

# **Analisis Diskriminan dalam Penelitian Ekonomi**

**Oleh**  
**Purwo Susongko, Inayah Adi Sari**

## **ABSTRAK**

Banyak dari objek studi ekonomi adalah mengamati adanya perbedaan kemudian mencari perbedaan yang jelas antar grup pada variabel dependen (variabel amatan), dan setelah ada perbedaan, variabel independen (variabel prediktor) manakah yang dominan memberikan perbedaan tersebut. Masalah-masalah tersebut sering ditemui terutama dalam bidang riset pemasaran, manajemen, sistem keuangan, pendidikan ekonomi dan bidang riset yang lain. Objek studi yang pertama dapat dianalisis dengan analisis kluster sedangkan untuk menjawab objek studi yang kedua dan ketiga dilaksanakan dengan analisis diskriminan. Oleh karenanya analisis diskriminan merupakan kelanjutan dari analisis kluster.

Kata kunci: analisis diskriminan, variabel amatan, variabel prediktor, fungsi diskriminan

## **1. Konsep Dasar Analisis Diskriminan**

Analisis diskriminan adalah analisis multivariat yang diterapkan untuk memodelkan hubungan antara satu variabel respon yang bersifat kategori (*non-metrik, nominal atau ordinal*, bersifat kualitatif) dengan satu atau lebih variabel prediktor yang bersifat kuantitatif (Tatham, Hair, Anderson, dan Black, 1998). Misalnya, variabel tak bebas merupakan pilihan 3 merek komputer pribadi (*personal komputer*), yaitu merek A, B, dan C dan variabel bebasnya merupakan penilaian atribut (*rating of attributes*) dari PC dengan menggunakan skala likert. Pemodelan antara variabel prediktor/bebas dan variabel respon (tak bebas) dinyatakan oleh satu fungsi. Fungsi ini dibentuk dengan memaksimumkan jarak antar kelompok, sehingga memiliki kemampuan untuk membedakan antar kelompok. Berdasarkan fungsi ini, pengamatan yang belum diketahui kelompoknya dapat ditentukan kelompoknya. Oleh karena itu, analisis diskriminan ini dapat dipergunakan sebagai metode pengklasifikasian. Dalam penerapannya, analisis ini mengasumsikan bahwa data pada masing-masing kategori/kelompok berdistribusi normal multivariat dan mempunyai struktur matrik varian-kovarian yang sama.

Pengklasifikasian adalah salah satu analisis statistika yang diperlukan jika ada beberapa kelompok kemudian ingin diketahui apakah kelompok-kelompok tersebut memang berbeda secara statistika. Kelompok-kelompok ini terjadi karena ada pengaruh satu atau lebih variabel lain yang merupakan variabel independen. Kombinasi linier dari variabel-variabel ini akan membentuk suatu fungsi diskriminan (Tatham et. al., 1998).

$$Z_{jk} = a + W_1 X_{1k} + W_2 X_{2k} + \dots + W_n X_{nk} \quad (1)$$

$Z_{jk}$  : Nilai diskriminan Z dari fungsi diskriminan j untuk obyek k

a : intersep

$W_i$  : Koefisien diskriminan untuk variabel independen ke-i

$X_{ik}$  : Nilai variabel ke-i untuk obyek ke-k

Fungsi diskriminan lain yang dapat digunakan antara lain fungsi diskriminan linier Fisher. Secara detail fungsi ini dijelaskan dalam Johson dan Winchern (1992). Nilai diskriminan Z dari (1) merupakan dasar untuk menentukan suatu obyek masuk kelompok yang mana dengan membandingkannya dengan rata-rata (centroid) dari nilai Z masing-masing kelompok. Jika ada dua kelompok, misalkan A dan B, maka

$$Z_{cu} = \frac{N_A Z_B + N_B Z_A}{N_A + N_B} \quad (2)$$

$Z_{cu}$  : Nilai Z kritis

$N_A$  : Jumlah obyek di dalam A

$N_B$  : Jumlah obyek di dalam B

$Z_A$  : centroid untuk A

$Z_B$  : centroid untuk B

Kelompokkan ke dalam A jika  $Z_n < Z_{ct}$

Kelompokkan ke dalam B jika  $Z_n > Z_{ct}$

### **Evaluasi Fungsi Klasifikasi**

Hal penting untuk menilai suatu prosedur klasifikasi adalah dengan menghitung peluang kesalahan klasifikasi. Ukuran ini dinamakan *Apparent Error Rates* (APER), yang didefinisikan dengan fraksi (proporsi) pengamatan pada sampel yang salah diklasifikasikan oleh fungsi klasifikasi. APER dihitung dengan terlebih dulu membuat tabel klasifikasi.

