

## KLASIFIKASI KELULUSAN MAHASISWA FAKULTAS SAINS DAN MATEMATIKA UNIVERSITAS DIPONEGORO MENGGUNAKAN *MULTIVARIATE ADAPTIVE REGRESSION SPLINE (MARS)*

Rizal Yuniarto Ghofar<sup>1</sup>, Diah Safitri<sup>2</sup>, Agus Rusgiyono<sup>3</sup>

<sup>1</sup>Mahasiswa Jurusan Statistika FSM UNDIP

<sup>2,3</sup>Staff Pengajar Jurusan Statistika FSM UNDIP

### ABSTRACT

Education is a top priority for today's society. The quality of education can be seen from the learning achievement. There are so many factors that influence learning achievement in this regard graduation, therefore, necessary to identify the most influential factors that will be used to improve the quality of education. This study was conducted to obtain a model that is capable of classifying the data Faculty of Science and Mathematics Diponegoro University Semarang graduation using Multivariate Adaptive Regression Spline (MARS) method. MARS is a nonparametric regression method that can be used for data of high dimension. To get the best MARS models, made possible combinations Basis Function (BF), Maximum Interaction (MI), and Minimum Observation (MO) by trial and error. The best model is the model that is used in combination with BF = 28, MI = 2, MO = 1 because it has the smallest GCV value that is equal to 0,17781. There are three variables that contribute to the MARS model of the variable GPA, majors and gender. As for the variable organization, part time, entry point, and scholarships do not contribute to the model. Obtained misclassification of 20,50%. Press's Q test value indicates that statistically MARS method has been consistent in classifying the data FSM Diponegoro University Semarang graduation.

### 1. PENDAHULUAN

Pengertian pendidikan menurut Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 adalah usaha sadar dan terencana untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukan dirinya, masyarakat, bangsa, dan negara. Jalur pendidikan dibagi menjadi 3 (tiga), yaitu pendidikan formal, pendidikan nonformal dan pendidikan informal. Pendidikan formal merupakan salah satu jalur pendidikan yang terstruktur dan berjenjang yang terdiri atas 3 (tiga) tahapan yaitu pendidikan dasar, pendidikan menengah, dan pendidikan tinggi (*Undang-Undang Republik Indonesia Nomor 20 Tahun 2003*).

Keputusan Menteri Pendidikan Nasional Republik Indonesia Nomor 232/U/2000 BAB III Pasal 5 menjelaskan bahwa beban studi program sarjana sekurang-kurangnya 144 SKS dan sebanyak-banyaknya 160 SKS yang dijadwalkan untuk 8 semester dan dapat ditempuh dalam waktu kurang dari 8 semester dan selama-lamanya 14 semester setelah pendidikan menengah. Pada penelitian ini, peneliti ingin mengklasifikasi data kelulusan mahasiswa ke dalam dua kategori yaitu  $\leq 48$  bulan untuk mahasiswa yang menempuh pendidikan S1 dengan lama studi kurang dari sama dengan 48 bulan dan lulus  $> 48$  bulan untuk

mahasiswa yang yang menempuh pendidikan S1 dengan lama studi lebih dari 48 bulan (<http://www.dikti.go.id/files/Lemkerma/kepmen232-2000.txt>).

Metode klasifikasi yang digunakan dalam penelitian ini adalah *Multivariate Adaptive Regression Spline* (MARS). Metode MARS merupakan pendekatan untuk regresi nonparametrik yang berguna untuk mengatasi data berdimensi tinggi dan menghasilkan prediksi variabel respon yang akurat dengan menghasilkan model yang kontinu pada *knot* berdasarkan nilai *Generalized Cross Validation* (GCV) terkecil (Friedman, 1991). Dari uraian di atas, peneliti tertarik mengangkat topik yang berjudul “Klasifikasi Kelulusan Mahasiswa Fakultas Sains dan Matematika Universitas Diponegoro Menggunakan *Multivariate Adaptive Regression Spline* (MARS)”.

