

PEMODELAN TERHADAP KELULUSAN SISWA MASUK KELAS AKSELERASI MENGGUNAKAN ANALISIS REGRESI LOGISTIK DAN *MULTIVARIATE ADAPTIVE REGRESSION SPLINE (MARS)*

Fentje Mandaku

Staf Dinas Pendidikan dan Olah Raga Provinsi Maluku

Hanok Mandaku

Dosen Program Studi Teknik Industri, Fakultas Teknik, Universitas Pattimura, Ambon

e-mail : mandakuh@yahoo.com

ABSTRAK

Analisis Regresi merupakan analisis yang digunakan untuk melihat hubungan dan pengaruh variabel prediktor terhadap variabel respon. Jika variabel respon dalam bentuk kategori dan terdapat banyak variabel predictor yang terdiri dari banyak kategori atau campuran keduanya maka analisis regresi yang dapat digunakan adalah analisis Regresi Logistik dan Multivariate Adaptive Regression Spline (MARS). Kedua metode tersebut dapat digunakan juga untuk pengklasifikasian sejumlah objek. Penelitian ini bertujuan untuk mendapatkan suatu model kuantitatif yang menjelaskan hubungan antara nilai rapor, nilai UN, pendidikan orang tua, pekerjaan orang tua, urutan kelahiran, jumlah saudara dengan kelulusan siswa masuk kelas akselerasi, sehingga dapat memudahkan orang tua siswa dalam pengambilan keputusan untuk merekomendasikan anaknya mengikuti seleksi kelas akselerasi. Pengujian ketepatan klasifikasi dilakukan dengan menggunakan uji Press's Q.

Hasil analisis data menunjukkan bahwa Regresi Logistik mempunyai tingkat akurasi klasifikasi sebesar 78,8 sedangkan MARS sebesar 77,7. Dengan demikian, model terbaik dalam memodelkan kelulusan siswa masuk kelas akselerasi adalah dengan menggunakan model MARS.

Kata Kunci: Akselerasi, Regresi Logistik, MARS, Press's Q

ABSTRACT

Regression Analysis is a statistical methodology that usually used for analyzing the relationship between a response and one or more predictor. When the response is categorical variable, then the regression methods that could be used are Logistic Regression and Multivariate Adaptive Regression Spline (MARS). The result of both modeling can be used to classify the objects. The aim of this research is to find a quantitative model for explaining factors that influenced the success of students in joining the acceleration class. Evaluation the accuracy of classification rate is done by implementing Press's Q statistic test.

The result of Logistic Regression shows that the classification accuracy is 74,8 where as MARS yields 77,7. Hence, MARS model is the best model for evaluating the success factor of student in joining the acceleration class.

Key Words: Acceleration, Logistic Regression, MARS, Press's Q

PENDAHULUAN

Perekrutan siswa dalam kelas akselerasi merupakan salah satu faktor yang sangat penting untuk diperhatikan, karena berkaitan dengan kelanjutan studi siswa itu kedepan. Depdiknas (2007) menyatakan bahwa perekrutan siswa dalam kelas akselerasi dilakukan dengan seleksi. Seleksi tersebut meliputi tes IQ dan tes Psikologi. Selanjutnya, proses multidimensional, artinya kriteria yang digunakan lebih dari satu (bukan sekedar inteligensia). Batasan yang digunakan adalah siswa yang mempunyai dimensi

kemampuan umum pada taraf cerdas ditetapkan skor IQ 130, dimensi kreativitas, dimensi komitmen dan dimensi kepribadian. Ketiga dimensi terakhir diukur lewat tes Psikologi.

Analisis Regresi adalah salah satu teknik statistik yang biasanya digunakan untuk menjelaskan hubungan antar variabel. Salah satu model regresi yang dapat digunakan untuk menjelaskan hubungan antar variabel respon yang meliputi 2 (dua) kategori dengan variabel prediktor yang terdiri dari variabel kontinyu, diskrit atau campuran antara keduanya adalah Regresi Logistik Dikhotomis (biner). Regresi Logistik ini dapat digunakan untuk mengklasifikasi sejumlah objek kedalam salah satu dari dua kelompok. Metode statistik lainnya yang dapat digunakan juga untuk pengklasifikasian objek dengan melibatkan variabel respon dan kategori dengan sejumlah variabel prediktor kontinyu ataupun kategori adalah metode *Multivariate Adaptive Regression Splines (MARS)*.

