

Metode Analisis Diskriminan dalam Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Indikator Indeks Pembangunan Manusia

Novi Afryanthi S.*, Muhammad Arif Tiro, & Ansari Saleh Ahmar

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Makassar, Indonesia

Keywords: Cluster analysis, Discriminant analysis, Human development index indicator.

Abstract:

Discriminant analysis is a method in multivariate statistical analysis that is related to objects which have separated into the defined group defined and see the accuracy of the formed group. In this research, cluster analysis is used for the first grouping, cluster analysis is a statistical analysis which aims to classify some objects based on the characteristics similarity among the object. Data for this study is HDI (Human Development Index) of indicator in south Sulawesi in 2016. The result of this research are 1st cluster (lower HDI indicator) which have 21 city/district and the 2nd cluster (higher HDI indicator) which have 3 city/district as the closeness value between the cluster that formed is 0.902 which shows the closeness between the cluster is high. Furthermore, the discriminant function that have formed explains that if the life expectancy increase, the HDI indicator in city/district in south Sulawesi province will decrease but if school expectation duration in school, average of duration in school, and parity of purchasing power is increasing, the HDI indicator in city/district in south Sulawesi will also increase.

1. Pendahuluan

Analisis Diskriminan merupakan teknik statistika untuk mengelompokkan individu-individu ke dalam kelompok-kelompok yang saling bebas dengan tegas berdasarkan segugus peubah bebas (Matjik, 2004). Analisis diskriminan adalah teknik statistika multivariat yang terkait dengan pemisahan atau alokasi/klasifikasi (*classification*) sekelompok objek ke dalam kelompok yang telah terlebih dahulu didefinisikan. Dalam tujuan pengenalan objek (observasi), metode ini mencoba menemukan suatu 'diskriminan' yang nilainya secara numeris sedemikian sehingga mampu memisahkan objek yang karakteristiknya telah diketahui. Sedangkan dalam tujuan klasifikasi objek, metode ini akan mensortir objek (observasi) kedalam 2 atau lebih kelas (Amir Hamzahan; Gatot Santosa; dan Wisnu Widiarto, 2002). Pengelompokan pada analisis diskriminan bersifat *mutually exclusive*, yaitu jika suatu objek telah dimasukkan dalam satu kelompok maka tidak dapat menjadi anggota kelompok lain. Setelah pengelompokan, analisis diskriminan juga berguna dalam melihat keakuratan pengelompokan yang terbentuk. Dalam penelitian ini, penulis akan menggunakan Analisis Diskriminan dalam Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Indikator IPM pada tahun 2016. Indeks Pembangunan Manusia (IPM) merupakan pendekatan untuk mengukur pencapaian pembangunan yang ditekankan pada kesejahteraan masyarakat.

Tinggi rendahnya IPM kabupaten/kota di Provinsi Sulawesi Selatan hanya ditunjukkan oleh Indeks Komposit, tetapi tidak ditunjukkan indikator mana yang dominan terhadap tinggi/rendahnya peringkat IPM. Padahal nilai dari tiap-tiap

* Corresponding author.

E-mail address: noviafryanthi@gmail.com



indikator pembentuk IPM perlu dilihat agar dapat diketahui pencapaian dari tiap-tiap indikator. Pengelompokan wilayah kabupaten/kota di Provinsi Sulawesi Selatan perlu dilakukan sebagai bahan perencanaan dan evaluasi sasaran program pemerintah untuk meningkatkan angka pembangunan manusia. Sehingga, untuk mengetahui pengelompokan dengan keakuratan yang tepat pada kabupaten/kota berdasarkan indikator IPM dapat digunakan analisis diskriminan.

2. Tinjauan Pustaka

2.1. Analisis Diskriminan

Analisis diskriminan adalah salah satu teknik statistika yang bisa digunakan pada hubungan dependensi (hubungan antarpeubah dimana sudah bisa dibedakan mana peubah respon dan mana peubah penjelas).

Asumsi Diskriminan

1. *Multivariate normality*, yang dalam hal ini berarti peubah independen seharusnya berdistribusi normal. Jika data tidak berdistribusi normal, hal ini akan menyebabkan masalah pada ketepatan fungsi (model) diskriminan. Regresi logistik (*logistic regression*) bisa dijadikan alternatif metode jika memang data peubah independen tidak berdistribusi normal secara nyata.
2. Matriks kovarians dari semua peubah independen seharusnya sama (equal).
3. Tidak ada korelasi antarpeubah independen. Jika dua peubah independen mempunyai korelasi yang kuat, maka dikatakan terjadi multikolinieritas. Seperti pada analisis regresi berganda, terjadinya multikolinieritas akan menyebabkan biasanya interpretasi dari model yang terbentuk.
4. Tidak adanya data sangat ekstrim (outlier) pada peubah independen. Jika ada data outlier dan proses pengolahan tetap dilakukan hal ini bisa berakibat berkurangnya ketepatan klasifikasi dari fungsi diskriminan.

