Volume x No x xxxxx (20xx)

e-ISSN: **2615-3270**

p-ISSN: **2615-3599**

Eigen Mathematics Journal

Homepage jurnal: http://eigen.unram.ac.id

Aplication of Multivariate Adaptive Regression Splines Method In Estimation of Long Study 0f Mathematics Students FMIPA University Of Mataram

***Widiatul Parida*[a], *Nurul Fitriyani*[b], *Lisa Harsyiah*[c]**

[a]Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mataram, Jalan Majapahit No. 62, Mataram, Indonesia, 83125, E-mail: widiatulparida2000@gmail.com
[b] Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mataram, Jalan Majapahit No. 62, Mataram, Indonesia, 83125, E-mail: uyu.statistika.its@gmail.com
[c]Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mataram, Jalan Majapahit No. 62, Mataram, Indonesia, 83125, E-mail: Lisa_harsyiah@unram.ac.id

A B S T R A C T

In the Mathematics Study Program FMIPA UNRAM there were 82,82% of students who graduated not on time from the 2014-2017 classes. Therefore, it is necessary to conduct an analysis to determine the factors that influence the length of the student's study using the Multivariate Adaptive Regression Splines (MARS) method. The MARS method is a nonparametric regression approach that is capable of processing high-dimensioanal data. This study aims to determine the factors that influence the length of the study and determine the accuaracy of classification using the MARS method. The results of this study indicate that the predictor variables that affect the length of study for student's are compulsory course grades of probability theory (X_7) *and real analysis I* (X_8) *and the accuracy of the classification of the length of study for Mathematics Study Program students based on the status of graduating on time or graduating not on time is 60%.*

Keywords: *Classification, Multivariate Adaptive Regression Splines, Study Length*

Diserahkan: xx-xx-xxxx; Diterima: xx-xx-xxxx; Doi: https://doi.org/10.29303/emj.xxx.x

1. Pendahuluan

Pendidikan merupakan salah dipenuhi oleh setiap orang. Selain itu, pendidikan merupakan wadah untuk meningkatkan kualitas sumber daya manusia agar salah satu tujuan pembangunan nasional tercapai (Parmadani dan Latifah, 2016). Pendidikan memegang peranan penting bagi masa depan bangsa. Di Indonesia perguruan tinggi sebagai salah satu jenjang pendidikan yang bertujuan melahirkan para sarjana yang handal dan memiliki keterampilan di bidangnya (Annur, dkk., 2015).

Universitas Mataram (UNRAM) merupakan salah satu perguruan tinggi negeri yang ada di Indonesia. UNRAM merupakan salah satu Universitas Negeri yang berlokasi di Jl. Majapahit No. 62, Gomong, Kec. Selaparang, Kota Mataram, Nusa Tenggara Barat, 83115. UNRAM memiliki lebih dari 50 jurusan di semua jenjang. Kampus ini terdiri dari 5 jenjang pendidikan, mulai dari jenjang Sekolah Vokasi (Diploma), Sarjana (S1), Profesi, Magister (S2), serta Doktor (S3). UNRAM merupakan salah satu kampus tertua di NTB yang didirikan sejak tanggal 1 Oktober 1962 hingga tahun 2022 ini. UNRAM telah memiliki 10 fakultas, salah

* Corresponding author.
Alamat e-mail: widiatulparida2000@gmail.com

satunya adalah Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) (Amin, 2014). Fakultas MIPA Universitas Mataram dibentuk pada tanggal 25 Agustus 2007 berdasarkan Surat Keputusan Rektor Universitas Mataram nomor 10. 147/H18/HK/2007. Fakultas MIPA yang terdiri dari lima program studi, yaitu Biologi, Fisika, Kimia, Matematika dan Ilmu Lingkungan.

Keputusan Menteri Pendidikan Nasional Republik Indonesia Nomor 232/U/2000 BAB III Pasal 5 menjelaskan bahwa beban studi program sarjana minimal 144 SKS dan maksimal 160 SKS yang dijadwalkan untuk 8 semester dan dapat ditempuh dalam waktu kurang dari 8 semester dan maksimal 14 semester setelah pendidikan menengah (Ghofar, 2014). Pada kenyataannya banyak mahasiswa yang mengalami kesulitan untuk menuntaskan studinya sesuai dengan standar waktu yang sudah ditentukan (Annur, dkk., 2015). Seperti yang terjadi di Program Studi Matematika FMIPA UNRAM masih banyak mahasiswa yang menyelesaikan masa studinya lebih dari 4 tahun (lulus tidak tepat waktu). Berdasarkan data akademik FMIPA Universitas Mataram menunjukkan bahwa rata-rata persentase jumlah mahasiswa yang tidak lulus tepat waktu dari tahun 2014 - 2017 sebesar 82,82% dari total mahasiswa Program Studi Matematika sebanyak 213 mahasiswa.