MARS merupakan salah satu model Regresi Nonparametrik Multivariat yang dikembangkan oleh Friedman tahun 1991. Metode ini merupakan pengembangan dari pendekatan *Recursive Partitioning Regression (PRR)* yang masih memiliki kelemahan dimana model yang dihasilkan tidak kontinyu pada knot. Proses pembentukan model *MARS* tidak mengasumsikan bentuk hubungan fungsional antara variabel respon dengan variabel prediktor, dan mempunyai bentuk fungsional yang fleksibel. Penentuan knot secara otomatis pada *MARS* menggunakan algoritma *forward stepwise* dan *backward stepwise* yang didasarkan pada nilai *Generalize d Cross Validation (GVC)* minimum.

Sehubungan dengan penelitian ini, maka variabel-variabel yang diduga berhubungan atau berpengaruh terhadap kelulusan siswa masuk kelas akselerasi terdiri dari faktor internal yang meliputi nilai UN untuk mata pelajaran Matematika, IPA dan Bahasa Indonesia, nilai rapor kelas VI untuk mata pelajaran Matematika, IPA dan Bahasa Indonesia, dan faktor eksternal yang meliputi jenis kelamin, pendidikan tertinggi ayah, pendidikan tertinggi ibu, pekerjaan ayah, pekerjaan ibu, urutan kelahiran, dan jumlah saudara.

Dengan demikian, penelitian ini bertujuan untuk menghasilkan model yang ideal terhadap faktor-faktor yang mempengaruhi siswa masuk kelas akselerasi menggunakan Analisis *Regresi Logistik* dan *MARS*, sehingga menjadi masukan bagi orang tua dalam merekomendasikan anaknya mengikuti tes ataukah tidak.

METODOLOGI PENELITIAN

1. Untuk mendapatkan model kelas akselerasi menggunakan analisis Regresi Logistik adalah sengan langkah-langkah sebagai berikut:
 - Mengidentifikasi variabel respon dan prediktor
 - Melakukan pengujian hipotesis
 - Menentukan model terbaik dari dua jenis data yang berbeda berdasarkan tingkat akurasi klasifikasi tertinggi
 - Melakukan *Uji Press's Q* untuk mengetahui signifikansi ketepatan klasifikasi
 - Menginterpretasi model
2. Untuk mendapatkan model kelas akselerasi menggunakan Regresi Nonparametrik *MARS* adalah dengan langkah-langkah sebagai berikut:
 - Mendefinisikan variabel respon dan prediktor
 - Melakukan *trial and error* untuk mendapatkan model terbaik dimana nilai *GCV* minimum
 - Pemilihan model terbaik dari dua jenis data berbeda dilakukan dengan melihat nilai akurasi klasifikasi tertinggi
 - Melakukan *Uji Press's Q* untuk mengetahui signifikansi ketepatan klasifikasi
 - Mendapatkan variabel yang berpengaruh signifikan dari pembentukan model *MARS*
3. Untuk memperoleh metode mana yang mempunyai performa terbaik, dilakukan tahapan sebagai berikut:
 - Mengambil nilai akurasi dan *NSR* pada klasifikasi untuk model terbaik Regresi Logistik
 - Mengambil nilai akurasi dan *NSR* pada klasifikasi untuk model terbaik *MARS*
 - Menentukan metode dengan performa terbaik dengan membandingkan nilai akurasi dan *NSR* pada dua metode tersebut

Proses pemodelan Regresi Logistik dilakukan dengan paket statistik *SPSS*, sedangkan pemodelan *MARS* menggunakan paket *MARS*.

