

**Perbandingan Hasil Klasifikasi Menggunakan Regresi logistik dan Analisis Diskriminan Kuadratik Pada Kasus Pengklasifikasian Jurusan Di SMA Negeri 8 Samarinda Tahun Ajaran 2014/2015**

*Comparison of Classification Results by Using Logistic Regression and Quadratic Discriminant Analysis in Classifying Programs at SMA Negeri 8 Samarinda Academic Year 2014/2015*

**Cristine Uli Artha<sup>1</sup>, Yuki Novia N<sup>2</sup>, Ika Purnamasari<sup>3</sup>**

<sup>1</sup>Mahasiswa Program Studi Statistika FMIPA Universitas Mulawarman

<sup>2,3</sup>Dosen Program Studi Statistika FMIPA Universitas Mulawarman

Email: cristinesiahaan@yahoo.co.id

**Abstract**

*Logistic Regression Analysis and Discriminant Analysis represent the statistical method for the classification of a number of object. In the case of classification especially if there's only two response categories, logistic regression is used more precisely if the assumption of multivariate normality in data cannot be fulfilled. The assumption of normality multivariate distribution and equality of variance covariance matrices represent the important matter in discriminant analysis for getting of high accuracy of classification. Discriminant analysis method that is used in inequality of variance covariance matrices is called quadratic discriminant analysis. The purpose of this study was to determine the classification results by using Logistic Regression and Quadratic Discriminant Analysis and compares the classification accuracy. The data that is used in the study is the average raport of the first and second semester of the class X at SMA Negeri 8 Samarinda academic year 2014/2015. Data consists of 190 students with two independent variables and four dependent variables. Based on research results, obtained results for the value of class accuracy is Logistic Regression 83.16% and Quadratic Discriminant Analysis 84,21%.*

*Keywords: Quadratic discriminant, classification, classification accuracy, logistic regression.*

**Pendahuluan**

Masalah klasifikasi sering kali ditemui di kehidupan sehari-hari, terkait dengan data sosial, data industri manufaktur, data pemasaran maupun data akademik. Mendeskripsikan dan memaparkan keunikan dari suatu pengelompokan tersebut merupakan hal yang menarik dan dapat memberikan ide-ide tertentu. Namun, jika pengelompokan menyangkut pengambilan tidak benar atau ada observasi-observasi tertentu yang salah dalam proses pengelompokan, maka akibatnya akan cukup fatal. Oleh karenanya, perlu dilakukan *review* pada proses klasifikasi.

Dalam ilmu statistika dikenal beberapa metode klasifikasi yang umumnya digunakan, diantaranya regresi logistik dan analisis diskriminan. Pada regresi logistik akan diperoleh suatu model logistik yang digunakan untuk menjelaskan hubungan antara variabel prediktor dan respon (yang bersifat dikotomis atau ada dua kategori/kelompok), serta untuk mengelompokkan obyek ke dalam salah satu dari dua kategori respon. Regresi logistik merupakan salah satu alat analisis yang merupakan bentuk khusus dari analisis regresi, yaitu variabel responnya merupakan data skala nominal atau ordinal sedangkan variabel prediktornya dapat berbentuk nominal, ordinal, ataupun rasio (Hosmer dan Lemeshow, 2000).

Kurt, Ture dan Kurum (2006) dalam penelitiannya menyatakan bahwa regresi logistik merupakan metode klasifikasi yang cukup baik,

setidaknya pada saat ada variabel prediktor berskala kuantitatif maupun kualitatif ataupun keduanya. Pada analisis diskriminan, memiliki kelebihan yaitu memberikan perhitungan yang lebih efisien (Sharma, 1996) sedangkan kelemahan dari analisis diskriminan adalah asumsi harus terpenuhi, dimana data harus memenuhi distribusi normal multivariat dan menghasilkan matriks kovarians yang sama untuk setiap populasi (Johnson dan Wichern, 2002). Dalam hal klasifikasi khususnya jika hanya ada dua kategori variabel respon, regresi logistik lebih tepat dipergunakan bila asumsi distribusi normal multivariat dalam data tidak terpenuhi (Hair dkk, 2010).

Menurut Rachmatin dan Sawitri (2010), terdapat beberapa kasus analisis diskriminan yang diketahui, yaitu analisis diskriminan linier, analisis diskriminan kuadratik, analisis diskriminan Fisher, dan analisis diskriminan nonparametrik. Setiap kasus analisis diskriminan memiliki penggunaan yang berbeda dalam menganalisis data. Analisis diskriminan linier digunakan jika data berdistribusi normal multivariat dan setiap kelompoknya memiliki matriks varians kovarians yang sama. Analisis diskriminan kuadratik digunakan jika data berdistribusi normal multivariat tetapi matriks varians kovariansnya tidak sama dalam setiap kelompok. Analisis diskriminan Fisher digunakan jika data tidak berdistribusi normal multivariat tetapi matriks varians kovariansnya sama dalam

setiap kelompok. Sedangkan analisis diskriminan nonparametrik digunakan jika data tidak berdistribusi normal multivariat dan matriks varians kovariansnya tidak sama setiap kelompoknya.