## 2. TINJAUAN PUSTAKA

### 2.1 Prestasi Belajar

Belajar merupakan serangkaian kegiatan jiwa raga untuk memperoleh suatu perubahan tingkah laku sebagai hasil dari pengalaman individu dalam interaksi dengan lingkungannya. Sedangkan prestasi belajar adalah perubahan yang terjadi sebagai akibat dari kegiatan belajar yang telah dilakukan oleh individu (Djamarah, 2002). Secara garis besar faktor-faktor yang mempengaruhi prestasi belajar dapat dikelompokkan menjadi dua, yaitu faktor dari luar (*eksternal*) dan faktor dari dalam (*internal*). Faktor *eksternal* bersumber dari luar individu yang bersangkutan yakni kondisi lingkungan di sekitar siswa. Faktor *eksternal* biasa disebut faktor ekstrinsik yang meliputi faktor lingkungan dan faktor instrumental. Faktor lingkungan mencakup lingkungan alam dan lingkungan sosial. Sedangkan faktor instrumental mencakup kurikulum/bahan pelajaran, guru/pengajar, sarana dan fasilitas serta administrasi/manajemen. Faktor *internal* merupakan faktor yang berasal dari dalam diri siswa yang meliputi dua aspek yaitu aspek fisiologis (yang bersifat jasmaniah) seperti kondisi fisik dan kondisi panca indra, dan aspek psikologis (yang bersifat rohaniyah) seperti tingkat kecerdasan/intelegensi, sikap, bakat, minat dan motivasi (Syah, 2009).

### 2.2 *Multivariate Adaptive Regression Spline* (MARS)

*Multivariate Adaptive Regression Splines* (MARS) merupakan metode dengan pendekatan regresi nonparametrik. Model MARS berguna untuk mengatasi permasalahan data berdimensi tinggi dan menghasilkan prediksi variabel respon yang akurat, dan menghasilkan model kontinu dalam *knot* berdasarkan nilai *Generalized Cross Validation* (GCV) terkecil. Data berdimensi tinggi yang dimaksud adalah data dengan ukuran  $3 \leq n \leq 20$ , dimana  $n$  adalah banyak variabel dan sampel data yang berukuran  $50 \leq N \leq 1000$ , dimana  $N$  untuk ukuran sampel (Friedman, 1991).

Pintowati dan Otok (2012) menyatakan ada tiga hal yang perlu diperhatikan dalam pemodelan MARS yaitu *knot*, *basis function* (BF) dan Interaksi. *Knot* dapat didefinisikan sebagai akhir dari sebuah garis regresi (*region*) dan awal dari sebuah garis regresi (*region*) yang lain. Metode MARS dapat menentukan *knot* secara otomatis oleh data dan menghasilkan model yang kontinu pada *knot*. Jarak minimum antar *knot* atau observasi minimum (MO) yang digunakan sebesar nol, satu, dua dan tiga. Sedangkan fungsi basis (BF) dapat didefinisikan sebagai selang antar *knot* yang berurutan. Friedman (1991) menyarankan maksimum fungsi basis

(BF) adalah dua sampai empat kali banyaknya variabel prediktor. Interaksi merupakan *cross product* atau hasil perkalian silang antar variabel satu dengan variabel yang lain yang saling berhubungan. Interaksi maksimum (MI) adalah satu, dua dan tiga dengan pertimbangan jika lebih dari tiga akan menghasilkan model yang sangat kompleks dan sulit untuk diinterpretasikan.

Model umum persamaan MARS menurut Friedman (1991) dapat ditulis sebagai berikut :

$$f(x) = \alpha_0 + \sum_{m=1}^M \alpha_m \prod_{k=1}^{K_m} [s_{km}(x_{v(k,m)} - t_{km})]_+ + \varepsilon \quad (1)$$

dengan,

$\alpha_0$  = konstanta regresi dari fungsi basis

$\alpha_m$  = koefisien dari fungsi basis ke- $m$

$M$  = maksimum fungsi basis (*nonconstant* fungsi basis)

$K_m$  = derajat interaksi

$S_{km}$  = bernilai  $\pm 1$

$x_{v(k,m)}$  = variabel prediktor

$t_{km}$  = nilai *knot* dari variabel prediktor  $x_{v(k,m)}$

Berdasarkan persamaan (1), model MARS dapat ditulis sebagai berikut :

$$f(x) = \alpha_0 + \sum_{m=1}^M \alpha_m B_m(x) + \varepsilon \quad (2)$$

dengan  $B_m(x) = \prod_{k=1}^{K_m} [s_{km}(x_{v(k,m)} - t_{km})]_+$

sehingga dapat dituliskan dalam bentuk matriks menjadi seperti berikut :

$$Y = B\alpha + \varepsilon \quad (3)$$

dengan,

$Y = (Y_1, \dots, Y_N)^T$ ,  $\alpha = (\alpha_0, \dots, \alpha_M)^T$ ,  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)^T$

