujian normalitas multivariat diperoleh hasil bahwa nilai p-value  $(0,7996) > \alpha$ , artinya jarak mahalnobis berdistribusi chi square sehingga data perkotaan berdistribusi normal multivariat. Secara visual dapat dikatakan bahwa data perkotaan berdistribusi normal multivariat karena pola yang terbentuk cenderung mengikuti garis lurus.

### **4.3. Uji Kesamaan Matriks Varian Kovarian**

Pengujian kesamaan matriks varian kovarian ini dilakukan untuk mengetahui jenis analisis diskriminan yang akan digunakan. Setelah dilakukan pengujian, didapatkan hasil bahwa nilai sig  $(0,000) < \alpha$ , artinya matriks varian kovarian pada daerah perdesaan dan daerah perkotaan tidak sama. Karena matriks varian kovarian pada daerah perdesaan dan daerah perkotaan tidak sama, maka analisis yang digunakan adalah analisis diskriminan kuadratik.

### **4.4. Pemilihan Variabel Pembeda**

Pemilihan variabel pembeda dilakukan untuk mengetahui variabel apa saja yang dapat membedakan antara desa/kelurahan yang berstatus perdesaan dan perkotaan. Setelah dilakukan pemilihan variabel pembeda, didapatkan variabel kepadatan penduduk per km<sup>2</sup>, persentase rumah tangga pertanian, jarak ke pasar terdekat, jarak ke Sekolah Menengah Pertama terdekat, jarak ke Sekolah Menengah Atas terdekat dan jarak ke Rumah Sakit terdekat yang dapat membedakan daerah perkotaan dan perdesaan.

### **4.5. Pendeteksian Pencilan**

Sebelum dilakukan analisis data, terlebih dahulu dilakukan pendeteksian pencilan untuk setiap kelompok. Desa/Kelurahan yang mempunyai nilai  $d^2 > \chi_{p,1-\alpha}^2$  dikatakan sebagai pencilan. Nilai  $\chi_{6,0,95}^2$  adalah 12,59. Dengan tingkat kepercayaan 95%, jumlah pencilan yang ditemukan pada desa/kelurahan yang berstatus perdesaan adalah sebesar 6,36% dan 6,45% untuk desa/kelurahan yang berstatus perkotaan. Terdapat 11 desa/kelurahan yang terdeteksi sebagai pencilan, yaitu Gogik, Gondoriyo, Jatijajar, Candigarón, Bejalen, Jombor, Karangtengah, Ngajaran, Bantal, Jumpang dan Lembu

untuk desa/kelurahan yang berstatus perdesaan. Sedangkan untuk desa/kelurahan yang berstatus perkotaan, terdapat 4 pencilan yaitu Wujil, Kupang, Gedangan dan Sruwen.

#### 4.6. Analisis Diskriminan Kuadratik Klasik

Langkah pertama yang dilakukan pada analisis diskriminan kuadratik klasik adalah menentukan vektor rata-rata dan matriks varian kovarian pada daerah perdesaan dan daerah perkotaan. Setelah mendapatkan vektor rata-rata dan matriks varian kovarian, selanjutnya dapat dibentuk fungsi diskriminan kuadratik klasik. Fungsi diskriminan kuadratik klasik untuk daerah perdesaan dilambangkan dengan  $\hat{d}_1^Q$ , sedangkan daerah perkotaan dilambangkan  $\hat{d}_2^Q$ . Fungsi diskriminan yang terbentuk dinyatakan dalam bentuk sebagai berikut:

$$\hat{d}_1^Q(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}}_1)' \mathbf{S}_1^{-1}(\mathbf{x} - \bar{\mathbf{x}}_1) - \frac{1}{2} \ln(|\mathbf{S}_1|) + \ln\left(\frac{173}{235}\right)$$

$$\hat{d}_2^Q(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}}_2)' \mathbf{S}_2^{-1}(\mathbf{x} - \bar{\mathbf{x}}_2) - \frac{1}{2} \ln(|\mathbf{S}_2|) + \ln\left(\frac{62}{235}\right)$$