Tabel 2.1. Klasifikasi actual dan predicted group

Actual group	Predicted Group	
	$\pi_1$	$\pi_2$
$\pi_1$	$n_{11}$	$n_{12} = n_{11} - n_{11}$
$\pi_2$	$n_{21} = n_2 - n_{22}$	$n_{22}$

$n_{11}$  = jumlah pengamatan dari  $\pi_1$  tepat diklasifikasi sebagai  $\pi_1$

$n_{12}$  = jumlah pengamatan dari  $\pi_1$  salah diklasifikasi sebagai  $\pi_2$

$n_{22}$  = jumlah pengamatan dari  $\pi_2$  tepat diklasifikasi sebagai  $\pi_2$

$n_{21}$  = jumlah pengamatan dari  $\pi_2$  salah diklasifikasi sebagai  $\pi_1$

$$APER = \frac{n_{12} + n_{21}}{n_1 + n_2} \quad (3)$$

## 2. Tujuan Analisis Diskriminan

Tujuan analisis diskriminan : (1) Membuat suatu fungsi diskriminan atau kombinasi linear, dari prediktor atau variabel bebas yang bisa mendiskriminasi atau membedakan kategori variabel tak bebas atau *criterion* atau kelompok, artinya mampu membedakan suatu objek (responden) masuk kelompok/kategori yang mana, (2) Menguji apakah ada perbedaan signifikansi antara kategori/kelompok, dikitikan dengan variabel bebas atau prediktor, (3) Menentukan prediktor/variabel bebas yang mana yang memberikan sumbangan terbesar terhadap terjadinya perbedaan antar-kelompok., (4) Mengklarifikasi/mengelompokkan objek/kasus atau responden ke dalam suatu kelompok atau kategori didasarkan pada nilai variabel bebas, (Pelanggan memilih komputer merek A? B? atau C?, ada 3 kelompok/kategori), (5) Mengevaluasi keakuratan klasifikasi (*the accuracy of clasification*).

Teknik analisis diskriminan dibedakan menjadi dua yaitu analisis diskriminan dua kelompok/kategori, kalau variabel tak bebas Y dikelompokkan menjadi dua. Diperlukan suatu fungsi diskriminan. Kalau variabel tak bebas dikelompokkan menjadi lebih dari dua kelompok disebut analisis diskriminan berganda (*multiple discriminant analysis*) diperlukan fungsi diskriminan sebanyak  $(K - 1)$  kalau memang ada K kategori. Misalnya hanya ada 2 merek A dan B cukup 1 fungsi diskriminan D, tetapi kalau ada 3 merek A, B, dan C diperlukan 2 fungsi diskriminan  $(3 - 1 = 2)$  katakan D1 dan D2.

Teknik analisis diskriminan dapat untuk menjawab pertanyaan berikut.

1. Dinyatakan dalam karakteristik demografi, bagaimana pelanggan yang loyal bisa dibedakan dengan yang tidak loyal?
2. Apakah peminum berat, medium, dan ringan dari soft drink berbeda kalau dinyatakan dalam konsumsi makanan beku?
3. Karakteristik psikografik apa yang bisa membantu membedakan/mendiskriminasi antara pembeli di pasar swalayan Hero yang sensitif terhadap harga dan yang tidak?
4. Apakah di berbagai segmen pasar berbeda di dalam kebiasaan menggunakan media promosi?
5. Dinyatakan dalam gaya hidup, apakah ada perbedaan antara pelanggan pasar tradisional dan pasra modern, kalau ada variabel apa saja yang bisa membedakan/mendiskriminasi secara signifikan?
6. Seorang direktur kredit suatu bank ingin mengetahui apakah seorang nasabah peminta kredit, merupakan nasabah yang jujur atau tidak jujur berdasarkan beberapa catatan tentang karakteristik/atribut pribadi dari nasabah tersebut.
7. Seorang direktur personalia suatu perusahaan ingin mengetahui apakah seorang sarjana yang mengajukan lamaran merupakan calon karyawan yang mempunyai komitmen pada perusahaan atau tidak berdasar beberapa data pribadi pelamar tersebut.
8. Seorang dokter penyakit dalam bisa menggunakan catatan tekanan darah, berat badan, kolestrol dan atribut tentang kesehatan lainnya, untuk membedakan pasien yang kemungkinan besar akan terkena serangan jantung dan yang tidak.

