

---

**PERBANDINGAN HASIL KLASIFIKASI ANALISIS DISKRIMINAN DAN  
REGRESI LOGISTIK RESPON BINER PADA PENGKLASIFIKASIAN  
KEBERHASILAN MAHASISWA POLITEKNIK  
(STUDI KASUS MAHASISWA TATA NIAGA POLBAN)**

Euis Sartika  
Staf Pengajar Jurusan Tata Niaga Politeknik Negeri Bandung  
E-mail:

**Abstrak**

Analisis diskriminan dan Regresi Logistik adalah analisis multivariat yang bertujuan untuk mengklasifikasikan suatu pengamatan ke dalam salah satu dari kelompok tertentu. Regresi logistik respon biner dalam penerapannya tidak memerlukan asumsi kenormalan atas variabel bebas. Variabel bebas dalam Regresi Logistik Biner bisa kontinu, diskrit dan dikotomis. Sedangkan Analisis Diskriminan memerlukan asumsi multivariat normal dan kesamaan struktur matriks varian kovarian antar grup, tetapi kedua asumsi ini tidak selalu dapat dipenuhi. Tujuan penelitian ini adalah memperoleh hasil klasifikasi analisis diskriminan dan regresi logistik respon biner serta membandingkan hasil klasifikasi tersebut dalam kasus keberhasilan mahasiswa Politeknik Negeri Bandung. Hasil penelitian ini menunjukkan apabila variabel penjelas berskala kategori, regresi logistik respon biner menghasilkan ketepatan klasifikasi yang lebih tinggi dibandingkan dengan analisis diskriminan. Peubah penjelas yang berpengaruh signifikan pada keberhasilan mahasiswa Polban bidang Tata Niaga adalah jenis kelamin, program studi, dan nilai IP semester satu.

Kata kunci : Analisis Diskriminan, Regresi Logistik Respon Biner.

**Abstract**

*Discriminant analysis and Logistic Regression are Multivariate analysis that aim to classification a research to one of the groupe. Logistic Regression binary respond in application didn't need normaly asumption for independent variable. Independent variable in Binary Logistic Regression can be continue, discrit and dikotomis. Hence, Discriminant Analysis need normaly multivariate asumption and the same of variant covariant matrices structure inside groupe. The aim of this research was to get comparing Discriminant analysis classification and Binary Respon Logistic Regression in succesive studying of student of Politeknik Negeri Bandung. The results showed that if the scale of independent variable was category, the Binary Logistic Regression was more acurate than the results of Discriminant Analysis. The explanatory variables that had a significant effect on the Polban student achievement were gender, study program, and the first semester GPA.*

*Key words: Discriminant Analysis, Binary Respon Logistic Regression*

## Pendahuluan

Analisis diskriminan adalah analisis multivariat yang digunakan untuk memperoleh model hubungan antara variabel respon yang bersifat dikotomis atau multikotomis maupun kualitatif dengan variabel penjelas yang bersifat kuantitatif (Hair *et al.*, 1998). Tujuan dari analisis diskriminan ini adalah untuk menghasilkan suatu aturan yang menggambarkan pemisahan beberapa kelompok pengamatan, serta mengklasifikasikan suatu pengamatan baru ke dalam salah satu dari kelompok yang telah diketahui. Dalam penerapannya, fungsi diskriminan ini mengasumsikan bahwa data pada masing-masing kelompok berdistribusi normal multivariat dan mempunyai struktur matriks varian kovarian yang sama (Hair *et al.*, 1998). Sedangkan Regresi logistik respon biner adalah bentuk khusus analisis regresi dengan respon biner dan variabel penjelas yang dapat terdiri dari data kontinu, kategori atau campuran antara keduanya. Persamaan regresi logistik respon biner ini tidak menghasilkan nilai pada variabel respon, namun menghasilkan nilai peluang kejadian pada variabel respon. Nilai peluang ini yang dipakai sebagai ukuran untuk mengklasifikasikan pengamatan. Regresi logistik respon biner tidak memerlukan asumsi distribusi multivariat normal dan kesamaan varian kovarian dalam masing-masing grup (Hosmer & Lemeshow dalam Wibowo, 2002). Sehingga metode ini cukup tahan (*robust*) untuk dapat diterapkan dalam berbagai skala/keadaan data (Hair *et al.*, 1998).