Masalah lama studi mahasiswa tersebut, perlu dilakukan analisis untuk mengetahui faktor-faktor yang mempengaruhi lama studi mahasiswa Program Studi Matematika FMIPA. Dalam ilmu statistika, banyak metode yang digunakan untuk mengetahui pengaruh variabel prediktor terhadap variabel responnya yang bersifat kategori, salah satunya adalah metode *Multivariate Adaptive Regression Splines* (MARS) (Anam, dkk., 2017). Metode MARS merupakan pendekatan untuk regresi nonparametrik yang berguna untuk mengatasi data berdimensi tinggi dan menghasilkan prediksi variabel respon yang akurat dan menghasilkan model yang kontinu pada *knot* berdasarkan nilai *Generalized Cross Validation* (GCV) terkecil (Friedman, 1991).

Kelebihan metode MARS adalah berguna untuk mengatasi masalah data yang berdimensi tinggi, yaitu data yang memiliki jumlah variabel prediktor sebesar $3 \leq n \leq 20$. Metode MARS merupakan pengembangan dari pendekatan *Recursive Partition Regression* (RPR) yang dikembangkan dengan metode *spline* sehingga model yang dihasilkan kontinu pada *knot* (Wicaksono, dkk., 2014). Menurut Nisa' dan Budiantara (2012), ada beberapa hal yang perlu diperhatikan dalam menggunakan metode MARS yaitu *knot*, *Basis Function* (BF) dan interaksi. Metode MARS sangat baik untuk klasifikasi karena menghasilkan sedikit kesalahan klasifikasi (Anam, dkk., 2017). Metode klasifikasi merupakan bagian dari analisis statistika yang bertujuan untuk memisahkan individu atau objek ke dalam suatu kelompok sehingga dapat diketahui suatu individu berada pada kelompok tertentu. Metode klasifikasi yang baik akan menghasilkan peluang kesalahan alokasi yang kecil (Johnson dan Wichern, 1992).

Berdasarkan uraian di atas, maka penulis melakukan penelitian untuk mengetahui karakteristik dan faktor-faktor yang berpengaruh terhadap lama studi mahasiswa Program Studi Matematika FMIPA Universitas Mataram angkatan 2014-2017 dengan menggunakan metode MARS. Penelitian menggunakan metode MARS diharapkan akan mendapatkan informasi mengenai faktor yang berpengaruh terhadap lama studi mahasiswa Program Studi Matematika FMIPA Universitas Mataram. Hasil tersebut dapat digunakan sebagai dasar dalam pengambilan kebijakan serta evaluasi untuk meningkatkan jumlah kelulusan mahasiswa Program Studi Matematika FMIPA Universitas Mataram di masa yang akan datang.

2. Landasan Teori

2.1. Regresi Nonparametrik

Regresi nonparametrik merupakan suatu metode yang digunakan untuk mengetahui pola hubungan antara variabel respon dan prediktor yang tidak diketahui bentuk fungsinya (Wulandari dan Budiantara, 2014). Dalam regresi nonparametrik, kurva regresi hanya diasumsikan mulus (*smooth*) dalam arti termuat pada suatu ruang fungsi tertentu sehingga mempunyai sifat fleksibilitas yang tinggi (Eubank, 1999).

Menurut Hardle (1990) model umum regresi nonparametrik dapat dituliskan sebagai berikut:

$$Y_i = m(X_i) + \varepsilon_i \quad , i = 1, 2, 3, \dots, n \qquad (2.1)$$

dengan,
Y_i : variabel respon pengamatan ke-i
X_i : variabel prediktor pengamatan komponen nonparametrik ke-i
m : fungsi regresi yang tidak diketahui
ε_i : residual ke-i

Menurut Wulandari dan Budiantara (2014), salah satu contoh regresi ialah regresi nonparametrik spline *truncated*. Model regresi spline *truncated* secara umum dapat ditulis dengan persamaan:

$$y_i = f(t_i) + \varepsilon_i \quad ; i = 1, 2, \dots, n \qquad (2.2)$$

dengan $f(t_i)$ merupakan fungsi spline berorde p dengan titik *knot* $k_1, k_2, \dots, k_r$ yang dapat ditulis dengan persamaan berikut:

$$f(t_i) = \sum_{j=0}^{p} \gamma_i \; t_i^j \; + \; \sum_{l=1}^{r} \gamma_{p+l} \; (t_i - k_l)_+^p \qquad (2.3)$$