Selama ini penjurusan siswa SMA Negeri 8 Samarinda berdasarkan nilai akademik yang diperoleh pada semester 1 dan semester 2 kelas X secara tuntas atau sama dengan Kriteria Ketuntasan Minimal (KKM) untuk semua mata pelajaran yang menjadi ciri khas jurusannya serta minat siswa terhadap jurusan yang ada. Penetapan jurusan SMA merupakan suatu sarana untuk memberikan kesempatan pada siswa dalam mengembangkan kompetensi sikap, kompetensi pengetahuan, dan kompetensi keterampilan siswa sesuai dengan minat, bakat atau kemampuan akademik dalam sekelompok mata pelajaran keilmuan. Mengingat kesesuaian pengklasifikasian siswa pada penjurusan merupakan hal yang penting untuk masa depan siswa tersebut, maka hal ini perlu untuk diperhatikan. Oleh karena itu perlu suatu metode statistik yang dapat digunakan untuk mengukur seberapa besar keakuratan penjurusan SMA Negeri 8 Samarinda. Beberapa metode statistik yang dapat digunakan untuk mengukur seberapa besar keakuratan penjurusan SMA Negeri 8 Samarinda yaitu analisis regresi logistik dan analisis diskriminan kuadrat.

Dari beberapa penelitian sebelumnya telah dilakukan penelitian untuk membandingkan metode klasifikasi diantaranya, Satriya (2011) membandingkan Regresi Logistik dan Analisis Jaringan Syaraf Tiruan, Firmansyah (2010) membandingkan Analisis Diskriminan dan Analisis Jaringan Syaraf Tiruan. Berdasarkan latar belakang tersebut, maka penulis tertarik untuk membandingkan metode klasifikasi menggunakan regresi logistik dan analisis diskriminan kuadrat.

**Regresi Logistik**

Regresi logistik merupakan salah satu metode klasifikasi yang sering digunakan. Regresi logistik merupakan suatu teknik analisis data dalam statistika yang bertujuan untuk mengetahui hubungan antara beberapa variabel dimana variabel responnya adalah bersifat kategorik, baik nominal maupun ordinal dengan variabel prediktornya dapat bersifat kategorik atau kontinu. Regresi logistik biner digunakan saat variabel respon merupakan variabel dikotomis (kategorik dengan dua macam kategori), sedangkan regresi logistik multinomial digunakan saat variabel respon adalah variabel kategorik dengan lebih dari dua kategori. Pada regresi logistik jika variabel responnya terdiri dari dua kategori misalnya  $Y=1$  menyatakan hasil yang diperoleh “sukses” dan  $Y=0$  menyatakan hasil

yang diperoleh “gagal” maka regresi logistik tersebut menggunakan regresi logistik biner (Fractal, 2003).

Metode regresi logistik memiliki teknik dan prosedur yang tidak jauh berbeda dengan metode regresi linear. Jika prosedur linear dalam mengestimasi nilai parameter sering menggunakan metode *Ordinary Least Squares* (OLS), maka untuk mengestimasi nilai parameter dalam regresi logistik adalah dengan menggunakan metode *Maximum Likelihood Estimation* (MLE). Untuk mencari persamaan logistiknya maka model yang digunakan adalah:

$$f(x) = \frac{e^{s_0 + \sum_{j=1}^p s_j x_j}}{1 + e^{s_0 + \sum_{j=1}^p s_j x_j}} \tag{1}$$

Dari persamaan (1) diperoleh  $1 - f(x)$  sebagai berikut:

$$\begin{aligned} 1 - f(x) &= 1 - \frac{e^{\beta_0 + \sum_{j=1}^p \beta_j x_j}}{1 + e^{\beta_0 + \sum_{j=1}^p \beta_j x_j}} \\ &= \frac{1 + e^{\beta_0 + \sum_{j=1}^p \beta_j x_j} - e^{\beta_0 + \sum_{j=1}^p \beta_j x_j}}{1 + e^{\beta_0 + \sum_{j=1}^p \beta_j x_j}} \\ &= \frac{1}{1 + e^{\beta_0 + \sum_{j=1}^p \beta_j x_j}} \end{aligned}$$

Sehingga  $\frac{f(x)}{1 - f(x)}$  sebagai berikut:

$$\frac{f(x)}{1 - f(x)} = e^{s_0 + \sum_{j=1}^p s_j x_j}$$

Jadi, persamaan logistiknya adalah:

$$\begin{aligned} g(x) &= \ln \left[ \frac{f(x)}{1 - f(x)} \right] \\ &= \ln e^{s_0 + \sum_{j=1}^p s_j x_j} \\ &= s_0 + \sum_{j=1}^p s_j x_j \end{aligned} \tag{2}$$

**Uji Signifikansi Parameter**

Pengujian keberartian parameter atau signifikansi parameter dalam regresi penting untuk dilakukan. Hal ini dikarenakan pengujian tersebut digunakan untuk menentukan apakah pengaruh variabel prediktor dalam model signifikan terhadap variabel respon. Pengujian dapat dilakukan secara simultan dan parsial.

