

Perbandingan Analisis Diskriminan Linier, Diskriminan Linier *Robust* dan Regresi Logistik Biner

(Studi Kasus Pada Penjurusan Bidang IPA / IPS Siswa Tingkat SMA Negeri 1 Bangorejo Banyuwangi)

Marino¹, I Made Tirta², Yuliani Setia Dewi³

¹Guru Matematika SMA Negeri 1 Bangorejo Banyuwangi, Dinas Pendidikan Banyuwangi
marinoyustari@yahoo.co.id

²Staf Pengajar Jurusan Magister Matematika FMIPA Universitas Jember,

³Staf Pengajar Jurusan Magister Matematika FMIPA Universitas Jember,

Abstrak

Penelitian ini membandingkan analisis diskriminan linier, diskriminan linier *robust* dan regresi logistik biner untuk mengelompokkan siswa SMA Negeri 1 Bangorejo kedalam kelompok IPA/IPS. Data yang digunakan adalah data nilai raport dan psikotes siswa kelas X semester 2 tahun pelajaran 2012-2013 SMAN 1 Bangorejo Banyuwangi. Data yang digunakan merupakan data terkontaminasi *outlier* sebesar 6,70%. Untuk mengetahui performa terhadap keberadaan *outlier*, maka dilakukan simulasi secara berulang-ulang mengaplikasikan analisis diskriminan linier, diskriminan linier *robust* dan regresi logistik biner dengan besar sampel bervariasi yaitu $n_1=40$, $n_2=80$, $n_3=120$ dan $n_4=120$ responden dan besar *outlier* yang bervariasi yaitu 5%, 10%, 15% dan 20%. Dari hasil simulasi ditunjukkan bahwa regresi logistik biner mempunyai ketepatan klasifikasi yang paling baik. Pengelompokan IPA atau IPS di SMA N. 1 Bangorejo dengan jumlah sampel keseluruhan (224 responden), dengan menggunakan analisis logistik biner mempunyai ketepatan klasifikasi sebesar 85,714%.

Kata Kunci: *outlier*, diskriminan linier, diskriminan linier *robust*, regresi logistik .

Abstract

This research is compare linear discriminant analysis, robust linear discriminant and biner logistic regression to classify the students of SMA N 1 Bangorejo into social/science group. Data used is the data from report book scores and psychotest of the tenth grade students at the second semester in education year 2012-2013 of SMA N 1 Bangorejo Banyuwangi, the data used contains outlier 6.70%. To know the performance to the outliers, then done many times the simulation is estimated using linear discriminant analysis, robust linear discriminant and biner logistic regression. Sample used are $n_1=40$, $n_2=80$, $n_3=80$ and $n_4=120$ respondents and outlier are varied from 5%, 10%, 15% and 20%. From the simulation shows that the biner logistic regression has the best classification accuracy. The classification to science or social at SMA N 1 Bangorejo with the whole samples (224 respondents), using logistic biner analysis has the classification accuracy 85,714%.

1 Pendahuluan

Sesuai kurikulum yang berlaku di seluruh Indonesia, maka siswa kelas X SMA yang naik ke kelas XI akan mengalami pemilihan jurusan/penjurusan. Penjurusan yang tersedia di SMA meliputi Ilmu Alam (IPA), Ilmu Sosial (IPS), dan Ilmu Bahasa. Dalam hal ini SMA Negeri 1 Bangorejo Banyuwangi hanya membuka dua jurusan yaitu

Jurusan IPA dan Jurusan IPS, sehingga siswa dikelompokkan dalam dua kelompok yaitu kelompok IPA atau IPS.

Menurut Gudono [3], bahwa peneliti yang ingin meneliti bangkrut/tidak bangkrut atau macet/lancar bisa menggunakan analisis diskriminan atau regresi logistik, tetapi karena analisis diskriminan sangat dipengaruhi oleh pengamatan *outlier* sehingga penaksirnya menjadi kurang tepat pada saat data telah terkontaminasi oleh *outlier* [5]. Agar analisis diskriminan tetap optimal dalam pengklasifikasian menggunakan penaksir *robust* disebut sebagai analisis diskriminan *robust* [1].

