

PERBANDINGAN METODE *MAXIMUM LIKELIHOOD ESTIMATION (MLE)* DENGAN *BAYESIAN* PADA REGRESI LOGISTIK MULTINOMIAL

Permadina Kanah Arieska

e-mail : belajar_statistika@yahoo.com

ABSTRAK

Regresi logistik multinomial klasik menggunakan metode Maximum Likelihood untuk mengestimasi parameter - parameter-nya. Sedangkan pada regresi logistik multinomial dengan menggunakan metode Bayesian, distribusi prior dipadukan dengan likelihood datanya untuk mendapatkan distribusi posterior. Distribusi posterior tersebut akan digunakan untuk mengestimasi parameter - parameter-nya. Penerapan dua metode ini dilakukan pada penelitian ini untuk mengklasifikasikan siswa SMAN 1 Grati Pasuruan pada 3 jurusan yang telah ditetapkan oleh pihak sekolah yaitu jurusan IPA, IPS dan Bahasa. Variabel prediktor yang digunakan ada 5 yaitu tuntas IPA, tuntas IPS, tuntas Bahasa, IQ dan minat. Dari kelima variabel prediktor tersebut ternyata variabel yang berpengaruh secara signifikan pada penjurusan siswa adalah variabel tuntas bahasa dan IQ. Misklasifikasi pada regresi logistik multinomial klasik lebih besar dibandingkan pada regresi logistik multinomial metode Bayesian. Misklasifikasi dengan metode pertama adalah sebesar 46,1% sedangkan dengan menggunakan metode kedua yaitu metode Bayesian, misklasifikasinya adalah sebesar 39,5%. Sehingga dalam penelitian ini, regresi logistik multinomial dengan metode Bayesian lebih baik dalam pemodelan jurusan di SMAN 1 Grati Pasuruan jika dibandingkan dengan regresi logistik Multinomial klasik.

Kata kunci: Regresi logistik Multinomial, Bayesian, Misklasifikasi, Jurusan

1. Pendahuluan

Regresi Logistik Multinomial merupakan salah satu metode yang diusulkan untuk mengatasi masalah penjurusan SMA. Alternatif pilihan sebagai nilai respon yang lebih dari dua yaitu IPA, IPS dan Bahasa, merupakan alternatif yang mengikuti distribusi multinomial. Sehingga nantinya siswa akan bisa dikelompokkan ke dalam 3 kelompok jurusan sesuai dengan faktor-faktor yang mempengaruhinya. Namun, kasus yang seringkali terjadi adalah bahwasanya probabilitas siswa dari suatu kelas tertentu untuk bisa masuk ke salah satu jurusan, tidaklah sama. Hal ini bisa disebabkan oleh faktor lain yang berbeda pada masing-masing kelas. Misal karena perbedaan guru yang mengajar, suasana kelas, metode pengajaran dan lain sebagainya. Sehingga, probabilitas seorang siswa untuk masuk ke salah satu jurusan memiliki distribusi tertentu. Hal ini bisa diatasi dengan metode Bayesian yang memperlakukan parameter-nya sebagai variabel.

Pendekatan algoritma komputasional *Markov Chain Monte Carlo (MCMC)* digunakan untuk mempermudah analisis, karena metode ini dapat menaksir distribusi marginal masing - masing parameter suatu variabel random yang berdimensi tinggi. Metode ini dapat digunakan walaupun bentuk distribusi posterior yang didapatkan sangat kompleks. Skenario sampling yang digunakan adalah *Gibbs Sampler*. *Gibbs Sampler* adalah suatu teknik simulasi untuk membangkitkan variabel random dari suatu distribusi tertentu secara langsung, tanpa perlu menghitung densitasnya. Madigan, Genkin, Lewis, dan Fradkin (2005) menjelaskan bahwa pada umumnya, distribusi Prior yang seringkali digunakan adalah distribusi Gaussian dengan mean 0 dan varian σ^2 .