LANDASAN TEORI

Regresi Logistik

Analisis Regresi Logistik adalah analisis yang digunakan untuk melihat hubungan antara variabel respon kategorik dengan variabel-variabel prediktor kategorik maupun kontinyu

dalam Regresi Logistik dapat berbentuk *dikhotomus* (biner) maupun *polykotomus* dengan skala data ordinal atau nominal (Agresti, 1990). Regresi Logistik dengan variabel respon berskala ordinal disebut Regresi Logistik Ordinal, sedangkan jika variabel respon berskala nominal dengan 2 (dua) kategori disebut Regresi Logistik Polikotomus. Regresi Logistik dapat digunakan untuk pengklasifikasian sejumlah objek ke dalam beberapa kelompok. Pada Regresi Logistik dikhotomus, respon Y hanya terdiri dari dua kategori (misalnya : 1 dan 0). Kondisi tersebut mengakibatkan respon Y berdistribusi Bernoulli. Distribusi Bernoulli untuk variabel random biner menghasilkan probabilitas sebagai berikut.

$$P(Y = y) = \pi(x)^y (1 - \pi(x))^{1-y} \quad (1)$$

Dimana $y=0,1$ adalah probabilitas kejadian yang diakibatkan oleh variabel predictor (x).

Regresi Logistik dikhotomus dapat disusun model yang terdiri dari banyak variabel prediktor yang dikenal sebagai model *multivariable*. Model Regresi Logistik *multivariable* dengan p variabel prediktor adalah:

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} \quad (2)$$

Jika model persamaan diatas ditransformasikan dengan transformasi logit, maka didapatkan fungsi logit sebagai berikut.

$$f(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] \quad (3)$$

Dengan mensubstitusikan persamaan (2) kedalam persamaan (3) maka diperoleh hasil sebagai berikut.

$$\ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (4)$$

Sedangkan *error* didefinisikan sebagai berikut.

$$y_i = \pi(x) + e_i \quad (5)$$

Jika $y_i = 1$, maka $e_i = 1 - \pi(x)$

Jika $y_i = 0$, maka $e_i = -\pi(x)$

Estimasi Parameter

Model Regresi Logistik dikhotomus dengan variabel respon bernilai 0 atau 1 dimana antar pengamatan diasumsikan saling bebas maka penduga parameter β diperoleh dengan metode *Maximum Likelihood Estimation (MLE)* dimana dengan metode ini parameter diestimasi dengan memaksimalkan fungsi turunan pertama. Estimasi varian dan kovarian diperoleh dari turunan kedua fungsi *log likelihood*. Pada dasarnya metode *MLE* memberikan nilai dugaan β dengan memaksimalkan fungsi *likelihood* (Hosmer & Lemeshow, 2000). Prinsip kerja pendugaan parameter dengan metode ini adalah jika terdapat fungsi *likelihood* $L(\beta_0, \beta_1, \dots, \beta_p)$ maka nilai $(\beta_0, \beta_1, \dots, \beta_p)$ akan diperoleh dengan memaksimalkan fungsi $L(\beta_0, \beta_1, \dots, \beta_p)$. Karena setiap pengamatan diasumsikan saling bebas maka fungsi *likelihood* merupakan fungsi kepadatan gabungan yaitu:

$$L(\beta) = \prod_{i=1}^n f(\beta, y_i) \quad (6)$$

Variabel respon dalam Regresi Logistik mengikuti sebaran Bernoulli dengan fungsi kepekatian peluang (*pdf*) adalah:

$$f(\beta, y_i) = [\pi(x_i)]^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (7)$$

Sehingga fungsi *likelihood* pada persamaan (6) menjadi:

$$L(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i}$$

Dimana $\pi(x)$ menunjukkan nilai $\pi(x)$ yang ke- x , dapat ditulis menjadi:

$$\ln[L(\beta)] = \sum_{i=1}^n [y_i (\beta_0 + \beta_1 x_{i1} + \beta_p x_{ip}) + \ln(1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_p x_{ip}))^{-1}] \quad (8)$$

Untuk mendapatkan nilai β yang memaksimalkan $L(\beta)$, maka perlu mendiferensialkan fungsi *log likelihood* terhadap parameter-parameternya dan membuat persamaan yang diperoleh = 0. Turunan pertama adalah:

$$\sum_{i=1}^n [y_i x_{ij} - x_{ij} \pi(x_i)] = 0; j = 1, 2, \dots, p \quad (9)$$

Hasil dari turunan pertama biasanya dimasukkan dalam sebuah vektor, yang disebut *vector gradient* ($\hat{\theta}$).

