

PENERAPAN METODE *MULTIVARIATE ADAPTIVE REGRESSION SPLINE* (MARS) UNTUK MENENTUKAN FAKTOR YANG MEMPENGARUHI MASA STUDI MAHASISWA FPMIPA UPI

Mardiah Annur, Jarnawi Afgani Dahlan, Fitriani Agustina

Departemen Pendidikan Matematika FPMIPA UPI

ABSTRAK Keterlambatan lulus dari program S1 menjadi salah satu permasalahan yang terjadi pada mahasiswa di sebagian besar perguruan tinggi, salah satunya terjadi di FPMIPA UPI. Salah satu upaya yang dapat dilakukan sebagai usaha meminimalisir banyaknya mahasiswa yang lulus terlambat adalah dengan cara menganalisis faktor yang mempengaruhi lama masa studi mahasiswa FPMIPA UPI.

Multivariate Adaptive Regression Spline (MARS), suatu kombinasi antara *Recursive Partitioning Regression* (RPR) dan metode *Spline*, mampu mengolah data berdimensi tinggi dan berukuran besar, serta mampu mengolah data dengan variabel respon berbentuk kontinu ataupun biner, sesuai dengan permasalahan klasifikasi alumni FPMIPA UPI. MARS akan membangun suatu model terbaik sebagai model klasifikasi yang melibatkan beberapa fungsi basis yang memuat variabel prediktor yang berpengaruh.

Berdasarkan analisis klasifikasi yang dilakukan diperoleh kesimpulan bahwa variabel prediktor yang paling berpengaruh terhadap masa studi mahasiswa FPMIPA UPI adalah jalur masuk. Sementara variabel prediktor yang berpengaruh lainnya yaitu IP semester pertama, jenis kelamin, dan program studi.

Kata kunci : Masa Studi, MARS

PENDAHULUAN

Latar Belakang

Perguruan Tinggi, sebagai salah satu jenjang pendidikan di Indonesia, bertujuan melahirkan para sarjana yang handal dan memiliki keterampilan di bidangnya. Menurut Guillory (2008), salah satu ukuran kesuksesan pembelajaran pada tingkat pendidikan tinggi adalah tingkat ketepatan masa studi. Waktu standar yang dibutuhkan untuk menyelesaikan studi program S1 adalah selama 4 tahun atau setara dengan 8 semester. Akan tetapi, pada kenyataannya banyak mahasiswa yang mengalami kesulitan untuk menuntaskan studinya sesuai dengan standar waktu studi yang telah ditentukan. Hal ini didukung oleh data statistik yang menunjukkan bahwa, salah satunya di FPMIPA UPI, rata-rata lama studi mahasiswa program S1 untuk waktu lulus Januari 2008 sampai dengan Agustus 2013 adalah sekitar 9,4 semester yang setara dengan 4,7 tahun.

Untuk mengantisipasi masalah lama masa studi mahasiswa tersebut, dapat dilakukan dengan analisis untuk mengetahui faktor-faktor yang mempengaruhi lama masa studi mahasiswa program S1 FPMIPA UPI. Penelitian sebelumnya yang dilakukan Permana (2011) menyatakan bahwa faktor-faktor yang diduga mempengaruhi lama masa studi mahasiswa adalah jenis kelamin, Indeks Prestasi (IP) semester satu, daerah asal, program studi, dan jalur masuk perguruan tinggi.

Permasalahan klasifikasi alumni FPMIPA UPI berdasarkan masa studinya dipengaruhi oleh banyak variabel prediktor, melibatkan variabel respon bernilai biner, dan memiliki data sampel berukuran besar. Karakteristik permasalahan ini sesuai dengan kemampuan dari suatu metode nonparametrik, yaitu *Multivariate Adaptive Regression Spline* (MARS). Metode MARS dapat mengatasi permasalahan data berdimensi tinggi, yaitu data yang memiliki banyak variabel prediktor sebesar $3 \leq n \leq 20$ dan data sampel berukuran besar yang memerlukan perhitungan rumit serta data dengan beberapa nilai input yang hilang (*missing input value*). Variabel respon yang diolah pada metode MARS pun dapat berbentuk kontinu ataupun biner (Kriener, 2007).

Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan sebelumnya, masalah yang dirumuskan dalam penelitian ini adalah:

1. Bagaimana hasil penerapan metode MARS pada kasus klasifikasi alumni FPMIPA UPI berdasarkan masa studi?
2. Bagaimana hasil interpretasi model klasifikasi MARS untuk permasalahan masa studi alumni FPMIPA UPI?

Batasan Masalah

Pada penelitian ini, analisis klasifikasi menggunakan data alumni FPMIPA UPI yang lulus pada Juni 2011 sampai Juni 2013 dan program komputer yang digunakan adalah *Salford Predictive Modeler* versi 7.0.

Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan penulisan skripsi ini adalah sebagai berikut:

1. Mengetahui hasil penerapan metode MARS pada kasus klasifikasi alumni FPMIPA UPI berdasarkan masa studi;
2. Mengetahui hasil interpretasi model klasifikasi MARS untuk permasalahan masa studi alumni FPMIPA UPI.

Manfaat Penulisan

Manfaat teoritis yang diharapkan dari penelitian ini adalah diharapkan dapat menambah wawasan mengenai metode klasifikasi MARS, khususnya dalam menentukan faktor-faktor yang mempengaruhi masa studi mahasiswa. Sementara manfaat praktis dari penelitian ini adalah dapat memahami penerapan metode MARS dalam bidang pendidikan, khususnya pengklasifikasian alumni berdasarkan masa studi serta dapat menjadi informasi strategis bagi program studi atau fakultas sehingga dapat meningkatkan upaya untuk mendorong dan mempercepat kelulusan mahasiswa.

PEMBAHASAN

Recursive Partitioning Regression Spline

Recursive partitioning regression (RPR) merupakan suatu metode nonparametrik yang dapat dipandang sebagai suatu prosedur regresi *stepwise* untuk mengaproksimasi fungsi yang tidak diketahui $f(\mathbf{x})$ yang berbentuk:

$$\hat{f}(\mathbf{x}) = \sum_{m=1}^M a_m B_m(\mathbf{x})$$

dengan a_m menyatakan koefisien fungsi basis ke- m , dan $B_m(\mathbf{x})$ menyatakan fungsi basis ke- m .

