

Analisis Diskriminan dan Regresi Logistik Multinomial pada Proses Pengklasifikasian Siswa

Wd. Rifqah Amalliah Ndangi^{1*}, Resmawan², Ismail Djakaria³

¹ Badan Pusat Statistik Provinsi Gorontalo, Jl. Aloi Saboe No. 117, Kabila, Kab. Bone Bolango, Gorontalo, Indonesia

²³ Jurusan Matematika, Fakultas MIPA, Universitas Negeri Gorontalo, Jl. Jenderal Sudirman No. 06, Kota Gorontalo 96128, Gorontalo, Indonesia

* Penulis Korespondensi. Email: rifqah_amalliah@yahoo.com

ABSTRAK

Penetapan jurusan siswa yang tidak sesuai dengan kemampuan yang dimiliki dapat menyebabkan capaian hasil belajar siswa tidak optimal. Kasus seperti ini sering terjadi karena tidak optimalnya proses pengklasifikasian siswa sesuai dengan kemampuan yang dimiliki. Beberapa analisis statistik telah banyak dikembangkan untuk membantu menyelesaikan masalah-masalah klasifikasi diantaranya analisis diskriminan dan regresi logistik multinomial. Kedua analisis tersebut dapat digunakan sebagai metode pengklasifikasian objek, sehingga keduanya dapat dibandingkan berdasarkan ketepatan pengelompokkannya. Artikel ini membandingkan analisis diskriminan dan analisis regresi logistik multinomial dalam pengklasifikasian siswa ke kelompok IPA, IPS, Bahasa atau Agama. Kriteria perbandingan didasarkan pada kesalahan klasifikasi yang dikenal dengan *Apparent Error Rate (APER)*. Data yang digunakan adalah nilai rata-rata raport, nilai baca alquran, nilai hasil tes, nilai wawancara dan nilai tes potensi akademik siswa. Kedua analisis menunjukkan hasil yang sama bahwa variabel yang secara signifikan mempengaruhi pengklasifikasian siswa adalah nilai rata-rata raport dan nilai tes potensi akademik. Ketepatan klasifikasi yang ditunjukkan pada kedua metode ini juga menunjukkan persentasi yang sama dengan nilai 53.60%. Hal ini menunjukkan bahwa kedua metode ini sama baiknya digunakan dalam proses pengklasifikasian siswa. Analisis regresi logistik hanya lebih mudah digunakan karena tidak mempertimbangkan asumsi yang harus dipenuhi, sementara analisis diskriminan harus mempertimbangkan dua asumsi yaitu data berdistribusi normal multivariat dan kesamaan matriks varians kovarians.

Kata Kunci:

Analisis Diskriminan, Regresi Logistik Multinomial, Klasifikasi Siswa

Diterima:
08-03-2019

Disetujui:
16-05-2019

Online:
01-07-2019

ABSTRACT

Determination of student majors that are not in accordance with their abilities can cause the achievement of student learning outcomes are not optimal. Cases like this often occur because the process of classifying students is not optimal according to their abilities. Several statistical analyses have been developed to help solve classification problems including discriminant analysis and multinomial logistic regression. Both of these analyzes can be used as methods of classifying objects so that both can be compared based on the accuracy of the grouping. This article compares discriminant analysis and multinomial logistic regression analysis in classifying students into science, social studies, language or religion. Student Classification The comparison criteria are based on a classification error known as the Apparent Error Rate (APER). The data used are the average value of report cards, the reading value of the Koran, the value of performance, the value of the interview and the test scores of students' academic potential. Both analyzes show the same results that the variables that significantly influence the classification of students are the average report cards and the test scores of academic potential. The accuracy of the classification shown in these two methods also shows the same percentage with a value of 53.60%. This shows that both of these methods are equally well used in the process of classifying students. Logistic regression analysis is only easier to use because it does not consider the assumptions that must be met, while discriminant analysis must consider two assumptions, namely multivariate normal distribution data and similarity of covariance variance matrices.