Langkah selanjutnya adalah menghitung nilai  $\hat{d}_1^Q$  dan  $\hat{d}_2^Q$  untuk setiap desa/kelurahan. Setelah didapatkan nilai  $\hat{d}_1^Q$  dan  $\hat{d}_2^Q$ , masing – masing desa/kelurahan dialokasikan ke dalam kelompok perkotaan atau perdesaan. Jika desa/kelurahan mempunyai nilai  $\hat{d}_1^Q < \hat{d}_2^Q$  maka desa/kelurahan tersebut termasuk dalam kelompok perkotaan. Pemisahan desa/kelurahan di Kabupaten Semarang dengan menggunakan fungsi diskriminan kuadratik klasik memberikan hasil 183 desa/kelurahan memiliki status perdesaan dan 52 desa/kelurahan memiliki status perkotaan.

#### 4.7. Analisis Diskriminan Kuadratik Robust

Salah satu tujuan dari tugas akhir ini adalah membandingkan kinerja antara diskriminan kuadratik klasik dan *robust*. Aplikasi analisis diskriminan kuadratik *robust* pada data Potensi Desa Kabupaten Semarang tahun 2011 sangat sesuai karena pada data tersebut ditemukan pencilan sebesar 6,39% untuk perdesaan dan 6,45% untuk perkotaan. Sehingga nantinya diharapkan pengelompokan yang dilakukan dengan diskriminan kuadratik *robust* menjadi lebih tepat.

Langkah yang dilakukan pada analisis diskriminan kuadratik *robust* sama dengan diskriminan kuadratik klasik. Pertama, menentukan vektor rata-rata dan matriks kovarian pada desa/kelurahan berstatus perdesaan dan perkotaan. Langkah selanjutnya adalah membentuk fungsi diskriminan kuadratik *robust*. Fungsi diskriminan kuadratik *robust* untuk daerah perdesaan dilambangkan dengan  $\hat{d}_1^{RQ}$ , sedangkan daerah perkotaan dilambangkan  $\hat{d}_2^{RQ}$ . Fungsi diskriminan yang terbentuk dinyatakan dalam bentuk sebagai berikut:

$$\hat{d}_1^{RQ}(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}}_{1MCD})' \mathbf{S}_{1MCD}^{-1}(\mathbf{x} - \bar{\mathbf{x}}_{1MCD}) - \frac{1}{2} \ln(|\mathbf{S}_{1MCD}|) + \ln\left(\frac{173}{235}\right)$$

$$\hat{d}_2^{RQ}(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}}_{2MCD})' \mathbf{S}_{2MCD}^{-1}(\mathbf{x} - \bar{\mathbf{x}}_{2MCD}) - \frac{1}{2} \ln(|\mathbf{S}_{2MCD}|) + \ln\left(\frac{62}{235}\right)$$

Langkah selanjutnya adalah menghitung nilai  $\hat{d}_1^{RQ}$  dan  $\hat{d}_2^{RQ}$  untuk setiap desa/kelurahan. Setelah didapatkan nilai  $\hat{d}_1^{RQ}$  dan  $\hat{d}_2^{RQ}$ , masing–masing desa/kelurahan dialokasikan ke dalam kelompok perkotaan atau perdesaan. Jika desa/kelurahan



mempunyai nilai  $\hat{d}_1^{RQ} < \hat{d}_2^{RQ}$  maka desa/kelurahan tersebut termasuk dalam kelompok perkotaan. Pemisahan desa/kelurahan di Kabupaten Semarang dengan menggunakan fungsi diskriminan kuadratik *robust* memberikan hasil 167 desa/kelurahan memiliki status perdesaan dan 68 desa/kelurahan memiliki status perkotaan.

