

**PENGAJIAN PEMBENTUKAN MODEL KLASIFIKASI
DALAM PENGELOMPOKKAN JURUSAN SISWA DI SMA
(Studi Kasus: Siswa SMA Negeri Siau Timur Kabupaten Siau Tagulandang
Biaro Propinsi Sulawesi Utara)**

**Nelda Ponto Guru SMA Siau Timur Kabupaten Siau Tagulandang
0852 4017 2909, E-mail; Neldaponto@gmail.com**

ABSTRACT

Modeling that involve categorical response variables give important role in the classification problem. Statistical analysis is applied to solve this problem are discriminant analysis and multinomial logistic regression. Implementation of both methods against student of senior high school of East Siau data produce multinomial logistic regression as best method for classify the students into Science Program, Social Program, and Language Program. Classification accuracy of model from resampling is 88.1% and of model validation from Tagulandang Senior High School is 70.6%. The variables give significantly effect in classification students to Science Program or Language Program are Mathematics, English, Chemistry, and German, whereas, classification students into Social Program or Language Program are Economy, English, German, and History.

Key words: *Discriminant analysis, classification, multinomial logistic regression*

A. PENDAHULUAN

Sekolah Menengah Atas (SMA) merupakan jenjang pendidikan menengah yang mengutamakan penyiapan siswa untuk melanjutkan pendidikan yang lebih tinggi dengan pengkhususan.¹ Perwujudan pengkhususan tersebut berupa penjurusan. Penjurusan dilakukan pada saat memasuki kelas XI yakni, penjurusan pada Ilmu Pengetahuan Alam (IPA), Ilmu Pengetahuan Sosial (IPS) dan Bahasa.

Penjurusan merupakan upaya strategis dalam memberikan fasilitas kepada siswa untuk menyalurkan bakat, minat dan kemampuan yang dimilikinya yang dianggap paling potensial untuk dikembangkan secara optimal. Sehingga menempatkan siswa pada jurusan tertentu secara tepat berarti memberikan peluang kepada siswa untuk dapat berhasil pada masa yang akan datang. Kekurangtepatan dalam penempatan jurusan dapat mengakibatkan prestasi belajar rendah.² Hal ini disebabkan karena adanya perbedaan individual antara siswa disekolah yaitu, meliputi perbedaan kemampuan kognitif, motivasi berprestasi, minat dan kreativitas dan dengan adanya perbedaan individu tersebut, maka fungsi pendidikan tidak hanya dalam proses belajar mengajar tetapi meliputi bimbingan konseling, pemilihan dan penetapan siswa sesuai dengan kapasitas individual yang dimiliki.³

Agar kesalahan dalam pemilihan dan penetapan jurusan di SMA dapat diminimalisasi maka perlu ada upaya dalam mencari model yang terbaik. Beberapa analisis statistik telah banyak dikembangkan untuk membantu menyelesaikan masalah-masalah dalam bidang pendidikan, di antaranya adalah analisis regresi logistik, analisis diskriminan, pohon klasifikasi dan *Artificial Neural Network* (ANN). Dalam penelitian ini, analisis yang digunakan adalah analisis diskriminan dan regresi logistik multinomial. Analisis diskriminan digunakan untuk mengklasifikasikan individu ke dalam salah satu dari dua

¹ Depdiknas. *Pedoman Umum Pengembangan Penilaian*. (Departemen Pendidikan Nasional. 2004)

² Subiyanto. *Evaluasi Pendidikan Ilmu Pengetahuan Alam*. (Jakarta: Depdiknas.1988).

³ Snow RE. 1986. *Individual Differences and the Design Of Educational Programs in Journal Of Psychology*

kelompok atau lebih.⁴ Sedangkan regresi logistik multinomial digunakan untuk memodelkan hubungan antara peubah respon dengan kategori lebih dari dua (*polytomous*) dengan peubah penjelas kategorik dan atau kontinu. Melalui metode regresi logistik multinomial akan dihasilkan peluang dari masing-masing kategori respon yang akan dijadikan sebagai pedoman pengklasifikasian suatu pengamatan akan masuk dalam respon kategori tertentu berdasarkan nilai peluang terbesar.⁵

Penelitian tentang analisis diskriminan dan regresi logistik multinomial banyak dilakukan antara lain oleh, Maulias klasifikasi penjurusan siswa SMK Negeri 1 Tual Maluku dengan pendekatan analisis diskriminan dan regresi logistik multinomial.⁶ Metode klasifikasi menggunakan fungsi diskriminan.⁷

Tujuan utama yang ingin dicapai dalam penelitian ini, yaitu (1) menerapkan metode analisis diskriminan dan multinomial logistik untuk klasifikasi, (2) mengevaluasi peubah yang konsisten muncul dari metode analisis diskriminan dan multinomial logistik dengan teknik resampling.

B. LANDASAN TEORI

Analisis Diskriminan

Analisis diskriminan (*Discriminant Analysis*) adalah salah satu metode analisis multivariat yang bertujuan untuk memisahkan beberapa kelompok data yang sudah terkelompokkan dengan cara membentuk fungsi diskriminan.⁸ Untuk melakukan analisis diskriminan ada dua asumsi yang harus diperhatikan.⁹ yaitu :

1. Sejumlah p peubah bebas menyebar mengikuti sebaran normal ganda.
2. Matriks peragam berdimensi $p \times p$ dari peubah-peubah bebas dalam setiap kelompok harus homogen.

⁴ Johnson RA, Wichern DW. *Applied Multivariate Statistical Analysis* Ed ke-4. (New Jersey: Hall.1998)

⁵ Hosmer DW, Lemeshow S. 2000. *Applied Logistic Regression*. (New York: John Wiley & Sons.2008)

⁶ Maulias SS. 2009. *Klasifikasi Penjurusan Siswa SMK Negeri 1 Tual Maluku Tenggara dengan Pendekatan Analisis Diskriminan dan regresi Logistik Multinomial*. [tesis]. Surabaya. Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember.