### 3. Melakukan Analisis Diskriminan

#### Merumuskan Masalah

Langkah pertama dalam analisis diskriminan ialah merumuskan masalah (*problem formulation*) dengan jalan mengenali :

- (1) Tujuannya (*the objective*). Memasukkan objek/elemen ke dalam kelompok atau kategori tertentu,
- (2) Variabel tak bebas (*dependent* atau *criterion*), terdiri dari dua kelompok /kategori atau lebih,
- (3) Variabel bebas, ada berapa? Variabel bebas atau prediktor.

Perlu diketahui bahwa analisis diskriminan hampir sama dengan analisis regresi berganda, hanya bedanya, variabel tak bebasnya berupa non-metrik atau kategori atau kelompok bersifat kualitatif (bisa berupa skala nominal atau ordinal). Untuk dua kelompok misalnya responden dikelompokkan menjadi dua pelanggan loyal dan tak loyal; karyawan jujur dan tak jujur; perusahaan akan maju dan akan bangkrut, dan lain sebagainya. Apabila data untuk variabel tak bebas berupa data metrik (interval atau rasio) harus diubah dahulu.

Misalnya data tingkat kepuasan : sangat puas = 5, puas = 4, netral (biasa saja) = 3, tak puas = 2, sangat tidak puas = 1. Kemudian ingin diubah menjadi 3 kelompok. Sangat puas = 5, puas = 4 dan 3, tidak puas = 2 dan 1. Sikap terhadap pusat belanja (*shopping center*), dimana 1 sangat tidak senang sampai dengan 7 = sangat senang, dikelompokkan menjadi 3, tidak senang = 1, 2, 3; senang = 4 dan sangat senang = 5, 6, 7.

Penentuan variabel bebas (prediktor) harus didasarkan pada teori atau hasil penelitian sebelumnya, akan tetapi untuk riset *exploratory* (riset tahap penjajagan), pengalaman peneliti diperlukan sebagai pegangan untuk memilih variabel tak bebas untuk dimasukkan dalam fungsi diskriminan.

Sampel responden kemudian dibagi menjadi dua bagian. Bagian sampel yang pertama disebut estimasi atau sampel analisis (*analysis sample*) dipergunakan untuk membuat estimasi (perkiraan) nilai koefisien fungsi diskriminan. bagian yang kedua disebut *holdout* atau sampel validasi (*validation sample*), dipergunakan untuk melakukan validasi (membuat valid atau sasih), fungsi diskriminan. Kalau sampel cukup besar bisa dibagi menjadi dua, bagian pertama untuk keperluan

estimasi (*estimation*) dan bagian kedua untuk validasi (*validation*). Apabila peran bagian pertama kemudian ditukar dengan peran bagian kedua, analisis diulangi, yang dipergunakan untuk estimasi kemudian untuk validasi, ini yang disebut *double cross validation*.

Seringkali terjadi, distribusi banyaknya objek atau kasus didalam sampel untuk estimasi dan validasi mengikuti distribusi, di dalam seluruh sampel (sebelum dibagi dua). Sebagai contoh, kalau seluruh sampel sebanyak 50% pelanggan loyal dan 50% pelanggan tak loyal, maka yang terdapat dalam sampel untuk estimasi dan validasi juga sama, masing-masing 50% loyal dan tak loyal. Perlu ditekankan bahwa validasi untuk koefisien diskriminan harus dilakukan berkali-kali, yang jelas tidak cukup hanya sekali. Banyaknya objek (kasus) dalam masing-masing bagian sampel tidak harus sama, bisa misalnya 25% dan 75% atau 40% dan 60%.