Outlier pada data multivariat umumnya menggunakan ukuran jarak mahalnobis, pengamatandideteksi sebagai outlier jika jarak mahalnobisnya adalah: $d_{MDi} = \sqrt{(x_i - \bar{x})^T S^{-1}(x_i - \bar{x})} > \sqrt{x_{p;\alpha}^2}$ dalam hal ini dan adalah vektor rata-rata dan matriks kovariansi. Notasi adalah nilai distribusi khi kuadrat dengan parameter [4]. *Outlier* yang disebabkan oleh variabel independen, dinamakan *outlier leverage*. *Outlier leverage* dideteksi dengan menggunakan jarak *robust* (d_{RD_i}) untuk setiap pengamatan ke-*i*, jarak *robust* didefinisikan pada persamaanberikut: dimanadanadalah vektor rata-rata dan matriks kovarians dari sebagian data yang mempunyai determinan matriks kovariansinya terkecil. Pendeteksian *outlier leverage* menggunakan jarak *robust* (d_{RD_i}) untuk setiap pengamatan ke-*i* dapat dituliskan sebagai berikut:

$$d_{RD_i} = \sqrt{(x_i - \bar{x}_{mcd})^T S_{mcd}^{-1}(x_i - \bar{x}_{mcd})},$$

dimana \bar{x}_{MCD} dan S_{MCD}^{-1} adalah vektor rata-rata dan matriks kovarians dari sebagian data yang mempunyai determinan matriks kovariansinya terkecil. Pendeteksian *outlier leverage* menggunakan jarak *robust* (d_{RD_i}) untuk setiap pengamatan ke-*i* dapat dituliskan sebagai berikut[8]:

$$laverage = \begin{cases} \text{jika } d_{RD} \leq C, \text{ maka pengamatan bukan outlier (diberi kode 0)} \\ \text{jika } d_{RD} > C, \text{ maka pengamatan merupakan outlier (diberi kode 1)} \end{cases}$$

dengan $C = \sqrt{x_{p;\alpha}^2}$, C dinyatakan sebagai nilai *cut-off*. *Outlier* berdasarkan *diagnostic plot* data pengamatan dibedakan menjadi empat tipe yaitu *bad leverage*, *outlier orthogonal*, pengamatan biasa dan *good leverage*. *Bad leverage* (terletak pada kuadran 1), merupakan suatu titik yang memiliki nilai jarak *robust* dan nilai jarak mahalnobis yang lebih besar dari nilai *cut-off*. Keberadaan titik *bad leverage* dapat merubah garis regresi sehingga dapat mempengaruhi hasil secara keseluruhan maka tindakan yang sebaiknya dilakukan adalah menghapus pengamatan [5].

2 Tinjauan Pustaka

2.1 Analisis Diskriminan Linier Biasa

Analisis diskriminan digunakan pada saat kondisi (1) sampel bisa dikelompokkan ke dalam dua atau lebih grup, (2) masing-masing grup tersebut memiliki fitur atau variabel yang bisa digunakan untuk meramal keanggotaan unit sampel ke dalam salah satu grup [3]. Asumsi Umum Analisis Diskriminan Linier yaitu (1) Berdistribusi normal multivariat, (2) matriks varians harus sama (*equal variances*) [4].

Rumus fungsi diskriminan untuk ini secara umum adalah:

$$\begin{aligned} y &= \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_{p1}, \\ &= (\bar{x}_1 - \bar{x}_2) S^{-1} x \end{aligned}$$

Dimana: \bar{x}_1 = rata – rata kelompok 1

\bar{x}_2 = rata – rata kelompok 2

S^{-1} = invers kovarian gabungan

2.2 Analisis Diskriminan Linier *Robust*

Konsep dasar analisis diskriminan linier *robust* yang menggunakan penaksir *fast-MCD* ialah dengan mengganti \bar{x} dan S pada rumus jarak *mahalanobis* yaitu $d_{MDi} = \sqrt{(x_i - \bar{x})^T S^{-1}(x_i - \bar{x})}$ dengan \bar{x}_{mcd} dan S_{mcd} yang merupakan vektor rata-rata dan matriks kovariansi dengan metode *fast-MCD*. Salah satu penaksir *robust* untuk vektor rata-rata dan matriks kovariansi adalah penaksir *Minimum Covariance Determinant (MCD)*. Analisis diskriminan linier *robust* dengan menggunakan penaksir *fast-MCD* ialah dengan mengganti \bar{x} dan S pada rumus diskriminan fungsi diskriminan dengan \bar{x}_{mcd} dan S_{mcd} yang merupakan vektor rata-rata dan matriks kovariansi dengan metode *fast-MCD* [5], rumus fungsi diskriminan linier *robust* masing-masing kelompok adalah