Walaupun kedua metode ini berbeda dalam asumsi atau metode estimasi serta inferensi parameternya. Tetapi hasil perbandingan klasifikasi antara analisis diskriminan dan regresi logistik respon biner dapat diketahui pada situasi bagaimana kedua metode tersebut lebih tepat digunakan. Salah satu contoh penerapannya adalah untuk melihat ketepatan klasifikasi keberhasilan mahasiswa Politeknik Negeri Bandung terhadap factor-faktor yang. Variabel respon dalam penelitian ini adalah keberhasilan mahasiswa Politeknik yang berupa data kategori (dua kategori atau biner) yakni berhasil atau tidak berhasil. Variabel penjelas adalah jenis kelamin, nilai NEM (Nilai Ebtanas Murni), nilai seleksi ujian Masuk Polban (Politeknik Negeri Bandung), asal daerah, nilai mata kuliah Matematika Bisnis, dan nilai IP (Indeks Prestasi) semester I. Penelitian dengan dua metode ini telah dilakukan oleh Wibowo (2002), dimana apabila variabel penjelas bersifat campuran antara kontinu dan kategori dalam studi kasus data kerusakan *wall tile*, analisis

diskriminan sebaiknya digunakan karena akan menghasilkan ketepatan yang lebih tinggi dibanding regresi logistik respon biner jika tidak ada pelanggaran asumsi. Apabila variabel penjelas bersifat kategori seperti contoh dalam studi kasus data faktor resiko dan kualitas penderita penyakit stroke, regresi logistik respon biner sebaiknya dipergunakan karena akan menghasilkan ketepatan klasifikasi yang lebih baik dibandingkan analisis diskriminan dan apabila variabel prediktor bersifat kontinu dalam studi kasus data demografi, analisis diskriminan sebaiknya dipergunakan karena akan menghasilkan ketepatan klasifikasi yang lebih tinggi dibanding regresi logistik respon biner meskipun ada pelanggaran asumsi, dengan catatan tidak banyak data yang outlier. Hasil penelitian lain, Kurniati E (2007) menunjukkan bahwa apabila variabel penjelas berskala kategori dalam kasus kejadian bayi dengan BBLR, regresi logistik respon biner menghasilkan ketepatan klasifikasi yang lebih tinggi dibandingkan dengan analisis diskriminan. Sedangkan faktor-faktor yang mempengaruhi keberhasilan mahasiswa Polban telah dilakukan oleh peneliti sebelumnya, antara lain : penelitian Binarto A (2010) menunjukkan bahwa salah satu faktor yang mempengaruhi keberhasilan mahasiswa Polban adalah nilai IP semester I. Metode yang digunakan adalah Regresi Logistik Biner. Untuk mendapatkan hasil yang lebih akurat, maka dalam penelitian ini digunakan dua metode yakni Regresi Logistik Biner dan Analisis Diskriminan dan membandingkan manakah yang lebih tepat.

## Perumusan Masalah

Permasalahan dalam penelitian ini adalah bagaimana hasil klasifikasi keberhasilan mahasiswa Politeknik bidang Tata Niaga Polban dengan menggunakan analisis diskriminan dan regresi logistik respon biner, berdasarkan peubah-peubah yang digunakan.

## Tujuan Penelitian

1. Membandingkan hasil klasifikasi keberhasilan mahasiswa Politeknik bidang Tata Niaga dengan menggunakan analisis diskriminan dan regresi logistik respon biner berdasarkan karakteristik peubah-peubah yang digunakan, baik peubah respon maupun peubah penjelas.
2. Mengetahui faktor-faktor signifikan yang mempengaruhi keberhasilan mahasiswa Polban bidang Tata Niaga.

## Tinjauan Pustaka

Faktor-faktor yang Mempengaruhi Hasil Belajar Mahasiswa:

Pendidikan Politeknik merupakan jalur pendidikan profesional yang membekali lulusannya dengan

keterampilan, pengetahuan teori dan sikap disiplin yang tinggi. Politeknik membekali mahasiswanya dengan kemampuan *hardskill* juga kemampuan *softskill*. Diharapkan alumni Politeknik menjadi tenaga profesional yang siap pakai dan kompeten di bidangnya. Pada proses belajar di politeknik, mahasiswa memperoleh 60% praktik dan 40% teori. Kondisi demikian membutuhkan adaptasi dan motivasi yang sangat tinggi dari mahasiswa agar dapat bertahan pada tiap semesternya sehingga mereka berhasil menyelesaikan pendidikan di politeknik tepat waktu. Salah satu yang menjadi tolok ukur keberhasilan studi mahasiswa adalah nilai indeks prestasi (IP), begitu juga di Politeknik Negeri Bandung (Polban). Nilai IP tiap semester menjadi salah satu faktor yang menentukan apakah mahasiswa tersebut dapat melanjutkan ke semester berikutnya atau tidak (*drop out*). Dalam aturan di Polban, derajat keberhasilan studi mahasiswa Polban dinyatakan dengan Indeks Prestasi Kumulatif (*IPK*). Seorang mahasiswa Polban dinyatakan lulus penuh pada suatu semester bila mempunyai nilai  $IP \geq 2.00$  dan jumlah mata kuliah dengan nilai D maksimum 7 SKS dan tanpa nilai E. Mahasiswa dinyatakan lulus percobaan pada suatu semester jika nilai  $IP \geq 2.00$  dan nilai  $D > 7$  SKS, tanpa nilai E atau  $1.70 \leq IP \leq 2.00$  dan nilai  $D \leq 7$  SKS tanpa nilai E. Mahasiswa akan dikeluarkan dari Politeknik (DO) dengan alasan akademik bila dua semester berturut-turut lulus percobaan sampai dengan semester IV, mempunyai nilai  $1.70 \leq IP < 2.00$  dan jumlah SKS nilai  $D > 7$  SKS, nilai IP di bawah 1.70, mempunyai nilai E pada semester I, II, III dan IV, melewati batas studi yang telah ditetapkan yakni sekurang-kurangnya enam semester dan selamalamanya 10 semester termasuk cuti dan mengulang, dan mempunyai status ketidakhadiran yang tidak diizinkan, yakni tidak hadir tanpa izin selama 38 jam dan tidak memenuhi syarat kelulusan pada semester VI setelah diberi kesempatan mengulang. Secara garis besar, faktor-faktor yang memengaruhi keberhasilan studi mahasiswa di perguruan tinggi dibagi menjadi dua faktor yakni faktor intelektual dan faktor nonintelektual (Munthe 1995 diacu Hidayati 2002). Faktor intelektual adalah kemampuan seseorang yang diperlihatkan melalui kecerdasan dan kepandaianya dalam berpikir dan berbuat. Faktor ini meliputi: bakat, kapasitas belajar, kecerdasan dan hasil belajar yang telah dicapai. Faktor non-intelektual adalah segala kondisi dari dalam dan dari luar dirinya atau lingkungan sekitar yang terkait dengan diri seseorang dalam