Fungsi basis merupakan kumpulan fungsi parametrik yang terdiri dari satu atau lebih variabel. Sedangkan *knot* adalah nilai variabel prediktor ketika *slope* suatu garis regresi mengalami perubahan (Febriyanti, dkk, 2013). Adapun fungsi basis B_m pada RPR didefinisikan sebagai berikut:

$$B_m = \prod_{k=1}^{K_m} H[s_k \cdot (X_{v(k,m)} - t_k)]$$

dimana H menyatakan *step function*, yang didefinisikan sebagai berikut:

$$H[\eta] = \begin{cases} 1, & \text{jika } \eta \geq 0 \\ 0, & \text{lainnya} \end{cases};$$

dengan K_m menyatakan derajat interaksi, s_{km} menyatakan tanda pada titik knot (bernilai +1 atau -1), $X_{v(k,m)}$ menyatakan variabel prediktor, dan t_{km} menyatakan nilai *knot* dari variabel prediktor $X_{v(k,m)}$.

Akan tetapi, RPR memiliki kelemahan yang mengurangi keefektifannya, diantaranya adalah model yang dihasilkan oleh RPR tidak kontinu pada batas regionnya (*lack of continuity*) dan juga model RPR tidak mampu mengidentifikasi fungsi $f(x)$ berbentuk linear atau aditif dan RPR cenderung sulit diinterpretasi apabila variabel prediktor terlalu banyak (Friedman, 1991).

Regresi Spline

Spline merupakan suatu pendekatan yang digunakan untuk mengatasi permasalahan yang ditimbulkan oleh data yang berpola naik atau turun secara tajam. Spline berupa fungsi *piecewise* (potongan) polinomial yang memiliki sifat tersegmen dan kontinu, dimana fungsi spline univariat berderajat q didefinisikan sebagai berikut:

$$s(x) = \sum_{i=0}^q a_i x^i + \sum_{j=1}^K b_j (x - t_j)_+^q,$$

dengan K knot, yaitu t_1, t_2, \dots, t_K . Adapun, *truncated power basis* $(x - t_k)_+^q$ didefinisikan sebagai:

$$(x - t_k)_+^q = \begin{cases} (x - t_k)^q; & x - t_k > 0 \\ 0 & ; x - t_k \leq 0. \end{cases}$$

Sehingga bentuk umum fungsi spline berderajat q dengan p variabel prediktor adalah:

$$\begin{aligned} s(\mathbf{x}) &= a_0 + \sum_{i=1}^q a_{1i} x_1^i + \sum_{j=1}^K b_{1j} (x_1 - t_j)_+^q + \sum_{i=1}^q a_{2i} x_2^i + \sum_{j=1}^K b_{2j} (x_2 - t_j)_+^q \\ &\quad + \dots + \sum_{i=1}^q a_{pi} x_p^i + \sum_{j=1}^K b_{pj} (x_p - t_j)_+^q \\ s(\mathbf{x}) &= a_0 + \sum_{l=1}^p \left[\sum_{j=1}^q a_l x_{li}^j + \sum_{k=1}^K b_{lk} (x_{li} - t_{lk})_+^q \right], \end{aligned}$$

dengan

$$(x_{li} - t_{li})_+^q = \begin{cases} (x_{li} - t_{li})^q; & x_{li} - t_{li} > 0 \\ 0 & ; x_{li} - t_{li} \leq 0. \end{cases}$$

Regresi spline merupakan salah satu metode regresi nonparametrik dimana bentuk kurva regresinya berupa fungsi spline. Secara umum, model regresi *spline* univariat dapat dituliskan sebagai berikut (Oktaviana & Budiantara, 2011):

$$y_i = a_0 + \sum_{j=1}^q a_j x^j + \sum_{k=1}^K b_k (x - t_k)_+^q + \varepsilon_i$$

dimana $q \geq 1$ serta a_j dan b_k bernilai real, untuk $i = 1, 2, \dots, n, j = 1, \dots, q$, dan $k = 1, 2, \dots, K$, serta ε adalah error random independen dengan mean nol dan varians σ^2 . Sedangkan untuk model regresi spline secara umum dengan p variabel prediktor adalah:

$$\begin{aligned}
y_i &= a_{0i} + \sum_{j=1}^q a_{1j} x_1^j + \sum_{k=1}^K b_{1k} (x_1 - t_k)_+^q + \sum_{j=1}^q a_{2j} x_2^j + \sum_{k=1}^K b_{2k} (x_2 - t_k)_+^q \\
&\quad + \dots + \sum_{j=1}^q a_{pj} x_p^j + \sum_{k=1}^K b_{pk} (x_p - t_k)_+^q + \varepsilon_i \\
y_i &= a_{0i} + \sum_{l=1}^p \left[\sum_{j=1}^q a_{lj} x_{li}^j + \sum_{k=1}^K b_{lk} (x_{li} - t_k)_+^q \right] + \varepsilon_i.
\end{aligned}$$

Multivariate Adaptive Regression Spline (MARS)

Multivariate Adaptive Regression Spline (MARS), yang diperkenalkan pertama kali oleh Friedman (1991), merupakan suatu pendekatan regresi nonparametrik yang dihasilkan dari kombinasi kompleks antara RPR dan pendekatan spline. Model MARS ini secara umum didefinisikan sebagai berikut:

$$\hat{f}(\mathbf{x}) = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} [s_k \cdot (X_{v(k,m)} - t_k)]_+ \quad (1)$$

dengan a_0 menyatakan konstanta, a_m koefisien fungsi basis ke- m , M menyatakan maksimum banyaknya fungsi basis *nonconstant*, s_k menyatakan tanda pada titik *knot* (bernilai +1 atau -1).

Selanjutnya, berdasarkan fungsi regresi nonparametrik, model MARS dinyatakan dalam persamaan berikut:

$$Y_i = a_{0i} + \sum_{m=1}^M a_{mi} \prod_{k=1}^{K_m} [s_k \cdot (X_{v(k,m)} - t_k)]_+ + \varepsilon_i \quad (2)$$

dengan Y_i adalah variabel respon dan ε merupakan *error* acak yang bersifat independen, memiliki *mean* sama dengan nol, dan varians σ^2 .