$$B = \begin{bmatrix} 1 & \prod_{k=1}^{K_1} [s_{k1}(x_{v1(k,1)} - t_{k1})]_+ & \dots & \prod_{k=1}^{K_M} [s_{kM}(x_{v1(k,M)} - t_{kM})]_+ \\ 1 & \prod_{k=1}^{K_1} [s_{k1}(x_{v2(k,1)} - t_{k1})]_+ & \dots & \prod_{k=1}^{K_M} [s_{kM}(x_{v2(k,M)} - t_{kM})]_+ \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \prod_{k=1}^{K_1} [s_{k1}(x_{vN(k,1)} - t_{k1})]_+ & \dots & \prod_{k=1}^{K_M} [s_{kM}(x_{vN(k,M)} - t_{kM})]_+ \end{bmatrix}$$

Persamaan (1) dapat dituliskan ke dalam bentuk sebagai berikut :

$$\hat{f}(x) = \alpha_0 + f_i(x_i) + f_{ij}(x_i, x_j) + f_{ijk}(x_i, x_j, x_k) + \dots \quad (4)$$

Persamaan (4) menunjukkan bahwa penjumlahan pertama meliputi semua fungsi basis untuk satu variabel, penjumlahan kedua meliputi semua fungsi basis untuk interaksi antara dua variabel, penjumlahan ketiga meliputi semua fungsi basis untuk interaksi antara tiga variabel dan seterusnya. Misalkan  $V(m) = \{v(k, m)\}_1^{K_m}$  adalah himpunan dari peubah yang dihubungkan dengan fungsi basis  $B_m$  ke- $m$ , maka setiap penjumlahan pertama pada persamaan (4) dapat dituliskan sebagai berikut :

$$f_i(x_i) = \sum_{i \in V(m)} \alpha_m B_m(x_i) \quad (5)$$

$f_i(x_i)$  merupakan penjumlahan semua fungsi basis untuk satu variabel  $x_i$  dan merupakan *spline* dengan derajat  $q=1$  yang mempresentasikan fungsi *univariate*. Setiap fungsi *bivariate* pada persamaan (4) dapat dituliskan sebagai berikut :

$$f_{ij}(x_i, x_j) = \sum_{(i,j) \in V(m)} \alpha_m B_m(x_i, x_j) \quad (6)$$

dengan  $B_m(x_i, x_j) = [s_{2m}(x_{i(2,m)} - t_{2m})]_+ [s_{2m}(x_{j(2,m)} - t_{2m})]_+$   
 Persamaan (6) merupakan penjumlahan semua fungsi basis dua variabel  $x_i$  dan  $x_j$ .  
 Untuk fungsi *trivariate* pada penjumlahan yang ketiga diperoleh dengan menjumlahkan semua fungsi basis untuk tiga variabel prediktor, yang dituliskan sebagai berikut :

$$f_{ijk}(x_i, x_j, x_k) = \sum_{(i,j,k) \in V(m)} \alpha_m B_m(x_i, x_j, x_k) \quad (7)$$

dengan

$$B_m(x_i, x_j, x_k) = [s_{3m}(x_{i(3,m)} - t_{3m})]_+ [s_{3m}(x_{j(3,m)} - t_{3m})]_+ [s_{3m}(x_{k(3,m)} - t_{3m})]_+$$

Persamaan (4) dikenal dengan dekomposisi ANOVA dari model MARS. Interpretasi model MARS melalui dekomposisi ANOVA adalah mempresentasikan variabel prediktor yang masuk ke dalam model, baik untuk satu variabel prediktor maupun interaksi lebih dari satu variabel prediktor

Penentuan *knot* pada metode MARS dapat ditentukan secara otomatis oleh data dan menghasilkan model yang kontinu pada *knot*. Penentuan *knot* pada MARS menggunakan *forward stepwise* dan *backward stepwise*. Untuk mendapatkan model dengan fungsi basis maksimum digunakan *forward stepwise*. Sedangkan *backward stepwise* digunakan untuk memenuhi prinsip parsimoni (model yang sederhana) dengan cara mengeliminasi fungsi basis yang berkontribusi kecil sampai tidak ada fungsi basis yang dapat dieliminasi. Ukuran kontribusi pada *backward stepwise* ditentukan dengan kriteria *Generalized Cross Validation* (GCV) (Pintowati dan Otok, 2012). Bentuk fungsi GCV minimum menurut Friedman (1991) dapat didefinisikan pada persamaan berikut :

$$GCV(M) = \frac{\frac{1}{N} \sum_{i=1}^N [y_i - \hat{f}_M(x_i)]^2}{\left[1 - \frac{\tilde{C}(M)}{N}\right]^2} \quad (8)$$

dengan,

N = ukuran sampel

$y_i$  = variabel respon

$x_i$  = variabel prediktor

M = banyaknya fungsi basis (*nonconstant* fungsi basis)