Keywords:

Discriminant Analysis, Multinomial Logistic Regression, Student Classification

Received:

2019-03-08

Accepted:

2019-05-16

Online:

2019-07-01

DOI: <http://dx.doi.org/10.34312%2Fjjom.v1i2.2100>

1. Pendahuluan

Penetapan jurusan siswa pada satuan pendidikan menengah merupakan suatu yang tak terelakkan dari dunia pendidikan. Hal ini diperlukan untuk mengarahkan siswa menekuni karir yang diinginkan dan sesuai dengan kemampuan. Pemilihan jurusan yang dilakukan siswa sering kali menjadi masalah karena tidak sesuai dengan kemampuan yang berdampak pada kurang optimalnya hasil belajar yang diperoleh.

Terdapat beberapa faktor perlu dipertimbangkan dalam pemilihan jurusan seperti hasil nilai tes yang diberikan dan minat siswa sebelumnya. Sekolah juga perlu menetapkan standar nilai agar dapat digunakan sebagai bahan pertimbangan dalam memutuskan apakah siswa tersebut layak dikelompokkan pada jurusan yang dipilih [1]. Beberapa faktor yang digunakan dalam ini antara lain nilai rata-rata raport, nilai baca alquran, nilai performance, dan nilai tes potensi akademik. Untuk meminimalisasi kesalahan dalam pemilihan dan penetapan jurusan maka perlu upaya untuk mencari model yang terbaik. Beberapa analisis statistik telah banyak dikembangkan untuk membantu menyelesaikan masalah-masalah klasifikasi di antaranya adalah analisis regresi logistic dan analisis diskriminan.

Analisis diskriminan digunakan untuk memilih objek ke dalam dua kelas atau lebih [2]. Analisis diskriminan bisa dikelompokkan ke dalam dua jenis yaitu metode diskriminan dengan dua kategori dan metode diskriminan dengan lebih dari dua kategori. Metode diskriminan dengan dua kategori disebut dengan *Two-Group Discriminant Analysis* sedangkan metode diskriminan dengan lebih dari dua kategori disebut *Multiple Discriminant Analysis* [3]. Analisis diskriminan ini termasuk dalam analisis multivariat dengan metode dependensi. Analisis dependensi/ketergantungan bertujuan untuk

menjelaskan atau meramalkan nilai variabel tak bebas berdasarkan lebih dari satu variabel bebas yang mempengaruhinya [4].

Analisis regresi logistik multinomial digunakan untuk menemukan model terbaik untuk mendeskripsikan hubungan antara variabel dependen dan variabel independen [5]. Dinamakan regresi logistik, karena dalam analisis regresi ini pembentukan modelnya didasarkan atas kurva logistik. Nilai yang dihasilkan persamaan regresi logistik merupakan peluang kejadian yang digunakan sebagai ukuran untuk pengklasifikasian [6].

Pembahasan tentang analisis diskriminan dan regresi logistik banyak dilakukan antara lain dilakukan oleh Maulidya, *et.al* [7] yang membahas perbandingan analisis diskriminan dan regresi logistik pada klasifikasi konsumen berdasarkan tempat berbelanja di wilayah Taman Sidoarjo. Penelitian ini menghasilkan metode terbaik berupa regresi logistik biner dengan tingkat ketepatan klasifikasi untuk data learning sebesar 88.00% dan data testing sebesar 80.00%. Artikel ini memberikan gambaran tentang perbandingan analisis diskriminan dan regresi logistik pada proses pengklasifikasian siswa dengan melibatkan variabel dependen 4 kategori, yaitu Jurusan Ilmu Pengetahuan Alam, Ilmu Pengetahuan Sosial, Bahasa dan Agama.

2. Metode

Penelitian ini melibatkan sampel seluruh siswa kelas X Madrasah Aliyah Negeri Model Gorontalo. Variabel yang digunakan berupa variabel dependen dan independent. Variabel dependen yang digunakan adalah ketersediaan jurusan yang dapat dipilih oleh siswa, yaitu jurusan Ilmu Pengetahuan Alam, Ilmu Pengetahuan Sosial, Bahasa, dan Agama. Adapun variabel independen yang digunakan antara lain nilai rata-rata rapor, nilai membaca alquran, nilai hasil tes, nilai hasil wawancara dan nilai tes potensi akademik.