Model MARS untuk variabel respon bernilai biner diadaptasi dengan pendekatan regresi logistik dan didefinisikan sebagai berikut (Friedman, 1991):

$$\log \left[\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right] = \hat{f}(\mathbf{x}), \quad (3)$$

dimana π menyatakan probabilitas bagi variabel respon Y bernilai terbesar ($Y = 1$) dan $\hat{f}(\mathbf{x})$ menyatakan model MARS dalam persamaan (3) Dengan demikian, model MARS untuk variabel respon biner adalah:

$$\log \left[\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right] = a_{0i} + \sum_{m=1}^M a_{mi} \prod_{k=1}^{K_m} [s_k \cdot (X_{v(k,m)} - t_k)]_+. \quad (4)$$

Algoritma MARS

1. *Forward Stepwise*

Forward stepwise bertujuan untuk memperoleh fungsi dengan jumlah fungsi basis maksimum. Langkah-langkah *forward stepwise* pada metode MARS adalah sebagai berikut (Zhang & Singer, 2010):

1. Memisalkan $B_0 = 1$, yang merupakan fungsi basis konstan, sebagai fungsi basis awal;
2. Menentukan pasangan fungsi basis $B_1 = (x_i - t)_+$ dan $B_2 = (t - x_i)_+$ sebagai kombinasi variabel prediktor x_i dan knot t_1 untuk selanjutnya ditambahkan pada model. Pada langkah 2 ini akan menghasilkan suatu kandidat model MARS yang memiliki nilai *average sum of square residual* (ASR) minimum, dimana ASR didefinisikan sebagai berikut:

$$A = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{f}_M(x_i))^2. \quad (5)$$

dengan N menyatakan ukuran sampel dan $\hat{f}_M(x_i)$ menyatakan nilai taksiran variabel respon pada M fungsi basis di x_i .

3. Langkah selanjutnya adalah memperluas model MARS dengan cara menambahkan perkalian fungsi basis yang dimiliki B_m dengan masing-masing fungsi basis baru kedalam model yang sudah ada sehingga akan menghasilkan beberapa kemungkinan kandidat model. Oleh karena itu, akan dipilih pasangan hasil kali yang menghasilkan model dengan nilai ASR terkecil;
4. Mengulangi langkah (3) hingga banyaknya fungsi basis dalam model lebih atau sama dengan maksimum banyaknya fungsi basis yang telah ditetapkan.

2. *Backward Stepwise* Algoritma MARS

Backward stepwise bertujuan untuk memperoleh model yang sesederhana mungkin (prinsip parsemoni). Proses ini dimulai pada model yang diperoleh pada *forward stepwise* yang memiliki M fungsi basis dengan langkah-langkah sebagai berikut (Zhang & Singer, 2010):

1. Menghapus salah satu fungsi basis *nonconstant* yang memiliki kontribusi terkecil, yaitu fungsi basis yang jika dihilangkan dari model sebelumnya akan menyebabkan terjadinya kenaikan ASR terkecil;
2. Mengulangi langkah 1, sampai model hanya mengandung fungsi basis konstan.

Selanjutnya, berdasarkan sekuens model yang ada, proses selanjutnya yaitu memilih model MARS terbaik dengan kriteria *Generalized Cross Validation* (GCV).

Penaksiran Parameter Model MARS

Model MARS dengan variabel respon biner yang didefinisikan pada persamaan (4) dapat dinyatakan kembali sebagai berikut:

$$\hat{f}(\mathbf{x}) = a_0 + \sum_{m=1}^M a_m B_m \quad (6)$$

dimana

$$B_m = \prod_{k=1}^{K_m} [S_k \cdot (x_{v(k,m)} - t_k)]_+,$$

dan B_m merupakan fungsi basis ke- m yang bergantung pada x_i .

Parameter model dalam persamaan (6) adalah koefisien-koefisien fungsi basis a_1, a_2, \dots, a_M yang akan diestimasi secara numerik dengan cara memaksimalkan likelihood dari data menggunakan *maximum likelihood estimator* (MLE) (Friedman, 1991).

Misalkan terdapat n observasi yang saling bebas dan Y_i merupakan variabel respon biner untuk observasi ke- i , $i = 1, 2, \dots, n$, maka fungsi densitas bersyarat dari Y_i diberikan nilai $x_{1i}, x_{2i}, \dots, x_{pi}$, yaitu :

$$f(Y_i | \mathbf{x}_i) = [\pi(\mathbf{x}_i)]^{y_i} [1 - \pi(\mathbf{x}_i)]^{1-y_i} = \left(\frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)} \right)^{y_i} (1 - \pi(\mathbf{x}_i)) \quad (7)$$

dengan diketahui bahwa

$$\pi(\mathbf{x}_i) = E(Y_i | \mathbf{x}_i) = \frac{\exp(a_0 + \sum_{m=1}^M a_m B_m)}{1 + \exp(a_0 + \sum_{m=1}^M a_m B_m)} \quad (8)$$

dan

$$= \frac{1}{1 + \exp(a_0 + \sum_{m=1}^M a_m B_m)} \quad (9)$$

maka diperoleh fungsi *likelihood* $L(\mathbf{a})$ sebagai berikut:

$$L(\mathbf{a}) = \prod_{i=1}^n \left(\exp \left(a_0 + \sum_{m=1}^M a_m B_m \right) \right)^{y_i} \left(\frac{1}{1 + \exp(a_0 + \sum_{m=1}^M a_m B_m)} \right) \quad (10)$$

Selanjutnya, akan ditentukan bentuk logaritma natural dari fungsi likelihood pada persamaan (10) yang dikenal sebagai fungsi *log-likelihood*, yaitu:

$$\ln L(\vec{a}) = \sum_{i=1}^N \left\{ y_i \left(a_{i0} + \sum_{m=1}^M a_m B_m \right) - \ln \left[1 + \exp \left(a_{i0} + \sum_{m=1}^M a_m B_m \right) \right] \right\}.$$

Estimator maksimum likelihood dari parameter a_{i0}, a_1, \dots, a_M diperoleh dari turunan parsial pertama fungsi log-likelihood terhadap parameter a_{i0}, a_1, \dots, a_M dan menyamakannya dengan nol yang disebut sebagai *persamaan likelihood*. Berdasarkan penurunan terhadap masing-masing parameter, diperoleh sebanyak $M + 1$ persamaan likelihood yang tidak linear dalam a_{i0}, a_1, \dots, a_M . Oleh karena itu, untuk menentukan atau memperoleh $\vec{a}_{i0}, \vec{a}_1, \dots, \vec{a}_M$ dipergunakan metode numerik. Metode numerik yang digunakan adalah metode Newton-Raphson.

Pemilihan Model MARS Terbaik

Model MARS terbaik adalah model yang memiliki nilai *Generalized Cross Validation* (GCV) minimum. Adapun GCV dirumuskan sebagai berikut (Friedman, 1991):

$$G(M) = \frac{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{f}_M(x_i))^2}{\left(1 - \frac{\bar{C}(M)}{N}\right)^2} \quad (11)$$

dimana $\bar{C}(M)$ dirumuskan:

$$\bar{C}(M) = C(M) + d \cdot M \quad (12)$$

dengan $C(M) = \text{tr} \left(\mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \right) + 1$, dan \mathbf{B} menyatakan matriks M fungsi basis *nonconstan* berukuran $M \times N$, serta d menyatakan nilai ketika setiap fungsi basis optimal, dimana $2 \leq d \leq 4$ adalah rentang nilai terbaik bagi d .