Sebagai contoh, analisis diskriminan dua kelompok (dua kategori) yang menjadi objek (elemen) penelitian ialah rumah tangga yang mengunjungi suatu tempat rekreasi terkenal. Rumah tangga dibagi dua, yaitu kelompok I yang dua tahun terakhir mengunjungi tempat rekreasi dengan simbol/lambang 1 dan yang tidak 2. Semuanya ada 42 rumah tangga sebagai sampel dibagi menjadi dua sampel yaitu sampel estimasi sebanyak 30 rumah tangga dan sampel validasi 12 rumah tangga.

Sampel estimasi 30 rumah tangga di mana ada 15 yang berkode/berlambang 1 dan 15 rumah tangga berkode 2. Sama halnya dengan sampel validasi ada 12 rumah tangga di mana ada 6 berkode 1 dan ada 6 berkode 2. Lihat Tabel 1 dan 2.

**Tabel 1**  
**Informasi Kunjungan Ke Tempat Rekreasi:**  
**Sampel untuk Estimasi**

Number	Resort Visit	Annual Family Income (\$ 000)	Attitude Toward Travel	Importance Attached to Family Vacation	Household Size	Age of Head of Household	Amount Spent on Family Vacation
1	1	50.2	5	8	3	43	M(2)
2	1	70.3	6	7	4	61	H(3)
3	1	62.9	7	5	6	52	H(3)
4	1	48.5	7	5	5	36	L(1)
5	1	52.7	6	6	4	55	H(3)
6	1	75.0	8	7	5	68	H(3)
7	1	46.2	5	3	3	62	M(2)
8	1	57.0	2	4	6	51	M(2)
9	1	64.1	7	5	4	57	H(3)
10	1	68.1	7	6	5	45	H(3)
11	1	73.4	6	7	5	44	H(3)
12	1	71.9	5	8	4	64	H(3)
13	1	56.2	1	8	6	54	M(2)
14	1	49.3	4	2	3	56	H(3)
15	1	62.0	5	6	2	58	H(3)
16	2	32.1	5	4	3	58	H(3)
17	2	36.2	4	3	2	55	L(1)
18	2	43.2	2	5	2	57	M(2)
19	2	50.4	5	2	4	37	M(2)
20	2	44.1	6	6	3	42	M(2)
21	2	38.3	6	6	3	45	L(1)
22	2	55.0	1	2	2	57	M(2)
23	2	46.1	3	5	3	51	L(1)
24	2	35.0	6	4	5	64	L(1)
25	2	37.3	2	7	4	54	L(1)
26	2	41.8	5	1	3	56	M(2)
27	2	57.0	8	3	2	36	M(2)
28	2	33.4	6	8	2	50	L(1)
29	2	37.5	3	2	3	48	L(1)
30	2	41.3	3	3	2	42	L(1)

**Tabel 2**  
**Informasi Kunjungan Ke Tempat Rekreasi:**  
**Sampel untuk Validasi**

Number	Resort Visit	Annual Family Income (\$ 000 )	Attitude Toward Travel	Importance Attached to Family Vacation	Household Size	Age of Head of Household	Amount Spent on Family Vacation
1	1	50.8	4	7	3	45	M(2)
2	1	63.6	7	4	7	55	H(3)
3	1	54.0	6	7	4	58	M(2)
4	1	45.0	5	4	3	60	M(2)
5	1	68.1	6	6	6	46	H(3)
6	1	62.1	5	6	3	56	H(3)
7	2	35.0	4	3	4	54	L(1)
8	2	49.6	5	3	5	39	L(1)
9	2	39.4	6	5	3	44	H(3)
10	2	37.0	2	6	5	51	L(1)
11	2	54.5	7	3	3	37	M(2)
12	2	38.2	2	2	3	49	L(1)

Karakteristik /atribut rumah tangga sebagai variabel bebas (prediktor) ialah penghasilan/pendapatan tahunan keluarga (dalam smu = satuan mata uang), sikap terhadap *travel* dengan skala 9 butir di mana 1 sangat tidak senang/suka sampai dengan 9 sangat senang/suka terhadap *travel*, pentingnya vakasi/liburan bagi keluarga (dimana 1 sangat tidak penting dan 9 sangat penting), banyaknya anggota rumah tangga (beberapa orang) dan umur kepala rumah tangga (tahun).