$\tilde{C}(M) = C(M) + dM$

$C(M) = \text{Trace}[\mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T] + 1$

d = nilai ketika setiap fungsi basis mencapai nilai optimal ( $2 \leq d \leq 4$ )

Cox dan Snell (1989) menyatakan bahwa jika variabel respon terdiri dari dua nilai, maka dapat dikatakan sebagai regresi dengan *binary response*. Sehingga didapatkan model probabilitas dengan persamaan sebagai berikut :

$$\pi(\mathbf{x}) = \frac{e^{\hat{f}(\mathbf{x})}}{1 + e^{\hat{f}(\mathbf{x})}} \quad (9)$$

$$\{1 - \pi(\mathbf{x})\} = \frac{1}{1 + e^{\hat{f}(\mathbf{x})}} \quad (10)$$

dengan  $\hat{f}(\mathbf{x}) = \text{logit } \pi(\mathbf{x})$

sehingga didapatkan  $\text{prob}(Y = 1) = \pi(\mathbf{x})$  dan  $\text{prob}(Y = 0) = 1 - \pi(\mathbf{x})$

Karena variabel respon Y bersifat biner (0 dan 1) dengan variabel prediktor X sebanyak n,  $\mathbf{x} = (x_1, \dots, x_n)$ , maka model MARS untuk klasifikasi dapat dituliskan sebagai berikut :

$$\text{logit } \pi(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha_0 + \sum_{m=1}^M \alpha_m \prod_{k=1}^{K_m} [s_{km}(x_{v(k,m)} - t_{km})]_+ \quad (11)$$

### 2.3 Apparent Error Rate (APER) dan Press's Q

Menurut Prasetyo (2012), sebuah sistem yang melakukan klasifikasi diharapkan dapat melakukan klasifikasi semua *data set* dengan benar, tetapi tidak dapat dipungkiri bahwa kinerja suatu sistem tidak bisa 100% benar sehingga sebuah sistem klasifikasi juga harus diukur kinerjanya. Pada umumnya, pengukuran klasifikasi dilakukan dengan matriks konfusi (*confusion matrix*). Jika  $y_0$  dan  $y_1$  merupakan subjek pengklasifikasian, maka bentuk matriks konfusi dapat dilihat pada tabel berikut :

**Tabel 1.** Matriks Konfusi untuk Klasifikasi Dua Kelas

| Hasil Observasi<br>(Actual Class) | Taksiran (Predicted Class) |          |
|-----------------------------------|----------------------------|----------|
|                                   | $y_0$                      | $y_1$    |
| $y_0$                             | $f_{00}$                   | $f_{01}$ |
| $y_1$                             | $f_{10}$                   | $f_{11}$ |

Menurut Johnson dan Wichern (1992), *Apparent Error Rate* (APER) adalah prosedur evaluasi yang digunakan untuk melihat kesalahan klasifikasi yang dilakukan oleh suatu fungsi klasifikasi. Nilai APER menunjukkan nilai proporsi sampel yang salah diklasifikasikan pada fungsi klasifikasi. Perhitungan besarnya nilai APER dapat dihitung dengan menggunakan rumus :

$$APER = \frac{f_{01} + f_{10}}{f_{00} + f_{01} + f_{10} + f_{11}} \times 100\% \quad (12)$$

Menurut Hair *et al* (2006), *Press's Q* merupakan statistik uji untuk mengetahui kestabilan dalam pengklasifikasian atau sejauh mana kelompok-kelompok tersebut dapat dipisahkan dengan menggunakan variabel yang ada. Perhitungan nilai *Press's Q* dapat dihitung dengan menggunakan rumus :

$$Press's Q = \frac{[N - (rK)]^2}{N(K-1)} \quad (13)$$

dengan,

$N$  = ukuran sampel

$r$  = banyaknya individu yang tepat diklasifikasikan

$K$  = banyaknya kelompok (kategori)

Nilai dari *Press's Q* membandingkan antara banyaknya ketepatan klasifikasi dengan ukuran sampel dan banyaknya kelompok. Nilainya dibandingkan dengan nilai kritis (tabel *chi-square* dengan derajat bebas 1). Jika nilai *Press's Q* melebihi nilai kritis, maka klasifikasi dapat dianggap sudah stabil dan konsisten secara statistik.