⁷ Purnomo H. 2003. *Metode Klasifikasi Menggunakan Fungsi Diskriminan*. [skripsi]. Bogor: Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor

⁸ Johnson & Wichern. op. cit,

⁹ Dillon W, Goldstein M. 1984. *Multivariate Analysis*. New York: Wiley.

Uji sebaran normal ganda dapat dilakukan dengan plot khi-kuadrat¹⁰. Setiap vektor pengamatan dihitung jarak Mahalanobisnya dengan persamaan:

$$d_i^2 = (\underline{x}_i - \underline{\bar{x}}_i)S_i^{-1}(\underline{x}_i - \underline{\bar{x}}_i)$$

dimana setiap d_i^2 akan menyebar khi-kuadrat dengan p derajat bebas, bila p menyatakan banyak peubah.

Plot khi kuadrat akan memeriksa apakah statistik d_i^2 mengikuti sebaran khi kuadrat, yaitu dengan mengurutkan d_i^2 dari yang terkecil ke yang terbesar ($d_1^2 \leq d_2^2 \leq \dots \leq d_n^2$), kemudian memplotkan d_i^2 dengan $\chi_p^2((i - 0.5)/n)$. Tebaran titik-titik yang membentuk garis lurus menunjukkan kesesuaian pola sebaran d_i^2 terhadap sebaran khi-kuadrat yang berarti data berasal dari sebaran normal. Jika asumsi normal ganda tidak terpenuhi maka dapat digunakan analisis diskriminan logistik sebagai solusinya.¹¹

Uji kehomogenan matriks peragam dilakukan menggunakan uji Box' M, statistik uji yang digunakan adalah :¹²

$$M = \prod_{i=1}^g \left(\frac{|S_i|}{|S_w|} \right)^{v_i/2}$$

Statistik M bernilai antara 0 dan 1, jika nilainya mendekati 0, maka telah cukup bukti untuk menolak H_0 pada taraf α atau berarti ada matriks peragam populasi normal ganda yang berbeda sedangkan jika nilainya mendekati 1 berarti belum cukup bukti untuk menolak H_0 pada taraf α .

Sebaran statistik uji M dapat di dekati dengan sebaran F tahapan pengujiannya adalah:¹³

$$\text{menghitung, } c_1 = \left[\sum_{i=1}^g \frac{1}{v_i} - \frac{1}{\sum_{i=1}^g v_i} \right] \left(\frac{2p^2 + 3p - 1}{6(p+1)(g-1)} \right) \text{ dan } c_2 = \frac{(p-1)(p-2)}{6(g-1)} \left[\sum_{i=1}^g \left(\frac{1}{v_i^2} \right) - \frac{1}{(\sum_{i=1}^g v_i)^2} \right],$$

$$\text{serta } a_1 = \frac{1}{2}(g-1)p(p-1), \quad a_2 = \frac{a_1 + 2}{|c_2 - c_1^2|}$$

¹⁰ Johnson & Wichern. *op. cit.*

¹¹ Cacoullos T. *Discriminant Analysis and Applications*. (New York and London: Academic Press.1973)

¹² Rencher AC. *Methods of Multivariate Analysis*. (New York: Wiley.2002)

¹³ *Ibid*

$$b_1 = \frac{1-c_1-a_1/a_2}{a_1}, \quad b_2 = \frac{1-c_1-2/a_2}{a_2}$$

Jika $c_2 > c_1^2$, $F = -2b_1 \ln M$ mendekati sebaran $F_{\alpha(a_1, a_2)}$ dan

Jika $c_2 < c_1^2$, $F = -2a_2 b_2 \ln M / a_2 (1 + 2b_2 \ln M)$ mendekati sebaran $F_{\alpha(a_1, a_2)}$. Untuk kedua kasus tersebut, tolak H_0 jika $F > F_{\alpha(a_1, a_2)}$. Jika asumsi kehomogenan matriks peragam yang tidak terpenuhi maka analisis yang dapat digunakan adalah analisis diskriminan kuadratik.¹⁴

Pembentukan Fungsi Diskriminan

Fungsi diskriminan, misalkan terdapat g kelompok populasi dengan masing-masing ukuran contoh n_i , $i = 1, 2, \dots, g$, vektor peubah acak populasi ke- i adalah $X_i = (X_{i1}, X_{i2}, \dots, X_{ip})$, dan baris ke- j adalah x_{ij} maka vektor rata-rata populasi ke- i dapat dinyatakan sebagai berikut:¹⁵

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$$

dan vektor rata-rata populasi adalah

$$\bar{x} = \frac{\sum_{i=1}^g n_i \bar{x}_i}{\sum_{i=1}^g n_i} = \frac{\sum_{i=1}^g \sum_{j=1}^{n_i} x_{ij}}{\sum_{i=1}^g n_i}$$

Misalkan B matriks peragam antar kelompok, W matriks peragam dalam kelompok, dengan matriks keragaman total $T = W + B$. Fungsi diskriminan disusun dengan memaksimalkan rasio antara ragam antar kelompok dengan ragam antar kelompok. Jika fungsi diskriminan dinyatakan dengan $y = a'x$ maka yang ingin dicari adalah a_i sehingga $\lambda = \frac{a'Ba}{a'Wa}$ maksimum. Nilai λ yang maksimum merupakan akar ciri terbesar dari matriks $W^{-1}B$ dan a merupakan vektor ciri yang sepadan.¹⁶

¹⁴ Gnanadesikan R. Methods for Statistical Data Analysis of Multivariate Observations. (New York: John Wiley & Sons.1977).

¹⁵ Johnson & Wichern. *op. cit.*

¹⁶ Sharma S. *Applied Multivariate Techniques*. (New York. John Wiley & Sons.1996)

Peranan relatif suatu fungsi diskriminan ke- r dalam memisahkan anggota-anggota kelompok diukur dari persentase relatif akar ciri yang berhubungan dengan fungsi diskriminan berikut :

$$Y_r = \frac{\lambda_r}{\sum_{m=1}^s \lambda_m} \times 100\%, \quad r = 1, 2, \dots, s$$

dengan $s = \min(p, g - 1)$.