Pengujian Signifikansi Model MARS

Pengujian signifikansi model MARS dilakukan untuk mengevaluasi kecocokan model (pengujian secara simultan) dan juga menguji signifikansi parameter (pengujian secara parsial) (Wicaksono, Wilandari, dan Suparti, 2014).

Adapun prosedur pengujian signifikansi model MARS secara simultan adalah sebagai berikut :

1. Menentukan hipotesis pengujian signifikansi model MARS secara simultan, yaitu:
 - Ho: $a_1 = a_2 = \dots = a_M = 0$ (model tidak signifikan);
 - H1: paling sedikit terdapat satu $a_j \neq 0, j = 1, 2, \dots, M$ (model signifikan);
2. Menentukan statistik uji F_{h1} yang digunakan dirumuskan sebagai berikut :

$$F_h = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 / M}{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n - M - 1)}$$

dimana n menyatakan ukuran sampel dan $F_{\alpha}(v_1, v_2)$ diperoleh dari tabel F dengan taraf signifikansi α serta $v_1 = M$ dan $v_2 = n - M - 1$;

3. Dengan mengambil taraf signifikansi α , kriteria pengujian model MARS adalah:

- Tolak H_0 jika $F_h > F_{\alpha}(v_1, v_2)$ atau jika $p\text{-value (Sig.)} < \alpha$;
- Terima H_0 pada keadaan sebaliknya.

Sedangkan prosedur pengujian signifikansi secara parsial adalah sebagai berikut :

1. Menentukan hipotesis pengujian signifikansi secara parsial, yaitu:

$H_0: a_j = 0$ (koefisien a_j tidak berpengaruh terhadap model);

$H_1: a_j \neq 0$, untuk setiap j , $j = 1, 2, \dots, M$ (koefisien a_j berpengaruh terhadap model);

2. Menentukan statistik uji t_h berdasarkan rumusan berikut (Neter, dkk, 1983):

$$t_h = \frac{a_j}{s_{a_j}}$$

dengan s_{a_j} adalah standar *error* dari a_j yang dirumuskan sebagai berikut :

$$s_{a_j} = \sqrt{\left(\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - M - 1} \right) C_j}$$

dimana C_j adalah elemen-elemen pada diagonal utama matriks $(\mathbf{B}^T \mathbf{B})^{-1}$ (Agwil, dkk, 2012). Sedangkan, nilai $t_{(\frac{\alpha}{2}, v_1)}$ diperoleh dari tabel t dengan taraf signifikansi α dan derajat kebebasan $v_1 = M$;

3. Dengan mengambil taraf signifikansi α , kriteria uji signifikansi parsial adalah:

- Tolak H_0 jika $t_h > t_{(\frac{\alpha}{2}, v_1)}$ atau jika $p\text{-value (Sig.)} < \alpha$;
- Terima H_0 pada keadaan sebaliknya.

Klasifikasi pada MARS

Model klasifikasi metode MARS di atas merupakan model dengan variabel respon biner Y (0 atau 1) yang mewakili dua kelas dan p variabel prediktor $\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{ip})$, yang didefinisikan dalam persamaan (4).

Analisis klasifikasi metode MARS melakukan prediksi nilai Y yaitu $\hat{y} = \hat{f}(x)$ dengan menggunakan model MARS sebagai model klasifikasi. Penentuan klasifikasi dilakukan dengan menggunakan titik potong (*cut off*) sebesar 0,5 dengan

ketentuan, apabila estimasi peluang melebihi 0,5; maka hasil prediksi adalah masuk ke kelas 1 ($Y = 1$), dan apabila estimasi peluang kurang dari atau sama dengan 0,5; maka hasil prediksi termasuk ke kelas 0 ($Y = 0$) (Meluarini, Safitri, dan Hoyyi, 2014).

STUDI KASUS

Sumber Data

Data penelitian meliputi data lama masa studi beserta lima faktor yang mempengaruhinya, yaitu jenis kelamin, program studi, jalur masuk, Indeks Prestasi (IP) semester pertama, dan daerah asal untuk alumni FPMIPA UPI yang lulus pada rentang waktu Juni 2011 sampai Juni 2013 yang diperoleh dari Biro Administrasi Akademik dan Kemahasiswaan (BAAK) UPI.