## 3. METODOLOGI PENELITIAN

### 3.1 Data dan Variabel Penelitian

Data yang digunakan pada penelitian ini adalah data primer dan data sekunder yang merupakan data kelulusan mahasiswa Fakultas Sains dan Matematika Universitas Diponegoro periode 130, 131, 132, dan 133.

Untuk variabel respon (Y) terdiri dari dua kategori, yaitu :

$Y = 0$ , untuk mahasiswa yang lulus  $\leq 48$  bulan

$Y = 1$ , untuk mahasiswa yang lulus  $> 48$  bulan

Sedangkan untuk variabel prediktor terdiri dari 7 komponen standar yaitu Jenis Kelamin ( $X_1$ ), IPK ( $X_2$ ), Jurusan ( $X_3$ ), Jalur Masuk ( $X_4$ ), Beasiswa ( $X_5$ ), Organisasi ( $X_6$ ), dan Kerja *Part Time* ( $X_7$ ).

### 3.2 Langkah-langkah Analisis

Langkah-langkah analisis yang digunakan pada penelitian ini adalah sebagai berikut :

1. Untuk mencapai tujuan pertama, dilakukan statistik deskriptif dengan bantuan *software* SPSS 16.0 dan Ms. Excel 2007 untuk mengetahui karakteristik lulusan mahasiswa FSM Undip Semarang.
2. Untuk mencapai tujuan kedua, dilakukan analisis MARS dengan bantuan *software* SPM 7.0 untuk pembentukan model MARS. Langkah-langkah analisisnya adalah sebagai berikut :
  - a. Pembentukan model MARS terbaik untuk data kelulusan mahasiswa FSM Undip Semarang dengan mengkombinasikan besarnya *Basis Function* (BF), *Maximum Interaction* (MI), dan *Minimum Observation* (MO). Dilakukan dengan cara *trial and error* dengan menentukan nilai maksimum BF, yaitu 2 sampai 4 kali banyaknya variabel prediktor yang digunakan, menentukan nilai MI yaitu 1,2, dan 3 dengan pertimbangan jika  $MI > 3$  akan menghasilkan model yang semakin kompleks, dan menentukan minimal banyaknya pengamatan setiap *knot* (MO) yaitu 0,1,2 dan 3.
  - b. Mendapatkan model MARS terbaik berdasarkan nilai GCV terkecil.
  - c. Mengklasifikasi ke dalam bentuk matriks konfusi (*confusion matrix*).
  - d. Mendapatkan nilai kesalahan klasifikasi dari nilai APER.
  - e. Menghitung nilai *Press's Q*.

## 4. PEMBAHASAN

### 4.1 Karakteristik Mahasiswa FSM Undip Semarang

Pada penelitian ini, statistik deskriptif bermanfaat untuk mengetahui karakteristik kelulusan mahasiswa FSM Undip Semarang berdasarkan lama studi  $\leq 48$  bulan atau  $> 48$  bulan. Data sampel jumlah mahasiswa FSM Undip Semarang sebanyak 200, terdiri dari 81 mahasiswa lulus  $\leq 48$  bulan dan 119 mahasiswa lulus  $> 48$  bulan, ditunjukkan pada Tabel 2.

**Tabel 2.** Ukuran Sampel dan Proporsi Kelompok Mahasiswa

| Lama Studi            | Mahasiswa | p (%) |
|-----------------------|-----------|-------|
| Lulus $\leq 48$ Bulan | 81        | 40.50 |
| Lulus $> 48$ Bulan    | 119       | 59.50 |

### 4.2 Klasifikasi Kelulusan Mahasiswa FSM Undip Semarang dengan Menggunakan Metode MARS

Untuk mengklasifikasi kelulusan mahasiswa FSM Undip Semarang, perlu dilakukan pembentukan model MARS terlebih dahulu. Pembentukan model dilakukan dengan cara *trial and error*, dengan mengkombinasikan beberapa kriteria model MARS yaitu fungsi basis (BF), maksimal interaksi antar variabel prediktor (MI) dan minimal pengamatan antar *knot* atau minimum observasi (MO) sampai diperoleh model MARS terbaik dengan nilai GCV minimum.