Sedangkan turunan kedua bentuk umumnya adalah:

$$\frac{\partial^2 \varphi(\beta_0, \beta_1, \dots, \beta_p)}{\partial \beta_j \partial \beta_k} = - \sum_{i=1}^n [x_{ij} x_{ik} \pi(x_i)(1 - \pi(x_i))] \quad (10)$$

Dimana j dan $k = 0, 1, \dots, p$

Hasilnya dimasukkan dalam matriks yang disebut matriks Hessian (\mathbf{H}). Sedangkan untuk memperoleh estimasi maksimum bagi parameter β maka digunakan metode iterasi Newton Raphson.

Pengujian Parameter

Model persamaan yang diperoleh perlu dilakukan pengujian signifikansi. Pengujian ini dimaksudkan untuk mengetahui apakah variabel yang terdapat dalam model memiliki kontribusi yang nyata bagi variabel respon. Pengujian yang dilakukan adalah:

a. Uji Serentak

Dilakukan untuk mengetahui signifikansi parameter β secara keseluruhan atau serentak.

Hipotesa untuk pengujian ini adalah:

$$H_0: \beta_0 = \beta_1 = \dots = \beta_p = 0$$

H_1 : Paling tidak ada satu $\beta_j \neq 0, j = 1, 2, \dots, p$

Dengan uji statistik:

$$G = -2 \ln \left[\frac{\left(\frac{n_1}{n}\right)^{n_1} \left(\frac{n_0}{n}\right)^{n_0}}{\prod_{i=1}^n (\pi_i)^{y_i} (1 - \pi_i)^{1 - y_i}} \right] \quad (11)$$

Dimana:

n_1 = Banyaknya siswa yang berada pada kelompok tidak lulus

n_0 = Banyaknya siswa yang berada pada kelompok lulus

n = Banyaknya pengamatan

Daerah penolakan: tolak H_0 apabila nilai $G > \chi^2_{(\alpha, 1)}$ dimana p merupakan banyaknya variabel prediktor dalam model atau $P\text{-value} < \alpha$.

b. Uji Individu

Dimaksudkan untuk memeriksa signifikansi parameter β secara individu. Hipotesanya adalah:

$$H_0: \beta_j = 0$$

$H_1: \beta_j \neq 0, j = 1, 2, \dots, p$

Dengan uji statistik :

$$W \text{ (Wald)} = \frac{\hat{\beta}_j}{S.E(\hat{\beta}_j)} \quad (12)$$

Daerah penolakan: tolak H_0 apabila $W_j^2 > \chi^2_{(\alpha, 1)}$ atau $P\text{-value} < \alpha$.

Interpretasi Model

Odds ratio (θ) merupakan salah satu ukuran tingkat resiko yang digunakan dalam menginterpretasi parameter. *Odds ratio* didefinisikan sebagai berikut:

$$\theta = \frac{\frac{\pi(1)}{1 - \pi(1)}}{\frac{\pi(0)}{1 - \pi(0)}} = \frac{\pi(1)[1 - \pi(0)]}{\pi(0)[1 - \pi(1)]} \quad (13)$$

$$= \exp(\beta_j)$$

Sehingga $\hat{\theta} = \exp(\hat{\beta}_j)$ maka $\ln \hat{\theta} = \hat{\beta}_j$ (14)

Pemodelan Regresi Nonparametrik Spline

Model Regresi *Spline* memberikan sebuah bentuk persamaan yang merepresentasikan bentuk *parametric polynomial piecewise* (Hastie dan Tibshirani, 1990). Pemodelan *parametric piecewise* menggunakan ide dasar yaitu fungsi f yang didekati oleh beberapa fungsi parametrik yang didefinisikan pada setiap region di dalam domain D . Setiap region dipisahkan oleh titik-titik *knots*, dan fungsi parametrik yang didefinisikan pada setiap region biasanya disebut sebagai fungsi basis. *Knots* merupakan akhir dari sebuah region dan awal dari sebuah region yang lain (Steinberg et al., 1999). Pemodelan