Analisis Diskriminan Linear atau *Linear Discriminant Analysis* (LDA) dapat digunakan sebagai metode klasifikasi maupun reduksi dimensi pada sebuah set data. Tujuan utama dari analisis diskriminan adalah untuk memperoleh kaidah matematis, yang dikenal dengan fungsi diskriminan, yang dapat digunakan untuk memisahkan kelompok objek yang berbeda. Misalkan pada data terdapat p peubah bebas dan j kelompok, dengan n_i adalah jumlah pengamatan untuk pengamatan ke- i . Suatu pengamatan baru akan ditempatkan pada salah satu kelompok berdasarkan fungsi diskriminan berikut:

$$D_i = b_0 + b_1x_{1j} + b_2x_{2i} + b_px_{pi}$$

dengan:

- D_i = skor diskriminan untuk pengamatan ke- i
 x_{pi} = nilai pengamatan ke- i peubah ke- p
 b_p = koefisien diskriminan peubah ke- p

2.2. Pengelompokan Awal

Analisis *cluster* merupakan suatu teknik analisis multivariat yang mempunyai tujuan untuk mengelompokkan data peubah atau observasi ke dalam *cluster* sedemikian rupa sehingga masing-masing *cluster* bersifat homogen sesuai dengan faktor yang digunakan untuk melakukan *clustering*. Karena yang diinginkan adalah untuk mendapatkan *cluster* yang sehomogen mungkin, maka yang digunakan sebagai dasar untuk *clustering* adalah kesamaan skor nilai yang dianalisis. Data mengenai ukuran kesamaan tersebut dapat dianalisis dengan analisis *cluster* sehingga dapat ditentukan siapa yang masuk *cluster* mana (Gudono, 2011 dalam Sofya Laeli, (2014))

3. Metode Penelitian

Jenis data yang digunakan dalam penelitian ini adalah data sekunder yang dipublikasikan oleh Badan Pusat Statistika (BPS) Provinsi Sulawesi Selatan yaitu data indikator indeks pembangunan manusia (IPM) Tahun 2016. Tahap awal dari analisis pengelompokan adalah merumuskan masalah dengan mendefinisikan peubah-peubah yang dipergunakan untuk dasar klasifikasi.

Teknik Analisis Data:

- 1) Tahap awal dari analisis pengelompokan adalah merumuskan masalah dengan mendefinisikan peubah-peubah yang dipergunakan untuk dasar klasifikasi.
- 2) Sampel yang diambil harus benar-benar bisa mewakili populasi.
- 3) Melakukan eksplorasi data.
- 4) Standarisasi data dan pengecekan asumsi.
Standarisasi data jika terdapat perbedaan nilai yang besar antar peubah yang dapat menyebabkan bias dalam analisis pengelompokan maka data asli perlu standarisasi. Misalnya, ada yang dalam satuan juta dan ada yang satuan puluhan atau bahkan lebih kecil. Perbedaan data yang besar akan menyebabkan perhitungan jarak menjadi tidak valid.
- 5) Melakukan penentuan kelompok awal menggunakan analisis *cluster*.
- 6) Pengujian asumsi analisis diskriminan.
- 7) Melakukan uji signifikan dari fungsi diskriminan.
- 8) Melakukan uji keeratan hubungan fungsi diskriminan.
- 9) Membentuk fungsi diskriminan.
- 10) Kesimpulan.

4. Hasil dan Pembahasan

4.1. Standarisasi Data

Sebelum melakukan pengolahan data dengan metode *cluster* dan diskriminan, terlebih dahulu dilakukan standarisasi peubah. Standarisasi digunakan ketika terdapat perbedaan satuan yang berbeda. Pada penelitian ini, data yang digunakan memiliki satuan yang berbeda maka dilakukan standarisasi data. Hasil standarisasi data Indikator dan IPM. Berdasarkan tabel dibawah, telah dilakukan standarisasi sehingga satuan dari setiap peubah telah sama. Selanjutnya, dapat dilanjutkan dengan melakukan pengelompokan peubah menggunakan *cluster* hirarki. Pada metode *cluster*, hal awal yang perlu dilakukan yaitu penentuan ukuran kemiripan antar objek.