memengaruhi kemampuan berfikir dan bertindak. Menurut Suryabrata (1989), hal-hal yang mempengaruhi proses belajar mengajar meliputi pengaruh dari dalam yaitu keadaan psikologis (kesehatan, kondisi pancaindra, dan gizi yang cukup) dan keadaan fisiologis (minat, kecerdasan, motivasi, dan kemampuan kognitif). Selanjutnya, pengaruh dari luar meliputi input lingkungan dan input instrumental.

### Analisis Diskriminan

Analisis diskriminan adalah analisis multivariat yang diterapkan untuk memodelkan hubungan antara satu variabel respon baik yang bersifat dikotomis atau multikotomis maupun kualitatif dengan variabel penjelas (*predictor*) yang bersifat kuantitatif (Hair *et al.*, 1998). Tujuan dari analisis diskriminan ini adalah untuk menghasilkan suatu aturan yang menggambarkan pemisahan beberapa kelompok (grup) pengamatan, serta mengklasifikasikan suatu pengamatan baru ke dalam salah satu dari kelompok (grup) yang telah diketahui. Jika  $\mathbf{x}$  adalah pengamatan baru yang tidak diketahui asalnya, maka dapat dihitung skor diskriminan linier yaitu:

$$W_{ij} = \mathbf{x}' S^{-1}_{pl} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j) + \frac{1}{2} (\bar{\mathbf{x}}_i + \bar{\mathbf{x}}_j) S^{-1}_{pl} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j) \dots (1)$$

Sehingga kriteria klasifikasi berdasarkan statistik  $W$  (*Wald-Anderson*) adalah:

Masukkan  $\mathbf{x}$  ke grup  $i$  jika  $W_{ij} > 0$  untuk semua  $j \neq i$ , dengan catatan  $W_{ij} = -W_{ji}$  (Morrison, 1978).

Keterangan:

$\mathbf{x}_i$  = vektor rata-rata sampel grup ke- $i$

$\mathbf{x}_j$  = vektor rata-rata sampel grup ke- $j$

$W_{ij}$  = Fungsi diskriminan yang akan menggolongkan individu ke grup  $i$  atau grup  $j$

Menurut Gazpert (1992), jika hanya ada dua grup maka kriteria klasifikasi berdasarkan

statistik  $W$  dengan tahapan sebagai berikut :

1. Golongkan  $\mathbf{x}$  ke dalam grup 1 jika  $W_{12} > 0$

2. Golongkan  $\mathbf{x}$  ke dalam grup 2 jika  $W_{12} \leq 0$

Salah satu cara untuk menghindari bias adalah membagi sampel menjadi dua bagian yaitu *training sample* yang digunakan untuk membentuk aturan klasifikasi dan *validation sample* yang digunakan untuk mengevaluasi fungsi klasifikasi (Rhencher, 1995). Hair *et al.* (1998) menyatakan, tidak ada ketentuan yang pasti dalam membagi sampel menjadi *training sample* dan *validation sample*. Proporsi pembagian sampel ini tidak harus sama besar untuk masing-masing bagian misalnya 60% dan 40%, 75% dan 25% atau yang lainnya. Wibowo (2002) membuat proporsi pembagian sampel sebesar 80% dan 20%, Suhaeni (2005) dalam penelitiannya mengelompokkan sampel sebesar 60% dan 40%. Proses validasi terhadap fungsi diskriminan yang terbentuk di *training*

sample harus dilakukan lebih dari satu kali, atau berkali-kali (Supranto, 2004).