Semua fungsi diskriminan yang terbentuk perlu diuji untuk mengetahui banyaknya fungsi yang dapat menjelaskan perbedaan-peubah-peubah penjelas di antara g kelompok.¹⁷ Adapun pengujian fungsi diskriminan dapat dilakukan dengan menggunakan statistik *V-Barlett* melalui pendekatan khi-kuadrat, sebagai berikut :

$$V_r = \left[N - 1 - \frac{1}{2}(p + g) \right] \ln \sum_{m=r}^s (1 + \lambda_m)$$

Jadi bila $V_r < \chi^2_{\alpha, (p-r+1)(g-2)}$ artinya fungsi diskriminan ke- r masih diperlukan untuk menerangkan perbedaan p -peubah diantara g -kelompok. Kriteria masuknya individu kedalam kelompok ke- i bila:¹⁸

$$\sum_{m=1}^r (y_m - \bar{y}_{im})^2 = \sum_{m=1}^r [a'_m(x_{ij} - \bar{x}_k)]^2 \leq \sum_{m=1}^r [a'_m(x - \bar{x}_k)]^2$$

Dari analisis diskriminan ini dapat pula digunakan untuk mencari peubah-peubah asal yang dianggap dominan untuk digunakan dalam membedakan antar kelompok.

Press Q merupakan tes statistik untuk mengukur kekuatan dari pengklasifikasian fungsi diskriminan statistik Q dihitung dengan :¹⁹

$$Press Q = \frac{[N - (n * g)]^2}{N(g - 1)}$$

¹⁷ Dillon W, Goldstein M. *op. cit.*

¹⁸ Gaspersz V. Teknik Analisis dalam Penelitian Percobaan. Ed ke-1. (Tarsito Bandung, 1992).

¹⁹ Hair JF, Anderson RE, Tatham RL, Black WC. *Multivariate Data Analysis with Readings*. (New Jersey: Prentice-Hall, 1995).

dengan : N = Jumlah contoh total

n = Jumlah klasifikasi yang benar

g = Jumlah grup/kelompok

Statistik Q kemudian dibandingkan dengan nilai kritis (nilai khi-kuadrat untuk derajat bebas 1 pada taraf α tertentu). Jika statistik Q lebih besar dari nilai kritis berarti persentase hasil klasifikasi yang dihasilkan memiliki kekuatan dalam mengklasifikasikan objek.

Regresi Logistik Multinomial

Regresi logistik multinomial merupakan perluasan dari regresi logistik dengan respon biner yang dapat menangani peubah respon dengan kategori lebih dari dua. Untuk model regresi dengan peubah respon berskala nominal tiga kategori digunakan kategori peubah hasil Y yang dikode 0, 1, dan 2. Peubah Y terparameterisasi menjadi dua fungsi logit. Sebelumnya perlu ditentukan kategori respon yang digunakan sebagai kategori pembanding terlebih dahulu. Pada umumnya digunakan $Y=0$ sebagai pembanding. Untuk membentuk fungsi logit, akan dibandingkan $Y=1$ dan $Y=2$ terhadap $Y=0$. Bentuk model regresi yang berupa fungsi peluang dengan p peubah bebas seperti pada persamaan berikut ini:²⁰

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}$$

Transformasi logit akan menghasilkan dua fungsi logit sebagai berikut, dengan menetapkan $x_0 = 1$.

$$\begin{aligned} g_1(x) &= \ln \left[\frac{p(Y=1|x)}{p(Y=0|x)} \right] \\ &= \beta_{10} + \beta_{11} x_1 + \beta_{12} x_2 + \dots + \beta_{1p} x_p \\ &= x' \beta_1 \\ g_2(x) &= \ln \left[\frac{p(Y=2|x)}{p(Y=0|x)} \right] \\ &= \beta_{20} + \beta_{21} x_1 + \beta_{22} x_2 + \dots + \beta_{2p} x_p \\ &= x' \beta_2 \end{aligned}$$

²⁰ Hosmer DW, Lemeshow S. *Applied Logistic Regression*. (New York: John Wiley & Sons.2000)

Berdasarkan kedua fungsi logit tersebut maka didapatkan probabilitas respon atau model regresi logistik multinomial dengan peubah respon berskala nominal tiga kategori sebagai berikut ²¹

$$\pi_0(x) = \frac{1}{1 + \exp g_1(x) + \exp g_2(x)}$$

$$\pi_1(x) = \frac{\exp g_1(x)}{1 + \exp g_1(x) + \exp g_2(x)}$$

$$\pi_2(x) = \frac{\exp g_2(x)}{1 + \exp g_1(x) + \exp g_2(x)}$$

Dalam menduga model logit dengan peubah responnya berskala kualitatif, teknik pendugaan parameter yang layak digunakan adalah metode kemungkinan maksimum. Prinsip dari metode kemungkinan maksimum memberikan nilai dugaan parameter suatu fungsi kemungkinan. Fungsi kemungkinan yang ingin dimaksimalkan adalah :²²

$$L(\beta) = \prod_{i=1}^n f(Y = y_i | x_i)$$

dengan n = banyaknya pengamatan

Pengujian Kesesuaian Model

Pengujian Kesesuaian model dilakukan untuk memeriksa pengaruh peubah-peubah penjelas dalam model. Pengujian dilakukan untuk masing-masing parameter model (β). Pengujian secara simultan dilakukan dengan menggunakan uji G yaitu uji nisbah kemungkinan (*likelihood ratio test*).