4.2. Proses Analisis Diskriminan

4.2.1. Pengujian Asumsi Diskriminan

a. Uji Normal Multivariat

Tabel 1 Uji Normal Multivariat

	Nilai Statistika	<i>p-value</i>
Kurtosis	4,464	0,903
Skewness	4,623	0,328

Berdasarkan tabel 1, didapatkan nilai signifikansi dari kurtosis dan skewness, nilainya lebih besar dari 0,05 ($p - value > \alpha = 0,05$), artinya data berdistribusi normal multivariat.

b. Uji Kesamaan Matriks Kovarians

Tabel 2 Uji Kesamaan Matriks Kovarians

	<i>K-Squared</i>	<i>df</i>	<i>p-value</i>
<i>Barlett Test</i>	0.453	1	0,500

Berdasarkan tabel 2, terlihat bahwa pada uji kesamaan matriks kovarians menggunakan uji *barlett test*, diperoleh nilai $p - value = 0,500 (> \alpha = 0,05)$, sehingga dapat disimpulkan bahwa matriks kovarians akan relatif sama untuk kedua kelompok.

c. Uji Multikolonieritas

Tabel 3 Uji Multikolonieritas

Peubah	Tolerance	VIF
X ₁ (AHH)	0,478	2,094
X ₂ (HLS)	0,167	6,002
X ₃ (RLS)	0,123	8,133
X ₄ (PDB)	0,482	2,076

Berdasarkan tabel 3, terlihat bahwa nilai toleransi setiap peubah lebih besar dari 0,10. Sementara itu, untuk nilai VIF dari setiap peubah lebih kecil dari 10,00. Sehingga dapat disimpulkan bahwa peubah-peubah tersebut tidak mengalami multikolonieritas.

4.2.2. Pembentukan Fungsi Diskriminan

Terdapat empat peubah yang terseleksi masuk dalam analisis, yaitu angka harapan hidup, harapan lama sekolah, rata-rata lama sekolah, dan paritas daya beli. Sebelum melanjutkan ke analisis diskriminan, terlebih dahulu akan dimunculkan rata-rata nilai peubah dari masing-masing kelompok yang terbentuk.

Tabel 4 Rata-Rata Nilai Peubah Dari Tiap Kelompok

Kelompok Indikator IPM	AHH	HLS	RLS	PDB
Rendah	68,383	12,755	7,187	9630,619
Tinggi	70,793	14,760	10,450	13711,667

Berdasarkan tabel 4, terlihat bahwa terdapat perbedaan rata-rata dari setiap peubah pada masing-masing kelompok. Nilai rata-rata setiap peubah untuk kelompok Indikator IPM rendah, rata-ratanya lebih kecil dibanding nilai rata-rata setiap peubah untuk kelompok Indikator IPM tinggi.

a. Uji Signifikan Diskriminan

Tabel 5 Uji Signifikan Diskriminan

Test of Function	Wilk's λ	Chi-square	df	p-value
1	0,176	34,687	4	0,001

Berdasarkan tabel 5, dapat dilihat bahwa nilai signifikan 0,001 ($p - value < \alpha = 0,05$), maka H₀ ditolak sehingga disimpulkan bahwa peubah-peubah yang diteliti memiliki perbedaan antar kelompok.

b. Uji Keeratan Hubungan Diskriminan

Tabel 6 Uji Keeratan Hubungan

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	4,365	100	100	0,902

Berdasarkan tabel 6, *Canonical Correlation* mengukur keeratan hubungan antara *diskriminant score* dan *group* (dalam hal ini, karena ada dua *cluster* yang terbentuk, maka ada dua *group*). Angka 0,902 menunjukkan keeratan yang tinggi, dengan ukuran skala asosiasi antara 0 sampai 1

c. Fungsi Diskriminan

Berdasarkan tabel 7, fungsi diskriminan pada prinsipnya membuat semacam persamaan regresi, dengan pembagian berdasarkan kode *cluster*. Adapun fungsi diskriminan yang terbentuk yaitu:

$$D = -0,2307 AHH + 0,1454 HLS + 1,7725 RLS + 0,0001 PDB$$

Berdasarkan fungsi diskriminan tersebut, dapat dijelaskan jika penambahan angka pada peubah Angka Harapan Hidup, maka nilai indikator IPM di kabupaten/kota di Provinsi Sulawesi Selatan akan menurun. Tetapi jika penambahan angka pada peubah Harapan Lama Sekolah, Rata-rata Lama Sekolah, dan Paritas Daya Beli, maka nilai indikator IPM di kabupaten/kota di Provinsi Sulawesi Selatan akan meningkat.