Penggunaan regresi logistik multinomial diterapkan pada penetapan jurusan di SMAN 1 Grati Pasuruan. Variabel independen seorang siswa diperhitungkan masuk ke jurusan tertentu antara lain nilai ketuntasan siswa pada setiap bidang di semester terakhir kelas X, IQ dan minat siswa. Model penetapan jurusan SMA ini akan dianalisis dengan menggunakan regresi logistik Multinomial dan Regresi logistik Multinomial dengan metode Bayesian. Dua metode ini akan dibandingkan untuk mengetahui misklasifikasi yang diperoleh.

2. Regresi Logistik Multinomial

Regresi logistik merupakan suatu pendekatan model matematis yang dapat digunakan untuk menyatakan hubungan dari beberapa variabel X terhadap variabel dependennya yang bersifat kategori (Kleinbaum, 1994). Perluasan regresi logistik dengan variabel dependen yang lebih dari dua kategori disebut dengan Regresi Logistik multinomial. Regresi Logistik multinomial banyak digunakan untuk mengidentifikasi hubungan antara variabel respon dan variabel prediktor. Pada variabel respon terdapat lebih dari 2 kategori, sehingga memiliki distribusi multinomial (Fahmeir dan Tutz, 1994).

Bila terdapat J kategori respon maka model yang didapatkan sebanyak $J - 1$. Dimisalkan terdapat tiga kategori respon Y yaitu 0, 1, dan 2, $Y = 0$ merupakan referensi/dasar, maka didapatkan model logit sebagai berikut :

Jika $\pi_j(\mathbf{x}_i) = P(Y=j|\mathbf{x}_i)$; $j = 0, 1, 2$; menyatakan probabilitas kategori respon ke- j pada p variabel prediktor yang dinyatakan dalam vektor \mathbf{x}_i maka secara umum probabilitas kategori respon ke- j pada p variabel prediktor yang dinyatakan dalam vektor \mathbf{x}_i adalah sebagai berikut :

$$\pi_j(\mathbf{x}_i) = P(Y = j | \mathbf{x}_i) = \frac{\exp(g_j(\mathbf{x}_i))}{\sum_{k=0}^{J-1} \exp(g_k(\mathbf{x}_i))} \quad (2.1)$$

yang mana $g_0(\mathbf{x}_i) = 0$, $i = 1, 2, \dots, p$. Untuk klasifikasi suatu pengamatan nilai $\pi_j(\mathbf{x}_i)$ pada persamaan di atas akan dijadikan pedoman pengklasifikasian. Suatu pengamatan akan masuk dalam respon kategori j berdasarkan nilai $\pi_j(\mathbf{x}_i)$ yang terbesar .

3. Distribusi Prior

Permasalahan utama dalam pendekatan Bayes adalah memilih distribusi prior $g(\theta)$ yang menunjukkan ketidakpastian tentang parameter θ yang tidak diketahui. Distribusi prior dapat diperoleh melalui data masa lalu yang telah ada dan distribusi prior ini biasa disebut dengan distribusi prior “data based” (DB). Jika data masa lalu tidak tersedia, distribusi prior dipilih berdasarkan kepercayaan peneliti, dan distribusi prior jenis ini disebut prior “non data based” (NDB).

Berkaitan dengan bentuk distribusi hasil identifikasi pola data, distribusi prior dibagi menjadi:

- Distribusi prior sekawan (conjugate prior) mengacu pada acuan analitis model terutama dalam pembentukan fungsi likelihoodnya. Sehingga dalam penentuan prior sekawan selalu difikirkan mengenai penentuan pola distribusi prior yang mempunyai bentuk sekawan dengan fungsi densitas pembangun fungsi likelihoodnya.
- Distribusi prior tidak sekawan (non conjugate prior) apabila pemberian prior pada suatu model tidak mengindahkan pola pembentuk fungsi likelihoodnya.