2.1. Analisis Diskriminan

Analisis diskriminan dilakukan dengan langkah-langkah sebagai berikut:

1. Uji Asumsi
 - a. Uji asumsi distribusi normal multivariat. Pengujian ini menggunakan plot jarak Mahalanobis d_j^2 dengan distribusi chi-square $\chi_p^2(\frac{j-1}{2})$. Untuk menilai normalitas dari kumpulan data didasarkan pada jarak umum kuadrat [2]:

$$d_j^2 = (x_j - \bar{x})' S^{-1} (x_j - \bar{x}), j = 1, 2, \dots, n$$

Plot menyerupai garis lurus dan lebih dari 50% menunjukkan data berasal dari populasi berdistribusi normal.

- b. Uji asumsi homogenitas matriks varians kovarians menggunakan uji Box's M. Uji ini merupakan uji matriks varians kovarians populasi secara keseluruhan. Untuk pendekatan χ_2 dihitung:

$$c_1 = \left[\sum_{i=1}^k \frac{1}{v_i} - \frac{1}{\sum_{i=1}^k v_i} \right] \left[\frac{2p^2 + 3p - 1}{6(p+1)(k-1)} \right]$$

$$\ln M = \frac{1}{2} \sum_{i=1}^k v_i \ln |S_i| - \frac{1}{2} \left(\sum_{i=1}^k v_i \right) \ln |S_{spl}|$$

$$u = -2(1 - c_1) \ln M$$

Kriteria tolak H_0 jika $u > \chi^2_{\left(\frac{1}{2}(k-1)p(p+1), \alpha\right)}$ [5].

2. Didalam mengevaluasi tingkat signifikansi uji statistik ada perbedaan kriteria statistika yang digunakan antara metode estimasi secara simultan dan estimasi dengan metode *stepwise*. Karena metode yang digunakan adalah *simultan* maka ada beberapa uji yang digunakan salah satunya adalah Wilk's Lambda. Adapun formula Wilks Lambda sebagai berikut:

$$\Lambda = \frac{SS_w}{SS_t} = \frac{\sum_{i=1}^r \sum_{j=1}^k (X_{i,j} - \bar{X}_{gj})^2}{\sum_{i=1}^r \sum_{j=1}^k (X_{i,j} - \bar{X}_{gj})^2 + \sum_{g=1}^g n_g (\bar{X}_{gi} - \bar{X}_j)^2}$$

3. Memilih metode pembentukan fungsi diskriminan. Metode yang dipilih dalam penelitian ini adalah metode secara menyeluruh atau simultan. Metode ini tepat digunakan karena peneliti menganggap bahwa secara teoritis semua variabel independen adalah tepat. Dengan alasan teoritis ini, peneliti ingin memasukkan semua variable independen di dalam analisis dan tidak tertarik untuk melihat hasil berdasarkan pada variabel independen yang mampu membedakan atau mendiskriminasikan kelompok dengan baik [3].
4. Mengestimasi fungsi Diskriminan. Fungsi diskriminan yang diperoleh digunakan sebagai model yang akan digunakan untuk menghitung skor diskriminan. Fungsi diskriminan linier untuk klasifikasi k populasi adalah sebagai berikut:

$$\hat{d}_i(x) = \bar{x}'_i S_{pooled}^{-1} x - \frac{1}{2} \bar{x}'_i S_{pooled}^{-1} \bar{x}_i + \ln(p_i)$$

untuk $i = 1, 2, \dots, g$; $p_i =$ probabilitas prior. Jika probabilitas prior tidak diketahui, maka $p_i = \frac{1}{g}$. Aturan pengelompokkannya yaitu alokasi x ke π_k jika skor diskriminan linear $\hat{d}_k(x) = (\text{maks } \hat{d}_1(x), \hat{d}_2(x), \dots, \hat{d}_g(x))$ [2].