#### **Mengestimasi Koefisien Fungsi Diskriminan**

Jika data sampel untuk estimasi sudah tersedia, maka dengan menggunakan program komputer seperti SPSS 10, koefisien fungsi diskriminan bisa dihitung. Ada dua pendekatan yaitu : *direct method* dan *Stepwise discriminant analysis*. *Direct method* meliputi estimasi koefisien fungsi deskriminan dimana seluruh variabel bebas (*predictor*) terlibat, maksudnya semua dimasukkan dalam analisis, secara simultan bersama-sama. Semua variabel diikutsertakan dalam analisis, tanpa memperhatikan *discriminating power*. Pendekatan ini tepat sekali, kalau berdasarkan penelitian (riset) sebelumnya atau model teoritis yang diketahui

peneliti menghendaki semua variabel harus diikutsertakan, walaupun ada kemungkinan *discriminating power*-nya lemah. Metode yang kedua yaitu *Stepwise discriminant analysis*, variabel bebas/prediktor diikutsertakan secara berurutan (*sequentially*), didasarkan pada kemampuannya untuk mendiskriminasi antar-kelompok. Pendekatan ini tepat sekali kalau memang penelitian akan melakukan seleksi variabel –variabel yang mana saja yang *discriminating power*-nya memang tinggi. Program komputer akan membantu melakukan perhitungan untuk menyeleksi. Dengan menggunakan program komputer, analisis diskriminan dua kelompok atau dua kategori berdasarkan data Tabel 1. dan 2. menghasilkan pengolahn data seperti terlihat pada Tabel 3.

**Tabel 3**  
**Hasil Analisis Diskriminan Dua Kelompok**

Group Means					
<i>Visit</i>	<i>Income</i>	<i>Travel</i>	<i>Vacation</i>	<i>H Size</i>	<i>Age</i>
1	60.52000	5.40000	5.80000	4.33333	53.73333
2	41.91333	4.33333	4.06667	2.80000	50.13333
Total	51.21667	4.86667	4.93333	3.56667	51.93333
Group Standar Deviations					
1	9.83065	1.91982	1.82052	1.23443	8.77062
2	7.55115	1.95180	2.05171	.94112	8.27101
Total	12.79523	1.97804	2.09981	1.33089	8.57395
Pooled With-inGroups Correlation Matrix					
	<i>Income</i>	<i>Travel</i>	<i>Vacation</i>	<i>H Size</i>	<i>Age</i>
Income	1.00000				
Travel	.19745	1.00000			
Vacation	.9148	.08434	1.00000		
H size	.08887	-.01681	.07046	1.00000	
Age	-.01431	-.19709	.01742	-.4301	1.00000
<i>Wilk's (U statistic) and univariate F ratio with 1 and 28 degrees of Freedom</i>					
<i>Variable</i>	<i>Wilk's</i>	<i>F</i>	<i>Significance</i>		
Income	.45310	33.80	.0000		
Travel	.92479	2.277	.1425		
Vacation	.82377	5.990	.0209		
H size	.65672	14.64	.0007		

Age	.95441	1.338	.2572
-----	--------	-------	-------

*Canonical Discriminant Functions*  
*Percent of Cumulative Canonical After Wilk's*

Function	Eigenvalue	Variance Percent	Correlation	Function $\lambda$	Chi-square	df	Significance
1*	1.7862	100.00	100.00	.8007	0.3589	26.130	5 .0001

---

\* marks the I canonical discriminant functions remaining in the analysis

*Standard Canonical Discriminant Function Coefficients*

	Func 1
Income	.74301
Travel	.09611
Vacation	.23329
H size	.46911
Age	.20922

*Structure Matrix:*

Pooled Within-group correlations between discriminating variables and canonical discriminant functions (variables ordered by size of correlation within function).