$(t_i - k_l)_+^p$ merupakan fungsi *truncated* (terpotong) yang dapat dijabarkan sebagai berikut:

$$(t_i - k_l)_+^p = \begin{cases} (t_i - k_l)_+^p & , t_i \ge k_l \\ 0 & , t_i < k_l \end{cases} \qquad (2.4)$$

Bila persamaan (2.3) disubstitusikan ke persamaan (2.2) akan menghasilkan model regresi nonparametrik *spline* sebagai berikut:

$$y_i = \sum_{j=0}^{p} \gamma_i \; t_i^j \; + \; \sum_{l=1}^{r} \gamma_{p+l} \; (t_i - k_l)_+^p \varepsilon_i \qquad (2.5)$$

2.2. Multivariate Adaptive Regression Splines

Multivariate Adaptive Regression Splines (MARS) merupakan pendekatan regresi nonparametrik yang menghasilkan pemodelan regresi yang fleksibel untuk data dengan variabel prediktor $3 \le k \le 20$ dan ukuran contoh $50 \le n \le 1000$. Model MARS merupakan perluasan hasil kali fungsi basis spline, dimana jumlah fungsi basis beserta parameter (derajat hasil kali, lokasi *knot*) ditentukan oleh data dengan menggunakan algoritma *recursive partitioning* yang dimodifikasi (Febriyanti, dkk., 2013). Metode MARS dapat digunakan untuk data yang berdimensi tinggi yaitu dengan variabel prediktor (X_v) sebanyak $3 \le v \le 20$ (Friedman, 1991). Menurut Pintowati dan Otok (2012), beberapa hal yang perlu diperhatikan dalam menggunakan model MARS yaitu:

a. *Knot*

Knot yaitu akhir dari sebuah regresi (*region*) dan awal dari sebuah garis regresi (*region*) yang lain. Di setiap titik *knot*, diharapkan adanya kontinuitas dari basis fungsi antar satu *region* dengan *region* lainnya.

b. Fungsi basis (FB)

Fungsi basis yaitu suatu fungsi yang digunakan untuk menjelaskan hubungan antara variabel respon dan variabel prediktor. Fungsi basis ini merupakan fungsi parametrik yang didefinisikan pada tiap *region*. Pada umumnya fungsi basis yang dipilih adalah berbentuk polinomial dengan turunan yang kontinu pada setiap *knot*. Friedman (1991), menyarankan banyaknya maksimum fungsi basis (FB) adalah 2-4 kali banyaknya variabel prediktornya. Banyaknya maksimum interaksi (MI) yang digunakan adalah 1, 2, atau 3. Jika MI > 3 akan dihasilkan model yang semakin kompleks dan model akan sulit untuk diinterpretasi. Minimum jarak antara *knot* atau minimum observasi (MO) yang digunakan sebesar 0, 1, 2, dan 3.

Menurut Friedman (1991) model umum MARS dapat ditulis:

$$f(x) = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{Km} \left[S_{km} . \left(x_{v(k,m)} - t_{km} \right) \right]_+ \qquad (2.6)$$

dengan,

a_0 = basis fungsi induk
a_m = koefisien dari fungsi basis ke-m
M = banyaknya maksimum fungsi basis
K_m = derajat interaksi pada fungsi basis ke-m
S_{km} = bernilai 1 jika x terletak di kanan titik knot dan bernilai -1 jika x terletak di kiri titik knot
$x_{v(k,m)}$ = variabel prediktor ke-v
t_{km} = nilai knot dari variabel prediktor $x_{vN(k,m)}$