**Regresi Logistik Respon Biner**

Analisis Regresi Logistik digunakan untuk memodelkan hubungan antara peubah respon yang dengan jenis peubah kualitatif, yaitu peubah berskala nominal atau ordinal dengan peubah-peubah penjelas (*predictor*) yang bisa terdiri dari peubah kualitatif maupun kuantitatif. Peubah respon dalam regresi logistik dapat berupa peubah dikhotom (*biner*) maupun poltom (*ordinal atau nominal*). Misalkan data hasil pengamatan mempunyai n peubah penjelas yang dinyatakan oleh vektor  $x' = (X_1, X_2, \dots, X_n)$  yang berpasangan dengan peubah respon Y yang bernilai 1 dan 0. Nilai peubah Y=1 menyatakan bahwa respon memiliki kriteria yang diharapkan (Mahasiswa Berhasil) dan Y=0 tidak memiliki kriteria yang diharapkan (Mahasiswa tidak Berhasil), maka peubah respon Y memiliki sebaran Bernoulli dengan parameter  $\pi(x)$  dan fungsi sebaran peluangnya adalah:

$$f(y_i) = \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} \dots\dots\dots(2)$$

Model Regresi Logistik antara  $\pi(x)$  dengan x adalah:

$$\pi(x) = \frac{\exp(g(x))}{1 + \exp(g(x))} \dots\dots\dots (3)$$

Fungsi regresi pada persamaan (3) berbentuk *curvilinear*. Dengan menggunakan transformasi logit, *curvilinear* tersebut akan menjadi fungsi linear. (Agresti 1996).

Model Logit mentransformasi masalah prediksi peluang dengan range (0,1) menjadi masalah prediksi log odds. Transformasi Logit dinyatakan dalam persamaan berikut : logit

$$(\pi(x)) = \ln \left[ \frac{\pi(x)}{1 - \pi(x)} \right] = g(x) \dots\dots\dots(4)$$

dimana

$$g(x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \dots\dots (5)$$

merupakan logit (Hosmer & Lemeshow 2000). Setelah diperoleh nilai dugaan  $\beta_0, \beta_1, \dots, \beta_n$ , maka dapat diperoleh penduga dari  $\pi(x)$  dengan persamaan:

$$\hat{\pi}(x) = \frac{\exp(\hat{g}(x))}{1 + \exp(\hat{g}(x))} \dots\dots\dots (6)$$

dimana  $\hat{g}(x) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_n X_n$

merupakan penduga Logit yakni fungsi linear dari peubah penjelas (Hosmer & Lemeshow 2000). Pengujian kesesuaian model secara simultan (melibatkan seluruh peubah penjelas) dilakukan dengan menggunakan uji nisbah kemungkinan (*likelihood ratio test*) atau uji-G. Uji-G digunakan untuk pengujian parameter  $\beta_i$  dengan hipotesis sebagai berikut:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \text{minimal ada satu } \beta \text{ yang tidak sama dengan nol}$$

Statistik uji yang digunakan adalah statistik G

$$G = -2 \ln \left[ \frac{\text{likelihood tanpa peubah penjelas}}{\text{likelihood dengan peubah penjelas}} \right] \dots\dots (7)$$

Statistik Uji-G ini secara teoritis mengikuti sebaran  $\chi^2$  dengan derajat bebas k. Kriteria keputusan yang diambil adalah menolak  $H_0$ , jika  $G_{hitung} > \chi^2_{\alpha(k)}$  (Hosmer & Lemeshow 2000). Apabila pada Uji G,  $H_0$  ditolak maka lanjutkan dengan Uji-Wald yang digunakan untuk menguji parameter  $\beta_i$  secara parsial. Hipotesis yang akan diuji adalah:

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0 \quad i=1,2,\dots,k$$

Statistik uji Wald adalah:

$$W = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \dots\dots\dots (8)$$

Secara teori, statistik W ini mengikuti sebaran normal baku. Dengan kriteria keputusan adalah menolak  $H_0$  jika  $|W| \geq Z_{\alpha/2}$  atau nilai  $p \leq \alpha$ . Pereduksian peubah dalam regresi logistik dikenal sebagai analisis regresi logistik bertatar (*stepwise logistic regression*), dimana langkah yang dilakukan adalah menambah dan mengurangi peubah-peubah penjelas satu demi satu dari model sampai didapatkan model dengan peubah-peubah penjelas yang mempunyai pengaruh signifikan. Sedangkan standar atau ukuran untuk melihat seberapa besar kecenderungan pengaruh peubah penjelas terhadap respon digunakan *Rasio Odds* (Hosmer & Lemeshow 2000). Interpretasi koefisien pada model regresi logistik dilakukan dengan melihat nilai rasio odds dan selang kepercayaan rasio oddsnya. Tanda positif dari koefisien pada model regresi logistik menunjukkan bahwa nilai rasio odds lebih dari satu. Begitupun sebaliknya, untuk tanda koefisien negatif, maka nilai rasio oddsnya kurang dari satu. Koefisien model logit dapat ditulis sebagai  $\beta = g(x+1) - g(x)$  yang menginterpretasikan bahwa perubahan nilai logit  $g(x)$  terjadi untuk setiap perubahan satu unit peubah penjelas X yang selanjutnya disebut log odds. Secara umum dapat dikatakan bahwa log odds adalah selisih

antara penduga logit yang dihitung pada dua nilai sembarang, misalnya  $x=a$  dan  $x=b$  dan dinotasikan sebagai berikut:

$\ln [\psi(a,b)] = g(x=a) - g(x=b) = \beta(a-b)$   
 Sedangkan penduga rasio odds dinyatakan sebagai  $\psi(a,b) = \exp[\beta(a-b)]$ . Sehingga apabila dimisalkan  $(a-b)=1$  maka diperoleh  $\psi(a,b) = \exp(\beta)$ . Interpretasinya adalah peluang untuk  $y=1$  pada  $x=1$  adalah  $\psi$  kali dibandingkan dengan  $x=0$ . Untuk lebih jelasnya, berikut disajikan tabel model logistik dengan satu peubah dikotomi.

Tabel-1.  
 Nilai-nilai Model Regresi Logistik dengan Peubah Penjelas Dikotomi

		Peubah Bebas X	
		x=1	x=0
Peubah Respon	Y=1	$\pi(1) = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}$	$\pi(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$
	Y=0	$1 - \pi(1) = \frac{1}{1 + e^{\beta_0 + \beta_1}}$	$1 - \pi(0) = \frac{1}{1 + e^{\beta_0}}$
Jumlah		1	1

Transformasi logitnya :

$$g(x) = \ln \left[ \frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad \dots(9)$$

Nilai odds pada  $y=1$  untuk  $x=1$  adalah  $\frac{\pi(1)}{1 - \pi(1)}$

dengan nilai log adalah  $g(1)$

Nilai odds pada  $y=1$  untuk  $x=0$  adalah  $\frac{\pi(0)}{1 - \pi(0)}$

dengan nilai log adalah  $g(0)$

Maka nilai rasio odds ( $\psi$ ) merupakan rasio dari odds untuk  $x=1$  dengan  $x=0$  yang dinotasikan sebagai:

$$\psi = \frac{\pi(1)/1 - \pi(1)}{\pi(0)/1 - \pi(0)} = g(1) - g(0) = \beta_1 \dots \dots \dots(10)$$

Jika  $\beta_1$  adalah selisih logit, maka  $\exp(\beta_1)$  adalah nilai rasio odds (Hosmer & Lemeshow 2000). Selang kepercayaan  $(1-\alpha)100\%$  untuk nilai rasio odds adalah:

$$\exp[\hat{\beta}_1 \pm Z_{1-\alpha/2} \hat{SE}(\hat{\beta}_1)] \dots \dots \dots (11)$$

Ukuran kebaikan model dalam regresi logistik adalah jika model mempunyai peluang salah klasifikasi minimal (Hosmer & Lemeshow 2000). Ketepatan dan kesalahan klasifikasi dapat dilihat dalam tabel klasifikasi. Tabel klasifikasi untuk peubah respon dikotom terdiri atas dua kolom nilai

dugaan dan dua baris nilai amatan. Dalam menentukan ketepatan klasifikasi (*correct classification*) terhadap amatan harus ditentukan terlebih dahulu nilai *cutpoint* (c). Nilai c yang biasa digunakan adalah 0.5. Jika peluang dugaan  $n(x) \geq c$ , maka nilai dugaan termasuk pada respon  $y=1$  dan sebaliknya adalah  $y=0$ . Tabel-2 memperlihatkan tabel klasifikasi secara umum.

Tabel-2.  
 Klasifikasi Respon

Amatan	Dugaan		Total	Ketepatan
	1	0		
1	a	B	$(a+b)=n_1$	$a/n_1$
0	c	D	$(c+d)=n_0$	$d/n_0$
Total	$(a+c)=n_1$	$(b+d)=n_0$	$(a+b+c+d)=n$	$(a+d)/n$
Kesalahan	$c/n_1$	$b/n_0$	$(b+c)/n$	

Ketepatan klasifikasi (*correct classification*) terdiri atas *specificity* dan *sensitivity*. Ketepatan klasifikasi dalam menduga kejadian bahwa respon tidak memiliki kriteria yang diharapkan ( $y=0$ ), atau prosentase nilai dugaan yang sama dengan nilai amatan untuk  $y=0$  disebut *specificity*. Dalam Tabel-2 nilai *specificity* dinyatakan dengan  $(d/n_0)100\%$ . Sedangkan prosentase nilai dugaan yang sama dengan nilai amatan untuk  $y=0$  disebut *sensitivity*. Dalam Tabel-2 nilai *sensitivity* dinyatakan dengan  $(a/n_1)100\%$ . Prosentase total nilai amatan yang secara tepat dapat diduga oleh model disebut *total correct classification*. Dalam Tabel-2 *total correct classification* dinyatakan dengan  $((a+d)/n)100\%$ . Kesalahan klasifikasi terdiri dari kesalahan positif dan kesalahan negatif. Prosentase kesalahan dalam menduga respon memiliki kriteria yang diharapkan ( $y=1$ ), padahal sebenarnya bernilai  $y=0$  disebut kesalahan positif. Dalam Tabel-2 kesalahan positif dinyatakan dengan  $(b/n_0)100\%$ . Prosentase kesalahan dalam menduga respon tidak memiliki kriteria yang diharapkan ( $y=0$ ), padahal sebenarnya bernilai  $y=1$  disebut kesalahan negatif. Dalam Tabel-2 kesalahan negatif dinyatakan dengan  $(c/n_1)100\%$ . Selanjutnya prosentase kesalahan klasifikasi total adalah  $((b+c)/n)100\%$  disebut *total misclassification rate*.