Uji G untuk pengujian parameter β_i dengan hipotesis :

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

$$H_1: \text{minimal salah satu } \beta_i \neq 0$$

Statistik uji yang digunakan adalah statistik uji :

$$G = -2 \ln \left[\frac{L_0}{L_1} \right]$$

Jika H_0 benar, statistik G ini mengikuti sebaran χ^2 dengan derajat bebas p , Kriteria keputusan yang diambil adalah menolak H_0 jika $G_{hitung} \geq \chi^2_{\alpha(p)}$. Seandainya H_0

²¹ Ibid

²² Ibid

ditolak, maka selanjutnya dilakukan *uji Wald* untuk menguji parameter β_i secara parsial. Hipotesis yang diujikan adalah :²³

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0, \text{ dimana } i = 1, 2, \dots, p$$

Sedangkan statistik *uji Wald* sebagai berikut :

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

Statistik *uji Wald* mengikuti sebaran normal baku, dengan β_i sebagai penduga dan $\hat{SE}(\hat{\beta}_i)$ sebagai penduga galat baku β_i . Kriteria keputusan adalah menolak H_0 jika $|W| \geq Z_{\frac{\alpha}{2}}$ atau nilai $P \leq \alpha$.²⁴

Pereduksian peubah

Pereduksian peubah dalam regresi logistik dikenal dengan *stepwise logistic regression*. *Stepwise logistic regression* terdiri dari seleksi langkah maju dan eliminasi langkah mundur.

Metode seleksi langkah maju prosedur dimulai dengan intersep, kemudian peubah penjelas dimasukkan satu persatu ke dalam model dan diuji dengan khi-kuadrat. Apabila peubah penjelas tidak signifikan atau tidak nyata pada nilai α yang ditentukan, maka peubah tersebut dikeluarkan dari model dan sebaliknya peubah yang nyata atau signifikan akan dimasukkan ke dalam model. Sedangkan dalam metode eliminasi langkah mundur, prosedur dimulai dengan model penuh yaitu memasukkan seluruh peubah penjelas ke dalam model, kemudian diuji satu persatu. Jika ditemukan peubah penjelas yang tidak nyata pada nilai α yang ditentukan maka peubah tersebut dikeluarkan dari model. Pada tiap prosesnya peubah yang memiliki nilai-p yang terbesar akan berakhir ketika peubah penjelas yang berada dalam model memiliki nilai-p kurang dari 0.05. Analisis akan selesai jika tidak ada lagi peubah yang dapat dieliminasi dari model.²⁵

²³ *Ibid*

²⁴ *Ibid*

²⁵ Garson. 2010. Logistic Regression: Statnotes. North Carolina State University. <http://faculty.chass.ncsu.edu/garson/PA765/Logistic.htm/>. Diakses 31 Januari 2012.

Interpretasi Koefisien

Setelah diperoleh model terbaik, dilakukan interpretasi koefisien yang diperoleh. *Rasio odds* dapat juga dipergunakan untuk memudahkan interpretasi model. *Rasio odds* adalah ukuran asosiasi yang memperkirakan berapa besar kemungkinan peubah-peubah penjelas terhadap peubah respon.²⁶ *Rasio odds* untuk $Y = j$ terhadap $Y = k$ yang dihitung pada dua nilai (misal $x = a$ dan $x = b$) adalah :

$$\begin{aligned}\Psi(a, b) &= \frac{P(Y = j | x = a)/P(Y = k | x = a)}{P(Y = j | x = b)/P(Y = k | x = b)} \\ &= \exp[\beta_i(a - b)]\end{aligned}$$

Sehingga jika $a - b = 1$ maka $\Psi = \exp(\beta_i)$.

Ukuran Ψ selalu positif dan umumnya digunakan sebagai pendekatan risiko nisbi (*relative risk*). Untuk $\Psi = 1$ berarti bahwa $x = a$ memiliki risiko yang sama dengan $x = b$ untuk menghasilkan $Y = j$. Bila $1 < \Psi < \infty$ berarti $x = a$ memiliki risiko lebih tinggi Ψ kali daripada $x = b$, dan sebaliknya untuk $0 < \Psi < 1$. Jika peubah penjelas kontinu maka interpretasi koefisien dugaan tergantung pada unit particular dari peubah penjelas. Untuk peubah penjelas kontinu diperlukan unit perubahan sebesar c , maka *rasio odds* diperoleh dengan $\exp(c\beta_i)$.

C. METODE PENELITIAN

Data yang digunakan dalam penelitian ini adalah data sekunder yang diambil dari dua SMA di Kabupaten Siau Tagulandang Biaro yaitu SMA Negeri Tagulandang yang berjumlah 170 siswa dan SMA Negeri Siau Timur yang berjumlah 252 siswa. Data SMA Negeri Siau Timur Tahun ajaran 2009/2010 dan 2010/2011 digunakan untuk pemodelan. Sedangkan yang digunakan untuk validasi model adalah data SMA Negeri Tagulandang. Peubah-peubah yang diamati dalam penelitian ini terdiri 14 peubah numerik, yaitu rata-rata nilai mata pelajaran Agama, PPKn, Bahasa Indonesia, Bahasa Inggris, Matematika, Fisika, Biologi, Kimia, Sejarah, Geografi, Ekonomi, Sosiologi, TIK, Bahasa Jerman serta

²⁶ Hosmer dan Lemeshow . *op.cit.*

6 peubah kategorik, yaitu: jenis kelamin, pendidikan ayah, pendidikan ibu, pekerjaan ibu, pekerjaan ayah, pendapatan.

Langkah-langkah analisis yang dilakukan adalah sebagai berikut:

Tahap I : Melakukan analisis diskriminan dan multinomial logit

1. Analisis diskriminan

a. Uji asumsi diskriminan

- 1) Menguji kenormalan ganda dengan menggunakan plot khi-kuadrat.
- 2) Menguji kehomogenan matriks peragam gabungan menggunakan uji Box'M dengan membandingkan signifikansinya terhadap taraf nyata (α). Jika tidak terpenuhi maka tidak dapat membuat fungsi diskriminan linear.

b. Melakukan proses diskriminan

- 1) Mencari fungsi diskriminan bertahap.
- 2) Menghitung peranan relatif dari fungsi diskriminan yang didapat dengan mengukur menggunakan persentase relatif akar ciri dari fungsi diskriminan yang terbentuk.
- 3) Menguji keterandalan fungsi diskriminan dengan uji V-Bartlett yang menggunakan pendekatan khi-kuadrat.
- 4) Menghitung ketepatan klasifikasi fungsi diskriminan.
- 5) Menguji keterandalan hasil klasifikasi dari proses analisis diskriminan dengan statistik Q .