5. Menghitung tingkat ketepatan pengelompokkan hasil prediksi fungsi diskriminan.
6. Menginterpretasikan hasil.

2.2. Analisis Regresi Logistik Multinomial

Analisis regresi logistik multinomial dilakukan dengan langkah-langkah sebagai berikut:

1. Mengestimasi parameter. Untuk menduga model logit dengan peubah responnya berskala kualitatif, teknik pendugaan parameter yang layak digunakan adalah metode *maksimum likelihood*. Hal ini bertujuan untuk mendapatkan model yang akan digunakan dalam melakukan pengklasifikasian [8]. Fungsi *maksimum likelihood* adalah :

$$l(\beta) = \prod_{i=1}^n \pi_0(x_i)^{y_{0i}} \pi_1(x_i)^{y_{1i}} \pi_2(x_i)^{y_{2i}}$$

dengan $\sum ij = 1$. Fungsi *log likelihood* adalah :

$$l(\beta) = \sum_{i=1}^n y_{1i} g_1(x_i) + y_{2i} g_2(x_i) - \ln (1 + e^{g_1(x_i)} + e^{g_2(x_i)})$$

2. Uji signifikansi parameter secara serentak dengan statistik uji *Gatau likelihood ratio* menggunakan hipotesis,

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

$$H_1 : \text{minimal ada satu } \beta_j \neq 0, \text{ dengan } j = 1, 2, \dots, p$$

Statistik Uji:

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

dengan

L_0 = likelihood tanpa peubah bebas

L_1 = likelihood dengan peubah bebas

Kriteria Uji :

Tolak H_0 jika $G > \chi^2_{(ab, \alpha)}$ atau p-value $< \alpha$.

3. Secara individual menggunakan uji Wald dengan hipotesis,

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0 ; \text{ dengan } j = 1, 2, \dots, p$$

Statistik Uji :

$$W = \left(\frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \right)^2$$

Kriteria Uji :

Tolak H_0 jika $W > \chi^2_{db, \alpha}$ atau p-value $< \alpha$.

4. Menghitung ketepatan klasifikasi. Kriteria perbandingan teknik klasifikasi didasarkan pada kesalahan klasifikasinya yang dikenal dengan *Apparent Error Rate* (APER) [2]. Adapun APER dihitung dengan membuat tabel klasifikasi yang diberikan pada Tabel 1.

Tabel 1. Matriks konfusi hasil klasifikasi

	Kelompok Hasil Klasifikasi		
		π_1	π_2
Kelompok Asli	π_1	n_{1C}	$n_{1M} = n_1 - n_{1C}$
	π_2	$n_{2M} = n_2 - n_{2C}$	n_{2C}

dengan

n_{1C} = jumlah item π_1 yang benar diklasifikasikan sebagai item π_1

n_{1M} = jumlah item π_1 yang salah diklasifikasikan sebagai item π_1

n_{2C} = jumlah item π_2 yang benar diklasifikasikan

n_{2M} = jumlah item π_2 yang salah diklasifikasikan

5. Interpretasikan hasil.

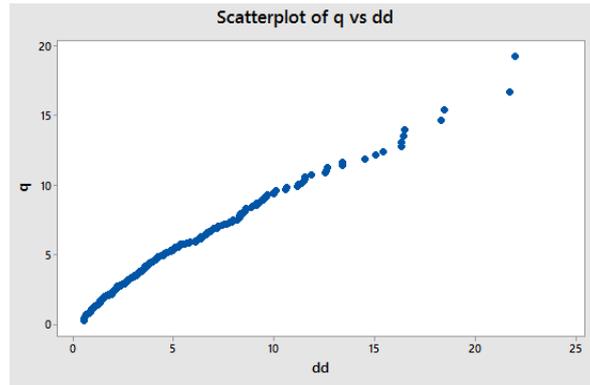
3. Hasil dan Pembahasan

3.1. Analisis Diskriminan

3.1.1. Uji Asumsi

a. Uji Distribusi Normal Multivariat

Pengujian asumsi distribusi normal multivariat pada data penjurusan menggunakan plot jarak Mahalanobis d_j^2 dengan distribusi chi-square $\chi_p^2 \left(\frac{j-1}{2} \right)$. Plot distribusi normal multivariate dapat dilihat pada Gambar 1.