**Metoda Penelitian**

Jenis penelitian ini merupakan penelitian terapan (*applied research*), yaitu penelitian yang dimaksudkan untuk menerapkan suatu teori dalam kaitannya dengan pemanfaatannya dalam bidang Pendidikan. Sampel yang digunakan dalam penelitian ini adalah *purposif* sampel yaitu mahasiswa Polban Tata Niaga angkatan 2009/2010 sampai dengan

2010/2011. Data yang digunakan adalah data sekunder, yang diperoleh dari dokumentasi dari Akademis Polban. Gambaran variabel-variabel yang akan diteliti sebagai berikut:

Tabel-3.  
Gambaran Variabel-variabel yang Akan Diteliti

No	Variabel	Label	Skala	Kategori
1.	Respon (Y)	Keberhasilan Mahasiswa Polban	Nominal	1 = Berhasil 0= Gagal
2.	Penjelas (X1)	Jenis Kelamin	Nominal	1= laki-laki 0=perempuan
3.	Penjelas (X2)	Asal Daerah	Nominal	1=Bandung 0=Luar Bandung
4.	Penjelas (X3)	Latar Belakang SLTA	Nominal	1=SMU 0=SMK
5.	Penjelas (X4)	Nilai Ebtanas Murni (NEM)	Kuantitatif	
6.	Penjelas (X5)	Nilai Seleksi Ujian masuk Polban	Kuantitatif	
7.	Penjelas (X6)	Nilai Matematika Semester I	Ordinal	1=D : 2=C 3=B : 4=A
8.	Penjelas (X7)	Nilai Indeks Prestasi Semester I	Ordinal	1=< 2,5 2=2,5-3 3=>3

Tahap-tahap analisis data dalam penelitian ini sebagai berikut:

1. Melakukan pengujian signifikansi terhadap variabel -variabel yang akan dianalisis yaitu : jenis kelamin, asal daerah, latar belakang SLTA,nilai Ebtanas murni, nilai seleksi ujian masuk, nilai matematika bisnis, dan nilai IP semester 1.
2. Melakukan pengujian asumsi multivariat normal dan kesamaan matriks varian kovarian data pada masing-masing grup respon.
3. Melakukan pembagian data menjadi dua bagian yaitu *training sample* dan *validation sample* dengan proporsi 80% dan 20%, 60 dan 40% . Pemilihan data untuk *training sample* maupun *validation sample* dilakukan secara acak.
4. Membentuk model klasifikasi dengan analisis diskriminan dan regresi logistik dengan menggunakan data *training sample*. Model klasifikasi yang telah diperoleh ini selanjutnya divalidasi dengan data *validation sample*. Jika ketepatan klasifikasi dari kedua sampel hampir

sama, maka fungsi diskriminan dan logistik respon biner dari *training sample* sudah valid.

**Hasil dan Pembahasan**

Hasil uji signifikansi secara parsial, dari tujuh variabel ternyata ada 2 variabel yang memberikan pengaruh secara signifikan terhadap keberhasilan mahasiswa Polban yaitu jenis kelamin, program studi, dan IP semester I. Hal ini ditunjukkan dari output SPSS dengan nilai signifikansi < 0,05. Sedangkan empat variabel lainnya tidak signifikan karena nilai signifikansi > 0,05. Pada penelitian ini yang akan dianalisis adalah data yang melibatkan semua variabel penjelas.

**Analisis Data**

Berdasarkan hasil pengujian asumsi untuk data, asumsi multivariat normal tidak dapat dipenuhi sedangkan matriks varian kovarian dapat dipenuhi. Hasil penerapan analisis Diskriminan dan Regresi Logistik respon Biner dengan proporsi data 80% dan 20% untuk *training* dan *validation sample* menghasilkan fungsi diskriminan dan regresi logistik respon biner seperti pada table-1.