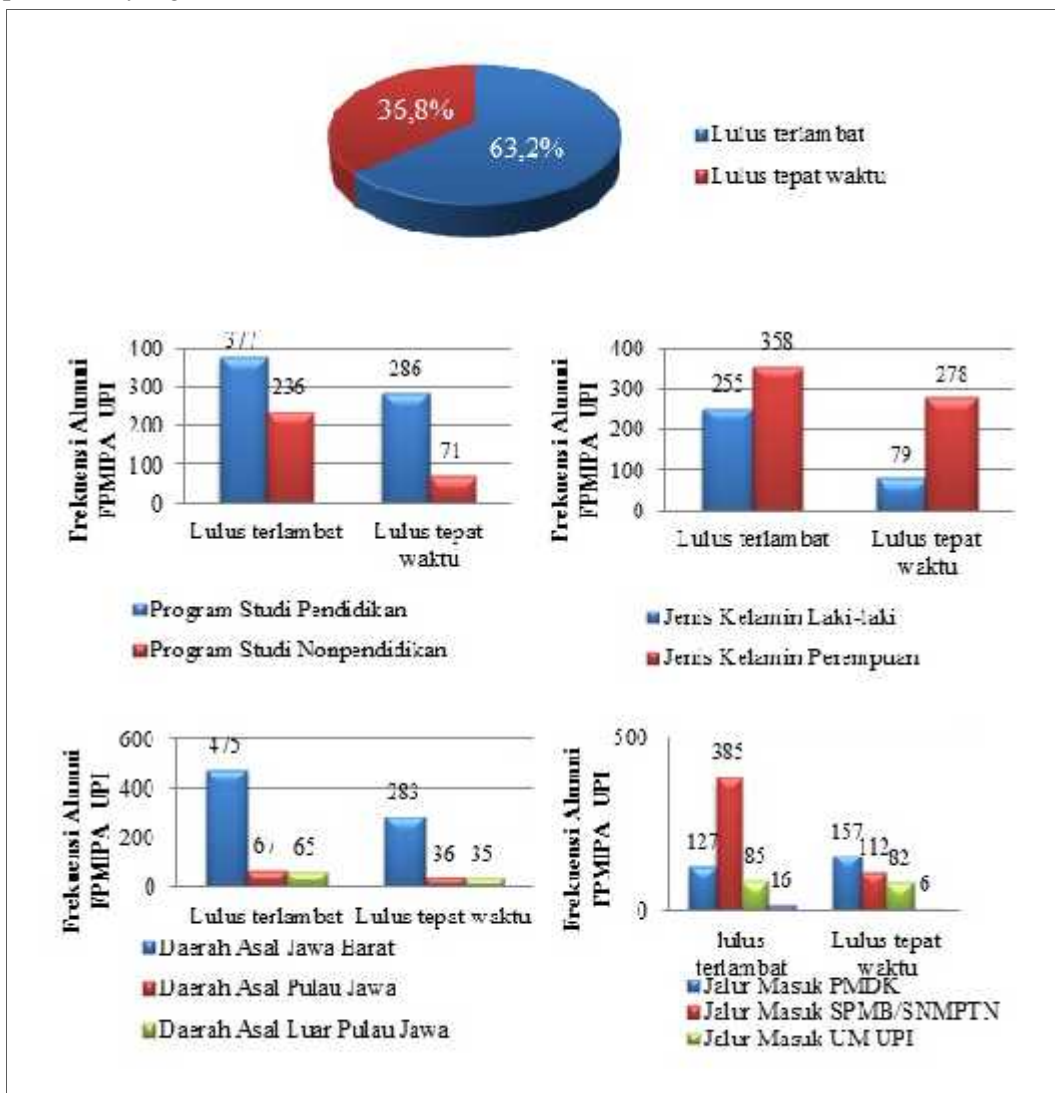
Variabel Penelitian

Tabel 1. Variabel Penelitian

Variabel Respon			
Kode	Variabel	Definisi Operasional	Skala Pengukuran
Y	Lama Masa Studi	<ul style="list-style-type: none"> • $Y = 1$: Alumni yang lulus 8 semester (lulus tepat waktu) • $Y = 0$: Alumni yang lulus > 8 semester (lulus terlambat) 	Nominal
Variabel Prediktor			
G	Jenis Kelamin	<ul style="list-style-type: none"> • $G = 1$: Alumni berjenis kelamin laki-laki • $G = 2$: Alumni berjenis kelamin perempuan 	Nominal
P	Program Studi	<ul style="list-style-type: none"> • $P = 1$: Alumni berasal dari program studi kependidikan • $P = 2$: Alumni berasal dari program studi nonkependidikan 	Nominal
JM	Jalur Masuk	<ul style="list-style-type: none"> • $J = 1$: Alumni diterima melalui jalur reguler PMDK • $J = 2$: Alumni diterima melalui jalur reguler SBMPTN • $J = 3$: Alumni diterima melalui jalur reguler UM UPI • $J = 4$: Alumni diterima melalui jalur Nonreguler 	Nominal
IP	IP Semester Pertama	Indeks Prestasi semester pertama adalah IP yang diperoleh mahasiswa setelah menyelesaikan semester pertama	Interval
D	Daerah Asal	<ul style="list-style-type: none"> • $D = 1$: Alumni berasal dari provinsi Jawa Barat • $D = 2$: Alumni berasal dari Pulau Jawa selain Jawa Barat • $D = 3$: Alumni berasal dari luar Pulau Jawa 	Nominal

Analisis Deskriptif

Berikut ini adalah *pie chart* data masa studi dan diagram batang variabel prediktor jenis kelamin, program studi, daerah asal, dan jalur masuk serta tabel 2 sebagai tabulasi data variabel prediktor IP semester pertama untuk data sampel penelitian yang berukuran 970.



Gambar 1. *Pie Chart* dan Diagram Batang Data Masa Studi Alumni FPMIPA UPI

Berdasarkan gambar 1 diperoleh informasi bahwa alumni FPMIPA UPI yang lulus pada rentang waktu Juni 2011 sampai dengan Juni 2013 mayoritas

menempuh program S1 lebih dari waktu standar 8 semester. Selain itu, sebagian besar alumni FPMIPA UPI merupakan lulusan dari program studi pendidikan. Hal ini dikarenakan, pada FPMIPA UPI, kuota yang disediakan bagi program studi kependidikan lebih banyak daripada program studi nonkependidikan. Untuk jenis kelamin, mayoritas alumni FPMIPA UPI berjenis kelamin perempuan. Sementara untuk jalur masuk, sebagian alumni FPMIPA UPI diterima melalui jalur reguler SBMPTN. Sementara untuk alumni yang lulus tepat waktu, sebagian besar diterima melalui jalur PMDK atau yang kini dikenal dengan nama SNMPTN, yaitu sebanyak 157 lulusan dan untuk alumni yang lulus terlambat sebagian besar diterima melalui jalur ujian tertulis yaitu SBMPTN, yaitu sebanyak 385 lulusan. Alumni FPMIPA UPI sebagian besar berasal dari kota atau kabupaten di provinsi Jawa Barat, yaitu sebanyak 758 lulusan. Sementara untuk daerah di Pulau Jawa selain Provinsi Jawa Barat dan daerah di luar Pulau Jawa berturut-turut ada sebanyak 103 dan 100 lulusan.

Tabel 2. Statistik IP Semester Pertama Alumni FPMIPA UPI

Statistik		IP Semester Pertama
N	Valid	970
	Missing	0
Mean		2,70
Median		2,75
Mode		2,80
Variance		0,33
Minimum		0,00
Maximum		4,00

Berdasarkan Tabel 2 diperoleh informasi bahwa keseluruhan data sampel valid tanpa data yang hilang (*missing value*) mempunyai rata-rata sebesar 2,70 dengan 50% alumni FPMIPA UPI mempunyai IP semester pertama di atas 2,75 dan 50% sisanya mempunyai IP semester pertama di bawah 2,75. Mayoritas alumni FPMIPA UPI mempunyai IP semester pertama sebesar 2,80. Varians data IP semester pertama yang menunjukkan sebaran data, adalah sebesar 0,33 dengan ketentuan nilai skala IP dalam rentang 0,00 – 4,00.

Analisis Uji Independensi

Analisis uji independensi bertujuan untuk mengetahui ada atau tidaknya hubungan antara variabel respon lama masa studi dengan setiap variabel prediktor yang mempunyai skala pengukuran nominal, yaitu variabel gender, program studi, jalur masuk, dan daerah asal. Teknik uji independensi yang digunakan adalah uji independensi nonparametrik *Chi-Square*. Berikut ini tahap-tahap analisis uji independensi *Chi-Square*.

- Perumusan hipotesis:
 - H₀: tidak ada hubungan antara variabel masa studi dan variabel prediktor;
 - H₁: terdapat hubungan antara variabel masa studi dan variabel prediktor.
- Statistik uji yang digunakan dalam uji independensi *Chi-Square* dirumuskan sebagai berikut:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

dengan χ^2 merupakan statistik berdistribusi *Chi-Square* dengan derajat kebebasan $v = (r - 1)(k - 1)$.