Berkaitan dengan penentuan masing-masing parameter pada distribusi prior tersebut dikelompokkan menjadi:

- Distribusi prior informatif mengacu pada pemberian parameter dari distribusi prior yang telah dipilih. Baik prior yang dipilih sekawan ataukah tidak, pemberian nilai parameter pada distribusi prior ini akan sangat mempengaruhi bentuk distribusi posterior yang akan didapatkan. Pemilihan prior informative ini didasarkan pada informasi dari data yang diperoleh. (Zhu, Mu dan Lu, Arthur, 2004)
- Distribusi prior noninformatif, pemilihannya tidak didasarkan pada data yang ada, atau prior yang tidak mengandung informasi tentang θ . (Tahir, Muhammad dan Hussain, Zawar, 2000)

4. Teorema Bayes

Besaran parameter $\theta = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)$ dalam sebuah populasi yang memuat variabel random X dapat disajikan dalam bentuk aturan probabilitas, dalam teorema Bayes akan disajikan sebagai berikut :

$$p(\theta|x) = \frac{l(x|\theta)p(\theta)}{p(x)} \quad (4.1)$$

Persamaan diatas sering dikatakan sebagai metode untuk update informasi prior parameter θ , $p(\theta)$ dengan menggunakan informasi sampel yang terdapat pada likelihood data, $l(x|\theta)$, untuk memperoleh informasi posterior, $p(\theta|x)$, yang akan digunakan dalam keputusan. Penyebut, $p(x)$ adalah suatu konstanta penormal (normalized constant). Sehingga posterior bisa dituliskan sebagai $p(\theta|x) \propto l(x|\theta)p(\theta)$. Persamaan ini menunjukkan bahwa distribusi posterior θ proporsional terhadap prior kali likelihood datanya. Sehingga distribusi probabilitas posterior terdiri atas semua informasi tentang parameter θ (Gelman, 1995)

5. Markov Chain Monte Carlo (MCMC)

Markov Chain Monte Carlo (MCMC) merupakan suatu metode simulasi yang menggunakan sifat rantai Markov untuk mendapatkan data sampel suatu variabel random berdasarkan skenario sampling yang mengikuti sifat markov. Metode MCMC tersebut digunakan untuk mempermudah analisis karena metode ini dapat

menaksir parameter – parameter suatu analisis suatu variabel random yang berdimensi tinggi sehingga metode ini dapat digunakan jika bentuk posterior yang didapatkan sangat kompleks.

6. Penetapan Jurusan pada Sekolah Menengah

Terdapat berbagai macam kriteria (multi criteria) pada penentuan jurusan sekolah di SMA. Berikut keterangan penentuan dan pelaksanaan penjurusan serta kriteria-kriteria yang diperlukan. Kriteria penjurusan program meliputi :

a. Nilai akademik,

Peserta didik yang naik kelas XI dan akan mengambil program tertentu yaitu : Ilmu Pengetahuan Alam (IPA) atau Ilmu Pengetahuan Sosial (IPS) atau Bahasa : boleh memiliki nilai yang tidak tuntas paling banyak 3 (tiga) mata pelajaran pada mata pelajaran-mata pelajaran yang bukan menjadi ciri khas program tersebut. Peserta didik yang naik ke kelas XI, dan yang bersangkutan mendapat nilai tidak tuntas 3 (tiga) mata pelajaran, maka nilai tersebut harus dijadikan dasar untuk menentukan program yang dapat diikuti oleh peserta didik. Ciri khas program IPA antara lain matematika, kimia, biologi, dan fisika. Ciri khas program IPS adalah ekonomi, geografi, sosiologi dan sejarah. Sedangkan ciri khas program bahasa adalah bahasa Indonesia, bahasa Inggris dan TI (Teknologi Informasi) (SK tentang kriteria kenaikan kelas nomor 421.3/163/431.070.05.3.01/2007)

b. Minat Peserta Didik

Untuk mengetahui minat peserta didik dapat dilakukan melalui angket/kuesioner dan wawancara, atau cara lain yang dapat digunakan untuk mendeteksi minat, dan bakat.

c. IQ

Variabel IQ juga digunakan untuk pertimbangan penjurusan. Dalam hal ini kemampuan seorang anak terkait daya analitis, kemampuan numerik dan yang lain bisa dilihat. Sehingga anak akan bisa di sarankan untuk masuk ke salah satu jurusan yang telah ditetapkan.