Regresi *Spline* diimplementasikan dengan membentuk kumpulan fungsi basis yang dapat mencapai pendekatan *spline* orde ke- q dan mengestimasi koefisien fungsi-fungsi basis tersebut menggunakan *least-square*. Untuk kasus data *univariat* dengan $v=1$, salah satu bentuk basis adalah sebagai berikut:

$$1, \{x^j\}_i^q, \{(x - t_k)_+^q\}_i^k \quad (15)$$

Dimana, $\{t_k\}_i^k$ adalah titik *knots*. Sedangkan Estimasi dari f secara umum didapatkan melalui *PLS* yaitu meminimumkan fungsi berikut (Eubank, 1988):

$$n^{-1} \sum_{i=1}^n [y_i - f(x_i)]^2 + \lambda \int_a^b [f^{(m)}(x)]^2 dx, \lambda \geq 0 \quad (16)$$

Multivariate Adaptive Regression Spline (MARS)

MARS merupakan suatu pendekatan analisis untuk memodelkan *Regresi Nonparametik Multivariate* dengan kombinasi yang kompleks dari *spline* dan *rekursif partisi*. Pendekatan ini dikenalkan oleh Friedman (1991). Model *MARS* digunakan untuk mengatasi kelemahan *RPR* yang menghasilkan model yang kontinyu pada *knots*. Penentuan *knots* secara otomatis pada *MARS* menggunakan algoritma *forward stepwise* dan *backward stepwise* yang didasarkan pada nilai *GCV* minimum.

Model *MARS* adalah sebagai berikut:

$$Y = \beta_0 + \sum_{m=1}^M \beta_m \mathbf{B}_m(X) + \varepsilon \quad (17)$$

dimana:

X adalah variabel prediktor

$\mathbf{B}_m(X)$ adalah Fungsi Basis

Setelah dilakukan modifikasi model *RPR* dan dikombinasikan dengan *spline* maka diperoleh model *MARS* sebagai berikut:

$$\hat{f}(x) = \alpha + \sum_{i=1}^v f_i(x_i) + \sum_{i,j=1}^v f_{ij}(x_i, x_j) + \sum_{i,j,k=1}^v f_{ijk}(x_i, x_j, x_k) + \dots \quad (18)$$

HASIL DAN PEMBAHASAN

Analisis Regresi Logistik

Pemodelan kelas akselerasi dengan menggunakan analisis Regresi Logistik dilakukan sesuai tahapan pada metodologi diatas. Tahapan awal adalah mengidentifikasi variabel respon dan prediktor. Variabel respon (Y) yaitu kelulusan siswa dalam kelas akselerasi yang dikategorikan menjadi LULUS dan TIDAK LULUS. Sedangkan variabel prediktor (X) terdiri atas: jenis kelamin (x_1), nilai rapor Matematika (x_2), nilai rapor IPA (x_3), nilai rapor Bahasa Indonesia (x_4), nilai UN Matematika (x_5), nilai UN IPA (x_6), nilai UN Bahasa Indonesia (x_7), pendidikan Ayah (x_8), pendidikan Ibu (x_9), pekerjaan Ayah (x_{10}), pekerjaan Ibu (x_{11}), urutan kelahiran (x_{12}), dan jumlah saudara (x_{13}). Selanjutnya analisis dilakukan secara *parsial* dan *multiple*. Pengujian dilakukan sebanyak dua kali meliputi pengujian pada semua variabel prediktor yang memuat data variabel x_2, x_3, x_4, x_5, x_6 dan x_7 adalah kontinyu dan data variabel x_2, x_3, x_4, x_5, x_6 dan x_7 adalah kategorik pada variabel prediktor nilai rapor dan UN.

Hasil pengujian secara parsial dengan data x_2, x_3, x_4, x_5, x_6 dan x_7 bertipe kontinyu menunjukkan bahwa dengan tingkat signifikansi 5%, maka terdapat lima variabel secara sendiri-sendiri signifikan berpengaruh yaitu jenis kelamin (x_1), nilai rapor Matematika (x_2), nilai rapor IPA (x_3), nilai rapor Bahasa Indonesia (x_4), dan nilai UN Matematika (x_5). Sedangkan hasil pengujian secara parsial bertipe kategori menunjukkan bahwa dengan tingkat signifikansi 5%, maka terdapat empat variabel yang secara sendiri-sendiri signifikan berpengaruh terhadap kelulusan siswa yaitu jenis kelamin (x_1), nilai rapor Matematika (x_2), nilai rapor IPA (x_3), nilai rapor Bahasa Indonesia (x_4).