**1. Pengujian Secara Simultan**

Bertujuan untuk mengetahui pengaruh variabel prediktor secara simultan terhadap variabel respon. Hipotesis yang digunakan adalah sebagai berikut:

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

$H_1$  : Paling sedikit ada satu  $\beta_j$  tidak sama dengan nol, dengan  $j=1, 2, \dots, p$

Statistik uji yang digunakan adalah uji G (*Likelihood Ratio Test*).  $H_0$  ditolak jika nilai  $G \geq t^2_{(r,db)}$  atau  $p-value \leq \alpha$

**2. Pengujian Secara Parsial**

Hasil pengujian secara individual akan menunjukkan apakah suatu variabel prediktor layak untuk masuk dalam model atau tidak. Hipotesis yang digunakan adalah sebagai berikut:

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0$$

Statistik uji yang digunakan adalah uji *Wald* dengan rumus sebagai berikut:

$$W = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \tag{4}$$

$H_0$  ditolak jika nilai  $W \geq z_{\alpha/2}$  atau  $p-value \leq \alpha$

**Uji Kecocokan Model**

Setelah penaksiran model dilakukan, maka langkah berikutnya adalah menentukan seberapa baik model tersebut cocok terhadap data atau seberapa dekat nilai-nilai dari model dengan nilai observasinya. Hipotesis yang digunakan adalah sebagai berikut:

$H_0$  : Tidak ada perbedaan antara hasil pengamatan dengan nilai dugaan.

$H_1$  : Ada perbedaan antara hasil pengamatan dengan nilai dugaan.

Statistik uji yang digunakan adalah uji Uji Hosmer dan Lemeshow dengan rumus sebagai berikut:

$$\hat{C} = \sum_{k=1}^g \frac{(O_k - n'_k f_k)^2}{n'_k f_k (1 - f_k)} \tag{5}$$

dengan:

$g$  = Jumlah grup (kombinasi dalam model serentak)

$n'_k$  = Jumlah pengamatan pada grup ke- $k$

$C_k$  = Jumlah kombinasi variabel prediktor

$O_k = \sum_{j=1}^{C_k} y_j$  = jumlah nilai variabel respon

pada  $C_k$  kombinasi variabel prediktor

$f_k = \sum_{j=1}^{C_k} \frac{m_j f(x_j)}{n_k}$  = rata-rata taksiran probabilitas

dimana  $m_j$  adalah banyaknya

subjek dengan  $C_k$  kombinasi variabel prediktor

$H_0$  ditolak jika nilai  $\hat{C} \geq t^2_{(r,db)}$  atau  $p-value \leq \alpha$

**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 dan Wichern, 2002).

**Asumsi Analisis Diskriminan**

Asumsi pada analisis diskriminan menurut Supranto (2004) adalah variabel independen (bebas) harus berdistribusi normal (*multivariate normality*) dan matriks peragam dari semua variabel independen (bebas) adalah sama.

**1. Distribusi Normal Multivariat**

Seperti pada banyak metode statistik lainnya, analisis diskriminan juga mensyaratkan data berdistribusi normal. Jika data berdistribusi tidak normal, hasil analisis dikhawatirkan menjadi bias. Menurut Johnson dan Wichern (2002), pengujian data berdistribusi multivariat normal dengan membuat *Q-Q plot*. Untuk menguji multivariat normal dengan menggunakan hipotesis sebagai berikut:

$H_0$  : Data berdistribusi multinormal

$H_1$  : Data tidak berdistribusi multinormal

Pengujian data berdistribusi multivariat normal dengan membuat *Q-Q plot* dari nilai sebagai berikut:

$$d_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) \tag{6}$$

dimana:

$d_i^2$  = jarak Mahalanobis *d-square*

$\mathbf{x}$  = skor nilai pengamatan

$\bar{\mathbf{x}}$  = rata-rata nilai pengamatan

$\mathbf{S}^{-1}$  = matriks invers varian-kovarian

$H_0$  ditolak jika nilai  $d_i^2 \leq t^2_{(r,db)}$ , dan nilai

$d_i^2 \leq t^2_{(r,db)}$  yang diharapkan kurang dari atau sama dengan 50% dari jumlah data dengan  $db = p$ , dimana  $p$  adalah jumlah variabel prediktor.

**2. Kesamaan Matriks Varians Kovarian**

Analisis diskriminan linier mengasumsikan bahwa matriks varians kovarians dari dua kelompok adalah sama. Pelanggaran pada asumsi ini akan mempengaruhi pengujian signifikansi dan hasil klasifikasi. Jika asumsi kesamaan matriks varians kovarians ditolak, dapat

digunakan fungsi diskriminan kuadratik untuk fungsi klasifikasi (Sharma, 1996). Uji yang digunakan untuk mengetahui kesamaan matriks varians kovarians adalah uji *Box's M*. Prosedur pada uji *Box's M* yaitu menggunakan pendekatan sebaran Chi-kuadrat dengan derajat bebas  $\frac{1}{2}(k-1)p(p+1)$ . Hipotesis matriks varians kovarians adalah sebagai berikut :

- $H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_k$
- $H_1 : \text{sedikitnya ada sepasang kelompok matriks varians kovarians yang berbeda antar kelompok}$