Berdasarkan fungsi regresi nonparametrik, model MARS dinyatakan dalam Persamaan berikut:

$$f(x) = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{Km} \left[S_{km} . \left(x_{v(k,m)} - t_{km} \right) \right]_+ + \varepsilon_i \qquad (2.7)$$

dengan

y_i = variabel respon ke-i
ε_i = *error*
i = $1, 2, \ldots n$

Dari model MARS pada persamaan (2.7) dalam bentuk matriks dapat ditulis sebagai berikut (Khishartini, dkk., 2014):

$$Y = Ba + \varepsilon \qquad (2.8)$$

dengan,

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}$$

$$a = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_M \end{bmatrix}$$

$$B = \begin{pmatrix} 1 \prod_{k=1}^{K_1} [S_{1m}.(x_{v1(1,m)} - t_{1m})] & \dots & \prod_{k=1}^{K_M} [S_{Mm}.(x_{v1(M,m)} - t_{Mm})] \\ 1 \prod_{k=1}^{K_1} [S_{1m}.(x_{v2(1,m)} - t_{1m})] & \ddots & \prod_{k=1}^{K_M} [S_{Mm}.(x_{v2(M,m)} - t_{Mm})] \\ \vdots & & \vdots \\ 1 \prod_{k=1}^{K_1} [S_{1m}.(x_{vN(1,m)} - t_{1m})] & \dots & \prod_{k=1}^{K_M} [S_{Mm}.(x_{vN(M,m)} - t_{Mm})] \end{pmatrix}$$

Selanjutnya estimasi modelnya diperoleh dari Persamaan (3.20) yaitu:

$$\hat{Y} = B\,\hat{a}$$

$$\hat{Y} = B\ (B^T\ B)^{-1}\ B^T Y \qquad (2.9)$$

Menurut Friedman (1991), model MARS untuk variabel respon bernilai biner, diadaptasi dengan pendekatan regresi logistik dan didefinisikan dalam Persamaan (2.10).

$$\log\left[\frac{\pi(x)}{1-\pi(x)}\right] = \hat{f}\ (x) \qquad (2.10)$$

dengan,
π = menyatakan probabilitas bagi variabel respon Y bernilai terbesar $(Y = 1)$
$\hat{f}\ (x)$ = menyatakan model MARS

2.3. Algoritma Multivariate Adaptive Regression Splines

Algoritma MARS dibagi menjadi dua yaitu:

a. *Forward Stepwise*

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

1) Memisalkan $B_0 = 1$, yang merupakan fungsi basis konstan, sebagai 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 memilki nilai *Mean Square Error* (MSE) minimum, MSE didefinisikan sebagai berikut:

$$MSE = \frac{1}{N}\sum_{t=1}^{N}(y_i - \hat{f}_M(x_i))^2 \qquad (2.11)$$

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

3) 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 mengasilkan beberapa kemungkinan kandidat model. Oleh karena itu, akan dipilih pasangan hasil kali yang menghasilkan model dengan nilai MSE terkecil;
4) Mengulangi langkah (3) hingga banyaknya fungsi basis dalam model lebih atau sama dengan maksimum banyaknya fungsi basis yang telah ditetapkan.

b. *Backward Stepwise*

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

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

2.4. Pemilihan Metode MARS Terbaik

Metode MARS terbaik adalah metode yang memiliki nilai *Generalized Cross Validation* (GCV) minium. GCV digunakan untuk mendapatkan *knot* yang optimum. Adapun GCV dirumuskan sebagai berikut (Friedman, 1991):

$$GCV\ (M) = \frac{\frac{1}{n}\sum_{i=1}^{n}\left[y_i - \hat{f}_M(x_i)\right]^2}{[n^{-1}tr(\boldsymbol{I} - \boldsymbol{H})]^2} \qquad (2.12)$$

dengan,
$\boldsymbol{I}$ = Matriks identitas
M = jumlah fungsi basis
x_i = variabel prediktor
y_i = variabel respon
n = banyaknya pengamatan
H = $[B(B^T B)^{-1} B^T]$

$$B = \begin{pmatrix} 1 \prod_{k=1}^{K_1} [S_{1m}.(x_{v1(1,m)} - t_{1m})] & \dots & \prod_{k=1}^{K_M} [S_{Mm}.(x_{v1(M,m)} - t_{Mm})] \\ 1 \prod_{k=1}^{K_1} [S_{1m}.(x_{v2(1,m)} - t_{1m})] & \ddots & \prod_{k=1}^{K_M} [S_{Mm}.(x_{v2(M,m)} - t_{Mm})] \\ \vdots & & \vdots \\ 1 \prod_{k=1}^{K_1} [S_{1m}.(x_{vN(1,m)} - t_{1m})] & \dots & \prod_{k=1}^{K_M} [S_{Mm}.(x_{vN(M,m)} - t_{Mm})] \end{pmatrix}$$

2.5. Pengujian Signifikansi Model MARS

Apabila telah ditemukan model MARS terbaik, maka dilakukan pengujian untuk mengecek signifikansi parameter untuk mengevaluasi kecocokan model. Pengujian dilakukan dengan menguji koefisien regresi secara simultan maupun secara parsial sebagai berikut (Wicaksono,dkk., 2014) :