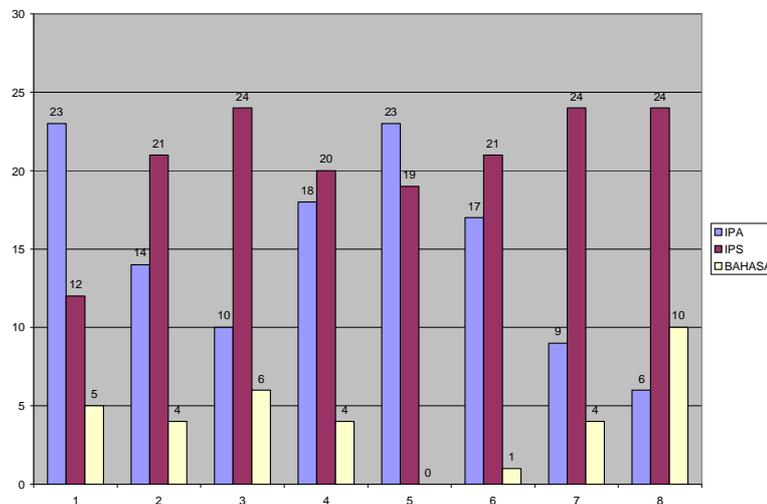
7. Metodologi

Penetapan jurusan di SMAN 1 Grati Pasuruan dilakukan pada saat siswa naik ke kelas XI. Penetapan jurusan tersebut berdasarkan pada nilai rata-rata UAS semester saat di kelas X. Disamping itu juga dilakukan tes IQ dan pemberian kuosioner untuk mengetahui minat dari masing-masing siswa. Jumlah siswa secara keseluruhan sebanyak 319 siswa yang nantinya akan menempati 1 kelas Bahasa, 3 kelas IPA dan 4 kelas IPS dengan rata-rata terdapat 40 siswa per kelasnya. Variabel respon terdiri atas 3 kategori. Kategori 0 jurusan IPA, kategori 1 jurusan IPS dan kategori 2 jurusan bahasa.

Maximum Likelihood merupakan metode yang akan digunakan untuk mengestimasi parameter. Setelah diperoleh parameter-parameter pembentuk model, maka akan dianalisa variabel yang signifikan dan dilakukan interpretasi terkait dengan variabel independen tersebut terhadap variabel responnya. Sedangkan metode yang kedua adalah regresi logistik Multinomial metode Bayesian. Dua metode ini akan dibandingkan misklasifikasinya. Misklasifikasi yang paling sedikit merupakan metode yang lebih baik digunakan.

8. Analisa Data

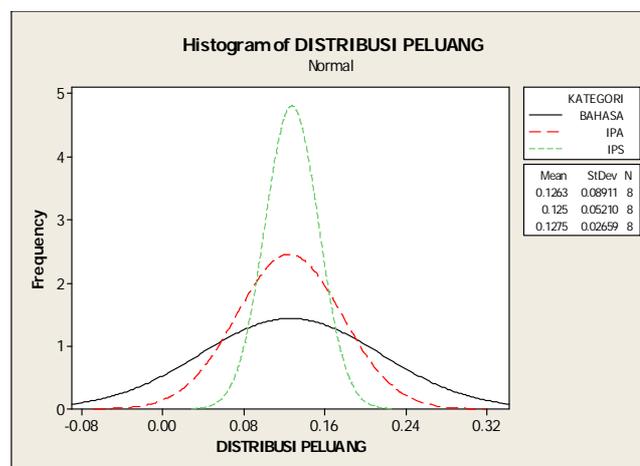
Jumlah kelas X yang ada pada SMAN 1 Grati Pasuruan adalah 8 kelas. Siswa kedelapan kelas tersebut akan ditempatkan pada tiga jenis jurusan ketika mereka naik ke kelas XI. Tiga jurusan tersebut adalah jurusan IPA sebanyak 3 kelas, jurusan IPS sebanyak 4 kelas dan jurusan bahasa sebanyak 1 kelas. Gambar 8.1 merupakan distribusi siswa kelas X tahun 2007/2008 yang menempati masing – masing jurusan pada kelas XI.