Statistik uji:

$$C = (1-u)M = (1-u) \sum_{i=1}^k (n_i - 1) \ln |S_i| - \sum_{i=1}^k (n_i - 1) \ln |S| \quad (7)$$

dengan:

$$u = \left[ \sum_{i=1}^k \frac{1}{(n_i - 1)} - \frac{1}{\sum_{i=1}^k (n_i - 1)} \right] \left[ \frac{2p^2 + 3p - 1}{6(p+1)(k-1)} \right] \quad (8)$$

**Analisis Diskriminan Kuadratik**

Analisis diskriminan bertujuan untuk membentuk fungsi diskriminan yang mampu membedakan kelompok. Analisis ini dilakukan berdasarkan suatu perhitungan statistik terhadap objek-objek yang telah diketahui dengan jelas pengelompokannya. 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 kovarians tidak sama. Jika asumsi kesamaan matriks varians kovarian ditolak, dapat digunakan fungsi diskriminan kuadratik untuk fungsi klasifikasi (Sharma, 1996).

Skor diskriminan kuadratik didefinisikan sebagai,

$$Q(x) = \ln p_i - \frac{1}{2} \ln |S_i| - \frac{1}{2} (x - \mu_i)^T S_i^{-1} (x - \mu_i) \quad (9)$$

dimana,

- $\mu_i$  = rata-rata populasi ke- $i$ ,  $i = 1, 2, \dots, g$
- $p_i$  = peluang prior
- $|S_i|$  = panjang matriks varians kovarians kelompok ke- $i$

Dalam penerapannya digunakan komponen sampel yaitu  $\bar{x}_i$  vektor rata-rata sampel dan  $S_i$  matriks kovarians sampel, sehingga dapat dibentuk fungsi diskriminan kuadratik yang didefinisikan:

$$Q_i(x) = \ln p_i - \frac{1}{2} \ln |S_i| - \frac{1}{2} (x - \bar{x}_i)^T S_i^{-1} (x - \bar{x}_i) \quad (10)$$

Dari fungsi diskriminan kuadratik pada Persamaan (10) dapat dibentuk aturan

pengelompokannya dengan mengalokasikan  $x$  ke dalam kelompok  $k$  jika,

$$Q_k(x) = \text{maks}(Q_1(x), Q_2(x), \dots, Q_g(x)) \quad (11)$$

Dua pendekatan umum yang dapat digunakan untuk mengestimasi probabilitas anggota  $p_i$  yaitu:

$$p_i = \frac{1}{g} \quad (12)$$

yang mana  $p_i$  diasumsikan sama untuk semua kelompok, dan untuk setiap kelompok dengan  $p_i$  diestimasi sebagai frekuensi relatif dari observasi pada setiap kelompok maka digunakan rumus sebagai berikut:

$$p_i = \frac{n_i}{n} \quad (13)$$

**Evaluasi Fungsi Klasifikasi**

Evaluasi fungsi klasifikasi baik analisis diskriminan kuadratik maupun regresi logistik dapat dilakukan dengan menghitung peluang kesalahan klasifikasi. Ukuran yang dapat digunakan adalah *Apparent Error Rate* (APER). Nilai APER menyatakan fraksi atau proporsi sampel yang salah diklasifikasikan oleh fungsi klasifikasi (Johnson dan Wichern, 2002).

Evaluasi fungsi klasifikasi dilakukan dengan cara membuat tabulasi antara actual group dan predicted group yang diperoleh dari fungsi klasifikasi dari analisis diskriminan kuadratik dan regresi logistik. Selanjutnya dihitung proporsi pengamatan yang salah diklasifikasikan. Diharapkan proporsi pengamatan yang salah diklasifikasikan tersebut bisa sekecil mungkin. Untuk menghitung nilai APER dapat dilakukan melalui tabel klasifikasi sebagai berikut:

Tabel 1. Tabel Klasifikasi

Actual membership	Predicted membership		Total
	$f_1$	$f_2$	
$f_1$	$n_{11}$	$n_{12}$	A
$f_2$	$n_{21}$	$n_{22}$	B
Total	C	D	E

Keterangan :

- $n_{11}$  = jumlah obyek dari  $f_1$  tepat diklasifikasikan sebagai  $f_1$
- $n_{12}$  = jumlah obyek dari  $f_1$  salah diklasifikasikan sebagai  $f_2$
- $n_{21}$  = jumlah obyek dari  $f_2$  salah diklasifikasikan sebagai  $f_1$
- $n_{22}$  = jumlah obyek dari  $f_2$  tepat diklasifikasikan sebagai  $f_2$

$$APER = \frac{n_{12} + n_{21}}{n} \tag{14}$$

**Metodologi Penelitian**

Data yang digunakan dalam penelitian ini merupakan data rata-rata nilai pelajaran IPA dan IPS pada rapor semester 1 dan semester 2 siswa kelas X SMA Negeri 8 Samarinda (sebanyak 190 siswa). Hasil analisis regresi logistik dengan bantuan *software* SPSS 16 dan hasil analisis diskriminan kuadrat dengan bantuan *software* Macro Minitab, SPSS 16 dan SAS 9.1.3 Variabel dalam penelitian ini meliputi:

$$Y = \text{Akreditasi Sekolah} \begin{cases} 1 = \text{IPA} \\ 2 = \text{IPS} \end{cases}$$