Gambar 1. Plot distribusi normal multivariat

Kriteria pemenuhan asumsi dilakukan secara visual yaitu plot menyerupai garis lurus yang menunjukkan bahwa data distribusi normal [2].

b. Uji Kesamaan Matriks Varians Kovarians

Matriks kovarian dari semua variabel independen seharusnya sama. Pengujian asumsi ini menggunakan uji Box's M yang ditunjukkan pada Tabel 2.

Tabel 2. Uji kesamaan matriks varians kovarians

Hasil Tes		
Box's M		82,49600
	Approx.	1,75800
F	df1	45,00000
	df2	7,228E4
	Sig.	0,00100

Hasil yang diperoleh pada Tabel 2 menunjukkan nilai Signifikan = 0,001 < α = 0,05, yang berarti bahwa H_0 ditolak. Dengan demikian, matriks varians kovarians antar kelompok dinyatakan berbeda.

3.1.2. Uji Vektor Rataan

Pengujian vektor rata-rata ini digunakan untuk menguji apakah terdapat perbedaan yang signifikan setiap variabel independen antar kelompok atau tidak. Uji ini menggunakan Wilks' Lambda yang ditunjukkan pada Tabel 3.

Tabel 3. Uji vektor rata-rata

Nilai	Wilks' Lambda	F.	df1	df2	Sig.	Ket
X_1	0,886	12,280	3	285	0,000	Signifikan
X_2	0,978	2,177	3	285	0,091	Tidak Signifikan
X_3	0,968	3,177	3	285	0,025	Signifikan
X_4	0,978	2,168	3	285	0,092	Tidak Signifikan
X_5	0,933	6,827	3	285	0,000	Signifikan

Hasil yang diperoleh pada Tabel 3 menunjukkan variabel yang secara signifikan mempengaruhi pengklasifikasian siswa antara lain nilai rata-rata raport, nilai hasil tes,

dan nilai tes potensi akademik. Adapun variabel yang tidak signifikan mempengaruhi pengklasifikasian siswa adalah nilai baca alquran dan nilai hasil wawancara.

3.1.3. Fungsi Diskriminan

Analisis diskriminan bertujuan membuat suatu fungsi diskriminan dari variabel independen yang bisa mendiskriminasi atau membedakan kategori variabel dependen. Fungsi diskriminan linier yang terbentuk dari data yang diperoleh antara lain:

- 1) $FD1 = 0,144X_1 + 0,011 X_2 + 0,028 X_3 + 0,005 X_4 + 0,046 X_5$
- 2) $FD2 = -0,080 X_1 - 0,033 X_2 + 0,085 X_3 - 0,020 X_4 + 0,061 X_5$
- 3) $FD3 = -0,015 X_1 + 0,082 X_2 + 0,018 X_3 - 0,104 X_4 + 0,000 X_5$

3.1.4. Hasil Klasifikasi

Hasil klasifikasi ini didasarkan pada skor diskriminan linier terbesar dibandingkan dengan titik tengah optimum *cutting score (m)*. Hasil yang diperoleh dapat dilihat pada Tabel 4.

Tabel 4. Hasil klasifikasi

		Kelompok Grup Prediksi				Total	
		Jurusan	IPA	IPS	Bahasa		Agama
Asli	Jumlah	IPA	133	10	0	1	144
		IPS	36	20	1	1	58
		Bahasa	35	11	1	0	47
		Agama	30	8	1	1	40
Asli	Persentase	IPA	92,40	6,90	0,00	0,70	100,00
		IPS	62,10	34,50	1,70	1,70	100,00
		Bahasa	74,50	23,40	2,10	0,00	100,00
		Agama	75,00	20,00	2,50	2,50	100,00
Ketepatan Klasifikasi						53,60	

Hasil yang ditampilkan pada Tabel 4 menunjukkan bahwa ketepatan pengklasifikasian analisis diskriminan sebesar 53,60%.