Pada pengujian secara *multiple* bertipe kontinyu, hasilnya adalah dengan tingkat signifikansi 10%, maka terdapat tiga variabel prediktor secara bersama-sama signifikan berpengaruh terhadap kelulusan, yaitu nilai rapor Matematika (x_2), pendidikan Ayah (x_8) dan jumlah saudara (x_{12}). Sedangkan hasil pengujian *multiple* bertipe kategori, hasilnya adalah dengan tingkat signifikansi 10%, maka terdapat empat variabel secara bersama-sama berpengaruh terhadap kelulusan yaitu jenis kelamin (x_1), nilai rapor Bahasa Indonesia (x_4), dan pendidikan Ayah (x_8).

Dari kedua tipe data, model terbaik dipilih berdasarkan nilai akurasi dan *Noise Signal Ratio (NSR)*. Dari hasil pengolahan data, ternyata data kategori adalah model terbaik dimana nilai akurasi lebih

besar dari data kontinyu (74,8 dan 70,1) dan *NSR* lebih kecil dari data kontinyu (0,28 dan 0,34). Sehingga, model Regresi Logistik yang memuat empat variabel prediktor yang signifikan adalah:

$$\hat{f}(x) = -4.217 + 0.997X_1(1) + 2.295X_2(1) + 0.058X_2(2) + 1.313X_2(3) + 0.657X_6(1) + 0.993X_6(2) + 2.278X_6(3) + 1.198X_9(1)$$

Model tersebut selanjutnya dapat diinterpretasi. Interpretasi dilakukan berdasarkan nilai *Odds Ratio* (9) dari model yang ditemukan. Interpretasi dilakukan terhadap variabel yang signifikan dalam model. Interpretasinya adalah sebagai berikut:

1. Variabel X_1 : peluang siswa perempuan lulus adalah sebesar 2,7 kali siswa laki-laki.
2. Variabel $X_2(1)$: siswa yang nilai rapor Bahasa Indonesia berada pada interval 9,01-10,00 berpeluang lulus sebesar 9,93 kali dari siswa yang nilainya berada pada interval 7,51-8,00.
3. Variabel $X_6(3)$: siswa yang nilai UN Bahasa Indonesia berada pada interval 8,51-9,00 berpeluang lulus masuk kelas akselerasi sebesar 9,75 kali dari siswa yang nilainya berada pada interval 8,00-8,50.
4. Variabel X_9 : siswa yang ayahnya berpendidikan sampai perguruan tinggi mempunyai peluang lulus sebesar 3,3 kali dari siswa yang ayahnya berpendidikan tidak sampai perguruan tinggi.

Pemodelan Dengan MARS

Pemodelan dengan MARS dilakukan secara *trial and error*, dengan melakukan kombinasi banyaknya *Fungsi Basis (FB)*, *Maksimum Interaksi (MI)* dan *Minimum Observasi (MO)*. Banyaknya Fungsi Basis yang digunakan adalah 2-4 kali banyaknya variabel prediktor yang diduga berpengaruh terhadap variabel respon. *Maximum Interaksi* adalah 1, 2, dan 3 yang sekaligus menggambarkan tingkat interaksi, sedangkan *Minimum Observasi* diambil mulai dari 0 sampai lebih dari setengah jumlah sampel data. Analisis dilakukan untuk dua macam data, yaitu kontinyu dan kategori seperti pada Regresi Logistik.

Dari hasil pengolahan, diperoleh nilai *GCV* minimum untuk ketiga $FB = 26, 39$ dan 52 . Maka selanjutnya nilai *GCV* minimum akan diperoleh dengan membandingkan ketiga nilai *GCV* minimum untuk masing-masing FB . Hasilnya diperoleh *GCV* minimum berada pada kombinasi dengan maksimum *Fungsi Basis* = 39, *Maksimum Interaksi* = 2 dan *Minimum Observasi* = 8,9. Selanjutnya dilakukan analisis dengan nilai $GCV = 0,209$.