2. Analisis multinomial logit

- a. Membentuk fungsi logit dari peubah penjelas, yang mana transformasi logit yang digunakan adalah :

$$L_n \left[\frac{\pi_j(x)}{\pi_k(x)} \right] = \beta_0 + \beta_1 x_{1j} + \dots + \beta_p x_{pj}$$

- b. Menyelesaikan dugaan parameter dengan metode Kemungkinan Maksimum.
- c. Melakukan pengujian parameter dengan statistik uji-G untuk melihat peran seluruh peubah penjelas di dalam model secara simultan.

- d. Melakukan pengujian parameter secara parsial dengan statistik *uji Wald* untuk melihat pengaruh masing-masing peubah penjelas terhadap peubah respon.
- e. Melakukan interpretasi koefisien model regresi logistik multinomial dengan *Rasio odds*.
- f. Menghitung ketepatan klasifikasi model logistik.

Tahap II : Mengevaluasi model klasifikasi dengan *resampling*

1. Melakukan penarikan contoh terhadap data awal yang berjumlah 252 pengamatan. Penarikan contoh dilakukan secara acak dan dilakukan *resampling* dengan pemulihan sebanyak 30 kali, dengan masing-masing contoh berukuran sama dengan data awal. Hal ini bertujuan untuk melihat kekonsistenan peubah-peubah yang signifikan terhadap jurusan siswa di SMA. Konsistensi peubah-peubah tersebut akan digunakan sebagai pertimbangan untuk memilih peubah bebas dalam membangun model klasifikasi pengelompokan jurusan siswa di SMA.
2. Pada setiap *resampling*, data yang terambil kemudian dianalisis menggunakan analisis logistik multinomial dan analisis diskriminan. Dengan demikian, akan diperoleh masing-masing 30 model untuk model logistik multinomial dan model diskriminan.
3. Mengidentifikasi peubah-peubah yang signifikan untuk masing-masing model. Peubah yang signifikan pada 10 model logistik multinomial atau lebih akan digunakan untuk membangun model klasifikasi, demikian pula halnya untuk model diskriminan.
4. Melakukan analisis logistik multinomial dan analisis diskriminan terhadap data awal, dengan peubah bebas yang terpilih pada langkah 3.
5. Mengevaluasi model logistik multinomial dan analisis diskriminan yang diperoleh pada langkah 4 dengan melihat ketepatan klasifikasi masing-masing model.
6. Memilih model klasifikasi terbaik berdasarkan hasil pada langkah 5. Model yang terbaik adalah model yang memiliki tingkat klasifikasi paling tinggi.

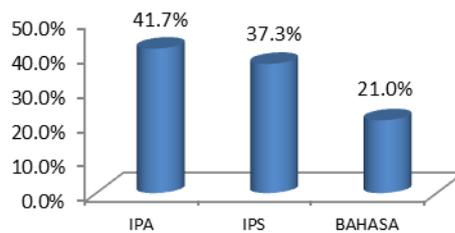
Tahap III : Melakukan validasi terhadap model terbaik

Validasi dilakukan dengan menerapkan model terbaik pada data SMA Negeri Tagulandang. Semakin tinggi tingkat ketepatan klasifikasi model yang dihasilkan pada tahap validasi, maka semakin baik kemampuan model dalam mengklasifikasikan siswa ke dalam jurusannya.

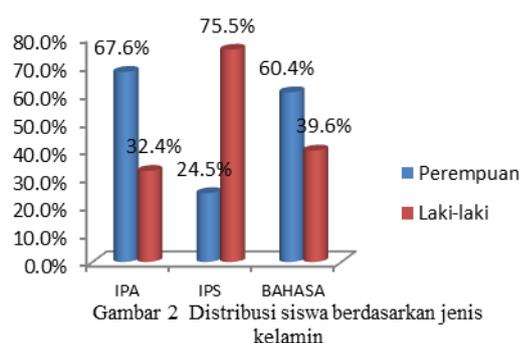
D. HASIL PENELITIAN DAN PEMBAHASAN

Karakteristik Siswa

Gambar 1 memperlihatkan Karakteristik siswa SMA Negeri Ulu Siau berdasarkan jurusan. Berdasarkan Gambar 1 umumnya siswa lebih memilih jurusan IPA daripada jurusan IPS dan Bahasa. Jurusan IPA memiliki persentase terbesar yaitu 41.7% (105 siswa), Jurusan IPS memiliki persentase 37.3% (94 siswa), dan jurusan Bahasa memiliki persentase 21.0% (53 siswa). Gambar 2 mendeskripsikan bahwa jurusan IPA dan Bahasa didominasi oleh perempuan yaitu jurusan IPA sebesar 67.6% (71 siswa) dan jurusan Bahasa sebesar 60.4% (32 siswa). Sebaliknya jurusan IPS didominasi oleh laki-laki yaitu sebesar 75.5% (71 siswa).



Gambar 1 Distribusi siswa berdasarkan jurusan



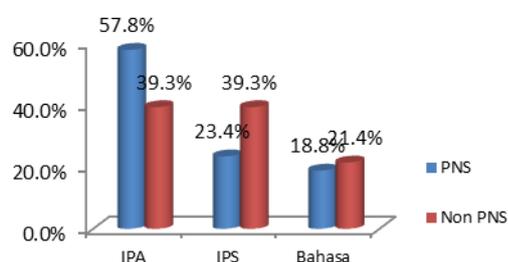
Tabel 2 menunjukkan persentase siswa berdasarkan jurusan yang diambil menurut pendidikan orang tua. Berdasarkan Tabel 2 mayoritas pendidikan orang tua siswa adalah Pendidikan Menengah, baik pendidikan ayah maupun pendidikan ibu, kemudian diikuti oleh Pendidikan Dasar dan Pendidikan Tinggi. Perbedaan latar belakang pendidikan orang tua siswa melahirkan persepsi yang berbeda tentang penjurusan. Hal ini mengindikasikan bahwa tinggi rendahnya tingkat pendidikan orang tua mempengaruhi pemilihan jurusan siswa di sekolah.