Gambar 8.1 Distribusi siswa kelas X pada tiap – tiap jurusan

Distribusi siswa yang masuk ke masing – masing jurusan memiliki pola yang berbeda – beda. Siswa yang paling banyak masuk jurusan IPA berasal dari kelas X.1 dan X.5. Sedangkan siswa yang masuk ke jurusan IPS banyak yang berasal dari kelas X.3, X.7 dan X.8. Untuk kelas X.2, X.4 dan X.6 memiliki komposisi siswa yang masuk IPA dan IPS yang hampir sebanding. Perbedaan ini bisa diakibatkan oleh pola pembelajaran guru yang berbeda antara kelas ganjil dan kelas genap.

Jika dipisahkan menurut tiap jurusan, maka distribusi peluang untuk bisa masuk ke tiap jurusan nampak pada Gambar 8.2. Pada jurusan IPS, siswa yang masuk jurusan ini, untuk masing-masing kelas hampir seragam dibandingkan dengan jurusan lain. Hal ini bisa dilihat dari nilai standar deviasi yang kecil. Secara fakta, jurusan IPS memang memiliki kelas yang lebih banyak dibandingkan dengan dua jurusan lainnya. Sehingga peluangnya lebih besar. Secara umum, rata – rata peluang untuk masuk ke jurusan tertentu hampir sama yaitu sebesar 0,12. Namun yang membedakan disini adalah nilai standar deviasinya. Sehingga memang sangat memungkinkan untuk menelaah kontribusi tiap kelas yang bisa masuk ke masing – masing jurusan. Dari Gambar 8.2 ini, nantinya akan dijadikan dasar penerapan regresi logistik multinomial dengan menggunakan metode Bayesian yang akan dibandingkan dengan regresi logistik multinomial klasik.



Gambar 8.2 Histogram Distribusi Peluang Tiap – Tiap Jurusan

8.1 Penentuan Model Regresi Logistik Multinomial

Pemilihan model regresi logistik multinomial akan dilakukan dengan pengujian terhadap parameter – parameternya. Dalam hal ini akan dilakukan terlebih dahulu regresi logistik multinomial univariabel untuk mengetahui signifikansi masing – masing variabel prediktor jika diregresikan dengan variabel responnya secara univariabel. Hasil regresi logistik multinomial univariabel ditunjukkan pada Tabel 8.1.

Tabel 8.1 Regresi Logistik Multinomial Univariabel

Variabel Prediktor	Chi Square	Sig.
Tuntas IPA	4,316	0,116
Tuntas Bahasa	13,491	0,001
IQ	9,538	0,008
Minat	1,527	0,822

Secara univariabel, dengan $\alpha = 5\%$, variabel prediktor yang signifikan adalah tuntas bahasa dan IQ. Variabel tuntas IPS tidak memiliki nilai Chi Square karena derajat kebebasannya adalah nol. Hal ini disebabkan, semua siswa memiliki nilai ketuntasan IPS yang lebih dari 3. Oleh karena itu, untuk mengetahui faktor – faktor yang mempengaruhi penjurusan siswa, berikut akan dilakukan analisis regresi multinomial secara multivariabel dengan menggunakan Backward Elimination dan Forward Entry. Kedua ini digunakan untuk memastikan variabel prediktor yang akan digunakan memang diyakini berpengaruh terhadap variabel respon. Karena terkadang variabel prediktor bernilai signifikan pada metode backward namun belum tentu dengan menggunakan Forward.

Dengan metode Forward menunjukkan bahwa variabel yang dimasukkan kedalam model adalah variabel tuntas Bahasa dan IQ. Hal ini juga terdapat pada metode Backward. Pada metode Backward, variabel

yang tidak masuk kedalam model adalah variabel tuntas IPA, tuntas Bahasa dan minat. Sehingga dari dua metode ini bisa disimpulkan bahwa variabel yang signifikan adalah tuntas Bahasa dan IQ. Berikut uji signifikansi melalui uji Serentak dan uji individu. Dalam pengujian serentak, uji signifikansi model dipergunakan *likelihood-ratio test* dengan hipotesis sebagai berikut :

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_5 = 0$$

$$H_1 : \text{minimal ada satu } \beta_k \neq 0 \quad ;$$

$$k = 1, 2, \dots, 5$$

Nilai $\chi^2_{(4;0,05)}$ adalah 24,119 dengan signifikansi sebesar 0,000 yang berarti didapatkan keputusan untuk tolak H_0 . Dari keputusan ini diketahui bahwa secara serentak model signifikan pada tingkat kepercayaan 95 %.