$$y_{1R}(x) = \bar{x}'_{1mcd} S_{mcd}^{-1} x - \frac{1}{2} \bar{x}'_{1mcd} S_{mcd}^{-1} \bar{x}_{1mcd} + \ln(p_1) \quad y_{2R}(x) = \bar{x}'_{2mcd} S_{mcd}^{-1} x - \frac{1}{2} \bar{x}'_{2mcd} S_{mcd}^{-1} \bar{x}_{2mcd} + \ln(p_2)$$

Prediksi responden akan termasuk dalam kelompok π_k jika skor diskriminan linier *robust* $y_{kR}(x) = \max\{y_{jR}(x); j = 1, 2\}$.

2.3 Regresi Logistik

Secara umum model probabilitas regresi logistik dengan melibatkan beberapa variabel prediktor (x) dapat diformulakan sebagai berikut :

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}}$$

Prediksi responden dari nilai $\pi(x) \geq 0,5$ maka diprediksikan masuk IPA sedangkan bila nilai $\pi(x) < 0,5$ maka diprediksikan masuk IPS.

Berdasarkan latar belakang di atas, maka permasalahan dalam penelitian ini adalah (1) Bagaimana pengaruh *outlier* terhadap model analisis diskriminan, analisis diskriminan *robust* dan regresi logistik biner?; (2) Bagaimana analisis diskriminan, analisis diskriminan *robust* dan regresi logistik biner dalam memprediksikan pengklasifikasian penjurusan IPA atau IPS siswa di SMA Negeri 1 Bangorejo?

3 Metode Penelitian

3.1 Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yaitu nilai raport dan nilai psikotes siswa kelas X (sepuluh) semester 2 tahun pelajaran 2012/2013 yang diperoleh dari Sekolah Menengah Atas Negeri 1 Bangorejo Banyuwangi. Variabel respon (Y) dalam penelitian terdiri dari dua kategori yaitu kategori 1 adalah jurusan IPA dan kategori 0 adalah jurusan IPS. Sedangkan variabel prediktor (X) adalah nilai raport semester 2 dan nilai psikotes siswa tahun pelajaran 2012/2013 nilai mata pelajaran sebagai syarat dalam pemilihan jurusan adalah nilai mata pelajaran matematika, fisika, kimia, biologi, sejarah, geografi, ekonomi, sosiologi dan nilai psikotes.

3.2 Langkah-langkah Penelitian

Melakukan pengumpulan data selanjutnya data diuji ada tidaknya *outlier* dengan metode MCD, data dipisahkan antara data yang mengandung *outlier* dan tidak, untuk simulasi dilakukan dengan menggunakan data yang tidak mengandung *outlier* diambil secara acak masing-masing sebanyak 5 (lima) kali sebesar $n_1=40$, $n_2=80$, $n_3=120$ dan $n_4=160$ responden selanjutnya dianalisa menggunakan metode yaitu analisis diskriminan linier, analisis diskriminan linier *robust* dan regresi logistik biner, dari data simulasi akan diambil secara acak sebesar 5%, 10%, 15% dan 20% responden selanjutnya diganti dengan data *outlier* (diambil secara acak), yang besarnya sesuai dengan responden yang diambil, selanjutnya diestimasi dengan menggunakan ketiga metode yaitu analisis diskriminan linier, analisis diskriminan linier *robust* dan regresi logistik biner selanjutnya dicari model yang paling tepat untuk pengelompokan IPA atau IPS.