Tabel 2 Distribusi siswa berdasarkan pendidikan orang tua

Program Studi	Pendidikan Ayah			Pendidikan Ibu		
	Dasar	Menengah	Tinggi	Dasar	Menengah	Tinggi
IPA	29	50	26	35	48	22
	27.6%	47.6%	24.8%	33.3%	45.7%	21.0%
IPS	45	42	7	31	53	10
	47.9%	44.7%	7.4%	33.0%	56.4%	10.6%
Bahasa	22	25	6	23	21	9
	41.5%	47.2%	11.3%	43.4%	39.6%	17.0%

Gambaran mengenai persentase untuk masing-masing jenis pekerjaan orang tua siswa dapat dilihat pada Gambar 3. Mayoritas pekerjaan orang tua siswa untuk jurusan IPA adalah PNS yaitu sebesar 57.8%, sedangkan jurusan IPS dan

Bahasa mayoritas adalah Non PNS yaitu masing-masing sebesar 39.3% dan 21.4%.



Gambar 3 Distribusi siswa berdasarkan pekerjaan orang tua

Deskripsi Nilai Rapor Menurut Program Studi

Berdasarkan laporan hasil prestasi belajar siswa SMA Negeri Siau Timur, diperoleh nilai rata-rata kelas dan simpangan baku untuk setiap mata pelajaran. Mata pelajaran yang menjadi ciri khas program studi IPA adalah Matematika, Fisika, Kimia, Biologi. Pelajaran ciri khas program studi IPS adalah Sejarah, Ekonomi, Sosiologi, Geografi dan pelajaran ciri khas program studi Bahasa adalah Bahasa Indonesia, Bahasa Inggris, Bahasa Jerman.

Rata-rata nilai rapor siswa jurusan IPA lebih tinggi dibandingkan jurusan IPS dan Bahasa pada hampir seluruh mata pelajaran kecuali Bahasa Jerman, yang selisihnya pun tidak jauh berbeda dengan siswa jurusan Bahasa. Hal ini menunjukkan bahwa siswa jurusan IPA tidak hanya menguasai mata pelajaran ciri khas IPA saja, namun mereka juga menguasai mata pelajaran lainnya. Meskipun demikian, simpangan baku untuk rata-rata rapor siswa jurusan IPA relatif lebih besar daripada jurusan lain, artinya nilai mata pelajaran untuk jurusan IPA cenderung lebih beragam dibandingkan jurusan IPS dan Bahasa.

Analisis Diskriminan

Pemeriksaan Asumsi Dasar Diskriminan

a. Asumsi kenormalan ganda

Hasil plot quantil khi-kuadrat mengikuti trend linier baik untuk kelompok IPA, IPS maupun Bahasa. Kelinieran tersebut terlihat dari hubungan regresi linier

antara jarak dan khi-kuadrat yang nyata pada taraf $\alpha = 0.05$, seperti yang terlihat pada Tabel 4.

Tabel 4 Nilai-p regresi linier tiap kategori

Jurusan Siswa	Nilai-p
IPA	.0001
IPS	.0001
Bahasa	.0001

Hasil tersebut menunjukkan bahwa hipotesis kenormalan ganda pada analisis diskriminan terpenuhi.

b. Asumsi kehomogenan ragam

Hasil pengujian Box's M menunjukkan, bahwa matriks peragam untuk ketiga kategori sudah bersifat homogen dengan nilai signifikansi sebesar $0.073 > \alpha = 0.05$, maka asumsi kehomogenan matriks peragam peubah penjelas terpenuhi. Selanjutnya akan dilakukan analisis diskriminan.

Pembentukan Fungsi diskriminan

Fungsi diskriminan dibentuk dengan menggunakan metode *stepwise discriminant*. Tabel 5 menunjukkan bahwa terdapat 5 peubah penjelas yang cukup mewakili dalam melihat perbedaan antara kelompok IPA, IPS dan Bahasa. Kelima peubah itu adalah Bahasa Inggris, Fisika Kimia, Ekonomi dan TIK.

Tabel 5 Koefisien fungsi diskriminan

Peubah	Fungsi	
	1	2
Bahasa Inggris	-.028	.214
Fisika	.219	-.023
Kimia	.084	-.051
Ekonomi	-.038	-.189
TIK	-.060	.030
(Constant)	-11.933	1.141

Peubah bertanda positif, artinya setiap kenaikan satu satuan nilai peubah maka akan memberikan skor yang makin tinggi bagi fungsi diskriminan. Sedangkan peubah bertanda negatif, artinya setiap kenaikan satu satuan nilai peubah maka akan memberikan skor yang makin rendah bagi fungsi diskriminan.

Peranan relatif suatu fungsi diskriminan dalam memisahkan anggota-anggota kelompok diukur dari persentase relatif akar ciri yang berhubungan dengan fungsi diskriminan itu. Dengan memperhatikan akar ciri pada fungsi pertama adalah 92.5 dan fungsi kedua adalah 7.5, artinya persentase relatif yang dapat dijelaskan oleh fungsi diskriminan pertama adalah 92.5%, sedangkan sisanya sebesar 7.5 dijelaskan oleh fungsi diskriminan kedua.

Untuk mengetahui apakah fungsi diskriminan yang terbentuk dapat menjelaskan perbedaan peubah diantara 3 kelompok atau cukup melibatkan satu atau dua fungsi diskriminan. Adapun uji yang digunakan adalah statistik V-Bartlett melalui pendekatan uji khi-kuadrat. Hasil V-Bartlett menunjukkan hasil yang signifikan, berarti diskriminan sisa setelah di terangkan oleh diskriminan satu masih bersifat nyata secara statistik, dengan demikian diskriminan kedua masih diperlukan untuk menerangkan perbedaan peubah.