- Kriteria Pengujian
 - Taraf signifikansi yang digunakan dalam penelitian ini sebesar 5% dan statistik $\chi^2_{(0,0);(r-1)(k-1)}$ kriteria pengujian adalah:
 - Tolak H₀, apabila $\chi^2 > \chi^2_{(0,0);(r-1)(k-1)}$ atau apabila nilai *p-value* (Sig.) < 0,05; dan
 - Terima H₀ pada keadaan lainnya.
- Hasil uji independensi *Chi-Square*:

Tabel 3. Hasil Uji Independensi Chi-Square

Variabel Respon	Variabel Prediktor	χ^2	<i>P-value</i>	Keputusan
Masa Studi	Jenis Kelamin	37,881	0,000	Tolak H ₀
	Program Studi	36,124	0,000	Tolak H ₀
	Jalur Masuk	96,913	0,000	Tolak H ₀
	Daerah Asal	0,383	0,826	Terima H ₀

Berdasarkan tabel 3, dapat disimpulkan bahwa terdapat hubungan antara variabel respon masa studi dengan masing-masing variabel prediktor jenis kelamin, program studi, dan jalur masuk yang signifikan secara statistik. Sedangkan untuk variabel prediktor daerah asal tidak terdapat hubungan yang signifikan secara statistik dengan variabel respon masa studi.

Analisis Klasifikasi Metode MARS

Pada pengklasifikasian ini, akan digunakan perbandingan antara data *training* dan data *testing* yaitu sebesar 50:50. Hal ini berarti model klasifikasi akan dibangun dari data *training* yang berukuran 50% dari data sampel dan pengujian model klasifikasi menggunakan data *testing* sebesar 50% dari data sampel.

1. Pemodelan MARS

Pemodelan MARS secara *trial and error* akan dilakukan dengan mengkombinasikan maksimum banyaknya fungsi basis (*max BF*), *Maximum Interaction* (MI), dan *Minimum Observation* (MO). Adapun aturan untuk pengkombinasian adalah sebagai berikut:

1. Menentukan kemungkinan maksimum banyaknya fungsi basis, yaitu 2 sampai 4 kali jumlah variabel prediktor yang digunakan. Pada penelitian ini maksimum banyaknya fungsi basis adalah 8, 12, dan 16;
2. Menentukan banyak maksimum interaksi (MI), yaitu 1, 2, dan 3;
3. Menentukan minimal jumlah pengamatan setiap *knot* (MO), yaitu 0, 1, 2 dan 3 (Friedman, 1991).

Pemodelan dilakukan dengan bantuan *software Salford Predictive Modeler* (SPM) versi 7.0 dan menghasilkan 36 model. Model MARS terbaik untuk masalah klasifikasi alumni berdasarkan masa studinya dibangun atas kombinasi $max BF = 16$, $MI = 3$, dan $MO = 1$, dengan nilai GCV model sebesar 0,16372, yang mempunyai bentuk persamaan sebagai berikut:

$$\hat{f}(x) = 0,249007 - 0,181434 \ B \ 1 + 0,411599 \ B \ 3 - 0,155581 \ B \ 5 + 0,217328 \ B \ 7 \quad (13)$$

dengan: $B \ 1 = (J \ \hat{t} \ (2, 4))$; $B \ 2 = (J \ \hat{t} \ (1, 3))$; $B \ 3 = \max(0, I \ - \ 2, 39)$; $B \ 5 = (G \ \hat{t} \ (1))$; dan $B \ 7 = (P \ \hat{t} \ (1)) \ B \ 2$.

Variabel prediktor yang berpengaruh pada model MARS yaitu variabel jalur masuk (*J*), IP semester pertama (*IP*), kelamin (*G*), dan program studi (*P*). Tingkat kepentingan dari masing-masing variabel prediktor yang berpengaruh tersaji pada tabel 4 berikut ini.

Tabel 4. Skor Variable Importance Model MARS Terbaik

Variabel	Skor Relative Variable Importance (%)
J	100,00
IP	98,55
G	34,02
P	32,03

Berdasarkan tabel 4 diperoleh informasi bahwa variabel yang mempunyai pengaruh paling dominan terhadap masa studi mahasiswa adalah variabel jalur masuk yang ditunjukkan dengan skor *variable importance*-nya sebesar 100%. Selain itu, variabel prediktor IP semester pertama, jenis kelamin, dan program studi juga berpengaruh terhadap masa studi dengan skor *variable importance* berturut sebesar 98,55%; 34,02%; dan 32,03%.

2. Pengujian Signifikansi Model MARS Terbaik

2.1 Pengujian Signifikansi Model MARS

- Rumusan hipotesis pengujian signifikansi model MARS secara simultan, yaitu:

$H_0: a_1 = a_2 = \dots = a_M = 0$ (model tidak signifikan);

H_1 : paling sedikit terdapat satu $a_j \neq 0$, $j = 1, 2, \dots, M$ (model signifikan);

- Statistik uji F_H yang digunakan, yaitu:
$$F_H = \frac{\frac{n}{l-1} \frac{(\sum_{i=1}^l \bar{y}_i - \bar{y})^2}{M}}{\frac{\sum_{i=1}^l (y_i - \bar{y}_i)^2}{n-M-1}}$$

dalam hal ini akan dihitung menggunakan *software* SPM v.7.0.

- Dengan mengambil taraf signifikansi $\alpha = 0,05$, kriteria pengujian model MARS adalah:
 - Tolak H_0 jika $F_H > F_{\alpha}(v_1, v_2)$ atau jika *p-value* (Sig.) $< 0,05$;
 - Terima H_0 pada kondisi sebaliknya.
- Keputusan: berdasarkan Tabel 5, diperoleh F_H sebesar 52,60160 dengan *p-value* sebesar 0,00000. Karena *p-value* $< 0,05$, maka H_0 ditolak pada taraf signifikansi 0,05.
- Kesimpulan: model MARS yang dimiliki adalah model yang signifikan.