Untuk interpretasi digunakan data keseluruhan yang diestimasi dengan analisis diskriminan linier, analisis diskriminan linier *robust*, regresi logistik biner yang sebelumnya dilakukan uji asumsi yaitu distribusi normal multivariat, uji asumsi homogenitas, uji vektor nilai rata-rata, uji asumsi *mutikolinieritas*, selanjutnya menentukan model dari ketepatan klasifikasi yang paling tepat.

4 Hasil dan Pembahasan

4.1 Performa Analisis Diskriminan, Analisis Diskriminan *Robust* dan Regresi Logistik Biner

Dalam menentukan model analisis diskriminan, analisis diskriminan *robust* dan regresi logistik biner dilakukan simulasi dari data yang diperoleh dari nilai raport semester 2 dan nilai psikotes siswa tahun pelajaran 2012-2013. Simulasi data dilakukan uji berulang-ulang masing-masing sebanyak 5 (lima) kali dengan besar sampel $n_1=40$, $n_2=80$, $n_3=120$ dan $n_4=160$ responden dan besar *outlier* yang bervariasi yaitu 5%, 10%, 15% dan 20%. Setelah dilakukan estimasi dengan menggunakan analisis diskriminan linier biasa, diskriminan linier *robust* dan regresi logistik biner maka didapatkan rata-rata hasil ketepatan klasifikasi pada Tabel 1 dan Tabel 2.

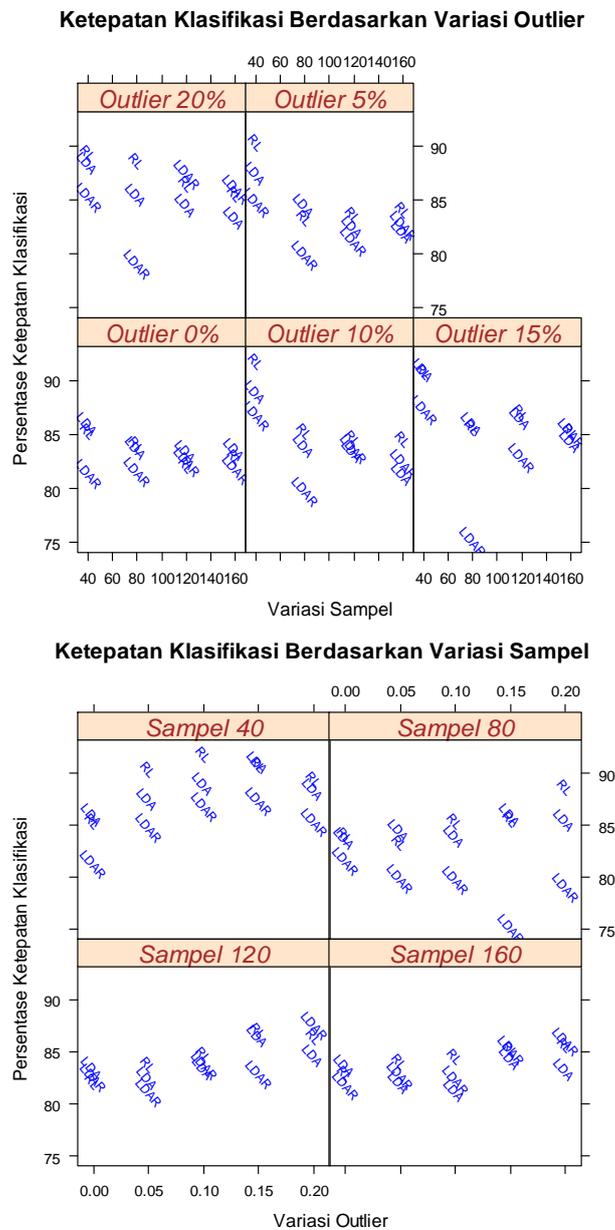
Tabel 1. Perbandingan Ketepatan Klasifikasi dari Berbagai Sampel Mengandung *Outlier*