Dari fungsi diskriminan yang terbentuk melalui analisis diskriminan bertahap, lalu dilakukan pengklasifikasian. Pengklasifikasian suatu objek pengamatan baru pada fungsi diskriminan linier, dilakukan dengan mengacu pada

konsep jarak bahwa pengklasifikasian suatu objek x dipilih dari jarak objek pengamatan x terhadap vektor rata-ratanya yang terdekat/terkecil pada masing-masing jurusan. Rata-rata kelompok (*group centroids*) dari jurusan siswa mempunyai nilai yang besarnya berbeda, yaitu dapat dilihat pada Tabel 6.

Tabel 6 Nilai Rata-rata Kelompok

Jurusan Siswa	Fungsi	
	1	2
IPA	1.433	-.053
IPS	-1.177	-.302
Bahasa	-.751	.641

Ketepatan Klasifikasi Fungsi Diskriminan

Pengklasifikasian kelompok asal siswa menunjukkan bahwa 76.2% siswa yang diteliti dapat diklasifikasikan dengan benar ke dalam jurusannya sedang sisanya mengalami salah klasifikasi.

Tabel 7 Hasil klasifikasi analisis diskriminan

Observasi	Prediksi			
	(%)			
	IPA	IPS	Bahasa	Benar
IPA	88	7	10	83.8%
IPS	4	69	21	73.4%
Bahasa	4	14	35	66.0%
% Keseluruhan				76.2%

Statistik Q dari hasil klasifikasi kebenaran yang sebesar 76.2% adalah 208.3 dan nilai kritis $\chi^2_{0.05(1)}$ adalah 3.84. Terlihat statistik Q lebih besar dari nilai kritis sehingga klasifikasi kebenaran yang didapat sebesar 76.2%, secara statistik sudah baik.

Analisis Regresi Logistik Multinomial

Hasil pendugaan model penuh dengan melibatkan 20 peubah penjelas menghasilkan nilai G sebesar 399.939 dan nilai $p = 0.000 < 0.05$, sehingga dapat ditarik kesimpulan bahwa ada satu atau lebih peubah penjelas yang berpengaruh terhadap pengelompokkan jurusan siswa di SMA. Selanjutnya, dilakukan pemilihan peubah yang signifikan dengan menggunakan eliminasi langkah mundur. Hasil setelah seleksi menghasilkan model logistik multinomial sebagai berikut:

$$\hat{g}_1(x) = -32.094 + 0.212Agama - 0.417B.Ingggris + 0.708Mat + 0.561Kim - 0.070Sej - 0.031Eko - 0.483B.Jer - 0.328JK$$

$$\hat{g}_2(x) = -4.707 + 0.070Agama - 0.147B.Ing + 0.047Mat + 0.011Kim - 0.222Sej - 0.166Eko - 0.307B.Jer - 0.047JK$$

Evaluasi Kebaikan Model

Berdasarkan Tabel 7 diperoleh total ketepatan klasifikasi analisis regresi logistik multinomial sebesar 83.7 %. Artinya model logistik multinomial mampu mengklasifikasikan siswa ke dalam jurusannya dengan benar sebesar 83.7% dari total siswa keseluruhan. Untuk masing-masing jurusan, model logistik mampu mengklasifikasikan siswa IPA dengan benar sebesar 95.2%, siswa IPS 79.8%, dan siswa Bahasa 67.9%. Hal ini menunjukkan bahwa 4.8% siswa IPA, 20.2% siswa IPS, dan 32.1% siswa Bahasa terklasifikasikan ke jurusan lain.

Tabel 8 Hasil prediksi multinomial logistik

Observasi	Prediksi			Benar (%)
	IPA	IPS	Bahasa	
IPA	100	4	1	95.2%
IPS	6	75	13	79.8%
Bahasa	3	14	36	67.9%
% Keseluruhan				83.7%

Pembangunan Analisis Model Diskriminan dan Logistik Multinomial dari Hasil *Resampling*

Untuk mengevaluasi kekonsistenan peubah-peubah yang masuk dalam model, dilakukan fitting model dengan *resampling* 30 kali menggunakan analisis diskriminan dan regresi logistik multinomial terhadap data siswa SMA Negeri Siau Timur yang berjumlah 252 pengamatan. Peubah dikatakan konsisten apabila jumlah kemunculannya pada model memberikan pengaruh yang signifikan sebanyak 10 kali atau lebih.

Analisis Diskriminan

Penentuan peubah penjelas berdasarkan kekonsistenan yang diperoleh melalui *resampling* menghasilkan tiga peubah penjelas yang berpengaruh terhadap pengelompokkan jurusan siswa (Tabel 8). Tingkat ketepatan klasifikasi yang dihasilkan dari model analisis diskriminan adalah 60.3% (Tabel 11).

Tabel 8 Fungsi Diskriminan

Peubah	Fungsi	
	1	2
Bahasa Indonesia	.055	.269
Biologi	.187	-.117
Sosiologi	-.023	-.081
(Constant)	-16.616	-5.462

Tabel 11 Hasil Klasifikasi Analisis Diskriminan

Observasi	Prediksi			Benar
	IP	IPS	Bahas	
	A	a		
IPA	68	12	25	64.8 %
IPS	10	60	24	63.3%
Bahasa	11	18	24	45.3%
% Keseluruhan				60.3%

Analisis Regresi Logistik Multinomial

Peubah-peubah penjelas dengan signifikansi yang konsisten diperoleh sebanyak 15 peubah. Pemodelan analisis regresi logistik multinomial dibangun menggunakan peubah-peubah tersebut, hasilnya dapat dilihat pada Tabel 12. Untuk mengelompokkan siswa ke jurusan IPA, peubah yang berpengaruh adalah mata pelajaran Agama, Bahasa Inggris, Matematika, Kimia, Bahasa Jerman, dan Pendidikan Ayah. Sedangkan untuk mengelompokkan siswa ke jurusan IPS, peubah-peubah yang mempengaruhinya adalah mata pelajaran Bahasa Inggris, Biologi, Sejarah, Geografi, Ekonomi, Bahasa Jerman, dan Jenis Kelamin.