Tabel-4.  
Koefisien Fungsi Diskriminan dan Regresi Logistik Respon Biner

Variabel	Analisis Diskriminan	Regresi Logistik Respon Biner
Jenis Kelamin	1,347	1,955
Program Studi	-5,44	2,992
Nilai IP semester 1	-1,919	3,792
Nilai Matematika	1,135	3,792
Asal Daerah	0,610	4,288
Nilai Rata-rata NEM	-0,263	4,451
Nilai Ujian Masuk	0,001	4,466
Konstanta	5,057	4,466

Tabel-5.  
Hasil Klasifikasi Analisis Diskriminan dan Regresi Logistik Respon Biner

Analisis Diskriminan		
Ketepatan		
Predicted Group		
Klasifikasi		
Tidak Berhasil		Berhasil
Training	Actual	Tidak Berhasil
	1	0
Sampel	Group	Berhasil
15	238	94,09%

Tidak Berhasil		Berhasil
Validation	Group	Tidak Berhasil
	1	0
Sampel	Actual	Berhasil
6	81	93,19%

Regresi Logistik Respon Biner Ketepatan		
Predicted Group Klasifikasi		
Tidak Berhasil	Berhasil	
Training	Actual 0	Tidak Berhasil 1
Sampel 0	Group 253	Berhasil 99,61%

Tidak Berhasil	Berhasil	
Validation	Group 0	Tidak Berhasil 1
Sampel 0	Actual 87	Berhasil 99,61%

Berdasarkan table-4, dapat dilihat bahwa *training sample* regresi logistik respon biner mempunyai nilai ketepatan klasifikasi lebih tinggi (99,61%) dibandingkan analisis diskriminan (94,09%). Begitu juga dengan *validation sample* regresi logistik respon biner menghasilkan nilai ketepatan klasifikasi lebih tinggi (99,61%) dibandingkan dengan analisis diskriminan (93,19%). Menurut Santoso (2002) dinyatakan bahwa penilaian valid atau tidaknya fungsi diskriminan maupun regresi logistik respon biner adalah jika nilai ketepatan klasifikasi antara *training* dan *validation sample* sama besar, artinya selisih antara keduanya seminimal mungkin. Berdasarkan hasil yang diperoleh, dapat ditunjukkan bahwa selisih nilai antara *training* dan *validation sample* sangat kecil pada analisis diskriminan adalah 0,9%. Sedangkan selisih nilai antara *training* dan *validation sample* pada regresi logistik respon biner adalah 0,00%. Sehingga dapat disimpulkan bahwa metode regresi logistik respon biner mempunyai hasil klasifikasi lebih baik dibandingkan analisis diskriminan. Hasil penerapan analisis diskriminan dan regresi logistik respon biner dengan proporsi data *training* dan *validation sample* adalah 60% dan 40% disajikan dalam table-2.

Tabel-6.  
Koefisien Fungsi Diskriminan dan Regresi Logistik Respon Biner

Variabel	Analisis Diskriminan	Regresi Logistik Respon Biner
Jenis Kelamin	-1,247	1,927
Program Studi	0,520	2,908
Nilai IP semester 1	1,136	3,590
Nilai Matematika	-1,088	3,918
Asal Daerah	0,796	3,981
Nilai Rata-rata NEM	1,103	3,983
Nilai Ujian Masuk	-0,001	3,983
Konstanta	-2,294	0,583

Tabel-7.  
Hasil Klasifikasi Analisis Diskriminan dan Regresi Logistik Respon Biner

Analisis Diskriminan Ketepatan		
Predicted Group Klasifikasi		
Tidak Berhasil	Berhasil	
Training	Actual 1	Tidak Berhasil 0
Sampel 15	Group 238	Berhasil 94,07%

Tidak Berhasil	Berhasil	
Validation	Group 3	Tidak Berhasil 0
Sampel 29	Actual 132	Berhasil 80,49%

Regresi Logistik Respon Biner Ketepatan		
Predicted Group Klasifikasi		
Tidak Berhasil	Berhasil	
Training	Actual 0	Tidak Berhasil 1
Sampel 0	Group 253	Berhasil 100%

Tidak Berhasil	Berhasil	
Validation	Group 3	Tidak Berhasil 0
Sampel 0	Actual 161	Berhasil 100%

Berdasarkan table-7, dapat dilihat bahwa *training sample* regresi logistik respon biner mempunyai nilai ketepatan klasifikasi lebih tinggi (100%) dibandingkan analisis diskriminan (94,07%). Begitu juga dengan *validation sample* regresi logistik respon biner menghasilkan nilai ketepatan klasifikasi lebih tinggi (100%) dibandingkan dengan analisis diskriminan (80,49%). Hasil yang diperoleh menunjukkan bahwa selisih nilai antara *training* dan *validation sample* sangat kecil pada analisis diskriminan adalah 13,58%. Sedangkan selisih nilai antara *training* dan *validation sample* pada regresi logistik respon biner adalah 0,00%. Sehingga dapat disimpulkan bahwa metode regresi logistik respon biner mempunyai hasil klasifikasi lebih baik dibandingkan analisis diskriminan.