	Func 1
Income	.82202
Hsize	.54096
Vacation	.34607
Travel	.21337
Age	.16354

*Unstandardized Canonical Discriminant Function Coefficients*

	Func 1
Income	.8476710E-01
Travel	.4964455E-01
Vacation	.1202813
H size	.4273893
Age	.245438E-01
(Constant)	-7.975476

*Canonical Discriminant Functions Evaluation at Group Means (Group Centroids)*

Group	Func 1
1	1.29118

2 -1.29118

*Classification Results for Cases Selected for Use in Analysis*

		<i>Predicted</i>		<i>Group</i>
<i>Membership</i>				
	<i>Actual Group</i>	<i>Number of cases</i>		
Group	1	15	12 80.0%	3
	2	15	0 .0%	15
100.0%				
Percent of grouped cases correctly classified : 90.00%				

*Classification Results for Cases Not Selected for Use in Analysis (Holdout Sample)*

		<i>Predicted</i>		<i>Group</i>
<i>Membership</i>				
	<i>Actual Group</i>	<i>Number of cases</i>		
Group	1	6	4 66.7%	2
	2	6	0 .0%	6
100.0%				
Percent of grouped cases correctly classified : 83.33%				

Koefisien fungsi diskriminan yang tidak dibakukan (*unstandardized*) juga disajikan. *Group centroids* atau rata-rata kelompok, ternyata mempunyai nilai yang sama besarnya dengan tanda yang berbeda  $D_a = 1,29$  dan  $D_b = -1,29$  di mana  $D_A = 1$  dan  $D_b = 2$ . oleh karena  $N_a = N_b = 15$ . Maka nilai pemisah (*cutting point*)  $D_o = (D_a + D_b)/2 = [1,29 + (-1,29)]/2 = 0$ , artinya kalau seorang

turis (responden) mempunyai skor atau nilai diskriminan, katakanlah  $D_i$  untuk responden ke  $i$ ,  $< 0$  (negative) masuk kelompok A atau kelompok 1 dan kalau  $D_i > 0$  (positif) masuk kelompok b atau kelompok 2.

Koefisien fungsi diskriminan yang baku (standar) semua positif, sebagai berikut: penghasilan = 0,74; travel = 0,10; liburan = 0,23; art (anggota rumah tangga) = 0,47 dan umur = 0,21. Ternyata yang pengaruhnya signifikan  $< \alpha = 0,05$ , lihat Tabel 4.3 ialah penghasilan, liburan dan art, dimana masing-masing nilai signifikannya: 0,000; 0,021, dan 0,001 semuanya  $< \alpha = 0,05$ . Variabel baku mempunyai rata-rata nol dan standar deviasi atau simpangan baku satu. Fungsi diskriminan yang baku, sebagai berikut.

$D = 0,74X_1 + 0,10X_2 + 0,23X_3 + 0,47X_4 + 0,21X_5$  dimana  $X_1$  = penghasilan,  $X_2$  = Travel,  $X_3$  = liburan (vakasi),  $X_4$  = anggota rumah tangga dan  $X_5$  = umur.

Semua koefisien fungsi diskriminan yang baku (standar) bertanda positif ini artinya semua variabel bebas sebanyak 5 buah yang tercantum dalam fungsi diskriminan  $D$ , mempunyai pengaruh yang positif, setiap mengalami kenaikan akan menaikkan *discriminant power*-nya. Oleh karena sudah dibuat baku, maka bisa disimpulkan bahwa semakin besar nilai koefisien dari suatu variabel bebas, semakin besar *discriminant power* dari variabel yang bersangkutan. Di dalam prakteknya peneliti akan menggunakan variabel yang pengaruhnya signifikan. Ternyata setelah dilakukan pengujian hipotesis hanya ada 3 dari 5 variabel bebas tersebut yang pengaruhnya memang signifikan, yaitu:

$X_1$  = Penghasilan,  $X_2$  = liburan (vakansi), dan  $X_4$  = anggota rumah tangga, jadi :

$D = 0,74$  (penghasilan) +  $0,23$  (liburan) +  $0,47$  (anggota rumah tangga), sudah cukup untuk melakukan pengelompokan.

### **Menilai Validitas Analisis Diskriminan**

Seperti telah dijelaskan sebelumnya, secara acak data dibagi menjadi dua sub-mapel, yaitu sampel untuk mengestimasi disebut *analysis sample* dan yang kedua untuk membuat validasi *validation sample* disebut juga *holdout sample*. Analisis sampel untuk mengestimasi koefisien fungsi diskriminan sedangkan validasi sampel digunakan untuk mengembangkan matrik klasifikasi.