Model klasifikasi yang disusun menggunakan analisis regresi logistik multinomial memiliki tingkat ketepatan klasifikasi sebesar 88.1% (Tabel 13). Ketepatan klasifikasi model ini lebih tinggi jika dibandingkan dengan model sebelum dilakukan *resampling*. Hal ini berarti bahwa model logistik multinomial cukup baik dalam mengklasifikasikan siswa ke dalam jurusannya dengan tepat sebanyak 88.1% dari total siswa.

Tabel 13 Ketepatan Klasifikasi Model Logistik multinomial

Observasi	Prediksi			Benar	(%)
	IPA	IPS	Bahas		
a					
IPA	99	5	1		94.3%
IPS	4	83	7		88.3%
Bahasa	2	11	40		75.5%
% Keseluruhan					88.1%

Perbandingan Hasil Klasifikasi Analisis Diskriminan dan Logistik Multinomial

Perbandingan hasil klasifikasi analisis diskriminan dan regresi logistik multinomial dalam mengklasifikasi siswa ke dalam jurusan IPA, IPS, dan Bahasa dapat dilihat dari tingkat ketepatan klasifikasi. Semakin besar persentase ketepatan klasifikasi suatu model maka semakin baik dan akurat model tersebut dalam mengklasifikasi jurusan siswa.

Tabel 14 Ketepatan klasifikasi

Model	Model	Model
	Sebelum <i>Resampling</i>	Sesudah <i>Resampling</i>
	(%)	(%)
Analisis Diskriminan	76.2	60.3
Regresi Logistik	83.7	88.1
Multinomial		

Berdasarkan Tabel 14, diperoleh informasi bahwa tingkat ketepatan klasifikasi yang dihasilkan model regresi logistik multinomial telah melebihi 80%,

baik pada model sebelum maupun sesudah *resampling*. Sementara tingkat ketepatan klasifikasi model analisis diskriminan sebesar 76.2% pada model sebelum *resampling*, dan 60.3% pada model sesudah *resampling*. Hal ini menunjukkan bahwa regresi logistik multinomial mampu memberikan tingkat ketepatan klasifikasi yang lebih tinggi dibandingkan model analisis diskriminan. Dengan demikian, regresi logistik multinomial merupakan model terbaik yang dapat digunakan untuk menduga siswa memilih jurusan di SMA.

E. KESIMPULAN DAN SARAN

Simpulan

Penerapan analisis diskriminan dan regresi logistik multinomial dalam mengelompokkan siswa SMA Negeri Siau Timur memberikan ketepatan klasifikasi masing-masing sebesar 76.2% dan 83.7%. Berdasarkan penentuan peubah dari proses *resampling* sebanyak 30 kali, diperoleh model terbaik yaitu regresi logistik multinomial dengan tingkat ketepatan klasifikasi sebesar 88.1%.

Pengelompokkan siswa ke dalam jurusan IPA atau Bahasa dipengaruhi oleh mata pelajaran Matematika, Bahasa Inggris, Kimia, dan Bahasa Jerman. Pengelompokkan siswa ke dalam jurusan IPS atau Bahasa dipengaruhi oleh mata pelajaran Ekonomi, Bahasa Inggris, Bahasa Jerman, dan Sejarah. Validasi model yang dilakukan dengan menggunakan data SMA Negeri Tagulandang memberikan ketepatan klasifikasi sebesar 70.6%. Hal ini menunjukkan bahwa kemampuan model regresi logistik multinomial sudah cukup baik dalam mengelompokkan siswa ke dalam jurusannya.

Saran

Penelitian lanjutan dapat dilakukan untuk studi kasus dengan mengambil sampel yang berbeda. Peneliti berikutnya dapat melakukan penelitian tentang penjurusan, apakah benar nilai Biologi merupakan salah satu penentu untuk jurusan IPS sebagaimana penulis temukan pada penelitian ini.

F. DAFTAR PUSTAKA

- Cacoullos T. 1973. *Discriminant Analysis and Applications*. New York and London: Academic Press.
- Depdiknas. 2004. Pedoman Umum Pengembangan Penilaian. Departemen Pendidikan Nasional.
- Dillon W, Goldstein M. 1984. *Multivariate Analysis*. New York: Wiley.
- Garson. 2010. Logistic Regression: Statnotes. North Carolina State University. <http://faculty.chass.ncsu.edu/garson/PA765/Logistic.htm/> [31 Januari 2012].
- Gaspersz V. 1992. Teknik Analisis dalam Penelitian Percobaan. Ed ke-1. Tarsito Bandung.
- Ghozali. 2006. Aplikasi Analisis Multivariate dengan program SPSS. Universitas Diponegoro.
- Gnanadesikan R. 1977. *Methods for Statistical Data Analysis of Multivariate Observations*. New York: John Wiley & Sons.
- Hair JF, Anderson RE, Tatham RL, Black WC. 1995. *Multivariate Data Analysis with Readings*. New Jersey: Prentice-Hall.
- Hosmer DW, Lemeshow S. 2000. *Applied Logistic Regression*. New York: John Wiley & Sons.
- Johnson RA, Wichern DW. 1998. *Applied Multivariate Statistical Analysis* Ed ke-4. New Jersey: Hall.
- Maulias SS. 2009. Klasifikasi Penjurusan Siswa SMK Negeri 1 Tual Maluku Tenggara dengan Pendekatan Analisis Diskriminan dan regresi Logistik Multinomial. [tesis]. Surabaya. Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember.
- Morrison DF. 1990. *Multivariate Statistical Methods* Ed ke-3. McGraw-Hill Publishing Company.
- Purnomo H. 2003. Metode Klasifikasi Menggunakan Fungsi Diskriminan. [skripsi]. Bogor: Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor.
- Rencher AC. 2002. *Methods of Multivariate Analysis*. New York: Wiley.
- Sharma S. 1996. *Applied Multivariate Techniques*. New York. John Wiley & Sons.
- Snow RE. 1986. *Individual Differences and the Design Of Educational Programs in Journal Of Psychology*.
- Subiyanto. 1988. Evaluasi Pendidikan Ilmu Pengetahuan Alam. Jakarta: Depdiknas.