Tabel 5. Output Pemodelan Data Menggunakan Metode MARS

```

=====
MARS Regression: Training Data
=====
W: 492.00          R-SQUARED: 0.30170
MEAN DEP VAR: 0.33740  ADJ R-SQUARED: 0.29596
UNCENTERED R-SQUARED = R-0 SQUARED: 0.53730
Parameter      Estimate    S.E.    T-Ratio    P-Value
-----
Constant      | 0.24901   0.04962   5.01790   0.00000
    
```

Basis Function 1	-0.18143	0.05209	-3.48312	0.00054
Basis Function 3	0.41160	0.04404	9.34599	0.00000
Basis Function 5	-0.15558	0.03825	-4.06704	0.00006
Basis Function 7	0.21733	0.05528	3.93107	0.00010

F-STATISTIC= 52.60160 S.E. OF REGRESSION = 0.39713
P-VALUE = 0.00000 RESIDUAL SUM OF SQUARES =
76.80750
[MDF,NDF] = [4, 487] REGRESSION SUM OF SQUARES =
33.18437

2.2 Pengujian Koefisien Regresi Model MARS

- Hipotesis pengujian signifikansi secara parsial, yaitu:
 $H_0: a_j = 0$ (koefisien a_j tidak berpengaruh terhadap model);
 $H_1: a_j \neq 0$, untuk setiap $j, j = 1, 2, \dots, M$ (koefisien a_j berpengaruh terhadap model);
- Menentukan statistik uji t_{H_1} berdasarkan rumusan: $t_{H_1} = \frac{a_j}{s_{a_j}}$,
- Mengambil taraf signifikansi $\alpha = 0,05$, kriteria pengujian model MARS adalah:
 - Tolak H_0 jika $t_{H_1} > t_{(\frac{\alpha}{2}, n-1)}$ atau jika $p\text{-value (Sig.)} < 0,05$;
 - Terima H_0 pada keadaan sebaliknya.
- Keputusan: karena seluruh estimator parameter, yaitu a_0, a_1, a_3, a_5 , dan a_j masing-masing mempunyai nilai $p\text{-value} < 0,05$ (Tabel 5) maka untuk seluruh estimator parameter, H_0 ditolak pada taraf signifikansi 0,05.
- Kesimpulan: konstanta dan koefisien fungsi basis 1, 3, 5, dan 7 mempunyai pengaruh yang signifikan terhadap model.

3. Interpretasi Model MARS Terbaik

Berikut ini adalah interpretasi untuk model MARS terbaik pada persamaan (13).

- i $B_1 = (J \text{ ti } (2, 4)) = \begin{cases} 1, & \text{jika } J = 2 \text{ atau } 4 \\ 0, & \text{jika } J = 1 \text{ atau } 3 \end{cases}$, dengan koefisien $-0,181434$, memiliki arti bahwa untuk setiap kali kondisi koefisien BF1 bermakna (BF1 bernilai 1) dipenuhi, hal tersebut akan menurunkan kemungkinan mahasiswa lulus tepat waktu sebesar $0,181434$.
- ii $B_3 = \max(0, I_1 - 2,39) = \begin{cases} (I_1 - 2,39), & \text{jika } I_1 > 2,39 \\ 0, & \text{jika } I_1 \leq 2,39 \end{cases}$, dengan koefisien $0,411599$, memiliki arti bahwa untuk setiap kali kondisi koefisien BF3 bermakna ($BF3 > 2,39$) dipenuhi, untuk dua mahasiswa yang mempunyai IP semester pertama berbeda sebesar satu satuan, maka mahasiswa yang memiliki IP semester pertama lebih besar akan memiliki kemungkinan lulus tepat waktu lebih tinggi sebesar $0,411599$.
- iii $B_5 = (G \text{ ti } (1)) = \begin{cases} 1, & \text{jika } G = 1 \\ 0, & \text{jika } G = 2 \end{cases}$, dengan koefisien $-0,155581$, memiliki arti bahwa untuk setiap kali kondisi BF5 bermakna dipenuhi, hal tersebut akan menurunkan kemungkinan mahasiswa lulus tepat waktu sebesar $0,155581$.
- iv $B_7 = (P \text{ ti } (1)) \quad B_2 = \begin{cases} 1, & \text{jika } P = 1, J = 1, 3 \\ 0, & \text{jika } P \text{ dan } J \text{ lainnya} \end{cases}$, dengan koefisien $0,217328$, memiliki arti bahwa untuk setiap kali kondisi koefisien BF7 bermakna dipenuhi, hal tersebut akan meningkatkan kemungkinan mahasiswa lulus tepat waktu sebesar $0,217328$.

Setiap interpretasi yang diberikan di atas diberikan dengan memandang fungsi basis lain yang terlibat di dalam model bersifat konstan.

4. Hasil Klasifikasi Metode MARS

Berikut ini adalah hasil klasifikasi yang diperoleh dengan *software* SPM versi 7.0.

Tabel 6. Hasil Klasifikasi Lama Masa Studi Mahasiswa FPMIPA UPI

Kelas Aktual	Kelas Prediktif		Total Aktual
	Kelas 0 (Lulus Terlambat)	Kelas 1 (Lulus Tepat Waktu)	
Kelas 0 (Lulus Terlambat)	297	29	326
Kelas 1 (Lulus Tepat Waktu)	82	84	166
Total Prediksi	379	113	492

Statistik uji yang digunakan untuk mengevaluasi hasil klasifikasi adalah statistik APER dan *Press's Q*. Adapun nilai APER untuk hasil klasifikasi di atas adalah:

$$\text{APER} = \frac{n_{1M} + n_{2M}}{n_1 + n_2} = \frac{2 + 8}{2 + 2 + 8 + 8} \times 100\% = 22,5610\%$$

Nilai APER sebesar 22,5610% menunjukkan persentase objek yang salah diklasifikasikan (*missclassification*). Karena nilai statistik APER masih di bawah 50% maka ketepatan hasil klasifikasi masih dapat diterima dan digunakan untuk mengklasifikasikan mahasiswa FPMIPA UPI yang baru termasuk pada kelompok tertentu (Wuensch, 2009).

Adapun untuk statistik *Press's Q*, diperoleh nilai sebagai berikut:

$$P \text{ 's } Q = \frac{[N - (n)]^2}{N(K - 1)} = \frac{[492 - (381 \times 2)]^2}{492 \times (2 - 1)} = 148,1707.$$

Nilai statistik *Press's Q* tersebut dibandingkan dengan nilai kritis distribusi *Chi-Square* dengan derajat bebas 1 dan taraf signifikansi = 0.05, yaitu $\chi^2_{(1;0.05)} = 3,84146$. Karena nilai statistik *Press's Q* lebih besar daripada nilai kritis *Chi-Square*, maka diperoleh kesimpulan bahwa pengklasifikasian ini stabil dan konsisten secara statistik.