Timbangan diskriminan, diestimasi dengan menggunakan analisis sampel dikalikan dengan nilai variabel prediktor didalam *holdout sample* untuk menghasilkan skor(nilai) diskriminan. Elemen sebagai objek penelitian (responden, seperti pelanggan) kemudian diputuskan untuk dimasukkan kedalam kelompok (group) tertentu berdasarkan skor(nilai) diskriminan dan suatu aturan keputusan yang tepat. Sebagai contoh, di dalam analisis diskriminan dua kelompok, suatu objek akan dimasukkan kedalam kelompok tertentu yang *centroid*-nya terdekat. *The hit ratio* atau persentase objek/kasus yang secara benar atau tepat diklasifikasikan. Kalau memang masuk A dimasukkan ke A, kalau memang masuk kelompok B dimasukkan ke B. Hit ratio dihitung dengan jalan menjumlahsluruh elemen yang berada pada diagonal pokok (utama) matrik klasifikasi dibagi dengan banyaknya objek/elemen.

Perlu sekali membandingkan persentase objek yang benar/tepat diklasifikasi berdasar analisis diskriminan, dengan persentase yang diperoleh secara kebetulan (by chance). Apabila kelompok mempunyai objek yang sama banyak, persentase klasifikasi yang tepat karena kebetulan ialah angka 1 dibagi dengan banyaknya kelompok. Seberapa jauh perbaikan bisa dicapai melalui cara kebetulan? Sebetulnya tidak ada petunjuk umum yang tersedia, meskipun beberapa ahli riset menyarankan bahwa akurasi klasifikasi tercapai dengan analisis diskriminan, kalau bisa 25% lebih besar daripada dengan cara kebetulan.

Kebanyakan program analisis diskriminan juga mengestimasi suatu metrik klasifikasi berdasarkan pada analisis sampel, yang pada umumnya memberikan hasil klasifikasi yang lebih apabila dibandingkan dengan metrik klasifikasi yang didasarkan pada *holdout sample*.

Tabel 3, mengenai contoh *vacation resort*, juga menunjukkan hasil klasifikasi berdasarkan pada *analysis sample*. *The hit ratio* yaitu persentase objek/kasus yang secara tepat diklasifikasi sebesar  $(12 + 15)/30 = 0,90$  atau 90%. Seseorang mungkin curiga, bahwa hit ratio digelembungkan, karena data yang dipergunakan untuk keperluan estimasi, juga dipergunakan untuk validasi membentuk analisis klasifikasi pada set data holdout yang bebas satu sama lain, menghasilkan metrik klasifikasi dengan sedikit lebih rendah hit ratio-nya =  $(4 + 6)/12 = 0,833$  atau 83,3% (lihat tabel 4.3). Dengan dua kelompok yang sama jumlah objeknya, secara

kebetulan kita akan mengharapkan hit ratio =  $\frac{1}{2} = 0,50$  atau 50 %. Jadi perbaikan yang dicapai melebihi 25% dibandingkan dengan cara kebetulan ( $83,3\% - 50\% = 33,3\% > 25\%$ ), sehingga validasi analisis diskriminan, dianggap memuaskan.

### **Contoh kasus lain pada analisis diskriminan 2 kelompok .**

Analisis diskriminan dua kelompok dipergunakan untuk mengakses atau menilai kekuatan setiap dimensi dari 5 dimensi untuk mengklasifikasi individu sebagai pengguna dan bukan pengguna TV. Prosedur tepat untuk pengguna ini, sebab sifat dari kelompok atau kategori yang sudah ditentukan sebelumnya dan skala interval dipergunakan untuk menghasilkan skor faktor setiap individu/objek.

Dua kelompok konsumen, masing-masing dengan jumlah responden yang sama besarnya yaitu 185 orang ,disebut sebagai kelompok engguna dan bukan pengguna TV. Persamaan diskriminasi untuk analisis diperkirakan dengan menggunakan sub-mapel sebanyak 370 orang. Dari sisa responden sebanyak 198 orang dipergunakan sebagai a validation sub-mapel in a cross – validation of the equation sebanyak 30 orang responden tidak diikutsertakan, sebab missing discriminant value.

Nilai canonical correlation untuk fungsi diskriminan sebesar 0,4291, signifikan pada  $P < 0,0001$   $\alpha = 1\%$ ,  $\alpha = 5\%$ . Jadi tetap signifikan pada nilai  $\alpha$  sebesar 1% dan 5%, sebab signifikan pada  $\alpha = 5\%$  Belem tentu signifikan pada  $\alpha$  sebesar 1%. Nilai eigen value sebesar = 0,2257.