1. Pengujian koefisien regresi simultan
   a. Rumusan Hipotesis :
   $H_0 : a_1 = 0$ (model tidak signifikan)
   $H_1 :$ minimal terdapat satu $a_m \neq 0$; $m = 10$ (model signifikan)
   b. Taraf signifikansi : α
   c. Statistika uji :

$$F_{hitung} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2/M}{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2/(N-M-1)} \quad (2.13)$$

   d. Daerah kritis :
   Tolak H_0 jika nilai $F > F_{\frac{\alpha}{2}(M;N-M-1)}$ atau P value $< \alpha$
2. Pengujian koefisien regresi parsial
   a. Rumusan hipotesis :
   $H_0 : a_m = 0$ (koefisien a_m tidak berpengaruh terhadap model)
   $H_1 : a_m \neq 0$; untuk setiap m, dimana $m = 10$ (koefisien a_m berpengaruh terhadap model)
   b. Taraf signifikansi : α
   c. Statistika uji :

$$t_{hitung} = \frac{\hat{a}_m}{Se(\hat{a}_m)} \quad (2.14)$$

   dengan $Se(\hat{a}_m) = \sqrt{var\ (\hat{a}_m)}$

$$Se(\hat{a}_m) = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{N-M-1} C_{jj}}$$

   C_{jj} adalah elemen-elemen pada diagonal utama matriks $(\boldsymbol{B^T B})^{-1}$
   d. Daerah kritis :
   Tolak H_0 jika $|t_{hitung}| > t_{\left(\frac{\alpha}{2},N-M\right)}$ atau P value < α.

2.6. *Variable Importance*

Kriteria yang digunakan untuk mengestimasi *variable importance* pada model MARS adalah *Residual Sum of Squares* (RSS). Variabel yang menyebabkan penurunan RSS yang lebih besar dianggap sebagai variabel yang lebih penting. Adapun RSS yang akan diturunkan apabila variabel terkait tidak dimasukkan ke dalam model, dirumuskan sebagai berikut (Oktora, 2016):

$$RSS = \sum_{i=1}^{n}\left(y_i - f(x_i)\right)^2 \quad (2.15)$$

dengan,
RSS = jumlah residual kuadrat
y_i = nilai ke-i yang akan diprediksi
$f(x_i)$ = nilai prediksi y_i
i = $1,2,3,\ldots n$

2.7. Klasifikasi pada MARS

Pada dasasrnya, klasifikasi dilakukan untuk melihat seberapa besar ketepatan pengelompokan sekumpulan data untuk digolongan dengan tepat pada kelompoknya. Menurut Cox dan Snell (1989) jika variabel respon terdiri dari dua nilai, maka dapat dikatakan sebagai regresi dengan *binary response*. Sehingga didapatkan model probabilitas dengan persamaan sebagai berikut.

$$\pi(x) = \frac{e^{\hat{f}(x)}}{1 + e^{\hat{f}(x)}} \quad (2.15)$$

$$\{1 - \pi(x)\} = \frac{1}{1 + e^{\hat{f}(x)}} \quad (2.16)$$

dengan $\hat{f}(x) = logit\ \pi(x)$
sehingga didapatkan $prob(Y = 1) = \pi(x)$ dan $prob(Y = 0) = 1 - \pi(x)$, karena variabel respon Y bersifat biner (0 dan 1) dengan variabel prediktor X sebanyak n, $x = (x_1, x_2, \ldots, x_n)$, maka model MARS untuk klasifikasi dapat dituliskan sebagai berikut:

$$logit\ \pi(x) = ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \hat{f}(x) \quad (2.17)$$

Menurut Hosmer dan Lemeshow (1989) jika variabel respon merupakan biner maka dapat dikatakan sebagai regresi respon biner. Untuk menentukan klasifikasi pada variabel respon biner (1 dan 0) dilakukan dengan menggunakan titik potong (*cut point*) sebesar 0,5 dengan ketentuan apabila estimasi peluang melebihi 0,5 maka hasil prediksi adalah masuk ke kelompok 1, apabila estimasi peluang kurang dari atau sama dengan 0,5 maka hasil prediksi adalah masuk ke kelompok 0.