Hasil yang ditunjukkan oleh *Likelihood Ratio Test* ini bisa dijelaskan masing – masing oleh variabel prediktor yang terpilih kedalam model seperti pada Tabel 8.2.

Tabel 8.2 Likelihood Ratio Test

Effect	Model Fitting Criteria	Likelihood Ratio Test		
	(-2 log likelihood of reduced model)	Chi Square	df	Sig.
Intercept	87,381	0,000	0	
Tuntas Bahasa	101,962	14,581	2	0,001
IQ	98,009	10,628	2	0,005

P value untuk tuntas bahasa dan IQ kurang dari $\alpha = 5\%$. Hal ini berarti variabel tuntas bahasa dan IQ merupakan variabel yang lebih baik dalam membentuk model dibandingkan dengan model yang hanya memasukkan intercept saja. Oleh karena itu akan dilakukan uji individu Wald Test (Hosmer dan Lemeshow, 2000). Uji ini dipergunakan untuk mengetahui variabel prediktor mana yang signifikan secara individu. Dengan dipergunakan hipotesis sebagai berikut :

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0 \quad ; k = 1, 2, \dots, 5$$

nilai statistik uji *W* untuk masing-masing parameter ditunjukkan pada Tabel 8.3.

Tabel 8.3 Nilai Statistik Uji *W*

Jurusan	Variabel Prediktor	Koefisien	Odds Ratio	P value
1 (IPS)	Intercept	-0,155		0,356
	[Tuntas_bahasa=0)	1,154	3,172	0,015*
	[Tuntas_bahasa=1)	.		
	[IQ=0)	0,635	1,887	0,000*
	[IQ=1)	.		
2 (Bahasa)	Intercept	-1,857		0,000*
	[Tuntas_bahasa=0)	0,635	1,888	0,004*
	[Tuntas_bahasa=1)	.		
	[IQ=0)	1,16	3,19	0,212
	[IQ=1)			

Keterangan : * signifikan pada $\alpha = 0,05$

Dari pengujian parameter ini maka model yang diperoleh dari regresi logistik multinomial adalah yang melibatkan 2 variabel prediktor yaitu tuntas bahasa dan IQ. Model regresinya sebagai berikut :

$$\pi(y = 0|x) = \frac{1}{1 + e^{-0.155 + 1.154 X_3 + 0.635 X_4} + e^{-1.857 + 0.635 X_3 + 1.16 X_4}}$$

$$\pi(y = 1|x) = \frac{e^{-0.155+1.154 X_3+0.635 X_4}}{1 + e^{-0.155+1.154 X_3+0.635 X_4} + e^{-1.857+0.635 X_3+1.16 X_4}}$$

$$\pi(y = 2|x) = \frac{e^{-1.857+0.635 X_3+1.16 X_4}}{1 + e^{-0.155+1.154 X_3+0.635 X_4} + e^{-1.857+0.635 X_3+1.16 X_4}}$$

Tabel 8.3 menunjukkan bahwa seorang siswa yang memiliki nilai tuntas bahasa yang kurang dari 3, lebih cenderung untuk memilih IPS daripada IPA. Begitupula untuk siswa yang IQ nya dibawah rata-rata (IQ<110) maka dia lebih cenderung untuk memilih IPS daripada IPA. Siswa ini juga lebih cenderung untuk memilih program bahasa dibandingkan dengan IPA. Atau bisa disimpulkan bahwa siswa yang memiliki IQ diatas rata – rata lebih cenderung memilih program IPA daripada memilih IPS ataukah bahasa.

8.2 Penentuan Model Regresi Logistik Multinomial Metode Bayesian

Pada metode Bayesian, hal yang perlu ditetapkan terlebih dahulu adalah masalah distribusi priornya. Prior ini akan digunakan untuk membentuk posteriornya. Chib & Charlin (1995) menggunakan Pseudoprior untuk membentuk fungsi posteriornya. Sesuai dengan namanya, Pseudoprior bukan prior yang sebenarnya namun pseudoprior mampu memberikan ketepatan pada density yang dimaksud, yang diperlukan nantinya untuk membentuk spesifikasi model.