Pada penelitian ini, digunakan 7 variabel prediktor untuk pendugaan variabel respon sehingga penulis menentukan maksimum fungsi basis (BF) sebesar 14, 21 dan 28, sesuai dengan yang disarankan oleh Friedman (1991) yaitu dua sampai empat kali banyaknya variabel prediktor. Untuk interaksi maksimal (MI) sebesar 1, 2 dan 3 dengan pertimbangan apabila lebih dari 3, maka akan terbentuk model yang semakin kompleks sehingga akan sulit untuk diinterpretasikan. Sedangkan untuk observasi minimum (MO) sebesar 0, 1, 2, dan 3. Langkah selanjutnya adalah mengkombinasikan secara *trial and error* terhadap nilai BF, MI dan MO.

**Tabel 3.** Penentuan Model MARS Terbaik

| No. Model | BF | MI | MO | GCV     | Ketepatan Klasifikasi (%) | APER  |
|-----------|----|----|----|---------|---------------------------|-------|
| 1         | 14 | 1  | 0  | 0,18220 | 75,00                     | 25,00 |
| 2         | 14 | 1  | 1  | 0,18220 | 75,00                     | 25,00 |
| 3         | 14 | 1  | 2  | 0,18220 | 75,00                     | 25,00 |
| 4         | 14 | 1  | 3  | 0,18220 | 75,00                     | 25,00 |
| 5         | 14 | 2  | 0  | 0,18076 | 77,50                     | 22,50 |
| 6         | 14 | 2  | 1  | 0,18076 | 77,50                     | 22,50 |
| 7         | 14 | 2  | 2  | 0,18076 | 77,50                     | 22,50 |
| 8         | 14 | 2  | 3  | 0,18076 | 77,50                     | 22,50 |
| 9         | 14 | 3  | 0  | 0,18183 | 78,50                     | 21,50 |
| 10        | 14 | 3  | 1  | 0,18183 | 78,50                     | 21,50 |
| 11        | 14 | 3  | 2  | 0,18183 | 78,50                     | 21,50 |
| 12        | 14 | 3  | 3  | 0,18183 | 78,50                     | 21,50 |
| 13        | 21 | 1  | 0  | 0,18220 | 75,00                     | 25,00 |
| 14        | 21 | 1  | 1  | 0,18220 | 75,00                     | 25,00 |
| 15        | 21 | 1  | 2  | 0,18220 | 75,00                     | 25,00 |
| 16        | 21 | 1  | 3  | 0,18220 | 75,00                     | 25,00 |
| 17        | 21 | 2  | 0  | 0,17880 | 77,50                     | 22,50 |
| 18        | 21 | 2  | 1  | 0,17880 | 77,50                     | 22,50 |
| 19        | 21 | 2  | 2  | 0,17880 | 77,50                     | 22,50 |
| 20        | 21 | 2  | 3  | 0,17880 | 77,50                     | 22,50 |
| 21        | 21 | 3  | 0  | 0,18231 | 78,50                     | 21,50 |
| 22        | 21 | 3  | 1  | 0,18231 | 78,50                     | 21,50 |
| 23        | 21 | 3  | 2  | 0,18231 | 78,50                     | 21,50 |
| 24        | 21 | 3  | 3  | 0,18231 | 78,50                     | 21,50 |
| 25        | 28 | 1  | 0  | 0,18220 | 75,00                     | 25,00 |
| 26        | 28 | 1  | 1  | 0,18220 | 75,00                     | 25,00 |
| 27        | 28 | 1  | 2  | 0,18220 | 75,00                     | 25,00 |
| 28        | 28 | 1  | 3  | 0,18220 | 75,00                     | 25,00 |
| 29        | 28 | 2  | 0  | 0,17800 | 77,50                     | 22,50 |
| 30*       | 28 | 2  | 1  | 0,17781 | 79,50                     | 20,50 |
| 31        | 28 | 2  | 2  | 0,17800 | 77,50                     | 22,50 |
| 32        | 28 | 2  | 3  | 0,17800 | 77,50                     | 22,50 |
| 33        | 28 | 3  | 0  | 0,18268 | 78,50                     | 21,50 |
| 34        | 28 | 3  | 1  | 0,18331 | 76,00                     | 24,00 |
| 35        | 28 | 3  | 2  | 0,18331 | 76,00                     | 24,00 |
| 36        | 28 | 3  | 3  | 0,18331 | 76,00                     | 24,00 |

\*) model terbaik

Model pada Tabel 3 merupakan hasil kombinasi BF, MI dan MO secara *trial and error*. Model terbaik yang diperoleh yaitu dengan mengkombinasikan BF=28, MI=2 dan MO=1. Model tersebut memiliki nilai GCV minimum sebesar 0,17781. Didapatkan persamaan model sebagai berikut :

$$Y = 0,604579 - 0,476434 * BF_2 + 0,341618 * BF_8 + 2,1431 * BF_{14} - 1,69627 * BF_{16} + 0,855259 * BF_{24} \quad (14)$$

dengan,

Subsets for X3 :