#### 4.8. Evaluasi Hasil Pemisahan

Cara untuk menilai kinerja dari pemisahan desa/kelurahan di Kabupaten Semarang menurut status daerah adalah dengan menghitung tingkat kesalahan atau probabilitas kesalahan klasifikasi. Kesalahan klasifikasi dihitung untuk masing-masing analisis diskriminan kuadratik klasik dan *robust*. Nilai APER untuk fungsi diskriminan kuadratik klasik adalah sebagai berikut:

$$\text{APER} = \frac{n_{1M} + n_{2M}}{n_1 + n_2} = \frac{10 + 20}{173 + 62} = 0,1277$$

Sedangkan nilai APER untuk fungsi diskriminan kuadratik *robust* adalah sebagai berikut:

$$\text{APER} = \frac{n_{1M} + n_{2M}}{n_1 + n_2} = \frac{15 + 9}{173 + 62} = 0,1021$$

Berdasarkan nilai APER tersebut, dapat dijelaskan bahwa dengan fungsi diskriminan kuadratik klasik dihasilkan proporsi salah pemisahannya sebesar 12,77% dan 10,21% untuk fungsi diskriminan kuadratik *robust*. Kinerja dari fungsi diskriminan kuadratik *robust* secara keseluruhan menghasilkan proporsi kesalahan pemisahan yang lebih kecil daripada fungsi diskriminan kuadratik klasik.

### 5. KESIMPULAN

Berdasarkan analisis dan pembahasan yang telah dilakukan, dapat diperoleh beberapa kesimpulan, yaitu:

1. Pemisahan desa/kelurahan di Kabupaten Semarang dengan analisis diskriminan kuadratik klasik memberikan hasil 183 desa/kelurahan memiliki status perdesaan dan 52 desa/kelurahan memiliki status perkotaan.
2. Pemisahan desa/kelurahan di Kabupaten Semarang dengan analisis diskriminan kuadratik *robust* memberikan hasil 167 desa/kelurahan memiliki status perdesaan dan 68 desa/kelurahan memiliki status perkotaan.
3. Penerapan analisis diskriminan kuadratik *robust* lebih sesuai dengan data Potensi Desa 2011 Kabupaten Semarang yang mengandung pencilan karena mampu memberikan ketepatan hasil pemisahan sebesar 89,79%, sedangkan dengan diskriminan kuadratik klasik memiliki ketepatan hasil pemisahan sebesar 87,23%.

### 6. DAFTAR PUSTAKA

- Badan Pusat Statistik, 2010, *Klasifikasi Perkotaan dan Perdesaan di Indonesia*, Badan Pusat Statistik, Jakarta.
- Badan Pusat Statistik, 2011, *Profil Desa Kabupaten Semarang*, Badan Pusat Statistik Kabupaten Semarang, Ungaran.
- Dajan, A., 1986, *Pengantar Metode Statistik*, LP3ES, Jakarta.
- Daldjoeni, N., 1998, *Geografi Kota dan Desa, Edisi kedua*, PT Alumni, Bandung.

- Giménez, E., Crespi, M., Garrido, S., and Gil, A.J., 2012, *Multivariate outlier detection based on robust computation of Mahalanobis distances*, International Journal of Applied Earth Observation and Geoinformation 16, 94-100.
- Hubert, M. and Driessen, K.V., 2004, *Fast and robust discriminant analysis*, Computational Statistics & Data Analysis 45. 301-320.
- Jin, J. and An, J., 2011, *Robust discriminant analysis and its application to identify protein coding regions of rice genes*, Mathematical Biosciences 232, 96-100.
- Johnson, R.A. and Wichern, D.W., 2007, *Applied Multivariate Statistical Analysis, Sixth Edition*, Pearson Education Ltd., London.
- Sadyohutomo, M., 2009, *Manajemen Kota dan Wilayah*, Bumi Aksara, Jakarta.
- Sharma, S., 1996, *Applied Multivariate Techniques*, John Wiley & Sons. Inc., Canada.