- X<sub>1</sub> = rata-rata nilai pelajaran IPA pada rapor semester 1 kelas X
- X<sub>2</sub> = rata-rata nilai pelajaran IPA pada rapor semester 2 kelas X
- X<sub>3</sub> = rata-rata nilai pelajaran IPS pada rapor semester 1 kelas X
- X<sub>4</sub> = rata-rata nilai pelajaran IPS pada rapor semester 2 kelas X

Adapun teknik analisis data dalam penelitian ini adalah :

1. Analisis statistika deskriptif.
2. Langkah-langkah yang dilakukan dalam regresi logistik:
  - a. Membentuk model awal regresi logistik.
  - b. Melakukan pengujian parameter regresi logistik.
  - c. Melakukan pengujian kesesuaian model.
  - d. Melakukan proses klasifikasi.
3. Langkah-langkah yang dilakukan dalam analisis diskriminan kuadrat:
  - a. Melakukan pengujian asumsi.
  - b. Menentukan fungsi diskriminan kuadrat.
  - c. Melakukan proses klasifikasi.
4. Evaluasi fungsi klasifikasi.

**Hasil dan Pembahasan**

**Analisis Deskriptif**

Data yang digunakan yaitu pada Lampiran 1 diperoleh dari SMA Negeri 8 Samarinda yang merupakan data Nilai Rata-rata rapor Siswa SMA Negeri 8 Samarinda kelas X Semester 1 dan semester 2 tahun ajaran 2014/2015. Data pada penelitian ini sebanyak 190 data, terdiri dari 4 Variabel Penjelas yaitu nilai rata-rata tiap semester dan 2 Variabel respon yaitu penjurusan IPA dan IPS. Penjurusan IPA sebanyak 120 data dan IPS sebanyak 70 data. Nilai-nilai tersebut ditunjukkan pada Tabel 2.

Berdasarkan Tabel 2, diketahui bahwa nilai rata-rata pelajaran IPS pada rapor semester 2 kelas X memiliki rata-rata tertinggi (X<sub>4</sub>) yaitu sebesar 81,86. Kemudian untuk rata-rata nilai pelajaran yang memiliki nilai minimum paling

besar adalah rata-rata nilai pelajaran IPS pada rapor semester 1 (X<sub>3</sub>) yaitu sebesar 76,75 .Selanjutnya untuk rata-rata nilai pelajaran yang memiliki nilai maksimum terkecil adalah rata-rata nilai pelajaran IPA pada rapor semester 1 (X<sub>1</sub>) yaitu sebesar 85,25.

Tabel 2. Analisis Statistika Deskriptif

Deskriptif	Rata-rata	Minimum	Maksimum
X <sub>1</sub>	78,33	71,75	85,25
X <sub>2</sub>	79,13	67	87
X <sub>3</sub>	81,17	76,75	87,75
X <sub>4</sub>	81,86	67	90

**Analisis Regresi Logistik**

**Kontruksi Model Regresi Logistik**

Dengan bantuan *software* SPSS diperoleh taksiran model regresi logistik sebagai berikut:

Tabel 3. Hasil Penaksiran Parameter untuk Variabel Prediktor

Variabel Prediktor		S.E
X <sub>1</sub>	0,271	0,106
X <sub>2</sub>	-0,767	0,162
X <sub>3</sub>	0,224	0,107
X <sub>4</sub>	-0,246	0,102
Konstan	40,409	12,916

Berdasarkan Tabel 3 maka diperoleh hasil penaksiran model regresi logistik sebagai berikut:

$$f(X) = \frac{e^{40,409+0,271X_1-0,767X_2+0,224X_3-0,246X_4}}{1+e^{40,409+0,271X_1-0,767X_2+0,224X_3-0,246X_4}}$$

**Uji Signifikansi Parameter**

**1. Uji Simultan**

Uji simultan dilakukan untuk mengetahui signifikansi parameter dari terhadap variabel respon yaitu secara bersama-sama (serentak).

Tabel 4. Uji Simultan

G	Df	-2 Log-likelihood
120,62	4	129,485

Berdasarkan Tabel 4 dan Persamaan 3 maka diperoleh nilai G adalah 120,62 dimana nilai nilai G (120,62) > <sup>2</sup><sub>(0.05,4)</sub> (9,488) maka dapat disimpulkan bahwa minimal ada satu variabel prediktor yang berpengaruh terhadap variabel respon.

**2. Uji Parsial**

Uji parsial berfungsi untuk mengetahui apakah ada pengaruh dari variabel prediktor terhadap variabel respon secara individu. Uji ini menggunakan statistik uji dari *Wald* sehingga bisa juga disebut sebagai uji *Wald*.