2.8. *Apparent Eror Rate* (APER)

Kesalahan klasifikasi merupakan kesalahan dari pengklasifikasian suatu observasi baru ke dalam suatu kelompok. Salah satu metode yang digunakan untuk menghitung probabilitas kesalahan klasifikasi adalah *Apparent Error Rate* (APER). APER merupakan perhitungan kinerja hasil klasifikasi yang dilakukan dengan menggunakan matriks konfusi (*confusin matric*). Persentase kesalahan klasifikasi dapat dihitung dari matriks yang menunjukkan nilai sebenarnya atau aktual dan nilai prediksi dari setiap kelompok, seperti terlihat pada tabel berikut (Rofiq, dkk., 2016).

Tabel 3.1 Klasifikasi Dua Kelompok

Aktual	Prediksi		
	π_1	π_2	Total
π_1	n_{11}	n_{12}	$n_{11}+n_{12}$
π_2	n_{21}	n_{22}	$n_{21}+n_{22}$
Total	$n_{11}+n_{21}$	$n_{12}+n_{22}$	N= $n_{11}+n_{12}+n_{21}+n_{22}$

Nilai APER dapat dihitung pada Persamaan (2.18):

$$\text{APER} = \frac{n_{12}+n_{21}}{N} \times 100\% \qquad (2.18)$$

3. Metode Penelitian

Data penelitian yang digunakan berupa data sekunder, yaitu data yang diperoleh atau dikumpulkan peneliti dari sumber yang telah ada. Data yang digunakan adalah data mahasiswa Program Studi Matematika angkatan 2014-2017 yang telah dicatat oleh Divisi Akademik FMIPA Universitas Mataram. Alat yang digunakan pada penelitian berupa perangkat *Microsoft Excel*, *Salford Predictive Modeler* (SPM) *7.0 trial* dan *Software* R Studio.

Variabel yang digunakan yaitu Lama Studi sebagai variabel respon, dengan variabel prediktor diantaranya Indeks Prestasi Kumulatif (X_1), Jenis Kelamin (X_2), Jalur Masuk (X_3), Nilai Mata Kuliah Wajib Struktur Aljabar (X_4), Nilai Mata Kuliah Wajib Kalkulus Lanjut (X_5), Nilai Mata Kuliah Wajib Program Linier (X_6), Nilai Mata Kuliah Wajib Teori Peluang (X_7) dan Nilai Mata Kuliah Wajib Analisis Real I (X_8).

Tahap analisis yang digunakan untuk mencapai tujuan penelitian sebagai berikut:

1. Mengkombinasikan besarnya Fungsi Basis (FB), Maksimu Interaksi (MI) dan Minimum Observasi (MO) dengan cara *trial and error* dengan langkah:
   a. Menentukan maksimum fungsi basis (BF) yaitu sebanyak 2 sampai 4 kali jumlah prediktor yang akan digunakan.
   b. Menentukan jumlah maksimum interaksi (MI) yaitu 1, 2 dan 3 dengan cara asumsi bahwa jika maksimum interaksi lebih besar dari 3, maka akan menghasilkan model yang semakin kompleks akan tetapi sulit untuk diinterpretasikan.
   c. Menentukan minimal observasi (MO) atau minimal jumlah pengamatan setiap *knot* yaitu 0, 1, 2 dan 3.
2. Menetapkan model terbaik dengan didasarkan pada nilai *Generalized Cross Validation* (GCV) minimal yang diproleh dari pengkombinasian antara fungsi basis (BF), maksimum interaksi (MI) dan minimum observasi (MO).
3. Melakukan pendugaan koefisien model $(a_1, a_2, \dots, a_m)$. Sehingga model yang diperoleh dapat dihitung dengan menggunakan model MARS.
4. Mengelompokkan fungsi basis berdasarkan variabel prediktor yang masuk dalam model.
5. Menginterpretasikan *variable importance* yang berpengaruh terhadap variabel respon.
6. Perhitungan ketepatan klasifikasi klasifikasi lama studi mahasiswa Matematika FMIPA UNRAM angkatan 2014-2017 berdasarkan lulus tepat waktu dan lulus tidak tepat waktu dilakukan menggunakan nilai *Apparent Error Rate* (APER).
7. Menarik kesimpulan.

4. Hasil dan Pembahasan

Langkah awal sebelum melakukan analisis MARS terlebih dahulu dilakukan standarisasi terhdap nilai pada setiap variabel menjadi nilai normal baku. Standarisasi dilakukan karena berbedanya skala nilai variabel respon dan beberapa variabel prediktor.

Selanjutnya, pemeriksaan multikolinearitas dilakukan untuk mengetahui ada tidaknya korelasi antar variabel independen dalam model regresi. Pemeriksaan ini dilakukan dengan melihat VIF, dengan nilai VIF masing-masing variabel prediktor sebagai berikut.