Tabel 7 Nilai Parameter

Peubah	Parameter
X ₁ (AHH)	-0,2307
X ₂ (HLS)	0,1454
X ₃ (RLS)	1,7725
X ₄ (PDB)	0,0001

5. Kesimpulan

1. Berdasarkan analisis pengelompokan awal, pengelompokan 24 kabupaten/kota di Provinsi Sulawesi Selatan berdasarkan indikator IPM menggunakan metode *cluster* hirarki, dapat dibentuk 2 *cluster*, dengan *cluster* 1 (kelompok Indikator IPM rendah) beranggotakan 21 kabupaten/kota yaitu Selayar, Bulukumba, Bantaeng, Jeneponto, Takalar, Gowa, Sinjai, Maros, Pangkep, Barru, Bone, Soppeng, Wajo, Sidrap, Pinrang, Enrekang, Luwu, Tana Toraja, Luwu Utara, Luwu Timur, dan Toraja Utara. Sedangkan *cluster* 2 beranggotakan 3 kabupaten/kota yaitu Makassar, Pare-pare, dan Palopo.
2. Ukuran keeratan hubungan antara 2 *cluster* yang terbentuk berada pada angka 0,902 yang menunjukkan keeratan yang tinggi, dengan ukuran skala asosiasi antara 0 sampai 1.
3. Berdasarkan analisis diskriminan, diperoleh fungsi diskriminan sebagai berikut:

$$D = -0,2307 AHH + 0,1454 HLS + 1,7725 RLS + 0,0001 PDB$$

Sehingga dapat disimpulkan jika penambahan angka pada peubah Angka Harapan Hidup, maka nilai indikator IPM di kabupaten/kota di Provinsi Sulawesi Selatan akan menurun. Tetapi jika penambahan angka pada peubah Harapan Lama Sekolah, Rata-rata Lama Sekolah, dan Paritas Daya Beli, maka nilai indikator IPM di kabupaten/kota di Provinsi Sulawesi Selatan akan meningkat. Atau kita juga bisa menyimpulkan sebagai berikut, semakin tinggi nilai Angka Harapan Hidup maka semakin rendah nilai indikator IPM dan semakin tinggi nilai Harapan Lama Sekolah, Rata-rata Lama Sekolah, dan Paritas Daya Beli maka nilai indikator IPM akan meningkat

References

- Abubakar, Warda. (2007). *Aplikasi Analisis Diskriminan pada Pengeluaran Daerah Di Provinsi Sulawesi Selatan. Makassar*. Skripsi Fmipa UNM.
- Annas, S., & Irwan. (2015). *Penerapan Analisis Diskriminan Dalam Pengelompokan Desa Miskin di Kabupaten Wajo*. Indonesian Journal of Fundamental Sciences, 1(1).
- Anuraga, G. (2015). Hierarchical Clustering Multiscale Bootstrap untuk Pengelompokan Kemiskinan di Jawa Timur. *Jurnal Statistika Universitas Muhammadiyah Semarang*, 3(1).
- BPS. (2017). *Indeks Pembangunan Manusia (IPM) Tahun 2016*. Badan Pusat Statistika Provinsi Sulawesi Selatan.
- Hamzahan, A., Santosa, G., & Widiarto, W. (2002). Klasifikasi Objek Dalam Visi Komputer Dengan Analisis Diskriminan. *Makara Journal of Technology*, 6(1), 146732.
- Laeli, S. (2014). *Analisis Cluster dengan Average Linkage Method dan Ward's Method untuk Data Responden Nasabah Asuransi Jiwa Unit Link*. Skripsi.
- Mattjik, A. A., dkk. (2004). *Modul Teori Pelatihan Analisis Multivariat*. Bandung: Departement Statistika. FMIPA IPB.
- Santoso, S. (2017). *Statistik Multivariat*. Jakarta: PT. Elaks Media Komputindo.
- Simamora, Bilson. (2005). *Analisis Multivariat Pemasaran*. Jakarta: PT. Gramedia Pustaka Utama.

- Siswadi,. & Budi, S. (1989). *Analisis Eksplorasi Data Peubah Ganda*. Bandung: Jurusan Matematika. FMIPA IPB.
- Soraya, Y. (2011). *Perbandingan Kinerja Metode Single LinkageMetode Complete Linkage , dan Metode K-Means dalam Analisis Cluster*. Skripsi.
- Tiro, M, A., Sukarna & Aswi. (2010). *Statistika Deskriptif Peubah Banyak*. Makassar: Andira Publisher.