Untuk memilih model terbaik dari kedua tipe data, dilakukan dengan kriteria nilai akurasi dan *NSR*, dimana model terbaik adalah model dengan nilai akurasi terbesar dan *NSR* terkecil. Dari hasil pengolahan, tipe data kontinyu merupakan model terbaik karena memiliki nilai akurasi sebesar 77,7 dan *NSR* 0,07, dibandingkan dengan tipe data kategori (akurasi=65,4 dan *NSR*=0,52). Sedangkan hasil pengujian *Press's Q* tipe data kontinyu sebesar 32,53 dan tipe data sedangkan data kategori sebesar 10,18. Nilai ini jika dibandingkan dengan nilai $\chi^2_{(\alpha=0,05;1)}$, ternyata lebih besar, sehingga disimpulkan baik data kontinyu maupun kategori adalah konsisten.

Selanjutnya dibuat Model MARS. Pada model tersebut, terdapat tujuh variabel prediktor yang masuk dalam model yaitu nilai rapor Matematika (x_2), nilai rapor IPA (x_3), nilai UN Matematika (x_5), nilai UN IPA (x_6), pendidikan Ibu (x_9), pekerjaan Ayah (x_{10}) dan urutan kelahiran (x_{12}). Modelnya adalah sebagai berikut:

$$\hat{f}(x) = 0,1 - 2,512FB_8 - 0,428FB_9 + 2,11FB_{26} + 0,154FB_{36} \quad (19)$$

Dengan:

$$FB_8 = \max(0, X_2 - 8,00) \cdot FB_7$$

$$FB_9 = (X_9=1) \cdot FB_2$$

$$FB_{26} = \max(0, X_2 - 8,00) \cdot FB_{25}$$

$$FB_{36} = (X_{12} = 1) \cdot FB_{35}$$

dan :

$$FB_7 = \max(0, 9,25 - X_5) \cdot FB_1$$

$$FB_2 = (X_{10} = 3) \cdot FB_1$$

$$FB_{25} = \max(0, 9,5 - X_5) \cdot FB_1$$

$$FB_{35} = \max(0, X_7 - 7,3) \cdot FB_1$$

Serta :

$$FB_1 = \max(0, X_4 - 7,83)$$

Model *MARS* tersebut dapat diinterpretasi sebagai berikut:

1. Koefisien FB_8 : untuk setiap kenaikan 1 angka pada FB_8 maka kecenderungan siswa lulus turun sebesar 2,512 dengan fungsi basis lainnya yang masuk dalam model dianggap konstan.
2. Koefisien FB_9 : untuk setiap kenaikan 1 angka pada FB_9 maka kecenderungan siswa lulus turun sebesar 0,428 dengan fungsi basis lainnya yang masuk dalam model dianggap konstan.
3. Koefisien FB_{26} : untuk setiap kenaikan 1 angka pada FB_{26} maka kecenderungan siswa lulus naik sebesar 2,11 dengan fungsi basis lainnya yang masuk dalam model dianggap konstan.
4. Koefisien FB_{36} : untuk setiap kenaikan 1 angka pada FB_{36} maka kecenderungan siswa lulus naik sebesar 0,154 dengan fungsi basis lainnya yang masuk dalam model dianggap konstan.

Perbandingan Hasil Klasifikasi Regresi Logistik dengan *MARS*

Perbandingan dimaksudkan untuk menentukan metode paling sesuai yang menghasilkan model terbaik. Perbandingan didasarkan pada tingkat akurasi dan *NSR*. Hasilnya adalah:

Tingkat Klasifikasi Regresi Logistik dan *MARS*

Metode	Sensitivity	Spesitivity	Accuracy	NSR
Regresi Logistik	85,9	52,8	74,8	0,28
<i>MARS</i>	97,2	38,9	77,7	0,07

Berdasarkan nilai akurasi dan *NSR*, maka model *MARS* merupakan metode terbaik dalam memodelkan kelulusan siswa masuk kelas akselerasi dengan tingkat akurasi sebesar 77,7% dan *NSR* sebesar 0,07.