Tabel 1 – Nilai VIF Variabel Prediktor

Variabel	VIF
(X_1)	3,413
(X_2)	1,098
(X_3)	1,094
(X_4)	1,992
(X_5)	2,347
(X_6)	2,139
(X_7)	1,600
(X_8)	1,779

Berdasarkan tabel 5.1 diperoleh nilai VIF masing-masing variabel prediktor < 10. Maka berdasarkan pengambilan keputusan dalam uji multikolinearitas dapat disimpulkan bahwa tidak terjadi gejala multikolinearitas dalam model regresi.

Untuk memodelkan data dengan menggunakan metode MARS dapat dibentuk berdasarkan kombinasi antara Fungsi Basis (FB), Maksimum Interaksi (MI) dan Minimum Observasi (MO). Fungsi basis digunakan untuk menjelaskan hubungan antara variabel respon dan variabel prediktor. Data yang digunakan pada pembentukan model MARS adalah data *training* sebanyak 90% atau 89 data. Variabel prediktor yang digunakan dalam penelitian ini berjumlah 8 variabel, sehingga banyaknya fungsi basis yang digunakan adalah 16, 24 dan 32.

Selanjutnya, maksimum interkasi (MI) menunjukkan banyaknya interaksi yang terjadi di dalam model, jumlah MI yang digunakan dalam penelitian ini adalah 1, 2 dan 3. Minimum observasi (MO) merupakan jarak minimum antara knot, MO yang digunakan dalam penelitian ini adalah 0, 1, 2 dan 3 sesuai dengan yang disarankan oleh Friedman. Pemodelan MARS dilakukan dengan cara *trial and error* dengan cara mengkombinasikan nilai Fungsi Basis (FB), Maksimum Interkasi (MI) dan Minimum Obsevasi (MO) sehingga didapatkan nilai terbaik berdasarkan nilai *Generalized Cross Validation* (GCV) minimum.

Pada Tabel 2 diberikan hasil pemodelan MARS dengan menggunakan kriteria GCV yang diselesaikan menggunakan bantuan *software Salford Predictive Modeler* (SPM) *7.0* sebagai berikut.

Tabel 2 – Hasil Pemodelan MARS dengan Kriteria GCV

No	BF	MI	MO	GCV
1	16	1	0	0,88265
2	16	1	1	0,81781
3	16	1	2	0,89580
4	16	1	3	0,89271
5	16	2	0	0,79558
6	16	2	1	1,43643
****7**	**16**	**2**	**2**	**0,77036**
8	16	2	3	0,83050
9	16	3	0	0,79429
10	16	3	1	1,01265
11	16	3	2	0,81088
12	16	3	3	0,82738
13	24	1	0	0,89014
14	24	1	1	0,82113
15	24	1	2	0,83989
16	24	1	3	0,82333
17	24	2	0	0,83794
18	24	2	1	0,99359
19	24	2	2	0,79581
20	24	2	3	0,83440
21	24	3	0	0,79858
22	24	3	1	0,83231
23	24	3	2	0,84794
24	24	3	3	0,96828
25	32	1	0	0,82786
26	32	1	1	0,82319
27	32	1	2	0,84427
28	32	1	3	0,91573
29	32	2	0	0,84483
30	32	2	1	0,90973
31	32	2	2	0,79517
32	32	2	3	0,83633
33	32	3	0	0,87752
34	32	3	1	0,82789
35	32	3	2	0,81962
36	32	3	3	0,96471

Keterangan:
** merupakan model terbaik

Berdasarkan Tabel 5.2 model MARS terbaik adalah model yang menghasilkan nilai GCV minimum. Nilai GCV dihasilkan dari mengkombinasikan jumlah fungsi basis (BF), maksimum interaksi (MI) dan minimum observasi (MO) secara *trial and error* sehingga dari 36 model yang terbentuk dihasilkan model terbaiknya yaitu dari kombinasi BF = 16, MI = 2 dan MO = 2 karena memiliki nilai GCV minimum sebesar 0,77036. Sehingga model MARS terbaik pada pendugaan lama studi mahasiswa Program Studi Matematika FMIPA UNRAM adalah sebagai berikut.

$$f(x) = -0{,}57679 + 0{,}150183^{*}BF_{10}$$

dengan,
$BF_1 = (ZX_8 + 1{,}60736)_+$
$BF_{10} = (ZX_7 + 2{,}08643)_+ * BF_1$

Dari model terbaik yang dihasilkan, dapat disimpulkan bahwa variabel-variabel prediktor yang mempengaruhi lama studi mahasiswa menggunakan metode MARS dengan nilai GCV minimum adalah Nilai Mata Kuliah Wajib Teori Peluang (X_7) dan Nilai Mata Kuliah Wajib Analisis Real I (X_8).