Tabel menunjukkan ringkasan koefisien canonical diskriminan yang telah dibakukan. Bagian terbesar varian dijelaskan oleh fungsi diskriminan, sebagai tambahan, seperti yang ditunjukkan oleh tabel the home orientation dimension memberikan kontribusi/sumbangan yang cukup besar didalam mengklasifikasikan individu/objek sebagai pengguna(users) dan bukan pengguna (non users) dari TV. Moral (morale), keamanan dan kesehatan(security and health) dan hormat (respect) juga memberikan kontribusi/sumbangan yang signifikan. Faktor sosial kelihatannya kecil sumbangannya.

Prosedur cross validation denagn menggunakan fungsi diskriminasi dari analisis sample memberikan dukungan kepada pendirian bahwa dimensi membantu peneliti dalam mendiskiminasikan antara pengguna dan bkan pengguan TV. Seperti tabel berikut menunjukkan fungsi diskriminan cukup berhasil di dalam

mengklasifikasi 75,76% dari objek/responden. Hasil analisis diskriminan ini menyarankan bahwa pertimbangan dimensi yang diidentifikasi akan membantu pemasaran memahami pasar yang dihadapi.

### Ringkasan Analisis Diskriminan (Hasil Print Out Computer)

#### *Standard Canonical Discriminant Function Coefficients*

Morale	.27798
Security and health	.39850
Home orientation	.77496
Respect	.32069
Social	.01996

#### *Classification Results for Cases Selected for Use in Analysis*

		<i>Predicted</i>		<i>Group</i>
<i>Membership</i>		Nonusers	Users	
<i>Actual Group</i>	<i>Number of cases</i>			
TV nonusers	77	56 72,7%	21 27,3%	
TV users	65	36,9%	63,1%	

Percent of grouped cases correctly classified : 63.31%

#### Classification Results for Cases Not Selected for Use in Analysis (Holdout Sample)

		<i>Predicted</i>		<i>Group</i>
<i>Membership</i>		Nonusers	Users	
<i>Actual Group</i>	<i>Number of cases</i>			
TV nonusers	108	85 78,7%	23 21,3%	
TV users	90	27,8%	72,2%	

Percent of grouped cases correctly classified : 75,76%.

### DAFTAR PUSTAKA

- Cox, D.R. dan Snell E.J. (1996), "*Analysis of Binary Data*", Chapman and Hall, London.
- Efron, B. (1975), "*The Efficiency of Logistic Regression Compared to Normal Discriminant Analysis*", Journal of the American Statistical Association, 70;892-898.
- Delury, G.E. (1973), *The 1973 World Almanac and Book of Facts*, New York.
- Farida, L. (2000), "*Analisis Statistik terhadap faktor Resiko dan Kualitas Hidup Penderita Stroke Berdasarkan Skala Neurologik Kanadian di RSUD Dr. Soetomo Surabaya*", Tugas Akhir, Statistika ITS.
- Hosmer, D.W. dan Lemeshow (1989), "*Applied Logistic Regression*," John Wiley, New York,
- Johnson, R.A. dan Winchern,D.W. (1992), "*Applied Multivariate Statistical Analysis*", Prentice Hall, New Jersey.
- Krzanowski, W.J. (1975), "*Discrimination and Classification using Both Binary and Continuous Variable*", Journal of the American Statistical Association, 70;782-352.
- Press, S.J. dan Wilson, S. (1978), "*Choosing Between Logistic Regression and Discrimination Analysis*", Journal of the American Statistical Association, 73;699-364.
- Santosa, S .(2003), *SPSS, Statistik Multivariate*, Elek Media Komputindo, Jakarta
- Sharma, S. (1996), "*Applied Multivariate Techniques*", John Wiley , New York.
- Tatham, R.L., Hair, J.F, Anderson, R.E., dan Black, W.C., (1998), "*Multivariate Data Analysis*", Prentice Hall, New Jersey.
- Yulia I.P.D. (1997)," *Analisis Statistika Terhadap Kerusakan Wall Tile di PD Sarana Bangunan Unit Pabrik Keramik dan Pengolahan Bahan Tulungagung*", Tugas Akhir, Statistika ITS.