SubSet1 = { "0", "3", "4" }

SubSet2 = { "2", "3", "4" }

SubSet3 = { "0", "3", "5" }

SubSet4 = { "1", "2", "3", "4" }

BF1 = max {0., X2 - 2,4};

BF2 = ( X3 is in SubSet1 ) \* BF1;

BF5 = ( X3 is in SubSet2 );

BF7 = ( X1 in ( "0" ) );

BF8 = ( X3 is in SubSet3 ) \* BF7;

BF14 = max {0., X2 - 3,46} \* BF5;

BF16 = max {0., X2 - 3,54};

BF17 = max {0., 3,54 - X2};

BF24 = ( X3 is in SubSet4 ) \* BF17;

Interpretasi dari koefisien-koefisien fungsi basis pada persamaan model MARS di atas adalah :

1. BF2 = ( X3 is in SubSet1 ) \* BF1

BF1 = max {0., X2 - 2,4}

Artinya, kontribusi BF2 bermakna jika mahasiswa berasal dari jurusan matematika atau fisika atau kimia (X3=0 atau X3=3 atau X3=4) dan IPK mahasiswa bernilai lebih dari 2,40 (X2-2,40). Tetapi jika kriteria tersebut tidak terdapat pada mahasiswa tersebut, maka koefisien BF2 tidak bermakna. Bermaknanya kontribusi BF2 mengurangi nilai Y sebesar 0,476434.

2. BF8 = ( X3 is in SubSet3 ) \* BF7

BF7 = ( X1 in ( "0" ) )

Artinya, kontribusi BF8 bermakna jika mahasiswa berasal dari jurusan matematika atau fisika atau biologi (X3=0 atau X3=3 atau X3=5) dan jenis kelamin mahasiswa adalah laki-laki (X1=0). Tetapi jika kriteria tersebut tidak terdapat pada mahasiswa tersebut, maka koefisien BF8 tidak bermakna. Bermaknanya kontribusi BF8 menambah nilai Y sebesar 0,341618.

3. BF14 = max {0., X2 - 3,46} \* BF5

BF5 = ( X3 is in SubSet2 )

Artinya, kontribusi BF14 bermakna jika IPK mahasiswa bernilai lebih dari 3,46 (X2-3,46) dan mahasiswa tersebut berasal dari jurusan teknik informatika atau fisika atau kimia (X3=2 atau X3=3 atau X3=4). Tetapi jika kriteria tersebut tidak terdapat pada mahasiswa tersebut, maka koefisien BF14 tidak bermakna. Bermaknanya kontribusi BF14 menambah nilai Y sebesar 2,1431.

4. BF16 = max {0., X2 - 3,54}

Artinya, kontribusi BF16 bermakna jika IPK mahasiswa bernilai lebih dari 3,54 (X2-3,54). Tetapi jika IPK mahasiswa bernilai kurang dari sama dengan 3,54 maka koefisien BF16 tidak bermakna sehingga nilainya adalah 0. Bermaknanya kontribusi BF16 mengurangi nilai Y sebesar 1,69627.

5. BF24 = ( X3 is in SubSet4 ) \* BF17

BF17 = max {0., 3,54 - X2}

Artinya, kontribusi BF24 bermakna jika mahasiswa berasal dari jurusan statistika atau teknik informatika atau fisika atau kimia (X3=1 atau X3=2 atau



X3=3 atau X3=4) dan IPK mahasiswa bernilai kurang dari 3,54 (3,54-X2). Tetapi jika kriteria tersebut tidak terdapat pada mahasiswa tersebut, maka koefisien BF24 tidak bermakna. Bermaknanya kontribusi BF24 menambah nilai Y sebesar 0,855259

Model MARS terbaik dipengaruhi oleh tiga variabel prediktor yaitu variabel IPK (X2), jurusan (X3) dan jenis kelamin (X1) yang secara signifikan memberikan kontribusi pada model. Sedangkan variabel organisasi (X6), *part time* (X7), jalur masuk (X4), dan beasiswa (X5) tidak memberikan kontribusi pada model.