KESIMPULAN

1. Pemodelan kelas akselerasi dengan menggunakan Regresi Logistik menghasilkan model terbaik adalah yang memuat variabel prediktor X_2, X_3, X_4, X_5, X_6 dan X_7 bertipe kategori sedangkan menggunakan *MARS* menghasilkan model terbaik adalah yang memuat variabel prediktor X_2, X_3, X_4, X_5, X_6 dan X_7 bertipe kontinyu, dimana:

X_2 : Nilai rapor Matematika

X_3 : Nilai rapor Bahasa Indonesia

X_4 : Nilai rapor IPA

X_5 : Nilai UN Matematika

X_6 : Nilai UN Bahasa Indonesia

X_7 : Nilai UN IPA

2. Pemodelan dengan Regresi Logistik menghasilkan tingkat akurasi klasifikasi sebesar 74,8 dengan empat variabel prediktor yang signifikan berpengaruh terhadap kelulusan siswa masuk kelas akselerasi yaitu Jenis Kelamin (X_1), Nilai Rapor Bahasa Indonesia (X_2), Nilai UN Bahasa Indonesia (X_6) dan Pendidikan Tertinggi Ayah (X_8). Model Regresi Logistik adalah sebagai berikut:

$$\hat{f}(x) = -4.217 + 0.997X_1(1) + 2.295X_2(1) + 0.058X_3(2) + 1.313X_4(3) + 0.657X_6(1) + 0.993X_6(2) + 2.278X_6(3) + 1.198X_8(1)$$

3. Pemodelan dengan *MARS* menghasilkan model sebagai berikut:

$$\hat{f}(x) = 0.1 - 2.512FB_8 - 0.428FB_9 + 2.11FB_{26} + 0.154FB_{36}$$

Dengan:

$$FB_8 = \max(0, X_2 - 8,00) \cdot FB_7$$

$$FB_9 = (X_9 = 1) \cdot FB_2$$

$$FB_{26} = \max(0, X_2 - 8,00) \cdot FB_{25}$$

$$FB_{36} = (X_{12} = 1) \cdot FB_{35}$$

dan :

$$FB_7 = \max(0, 9,25 - X_5) \cdot FB_1$$

$$FB_2 = (X_{10} = 3) \cdot FB_1$$

$$FB_{25} = \max(0, 9,5 - X_5) \cdot FB_1$$

$$FB_{35} = \max(0, X_7 - 7,3) \cdot FB_1$$

Serta :

$$FB_1 = \max(0, X_4 - 7,83)$$

4. Variabel yang memberikan kontribusi pada pemodelan kelas akselerasi dengan MARS adalah meliputi interaksi tiga variabel yaitu untuk FB_9 adalah nilai rapor Matematika (x_2), nilai UN Matematika (x_5) dan nilai rapor IPA (x_3). Untuk FB_9 adalah pekerjaan Ibu (x_{11}), pekerjaan Ayah (x_{10}), dan nilai rapor IPA (x_3). Untuk FB_{26} adalah nilai rapor Matematika (x_2), nilai UN Matematika (x_5) dan nilai rapor IPA (x_3). Sedangkan untuk FB_{26} adalah urutan kelahiran (x_{12}), nilai UN IPA (x_6) dan nilai rapor IPA (x_3). Tingkat Akurasi Klasifikasi adalah sebesar 77,7.

DAFTAR PUSTAKA

- Agresti, A., (1990), "Categorical Data Analysis", John Wiley and Son, Florida.
- Akbar, R., (2004) "Akselerasi", PT Gramedia Widya Sarana Indonesia, Jakarta.
- Breiman, L., Freidman, J. H., Olshen, R. A., and Stone, C. J., (1984), "Classification and Regression Tree", Wadsworth, Beelmont, C. A.
- Eubank, R., L., (1988), "Spline Smoothing and Nonparametric Regression", Marcel Dekker, New York.
- Friedman, J. H., (1991), "Multivariate Adaptive Regression Splines, The Annals of Statistics", Vol. 19, No. 1. (Mar., 1991), hal 1-67.
- Hair, J. F. JR., Anderson, R. E., Tatham, R. L. and Black, W. C., (2006), "Multivariate Data Analysis", Fifth Edition, Prentice-Hall, International, Inc.
- Steinberg, D., Colla, P. L. and Kerry, M., (1999), "MARS user Guide", San Diego, CA: Salford System