Tabel 5. Uji Parsial Variabel X dengan Variabel Y

Variabel Prediktor	Wald	P-Value	Keputusan
X <sub>1</sub>	0,271	6,489	0,011 H <sub>0</sub> ditolak
X <sub>2</sub>	-0,767	22,495	0,000 H <sub>0</sub> ditolak
X <sub>3</sub>	0,224	4,389	0,036 H <sub>0</sub> ditolak
X <sub>4</sub>	-0,246	5,819	0,016 H <sub>0</sub> ditolak
Konstan	40,409	9,788	0,002

Berdasarkan Tabel 5 menunjukkan bahwa semua variabel prediktor memiliki nilai probabilitas (*p-value*) lebih kecil dari  $\alpha$ , maka variabel prediktor berpengaruh terhadap variabel respon. Sehingga model regresi logistik terbaik untuk memprediksi pengklasifikasian jurusan dengan menggunakan analisis regresi logistik adalah

$$f(X) = \frac{e^{40,409+0,271X_1-0,767X_2+0,224X_3-0,246X_4}}{1+e^{40,409+0,271X_1-0,767X_2+0,224X_3-0,246X_4}} \quad (15)$$

**Uji Kecocokan Model**

Uji ini berfungsi untuk menilai kesesuaian model regresi logistik dengan membandingkan hasil pengamatan dengan nilai dugaan. Statistik uji yang digunakan untuk menentukan kecocokan model disebut juga sebagai statistik kecocokan model (*Goodness of Fit*).

Tabel 6. Uji Kecocokan Model

Chi-square	Df	p-value
14,191	8	0,077

Berdasarkan Tabel 6 dapat diketahui bahwa nilai  $\hat{C} = 14,191 < t^2_{(0,05,8)} = 15,507$  dan nilai *p-value* sebesar  $0,077 > (0,05)$  maka diputuskan H<sub>0</sub> gagal ditolak. Sehingga dapat disimpulkan bahwa tidak ada perbedaan antara hasil pengamatan dengan nilai dugaan atau model regresi logistik tersebut layak untuk digunakan.

**Proses Klasifikasi**

Setelah diperoleh model regresi logistik pada Persamaan (15) maka dapat dilakukan proses klasifikasi. Misalkan untuk siswa pertama dengan nilai yaitu X<sub>1</sub>=77,25, X<sub>2</sub>=81,25, X<sub>3</sub>=81,25, dan X<sub>4</sub>=82. Selanjutnya nilai tersebut disubstitusikan ke dalam model regresi logistik, apabila nilai regresi logistik lebih besar dari nilai *cut-off* 0,5 atau  $f(x) \geq 0,5$ , maka siswa tersebut masuk ke dalam kelompok IPS sedangkan apabila nilai regresi logistik lebih kecil dari nilai *cut-off* 0,5 atau  $f(x) < 0,5$ , maka siswa tersebut masuk ke dalam kelompok IPA. Pada Tabel 7 diberikan contoh 10 siswa pertama yang diklasifikasikan untuk masing-masing kelompok.

Tabel 7. Hasil Klasifikasi dengan Regresi Logistik

No	Nilai f(x)	Jurusan Awal	Hasil Klasifikasi
1	0,04988	IPA	IPA
2	0,05121	IPA	IPA
3	0,01162	IPA	IPA
4	0,00609	IPA	IPA
5	0,06413	IPA	IPA
6	0,01093	IPA	IPA
7	0,20821	IPA	IPA
8	0,02554	IPA	IPA
9	0,02883	IPA	IPA
10	0,00682	IPA	IPA

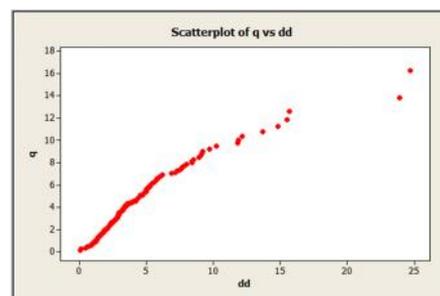
**Analisis Diskriminan**

**Uji Asumsi**

Sebelum dilakukan analisis diskriminan data harus memenuhi asumsi distribusi normal multivariat dan homogenitas matriks kovarians sebagai berikut:

**1. Uji Normal Multivariat**

Uji normal multivariat digunakan untuk mengetahui apakah data keempat variabel prediktor berdistribusi normal multivariat atau tidak.



Gambar 1. Plot uji multinormal data

Hasil uji normal multivariat dengan menggunakan *software* Minitab 16 diperoleh bahwa ada lebih dari 50%, yaitu 59,4737% dengan nilai  $t^2_{(0,5,4)} = 3,36$  yang berarti bahwa data telah memenuhi asumsi kenormalan (multinormal). Selain itu, dapat pula dilihat berdasarkan Q-Q *plot* pada Gambar 1, terlihat bahwa *plot* cenderung membentuk garis lurus, sehingga dapat disimpulkan bahwa data berdistribusi multinormal.

**2. Uji Kesamaan Matriks Varians Kovarian**

Uji Kesamaan matriks varians kovarians antara kelompok 1 dan kelompok 2 dilakukan menggunakan SPSS yaitu dengan menggunakan statistic *Box-M*.

Tabel 8. Uji Kesamaan Matriks Varians Kovarian

Box's M	p-value
19,972	0,035

Berdasarkan Tabel 8 diketahui bahwa nilai  $C = 19,972 \geq t_{(0,05,10)}^2 = 18,3070$  atau nilai  $p$ -value 0,035 (0,05) maka  $H_0$  ditolak. Sehingga dapat disimpulkan bahwa matriks varians kovarians antar kelompok adalah berbeda.