### Simpulan dan Saran

Berdasarkan analisis dan pembahasan pada bab sebelumnya, maka dapat disimpulkan yaitu apabila variabel penjelas berskala kategori dalam kasus keberhasilan mahasiswa Polban, regresilogistik respon

biner menghasilkan ketepatan klasifikasi yang lebih tinggi dibandingkan dengan analisis diskriminan. Variabel-variabel yang mempengaruhi keberhasilan mahasiswa Polban adalah jenis kelamin, program studi, dan nilai IP semester satu. Hal ini ditunjukkan dengan nilai signifikansi ketiga variabel penjelas tersebut, nilai probabilitasnya kurang dari 0,05.

#### Saran

Diharapkan ada penelitian lanjutan yang berkaitan dengan pengklasifikasian data dengan menggunakan metode lain. Begitu pula dengan karakteristik variabel yang digunakan, hendaknya digunakan variabel tambahan yang berkaitan dengan latar belakang mahasiswa, pendidikan orang tua, pendapatan orang tua, dan lain sebagainya. Variabel dari segi proses pembelajaran, kemampuan dosen, keberadaan modul, sarana dan prasarana.

#### Daftar Pustaka

- [1] Agresti, A (1996),. *An Introduction to Categorical Data Analysis*, John Wiley and Sons Inc
- [2] Anonim. 2004. *Estimation and Hypothesis Testing for Logistic Regression* . <http://courses.washington.edu/b515/l13.pdf>
- [3] Anonim.2005.*DiscriminantFunctionAnalysis* .<http://www.statsoft.com> ,
- [4] Binarto A, Sartika E, 2010, *Analisis Faktor-faktor yang Mempengaruhi Keberhasilan Mahasiswa Politeknik ( Studi Kasus Mahasiswa Tata Niaga)* , Hasil Penelitian Pemula, Polban.
- [5] Fox, J. 2005. *Maximum Likelihood Estimation of the Logistic Regression Model*.[www.socserv.mcmaster.ca/jfox/Courses/UCLA/logistic-regression-notes.pdf](http://www.socserv.mcmaster.ca/jfox/Courses/UCLA/logistic-regression-notes.pdf).
- [6] Gazpert (1992), *Teknik Analisis dalam Penelitian Percobaan* . Tarsito . Bandung
- [7] Garson, G.D. 2006. *Logistic Regression*. <http://www2.chass.ncsu.edu/garson/pa765/logistic.htm>.
- [8] Hair, J.F. et al. 1998. *Multivariate Data Analysis*. Prentice Hall. New Jersey.
- [9] Hosmer, D.W, dan Lemeshow, S. (2000), *Applied Logistic Regression*, New York, John Wiley & Sons
- [10] Hidayati M, 2002. Penelusuran Faktor-faktor yang Mempengaruhi Prestasi Akademik Mahasiswa Semester I Universitas IBN Khaldun Bogor. Skripsi. Jurusan Fakultas Mipa Statistika IPB.
- [11] Institut Teknologi Bandung 1997 , *Buku Panduan Politeknik Negeri Bandung*.
- [12] Kurniati,E .2007 . *Perbandingan Hasil Klasifikasi Analisis Diskriminan dan Regresi Logistik Respon Biner Pada Pengklasifikasian Kejadian Bayi dengan Berat Badan Lahir Rendah (BBLR)*.Skripsi F.MIPA Universitas Bengkulu.
- [13] Larsen, P.V. 2006. *Logistic Regression*. <http://statmaster.sdu.dk/courses/st111/module14/module.pdf> .
- [14] Morrison, D.F. 1978. *Multivariate Statistical Methods*. International Student Edition, McGrawhill.
- [15] Rhencher, A.C. 1995. *Method of Multivariate Analysis* . John Wiley and Sons. New York.
- [16] Ridgeway, G. 2004. *Maximum Likelihood and Logistic Regression* . [www.i-pensieri.com/gregr/ModernPrediction/L2logistic.pdf](http://www.i-pensieri.com/gregr/ModernPrediction/L2logistic.pdf) .
- [17] Suryabrata S. 1995. *Proses Belajar Mengajar di Perguruan Tinggi*, Yogyakarta Andi Offset.
- [18] Santoso, S. 2002. *SPSS Statistik Multivariat*. Gramedia. Jakarta.
- [19] Supranto, J. 2004. *Analisis Multivariat: Arti dan Interpretasi* . Rineka Cipta. Jakarta.
- [20] Suhaeni, C. 2005. *Analisis Diskriminan Untuk Prediksi Indeks Prestasi Mahasiswa Pada Semester Pertama*. Skripsi FMIPA Universitas Bengkulu.
- [21] Wibowo, W. 2002. *Perbandingan Hasil Klasifikasi Analisis Diskriminan dan Regresi Logistik Pada Pengklasifikasian Data Respon Biner*. Kappa, Vol 3, No 1:36-45.[www.fmipa.its.ac.id/isi%20mipajurnal/jurnal/KAPPA%20\(2002\)%20Vol.3,%20No.1,%2036-45.pdf](http://www.fmipa.its.ac.id/isi%20mipajurnal/jurnal/KAPPA%20(2002)%20Vol.3,%20No.1,%2036-45.pdf)