Untuk menentukan pseudoprior, hal yang dilakukan pertama kali adalah meregresikan masing – masing variabel independen dan variabel responnya. Dalam penelitian ini, variabel yang signifikan adalah variabel tuntas bahasa dan IQ. Pada lampiran 3 terlihat bahwa parameter untuk tuntas bahasa berdistribusi normal (1,096, 0,317) dan untuk variabel IQ juga berdistribusi normal (0,557, 0,255). Parameter – parameter ini akan dimasukkan ke program winbugs. Pada program ini, parameter untuk variabel independen yang pertama yaitu IQ dinamakan dengan beta dan parameter untuk variabel tuntas bahasa dinamakan Gamma. Sedangkan intercept nya dalam hal ini disebut sebagai alpha. Hasil running program winbugs dengan update 50.000 adalah sebagai berikut :

Tabel 8.4 Node Statistics dari model pada Software Winbugs

Node	Mean	SD	MC error	2,50%	Median	97,50%	Start	Sample
alpha[2]	1.644	0.3568	0.0138	0.9669	1.642	2.354	4001	46000
alpha[3]	-0.1378	0.5498	0.0202	-1.27	-0.1175	0.8895	4001	46000
beta[2,2]	-0.6317	0.2691	0.0071	-1.16	-0.63	-0.119	4001	46000
beta[2,3]	-1.158	0.4188	0.0096	-1.988	-1.153	-0.3568	4001	46000
gamma[2,2]	-1.164	0.3188	0.0115	-1.806	-1.166	-0.5579	4001	46000
gamma[2,3]	-0.5851	0.5248	0.0182	-1.58	-0.6044	0.5056	4001	46000

Pengujian parameter – parameter pada program Winbugs menggunakan *Highest Posterior Distribution (HPD)*. Pada tabel 8.4 diatas bisa dilihat pada nilai antara 2.5% dan 97.5%. Jika mengandung nilai nol maka parameter tidak signifikan. Dengan hipotesis sebagai berikut :

Uji hipotesis :

H_0 : Tidak ada parameter yang signifikan

H_1 : Minimal terdapat satu parameter yang signifikan

Dari Tabel 8.4, terlihat bahwa parameter yang signifikan adalah alpha[2], beta[2,2], beta[2,3], dan gamma[2,2].

Sehingga model yang diperoleh untuk regresi logistik multinomial Bayesian ini adalah :

$$\pi(y = 0|x) = \frac{1}{1 + e^{-0.1378+1.158 X_3+0.5851 X_4} + e^{-1.644+1.164 X_3+0.6317 X_4}}$$

$$\pi(y = 1|x) = \frac{e^{-0.1378+1.158 X_3+0.5851 X_4}}{1 + e^{-0.1378+1.158 X_3+0.5851 X_4} + e^{-1.644+1.164 X_3+0.6317 X_4}}$$

$$\pi(y = 2|x) = \frac{e^{-1.644+1.164 X_3+0.6317 X_4}}{1 + e^{-0.1378+1.158 X_3+0.5851 X_4} + e^{-1.644+1.164 X_3+0.6317 X_4}}$$

8.3 Perbandingan Misklasifikasi Model Regresi Logistik Multinomial Klasik dengan Regresi Logistik Multinomial metode Bayesian.

Untuk mengetahui, metode regresi logistik multinomial klasik dengan metode Bayesian, akan dilakukan perhitungan misklasifikasi pada kedua metode tersebut. Pada Tabel 8.5 terlihat bahwa misklasifikasi pada model regresi logistik multinomial yang terjadi cukup besar yaitu sekitar 46,1%. Sedangkan dengan menggunakan

regresi logistik Multinomial bayesian, misklasifikasinya lebih kecil dibandingkan dengan metode pertama, yaitu sebesar 39,5%.