Besarnya Sampel	Outlier 5%			Outlier 10%			Outlier 15%			Outlier 20%		
	LDA Biasa	LDA Robust	R. Log	LDA Biasa	LDA Robust	R. Log	LDA Biasa	LDA Robust	R. Log	LDA Biasa	LDA Robust	R. Log
40	87,500	85,000	90,500	89,000	87,000	92,000	91,000	87,500	91,000	88,500	85,500	89,500
80	84,500	80,117	83,500	84,000	80,000	85,500	86,000	75,250	85,750	85,500	79,250	88,750
120	82,500	81,333	83,833	83,500	83,833	84,833	86,500	83,000	87,167	84,583	87,666	86,667
160	82,125	82,875	84,250	81,375	82,500	84,750	84,375	85,375	85,500	83,375	86,250	85,750

Tabel 2. Perbandingan Ketepatan Klasifikasi dari Berbagai Sampel Tidak Mengandung *Outlier*

Besarnya Responden	Tidak ada <i>Outlier</i>		
	LDA Biasa	LDA <i>Robust</i>	Regresi Log
40	85,500	81,500	85,500
80	83,500	81,750	84,250
120	83,667	82,667	82,333
160	83,375	82,000	83,000

Berdasarkan Tabel 1 dan Tabel 2 dapat dibuat Grafik 1



Grafik 1 Grafik Ketepatan Klasifikasi

Dari Tabel 1, Tabel 2 dan Grafik 1 Perbandingan ketepatan klasifikasi ketiga metode terhadap *outlier* yaitu:

- Analisis regresi logistik biner secara umum mempunyai ketepatan klasifikasi yang lebih baik dibanding analisis diskriminan linier biasa dan analisis diskriminan linier *robust* pada saat data yang mengandung outlier.
- Analisis diskriminan linier *robust* mempunyai ketepatan klasifikasi lebih besar jika dibandingkan dengan analisis diskriminan linier biasa untuk data yang besar (160 responden) dengan besar *outlier* yang semakin besar.

Analisis diskriminan linier biasa mempunyai ketepatan klasifikasi yang cenderung lebih baik dibanding analisis diskriminan *robust* maupun regresi logistik biner untuk data yang tidak mengandung outlier.

4.2 Analisis Data Sampel Keseluruhan n = 224

Dalam memprediksikan pengklasifikasian penjurusan IPA atau IPS siswa di SMA Negeri 1 Bangorejo akan dilakukan estimasi dengan metode analisis diskriminan linier biasa, diskriminan linier *robust* dan regresi logistik biner.

4.2.1 Pengujian Asumsi Analisis Diskriminan Linier

Asumsi pada analisis diskriminan antara lain: (1) asumsi tidak ada *outlier* untuk menguji keberadaan *outlier* menggunakan metode MCD pada penelitian ini data yang dianalisis mengandung *outlier* sebesar 6,70%; (2) asumsi distribusi normal multivariat untuk menguji distribusi normal multivariat menggunakan plot jarak mahalanobis dan khi-kuadrat pencarian titik yang mendekati garis lurus dan lebih dari 50% yaitu sebesar 55,357% jadi data penjurusan siswa memenuhi asumsi distribusi multivariat normal; (3) asumsi homogenitas uji homogenitas digunakan uji Box's M nilai statistik Box's M sebesar 90,020 dengan pendekatan khi-kuadrat pada taraf signifikansi 5% sebesar 61,656 nilai uji Box's M lebih besar dari pada nilai sebaran khi-kuadrat, sehingga matriks varians kovarians grup adalah heterogen. Menurut Mattjik dan Sumertajaya [6] sangat sulit sekali untuk dapat memenuhi persyaratan *equal variances* dan *multivariate normal distribution* yang dalam praktek tidak pernah diuji. Berdasar pada hal diatas maka dalam karya ilmiah ini, penulis dalam mengestimasi data akan mengabaikan/melanggar asumsi tersebut; (4) asumsi vektor nilai rata-rata digunakan untuk mengetahui apakah terdapat perbedaan yang signifikan setiap variabel bebas antar kelompok pada taraf signifikansi 5% yang menyertakan seluruh variabel menggunakan MANOVA dengan uji Wilk's Lamda, nilai uji statistik Wilk's Lamda sebesar 0,467 dengan nilai F maksimum sebesar 20,829 dan taraf signifikansi di bawah 5% yaitu $2,2 \times 10^{-16}$ menunjukkan bahwa ada perbedaan yang signifikan antar kelompok/ jurusan untuk setiap variabel bebas.