3.2. Analisis Regresi Logistik Multinomial

3.2.1. Estimasi Parameter

Langkah awal yang dilakukan dalam analisis ini adalah estimasi parameter. Metode yang digunakan dalam mengestimasi parameter adalah maksimum likelihood. Model regresi logistik multinomial yang diperoleh dari data adalah sebagai berikut:

- 1) $g_1x = -7,331 + 0,042 X_1 - 0,007 X_2 + 0,050 X_3 - 0,013 X_4 + 0,052 X_5$
- 2) $g_2x = 9,030 - 0,093 X_1 - 0,013 X_2 + 0,020 X_3 - 0,021 X_4 + 0,000 X_5$
- 3) $g_3x = 4,389 - 0,077 X_1 - 0,026 X_2 + 0,045 X_3 - 0,014 X_4 + 0,033 X_5$

3.2.2. Uji Simultan

Uji simultan merupakan uji signifikansi yang digunakan untuk mengetahui variabel yang secara simultan mempengaruhi pembentukan model regresi logistik multinomial. Uji simultan ini menggunakan uji rasio likelihood yang dapat dilihat pada Tabel 5.

Tabel 5. Uji simultan

Effect	Kriteria Model Sesuai		Uji Rasio Likelihood		
	-2 log likelihood	Chi-Square	df	Sig.	Ket.
Intercept	688,497	28,473	3	0,000	
X_1	682,037	22,012	3	0,000	Signifikan
X_2	662,150	2,125	3	0,547	Tidak Signifikan
X_3	667,938	7,914	3	0,048	Signifikan
X_4	660,626	0,601	3	0,896	Tidak Signifikan
X_5	668,022	7,998	3	0,046	Signifikan

Hasil yang diperoleh pada Tabel 5 menunjukkan variabel independen yang signifikan mempengaruhi pembentukan model regresi logistik multinomial antara lain nilai rata-rata raport, nilai hasil tes, dan nilai tes potensi akademik. Adapun variabel yang tidak signifikan mempengaruhi model adalah nilai baca alquran dan nilai wawancara.

3.2.3. Uji Parsial

Uji parsial merupakan uji signifikansi setiap parameter dalam model dengan menggunakan Uji Wald. Uji ini dilakukan untuk mengetahui apakah setiap variabel independen dapat diandalkan untuk membangun model atau tidak dalam proses klasifikasi jurusan.

Tabel 6. Uji parsial

Jurusan		Wald	Sig.	Exp(B)	Ket.
IPA	Intercept	3,644	0,056		
	X_1	1,044	0,307	1,043	Tidak Signifikan
	X_2	0,164	0,685	0,993	Tidak Signifikan
	X_3	6,779	0,009	1,052	Signifikan
	X_4	0,280	0,597	0,987	Tidak Signifikan
	X_5	4,366	0,037	1,053	Signifikan
IPS	Intercept	5,029	0,025		
	X_1	0,701	0,030	0,991	Signifikan
	X_2	0,507	0,477	0,987	Tidak Signifikan
	X_3	0,945	0,331	1,020	Tidak Signifikan
	X_4	0,596	0,440	0,979	Tidak Signifikan
	X_5	0,000	0,996	1,000	Tidak Signifikan
Bahasa	Intercept	1,030	0,310		
	X_1	3,014	0,083	0,925	Tidak Signifikan
	X_2	1,763	0,184	0,974	Tidak Signifikan
	X_3	3,673	0,055	1,046	Tidak Signifikan
	X_4	0,261	0,609	0,986	Tidak Signifikan
	X_5	1,269	0,260	1,034	Tidak Signifikan

Berdasarkan hasil yang diperoleh pada Tabel 6, dapat ditunjukkan bahwa terdapat 2 variabel independen yang membedakan siswa yang memilih jurusan IPA dan Agama, yaitu nilai hasil tes dan nilai tes potensi akademik. Nilai hasil tes dan nilai tes potensi akademik siswa jurusan IPA cenderung lebih tinggi dibanding jurusan Agama.

Adapun yang membedakan siswa jurusan IPS dan Agama hanya satu variabel independen, yaitu nilai rata-rata raport. Nilai rata-rata raport jurusan IPS cenderung lebih rendah dibanding jurusan Agama.

Selanjutnya tidak terdapat variabel independen yang membedakan siswa pada jurusan Bahasa dan Agama.