Tabel 8.5 Tabel Misklasifikasi Regresi Logistik Multinomial Klasik

Observasi	Prediksi			Percent correct
	0	1	2	
0	72	48	0	60,00%
1	65	100	0	60,61%
2	9	25	0	0,00%
Overall Percentage	45,80%	54,20%	0,00%	53,9%

Pada regresi logistik multinomial klasik, penyumbang misklasifikasi yang paling besar adalah penjurusan terhadap siswa bahasa. Prediksi yang diberikan adalah 0% padahal ada sekitar 34 siswa yang masuk jurusan bahasa. Namun pada regresi logistik multinomial metode Bayesian, misklasifikasi ini bisa diminimalkan. Pada Tabel 8.6 menunjukkan bahwa prediksi siswa yang masuk jurusan Bahasa adalah 55,88%. Sehingga dalam hal ini, regresi logistik multinomial metode Bayesian lebih baik digunakan untuk memodelkan pelaksanaan penjurusan siswa di SMAN 1 Grati Pasuruan.

Tabel 8.6 Tabel Misklasifikasi Regresi Logistik Multinomial Bayesian

Observasi	Prediksi			Percent correct
	0	1	2	
0	72	48	0	60,00%
1	63	102	0	61,81%
2	9	6	19	55,88%
Overall Percentage	45,14%	48,9%	5,96%	60,5%

9. Kesimpulan

Kesimpulan dari pembahasan yang telah dilakukan adalah :

1. Dari kelima variabel prediktor yang digunakan, ternyata terdapat hanya dua variabel prediktor yang signifikan. Dua variabel prediktor tersebut adalah tuntas bahasa dan IQ. Siswa yang memiliki nilai tuntas bahasa yang kurang dari 3, lebih cenderung untuk memilih IPS daripada IPA. Begitupula untuk siswa yang IQ nya dibawah rata-rata maka dia lebih cenderung untuk memilih IPS daripada IPA. Siswa ini juga lebih cenderung untuk memilih program bahasa dibandingkan dengan IPA.
2. Faktor – faktor yang mempengaruhi penjurusan siswa adalah tuntas bahasa dan IQ. Nilai parameter pada regresi logistik multinomial metode Bayesian hampir sama dengan metode klasik. Dalam hal ini, siswa yang memiliki nilai tuntas bahasa kurang dari 3, lebih memilih program Bahasa dibandingkan dengan IPA. Atau bisa disimpulkan bahwa siswa yang memiliki IQ diatas rata – rata lebih cenderung memilih program IPA daripada memilih IPS ataukah bahasa.
3. Misklasifikasi pada regresi logistik multinomial klasik lebih besar dibandingkan pada regresi logistik multinomial metode Bayesian. Misklasifikasi dengan metode pertama adalah sebesar 46,1% sedangkan dengan menggunakan metode kedua yaitu metode Bayesian, misklasifikasinya adalah sebesar 39,5%.

Daftar Pustaka

- Carlin dan Chib. (1995). *Bayesian Model Choice via Markov Chain Monte Carlo Methods*, Journal of the Royal Statistical Society, Series B, Volume 57, 473-484.
- Depdiknas (2007). SK kriteria Kenaikan Kelas bagi Siswa Kelas X dan XI tahun 2006/2007. Grati Pasuruan.
- Fahmeir,L dan Tutz, G. (1994), *Multivariate Statistical Modelling Based on Generalized Linier Models*, Springer-Verlag.
- Hosmer, D. W dan Lemeshow. (2000), *Applied Logistik Regression*, second edition, John Wiley and Sons. Inc, New York.
- Kleinbaum, David G. (1994). *Logistic Regression, a self Learning Text*. Springer Verlag new york.

- Madigan, David. Genkin, Alexander. Lewis, David D dan Fradkin, Dmitry (2005). *Bayesian Multinomial Logistic Regression for Author Identification*, Rutger university.
- Madigan, David. Genkin, Alexander. Lewis, David D dan Fradkin, Dmitry (2005). *Bayesian Multinomial Logistic Regression for Author Identification*, Rutger university, Amerika Serikat.
- Tahir, Muhammad dan Hussain, Zawar. (2000). *Comparison of Non Informative Priors for Number of Defects (Poisson) Model*. Department of Statistics, Government Degree College, Pindhi Gheb, Pakistan.
- Zhu, Mu dan Lu, Arthur. (2004). *The Counter – Intuitive Non Informative Prior for the Bernoulli Family*. Journal of Statistics Education Volume 12, Nomor 2, University of Waterloo.