4.2.2 Analisis Diskriminan Linier Biasa Data Keseluruhan

Fungsi Analisis Diskriminan dengan Data Sampel Keseluruhan

$Z = 0,127x_1 + 0,121x_2 + 0,130x_3 + 0,049x_4 + 0,047x_5 - 0,027x_6 - 0,023x_7 - 0,061x_8 + 0,109x_9$
 Nilai *Optimum Cutting Score* (Z_{cu}) sebesar 39,971 memprediksi responden yang nilai diskriminan (Z) yang kurang dari Z_{cu} masuk dalam kelompok IPS sedangkan yang nilai diskriminan (Z) yang lebih besar dari Z_{cu} masuk dalam kelompok IPA.

Tabel 3 Hasil Pengklasifikasian Analisis Diskriminan Linier dari Data Keseluruhan

		Keanggotaan yang diprediksi		
		IPS	IPA	Jumlah
Keanggotaan yang kenyataan	IPS	103	6	109
	IPA	28	87	115
	Jumlah	131	93	224

Total persentase ketepatan klasifikasi sebesar 84,821% atau persentase misklasifikasi sebesar 15,179%.

4.2.3 Analisis Diskriminan Linier Robust

Fungsi analisis diskriminan linier *robust* dengan sampel yang mengandung *outlier* sebagai berikut:

fungsi analisis diskriminan linier *robust* kelompok IPA

$$y_1(x) = -5,417x_1 + 14,660x_2 + 22,999x_3 + 26,541x_4 + 6,177x_5 + 3,874x_6 + 8,807x_7 + 9,049x_8 + 4,693x_9 - 3766,586$$

fungsi analisis diskriminan linier *robust* kelompok IPS

$$y_2(x) = -5,596x_1 + 14,274x_2 + 22,543x_3 + 25,823x_4 + 6,034x_5 + 3,986x_6 + 9,009x_7 + 9,114x_8 + 4,488x_9 - 3635,699$$

Klasifikasi responden bila nilai diskriminan *robust* yang lebih besar nilai $y_1(x)$ terklasifikasi masuk IPA sedangkan bila lebih besar nilai $y_2(x)$ terklasifikasi masuk IPS.

Tabel 4 Hasil Pengklasifikasian Analisis Diskriminan Linier Robust

		Keanggotaan yang diprediksi		
		IPS	IPA	Jumlah
Keanggotaan yang kenyataan	IPS	95	14	109
	IPA	19	96	115
	Jumlah	114	110	224

Total persentase ketepatan klasifikasi sebesar 85,268% Sedangkan persentase misklasifikasi sebesar 14,732%.

4.2.4 Analisis Regresi Logistik Biner

Asumsi untuk analisis logistik biner adalah tidak terjadi *multikolinieritas* antar variabel.

Uji *Multikolinieritas* menggunakan $VIF_j = \frac{1}{(1-R_j^2)}$, dimana VIF_j adalah VIF untuk variabel ke-j, R_j^2 adalah koefisien dari model determinan variabel ke-j.

Tabel 5 Nilai koefisien determinasi dan VIF

No Urut	Variabel dependen	R ²	VIF
1	y	0,4670	-
2	x ₁	0,5580	2,262512
3	x ₂	0,4307	1,756581
4	x ₃	0,2634	1,357660
5	x ₄	0,5724	2,338565

6	x_5	0,1245	1,142237
7	x_6	0,3837	1,622500
8	x_7	0,1075	1,120431
9	x_8	0,2193	1,280896
10	x_9	0,2552	1,342574

Berdasarkan Tabel 5 dari nilai VIF untuk semua variabel (x_i) kurang dari 10 maka dapat dianggap tidak signifikansi atau tidak ada *multikolinieritas* antar variabel.

Model Analisis Regresi Logistik Biner.