3.2.4. Hasil Klasifikasi

Peluang ketepatan klasifikasi jurusan dapat dilihat pada Tabel 7 sebagai berikut:

Tabel 7. Hasil klasifikasi

Observasi	Prediksi				Persentase Kebenaran Prediksi
	IPA	IPS	Bahasa	Agama	
IPA	133	11	0	0	92,40
IPS	36	20	1	1	34,50
Bahasa	35	11	1	0	2,10
Agama	30	8	1	1	2,50
Total Persentase	81,00	17,30	1,00	0,70	53,60

Berdasarkan Tabel 7, diperoleh hasil klasifikasi dalam analisis regresi logistik sebesar 53.63% yang terklasifikasi dengan benar dalam pemilihan jurusan.

3.3. Perbandingan Hasil Klasifikasi

Analisis diskriminan dan regresi logistik multinomial merupakan dua metode yang dapat digunakan sebagai metode pengklasifikasian objek, sehingga kedua metode tersebut dapat dibandingkan berdasarkan ketepatan pengelompokkannya. Perbandingan hasil klasifikasi dapat dilihat pada Tabel 8.

Tabel 8. Perbandingan hasil klasifikasi

Jurusan	Persentase Ketepatan Klasifikasi	
	Analisis Diskriminan	Regresi Logistik Multinomial
IPA	92,40	92,40
IPS	34,50	34,50
Bahasa	2,10	2,10
Agama	2,50	2,50
Keseluruhan	53,60	53,60

Hasil yang diperoleh pada Tabel 8 menunjukkan bahwa ketepatan klasifikasi jurusan untuk kedua metode sama. Ketepatan klasifikasi jurusan IPA sebesar 92,40%, IPS sebesar 34,50%, Bahasa sebesar 2,10% dan Agama 2,50%. Secara keseluruhan ketepatan klasifikasi untuk kedua metode yaitu sebesar 53,60%. Jadi dapat disimpulkan bahwa kedua metode tersebut sama baiknya digunakan dalam pengklasifikasian jurusan.

4. Kesimpulan

Penelitian ini menunjukkan bahwa analisis diskriminan dan regresi logistik multinomial dapat digunakan sama baiknya dalam pengklasifikasian siswa, dengan persentase ketepatan hasil klasifikasi analisis masing-masing sebesar 53,60%. Kedua metode juga

menghasilkan identifikasi yang sama bahwa faktor-faktor yang secara signifikan mempengaruhi pengklasifikasian siswa adalah nilai rata-rata raport dan nilai tes potensi akademik.

Referensi

- [1] Zulkifli, J.M., 2014, Pendekatan Regresi Logistik Multinomial Pada Klasifikasi Pemilihan Jurusan Siswa SMA Negei 5 Malang, *Jurnal Mahasiswa Statistika*, Vo. 2, No. 5: 349 - 352
- [2] Johnson, R.A. & Wichern, D.W., 2002, *Applied Multivariate Statistical Analysis* Ed ke- 5, New Jersey: Hall
- [3] Widarjono, A., 2015, *Analisis Multivariat Terapan edisi kedua*, Yogyakarta: UPP STIM YKPN
- [4] Supranto, 2004, *Analisis Multivariat: Arti dan Interpretasi*, Jakarta: PT Asdi Mahasatya
- [5] Hosmer, D.W. & Lemeshow, S., 2002, *Applied Logistic Regression Second Edition*, United States of American: John Wiley & Sons Inc
- [6] Qudratullah, M.F., 2011, Misklasifikasi Mahasiswa Baru F Saintek UIN Sunan Kalijaga Jalur Tes Tertulis dengan Analisis Regresi Logistik, *Jurnal Matematika Murni dan Aplikasi*, Vol. 1, No. 4, Program Studi Matematika, UIN Sunan Kalijaga Yogyakarta
- [7] Maulidya, Sutanto, H.T., & Oktaviarina, A., 2014, Perbandingan Analisis Diskriminan dan Regresi Logistik: Studi Kasus Klasifikasi Konsumen Berdasarkan Tempat Berbelanja di Wilayah Taman Sidoarjo, *Jurnal MATHunesa*, Vol. 3, No.1, Jurusan Matematika, Universitas Negeri Surabaya
- [8] Rencher, A.C., 2002, *Methods of Multivariate Analysis Second Edition*, New York: John Wilwy and Sons