**Fungsi Diskriminan Kuadrat**

Nilai peluang awal dalam analisis diskriminan kuadrat akan digunakan dalam pembentukan fungsi diskriminan kuadrat untuk mengklasifikasikan siswa/i ke dalam kelompok jurusan yang sudah ada yaitu IPA dan IPS. Berdasarkan Persamaan (13) maka diperoleh nilai peluang awal (*prior probabilities*) masing-masing kelompok sebesar 0,632 untuk kelompok IPA dan 0,368 untuk kelompok IPS. Berdasarkan Persamaan (10) maka fungsi diskriminan kuadrat yang terbentuk yaitu,

$$Q_{IPS}(x) = \ln 0,368 - \frac{1}{2} \ln 855,5 - \frac{1}{2} (x - \bar{x}_{IPS})^T S_{IPS}^{-1} (x - \bar{x}_{IPS})$$

$$Q_{IPA}(x) = \ln 0,632 - \frac{1}{2} \ln 347,93 - \frac{1}{2} (x - \bar{x}_{IPA})^T S_{IPA}^{-1} (x - \bar{x}_{IPA})$$

Setelah terbentuknya fungsi diskriminan kuadrat, maka dapat dihitung skor diskriminan kuadrat untuk masing-masing siswa. Skor diskriminan kuadrat akan digunakan dalam mengklasifikasikan siswa ke dalam kelompok IPA atau IPS.

**Proses Klasifikasi**

Setelah diperoleh fungsi diskriminan maka dapat dilakukan proses klasifikasi dengan fungsi diskriminan kuadrat. Misalkan untuk siswa pertama dengan nilai  $X_1 = 77,25$ ,  $X_2 = 81,25$ ,  $X_3 = 81,25$ , dan  $X_4 = 82$ . Selanjutnya nilai tersebut disubstitusikan ke dalam fungsi diskriminan kuadrat, apabila nilai  $Q_{IPA} > Q_{IPS}$ , maka siswa tersebut masuk ke dalam kelompok IPA sedangkan apabila nilai  $Q_{IPA} < Q_{IPS}$ , maka siswa tersebut masuk ke dalam kelompok IPS. Hasil klasifikasi terdapat pada Tabel 9.

Tabel 9 Hasil Klasifikasi Dengan Analisis Diskriminan Kuadrat

No	$Q_{IPA}$	$Q_{IPS}$	Jurusan Awal	Hasil Klasifikasi
1	-4,98	-8,6	IPA	IPA
2	-4,14	-7,84	IPA	IPA
3	-4,19	-8,72	IPA	IPA
4	-4,76	-10,1	IPA	IPA
5	-4,62	-8,3	IPA	IPA
6	-4,3	-9,01	IPA	IPA
7	-4,46	-6,6	IPA	IPA
8	-5,76	-8,57	IPA	IPA
9	-3,99	-8,05	IPA	IPA
10	-4,21	-9,09	IPA	IPA

**Evaluasi Fungsi Klasifikasi**

Evaluasi fungsi klasifikasi baik analisis diskriminan kuadrat maupun regresi logistik dapat dilakukan dengan menghitung peluang kesalahan klasifikasi. Ukuran yang dapat digunakan adalah *Apparent Error Rate* (APER) berdasarkan Persamaan (14). Dari masing-masing bagian tersebut dihitung nilai APER-nya sehingga diperoleh ketepatan klasifikasinya.

**1. Ketepatan Klasifikasi Regresi Logistik**

Pada Tabel 10 diberikan jumlah obyek yang tepat diklasifikasikan dan yang salah untuk masing-masing kelompok.

Tabel 10 Klasifikasi Analisis Regresi Logistik

Jurusan	Jurusan yang diprediksi oleh model		Total
	IPA	IPS	
Jurusan	IPA	107	120
	IPS	19*	70
Total	126	64	190

Setelah diketahui ketepatan klasifikasi untuk tiap-tiap kelompok maka ketepatan hasil klasifikasi secara total dapat diketahui dengan menghitung nilai APER. Berdasarkan Persamaan (14) maka diperoleh nilai APER sebagai berikut :

$$APER = \frac{13 + 19}{190} = \frac{32}{190} = 0,1684$$

Berdasarkan nilai APER tersebut, dapat dijelaskan bahwa dengan fungsi diskriminan kuadrat dihasilkan proporsi salah klasifikasi sebesar 16,84% sehingga ketepatan klasifikasi yang dihasilkan adalah 83,16%.

**2. Ketepatan Klasifikasi Analisis Diskriminan Kuadrat**

Pada Tabel 11 diberikan jumlah obyek yang tepat diklasifikasikan dan yang salah untuk masing-masing kelompok. Setelah diketahui ketepatan klasifikasi untuk tiap-tiap kelompok maka ketepatan hasil klasifikasi secara total dapat diketahui dengan menghitung nilai APER. Berdasarkan Persamaan (14) maka diperoleh nilai APER sebagai berikut :

$$APER = \frac{12 + 18}{190} = \frac{30}{190} = 0,1579$$

Berdasarkan nilai APER tersebut, dapat dijelaskan bahwa dengan fungsi diskriminan kuadrat dihasilkan proporsi salah klasifikasi sebesar 15,79% sehingga ketepatan klasifikasi yang dihasilkan adalah 84,21%.