Hasil pengujian model regresi logistik dengan variabel dependen penjurusan IPA dan IPS (Y) diperoleh model peluang untuk masuk dalam kelompok IPA sebagai berikut:

$$\pi(x) = \frac{\exp(-126,307 + 0,325x_1 + 0,396x_2 + 0,332x_3 + 0,479x_4 + 0,102x_5 - 0,076x_6 - 0,084x_7 - 0,146x_8 + 0,217x_9)}{1 + \exp(-126,307 + 0,325x_1 + 0,396x_2 + 0,332x_3 + 0,479x_4 + 0,102x_5 - 0,076x_6 - 0,084x_7 - 0,146x_8 + 0,217x_9)}$$

$\pi(x) \geq 0,5$ maka diprediksikan masuk IPA sedangkan bila nilai $\pi(x) < 0,5$ maka diprediksikan masuk IPS.

Tabel 6 Hasil Pengklasifikasian Regresi Logistik dari Data yang Mengandung *Outlier*

		Keanggotaan yang diprediksi		
		IPS	IPA	Jumlah
Keanggotaan yang kenyataan	IPS	97	12	109
	IPA	20	95	115
Jumlah		117	107	224

Total persentase ketepatan klasifikasi sebesar 85,714% dan total persentase misklasifikasi sebesar 14,286%.

Dalam memprediksikan pengklasifikasian penjurusan IPA atau IPS siswa di SMA Negeri 1 Bangorejo dengan jumlah data sampel keseluruhan, dengan menggunakan metode analisis diskriminan linier, analisis diskriminan linier *robust* dan regresi logistik biner diperoleh ketepatan klasifikasi untuk regresi logistik yang paling besar yaitu sebesar 85,714%.

5 Penutup

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan yang diperoleh maka dapat diambil suatu kesimpulan sebagai berikut.

- 5.1.1 Berdasarkan uji simulasi data yang dilakukan berulang-ulang dengan jumlah sampel dan besar *outlier* yang berbeda, analisis regresi logistik biner mempunyai ketepatan klasifikasi yang lebih baik dibanding analisis diskriminan linier maupun analisis diskriminan linier *robust* pada saat data yang mengandung outlier. responden) dengan besar *outlier* yang semakin besar.
- 5.1.2 Pengelompokan jurusan IPA atau IPS di SMA Negeri 1 Bangorejo dengan jumlah data 224 responden terkontaminasi *outlier* 6,70% menggunakan analisis logistik biner mempunyai ketepatan klasifikasi lebih baik dan tidak banyak asumsi yang dilakukan, ketepatan klasifikasi yang diperoleh sebesar 85,714%.

5.2 Saran

Penelitian ini dapat dikembangkan dengan banyaknya kelompok lebih dari 2 kelompok/jurusan dan menggunakan variabel yang sesuai dengan aturan ataupun kebutuhan sekolah.

Daftar Pustaka

- [1] Budyanra. 2010. *Ketepatan Pengklasifikasian Fungsi Diskriminan Linier Robust Dua Kelompok Dengan Metode Fast Minimum Covariance Determinant (FAST-MCD)*, Tesis Program Pascasarjana Universitas Padjadjaran Bandung.
- [2] Filzmoser, P., 2004. *A Multivariate Outlier Detection Method*, Departement of Statistics and Probability Theory, Vienna, Austria
- [3] Gudono. 2012. *Analisis Data Multivariat*, Jogjakarta: Fakultas Ekonomi dan Bisnis UGM.
- [4] Hair, J. F. Jr., Anderson, R. E., Tatham, R. L. and Black W. C. 2006. *Multivariate Data Analysis*. 5th ed. New Jersey: Prentice Hall
- [5] Hubert M. dan Driessen K. V. 2002. "Fast and Robust Discriminant Analysis", *Computational Statistics and Data Analysis*, Vol. 45, hal. 301-320
- [6] Mattjik A. A., dan Sumertajaya. I. M. 2011. *Sidik Peubah Ganda Dengan menggunakan SAS*, Bogor: IPB Press.
- [7] Maja P., Mateja B. dan Sandra. T. 2004. *Comparason of Logistic Regression and Linear Discriminant Analisis: A Simulation Study*, Metodološki zvezki, Vol. 1, No. 1, 143-161
- [8] Mukti K., Marisi A., dan Muhlasah N. M. 2012. *Mendeteksi Outlier Dengan Metode Minimum Covariance Determinant*, [Portal Jurnal Ilmiah Universitas Tanjungpura](#), Vol. 01, No. 1, hal 31 – 40.