KESIMPULAN DAN SARAN

Kesimpulan

1. Analisis klasifikasi masa studi mahasiswa FPMIPA UPI menggunakan metode MARS menyatakan bahwa empat variabel prediktor yang berpengaruh terhadap masa studi mahasiswa FPMIPA UPI, yaitu jalur masuk, IP semester pertama, jenis kelamin, dan program studi, dimana variabel prediktor yang paling berpengaruh adalah jalur masuk. Disamping itu, model klasifikasi alumni FPMIPA UPI berdasarkan lama masa studi tersebut dapat digunakan untuk mengklasifikasikan mahasiswa FPMIPA UPI baru dengan hasil klasifikasi bersifat stabil;
2. Model klasifikasi MARS terbaik untuk data masa studi alumni FPMIPA UPI dihasilkan dari kombinasi $max BF = 16$, $MI = 3$, dan $MO = 1$, dengan persamaan sebagai berikut:

$$\hat{f}(x) = 0,249007 - 0,181434 \quad B \ 1 + 0,411599 \quad B \ 3 - 0,155581 \\ B \ 5 \\ + 0,217328 \quad B \ 7$$

dimana: $B_1 = (J_{in}(2,4))$; $B_2 = (J_{in}(1,3))$; $B_3 = \max(0, I_{in} - 2,39)$; $B_5 = (G_{in}(1))$; dan $B_7 = (P_{in}(1))$ B_2 .

Secara umum, model MARS di atas memberikan gambaran bahwa:

- i. Kemungkinan lulus tepat waktu bagi mahasiswa yang diterima melalui jalur SBMPTN atau jalur Nonreguler cenderung lebih rendah daripada mahasiswa yang diterima melalui jalur masuk PMDK atau UM UPI;
- ii. Kemungkinan lulus tepat waktu cenderung lebih tinggi bagi mahasiswa yang memperoleh IP semester pertama lebih dari 2,39, dimana apabila terdapat dua mahasiswa yang mempunyai IP semester pertama lebih dari 2,39 dan keduanya berbeda sebesar satu satuan, maka mahasiswa yang memiliki IP semester pertama lebih besar akan memiliki kemungkinan lulus tepat waktu lebih tinggi sebesar 0,411599;
- iii. Kemungkinan lulus tepat waktu bagi mahasiswa laki-laki cenderung lebih rendah daripada mahasiswa perempuan;
- iv. Sedangkan kemungkinan lulus tepat waktu bagi mahasiswa program studi kependidikan yang diterima melalui jalur PMDK atau UM UPI akan mempunyai nilai yang lebih tinggi daripada mahasiswa dari kondisi lainnya;

dimana interpretasi pada masing-masing poin (poin i sampai dengan poin iv) memandang faktor lainnya dalam keadaan konstan.

Saran

Untuk pengelola fakultas, model klasifikasi MARS untuk mahasiswa FPMIPA UPI berdasarkan lama masa studinya dapat menjadi masukan positif dalam mempertimbangkan kebijakan yang bertujuan meminimalisir jumlah mahasiswa yang lulus terlambat. Sedangkan untuk penelitian lebih lanjut, dapat melakukan penelitian serupa dengan melibatkan variabel prediktor lain seperti nilai Ujian Nasional SMA, kondisi ekonomi keluarga, skor IQ, dan lain sebagainya. Selain itu, penelitian serupa dapat dilakukan untuk ruang lingkup yang lebih luas, misalnya menganalisis klasifikasi alumni untuk lingkungan se-universitas sehingga akan menghasilkan suatu keputusan yang lebih umum dan mencakup ranah yang lebih luas. Serta, dapat dilakukan penelitian mengenai perbandingan klasifikasi menggunakan metode MARS dengan metode klasifikasi lainnya sehingga dapat diketahui metode manakah yang menghasilkan pengklasifikasian lebih baik.

REFERENSI

- Agwil, W., Rahmi, I., & Yozza, H. (2012). *Prediksi Luas Area Kebakaran Hutan Berdasarkan Data Meteorologi dengan Menggunakan Pendekatan Multivariate Adaptive Regression Spline (MARS)*. (Tugas Akhir). Universitas Andalas, Padang.
- Febriyanti A., dkk. (2012). Penerapan Metode *Multivariate Adaptive Regression Spline* (MARS) untuk Mengidentifikasi Komponen yang Berpengaruh terhadap Peringkat Akreditasi Sekolah (Kasus SMA/MA di Provinsi Sumatera Barat). *Jurnal Matematika UNAND*, 2(2), hlm. 44-53.
- Friedman, J. H. (1991). Multivariate Adaptive Regression Spline. *The Annals of Statistics*, 19(1), hlm. 1-67.
- Guillory, C. W. (2008). *A Multilevel Discrete Time Hazard Model of Retention Data in Higher Education*. (Disertasi). Louisiana State University, Louisiana.
- Kriner, M. (2007). *Survival Analysis with Multivariate Adaptive Regression Spline*. (Disertasi). Universitas Munchen, Munich.
- Merluarini, B., Safitri, D., & Hoyyi, A. (2014). Perbandingan Analisis Klasifikasi Menggunakan Metode *K-Nearest Neighbor* (K-NN) dan *Multivariate Adaptive Regression Spline* (MARS) pada Data Akreditasi Sekolah Dasar Negeri di Kota Semarang. *Jurnal Gaussian*, 3(3), hlm. 313-322.
- Neter, J., William, W. & Michael, H. K. (1983). *Applied Linear Regression Models*. Homewood : Richard D. Irwin, Inc.
- Oktaviana, D. & Budiantara, I. N. (2011). *Regresi Spline Birespon untuk Memodelkan Kadar Gula Darah Penderita Diabetes Melitus*. [Online]. Diakses dari <http://digilib.its.ac.id/public/ITS-paper-19523-1307100068-Paper.pdf>. [16 Januari 2015].
- Permana, H. (2011). *Klasifikasi dengan Metode CHAID (Chi-Squared Automatic Interaction Detection) dan Penerapannya pada Klasifikasi Alumni FMIPA UNY*. (Skripsi). Universitas Negeri Yogyakarta, Yogyakarta.
- Wicaksono, W., Wilandari, Y., & Suparti. (2014). Pemodelan Multivariate Adaptive Regression Splines (MARS) pada Faktor-faktor Resiko Angka

- Kesakitan Diare (Studi Kasus: Angka Kesakitan Diare di Jawa Tengah, Jawa Timur, dan Daerah Istimewa Yogyakarta Tahun 2011). *Jurnal Gaussian*, 3(3), hlm. 253-262.
- Wuensch, K. (2009). *Binary Logistic with PASW/SPSS*. [Online]. Diakses dari <http://core.ecu.edu/psyc/wuenschk/MV/Multreg/Logistic-SPSS.doc>. [15 Juni 2015].
- Zhang, H. & Singer, B. H. (2010). *Recursive Partitioning and Applications 2nd Edition*. New York: Springer.