Tabel 11 Klasifikasi Analisis Diskriminan

Kuadratik				
		Jurusan yang diprediksi oleh model		Total
		IPA	IPS	
Jurusan	IPA	108	12*	120
	IPS	18*	52	70
Total		126	64	190

**Perbandingan Hasil Klasifikasi**

Berdasarkan analisis data yang telah dilakukan maka perbandingan ketepatan klasifikasi antara analisis regresi logistik dan analisis diskriminan kuadratik diberikan pada Tabel 12.

Tabel 12. Ketepatan Klasifikasi Kedua Metode

	Analisis Regresi Logistik	Analisis Diskriminan Kuadratik
APER	16,84 %	15,79 %
Ketepatan	83,16 %	84,21 %

Berdasarkan nilai ketepatan klasifikasi, pada analisis diskriminan kuadratik mempunyai ketepatan yang lebih tinggi. Dengan demikian metode analisis diskriminan kuadratik merupakan metode terbaik dalam pengklasifikasian jurusan di SMA Negeri 8 Samarinda.

**Kesimpulan**

Berdasarkan hasil penelitian yang diperoleh, maka penulis dapat menarik kesimpulan sebagai berikut :

1. Hasil pengklasifikasian jurusan SMA Negeri 8 Samarinda tahun ajaran 2014/2015 dengan menggunakan regresi logistik diperoleh hasil yaitu, untuk pengklasifikasian IPA dari 120 siswa yang diklasifikasikan ke Jurusan IPA, 107 siswa tepat diklasifikasikan ke jurusan IPA dan sisanya 13 siswa tidak tepat diklasifikasikan ke Jurusan IPA. Sedangkan untuk pengklasifikasian IPS dari 70 siswa yang diklasifikasikan ke Jurusan IPS, 51 siswa tepat diklasifikasikan ke jurusan IPS dan sisanya 19 siswa tidak tepat diklasifikasikan ke Jurusan IPS.
2. Hasil pengklasifikasian jurusan SMA Negeri 8 Samarinda tahun ajaran 2014/2015 dengan menggunakan analisis diskriminan kuadratik diperoleh hasil yaitu, untuk pengklasifikasian IPA dari 120 siswa yang diklasifikasikan ke Jurusan IPA, 108 siswa tepat diklasifikasikan ke jurusan IPA dan sisanya 12 siswa tidak tepat diklasifikasikan ke Jurusan IPA. Sedangkan untuk pengklasifikasian IPS dari 70 siswa yang diklasifikasikan ke Jurusan IPS, 52 siswa tepat diklasifikasikan ke

jurusan IPS dan sisanya 18 siswa tidak tepat diklasifikasikan ke Jurusan IPS.

3. Dengan menggunakan regresi logistik dan analisis diskriminan kuadratik diperoleh ketepatan klasifikasi masing-masing teknik analisis sebesar 83,16% dan 84,21%. Berdasarkan ketepatan klasifikasi kedua teknik analisis, dapat disimpulkan bahwa analisis diskriminan kuadratik lebih baik daripada regresi logistik dalam menyelesaikan masalah pengklasifikasian jurusan di SMA Negeri 8 Samarinda.

**Daftar Pustaka**

Firmansyah, Yudhi. 2010. *Perbandingan Analisis Diskriminan Mahalanobis dan Analisis Jaringan Syaraf Tiruan Dalam Menyelesaikan Masalah Pengelompokkan*. Skripsi tidak diterbitkan. Samarinda: Jurusan Statistika Fakultas MIPA Universitas Mulawarman.

Fractal. 2003. *Comparative Analysis of Classification Techniques*, A Fractal White Paper.

Hair, Jr. F. J., W. C Black, B. J Barry, dan R. E Anderson. 2010. *Multivariate Data Analysis*, 7<sup>th</sup> edn. New York: Pearson.

Hosmer D.W, dan S. Lemeshow. 2000. *Applied Logistic Regression*. New York: John Wiley & Sons, Inc.

Johnson, R. A., dan D. W Winchern. 2002. *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.

Kurt, I., M. Ture, dan A. T Kurum. 2006. *Comparing Performances of Logistic Regression Classification and Regression Tree, and Neural Network for Predicting Coronary Artery Disease, Expert Systems with Applications Vol 34, No. 1*.

Rachmatin, D dan K. Sawitri. 2010. Penerapan Prosedur Lachenbruch Pada Kasus Quadratic Discriminant Analysis, Bandung.

Satriya, Andi M. Ade. 2011. *Perbandingan Hasil Klasifikasi Regresi Logistik dengan Analisis Jaringan Syaraf Tiruan Pada Kasus Pengklasifikasian Jurusan Di SMA Negeri 5 Samarinda Tahun Ajaran 2010/2011*. Skripsi tidak diterbitkan. Samarinda: Jurusan Statistika Fakultas MIPA Universitas Mulawarman.

Sharma, S. 1996. *Applied Multivariate Techniques*. New York: John Wiley & Sons, Inc.

Supranto, J. 2004. *Analisis Multivariat Arti dan Interpretasi*. Jakarta : PT. Rineka Cipta.