**Tabel 4.** Matriks Konfusi untuk Klasifikasi Kelulusan Mahasiswa

| Hasil Observasi<br>(Actual Class) | Taksiran (Predicted Class) |                | Total Aktual |
|-----------------------------------|----------------------------|----------------|--------------|
|                                   | y <sub>0</sub>             | y <sub>1</sub> |              |
| y <sub>0</sub>                    | 58                         | 23             | 81           |
| y <sub>1</sub>                    | 18                         | 101            | 119          |
| Total Prediksi                    | 76                         | 124            | 200          |

$$APER = \frac{23 + 18}{58 + 23 + 18 + 101} \times 100\% = 20,50\%$$

$$ketepatan\ klasifikasi = 100\% - 20,50\% = 79,50\%$$

Sehingga dapat disimpulkan bahwa nilai APER dari klasifikasi kelulusan mahasiswa FSM Undip Semarang dengan menggunakan metode MARS sebesar 20,50% dan nilai ketepatan klasifikasinya sebesar 79,50%. Selanjutnya dilakukan uji *Press's Q* untuk menguji kestabilan pengklasifikasian model.

$$Press's\ Q = \frac{[200 - (159 * 2)]^2}{200(2 - 1)} = 69,62$$

Nilai uji *Press's Q* yang didapatkan sebesar 69,62 kemudian dibandingkan dengan nilai chi-square dengan derajat bebas 1 ( $\chi^2_{(1;0,05)} = 3,841$ ). Nilai *Press's Q* yang didapatkan lebih besar dibandingkan dengan  $\chi^2_{(1;0,05)}$ , sehingga dapat dikatakan keakuratan klasifikasi kelulusan mahasiswa menggunakan model MARS sudah konsisten secara statistik.

## 5. KESIMPULAN

Kesimpulan yang dapat diambil dari hasil dan pembahasan adalah sebagai berikut :

1. Data kelulusan mahasiswa menunjukkan bahwa dari 200 sampel mahasiswa FSM Undip Semarang terdapat 40,50% mahasiswa lulus dengan lama studi  $\leq$  48 bulan dan 59,50% mahasiswa lulus dengan lama studi  $>$  48 bulan.
2. Model MARS yang digunakan untuk klasifikasi kelulusan mahasiswa FSM Undip Semarang adalah model MARS dengan kombinasi BF=28; MI=2; MO=1 karena memiliki nilai GCV paling kecil yaitu sebesar 0,17781. Terdapat tiga variabel yang berpengaruh terhadap model yaitu variabel IPK, jurusan dan jenis kelamin. Sedangkan variabel organisasi, *part time*, jalur masuk, dan beasiswa tidak memberikan kontribusi pada model. Didapatkan ketepatan klasifikasi sebesar 79,50% dan nilai APER sebesar 20,50%. Nilai uji *Press's Q* menunjukkan bahwa secara statistik metode MARS sudah konsisten dalam mengklasifikasi data kelulusan mahasiswa FSM Undip Semarang.

## DAFTAR PUSTAKA

- Cox, D.R. and Snell, E.J. 1989. *Analysis of Binary Data*. Second Edition. Chapman and Hall, London.
- Djamarah, S.B. 2002. *Psikologi Belajar*. Jakarta: PT. Rineka Cipta.
- Friedman, J.H. 1991. *Multivariate Adaptive Regression Splines*. The Annals of Statistics, Vol. 19 No. 1.
- Hair, J. F. JR., Anderson, R. E., Tatham, R. L. and Black, W. C., (2006). *Multivariate Data Analysis. Sixth Edition*, Pearson Education Prentice Hall, Inc.
- Johnson, R.A. and Wichern, D.W. 1992. *Applied Multivariate Statistical Analysis*. Prentice Hall, New Jersey.
- Kementrian Pendidikan Nasional RI. 2000. *Pedoman Penyusunan Kurikulum Pendidikan Tinggi dan Penilaian Hasil Belajar Mahasiswa*. <http://www.dikti.go.id/files/Lemkerma/kepmen232-2000.txt> [diakses 23 April 2014].
- Pintowati, W. dan Otok, B.W. 2012. *Pemodelan Kemiskinan di Propinsi Jawa Timur dengan Pendekatan Multivariate Adaptive*. JURNAL SAINS DAN SENI ITS Vol.1, No.1.
- Prasetyo, E. 2012. *Data Mining Konsep dan Aplikasi menggunakan MATLAB*. Yogyakarta: Penerbit Andi Yogyakarta.
- Syah, M. 2009. *Psikologi Belajar Edisi Revisi 9*. Jakarta: Rajawali Pers.
- Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional*. <http://books.google.com/books?isbn=9791043639> [diakses 29 April 2014].