Setelah didapatkan model MARS terbaik maka dilakukan Pengujian signifikansi model MARS untuk mengevaluasi kecocokan model.

Pengujian dilakukan dengan menguji koefisien regresi secara simultan maupun secara parsial sebagai berikut:

1. Pengujian koefisien regresi simultan
   Hipotesis yang digunakan sebagai berikut:
   $H_0 : a_1 = 0$ (model tidak signifikan)
   H_1 :minimal terdapat satu $a_m \neq 0$; $m = 10$ (model signifikan)
   Statistika uji:
   $F_{hitung} = 29{,}4208$
   $F_{\frac{\alpha}{2}(M;N-M-1)} = F_{0,025(1;89-1-1)} = 5{,}2023$
   Kesimpulan:
   Berdasarkan hasil perhitungan di atas maka H_0 ditolak karena $F_{hitung} > F_{tabel}$. Maka dapat disimpulkan model signifikan.
2. Pengujian koefisien regresi parsial
   Hipotesis yang digunakan sebagai berikut:
   $H_0 : a_m = 0$ (koefisien a_m tidak berpengaruh terhadap model)
   $H_1 : a_m \neq 0$; untuk setiap m, dimana $m = 10$ (koefieien a_m berpengaruh terhadap model)
   Statistika uji:
   $t_{hitung}\left(a_{m\,10}\right) = 5{,}4241$
   $t_{\left(\frac{\alpha}{2};N-M\right)} = t_{(0,025;\,88)} = 1{,}98$
   Kesimpulan:
   Berdasarkan nilai t_{hitung} yang diperoleh maka keputusan tolak H_0, karena nilai untuk semua $|t_{hitung}| > t_{tabel}$ sehingga kesimpulannya semua koefisien a_m berpengaruh terhadap model.

Interpretasi MARS pada persamaan terbaik tersebut sebagai berikut.
$BF_{10} = (ZX_7 + 2{,}08643)_+ * BF_1$
$BF_1 = (ZX_8 + 1{,}60736)_+$
Dengan koefisien 0,150183 artinya bahwa setiap kenaikan nilai mata kuliah wajib teori peluang dan analisis real I dalam BF_{10} sebesar satu satuan akan meningkatkan lama studi mahasiswa sebesar 0,150183 pada FMIPA UNRAM dengan nilai mata kuliah wajib teori peluang lebih dari -2,08643 dan analisis real I lebih dari -1,60736.

Tingkat kepentingan variabel prediktor (*variable importance*) yang berpengaruh tersaji pada tabel 3 sebagai berikut.

Tabel 3 - *Variable importance*

Nama Variabel	*Variable Importance*
Nilai Mata Kuliah Wajib Teori Peluang	87,8307
Nilai Mata Kuliah Wajib Analsis Real I	85,5814
Jenis Kelamin	0,00000
Jalur Masuk	0,00000
Mata Kuliah Wajib Struktur Aljabar	0,00000
Nilai Mata Kuliah Wajib Kalkulus Lanjut	0,00000
Nilai Mata Kuliah Wajib Program Linier	0,00000
IPK	0,00000

Berdasarkan tabel 3 menunjukkan bahwa variabel yang memberikan pengaruh dominan terhadap masa studi mahasiswa Program Studi Matematika FMIPA UNRAM adalah Nilai Mata Kuliah Wajib Teori Peluang (X_7) dengan skor 87,8307%. Kemudian variabel Nilai Mata Kuliah Wajib Analisis Real I (X_8) dengan skor 85,5814%. Sedangkan variabel lainnya tidak memberikan pengaruh apapun karena skornya adalah 0%.

Dari model MARS terbaik yang diperoleh, maka akan dilakukan klasifikasian untuk mengetahui seberapa baik model tersebut. Prosedur umum yang digunakan untuk menghitung ketepatan klasifikasi dengan menggunakan alat ukur yang bernama *Apparent Error Rate* (APER). Nilai APER menyatakan representasi proporsi sampel yang salah diklasifikasikan oleh fungsi klasifikasi (Johnson dan Wichern, 1992).

Data yang digunakan untuk menghitung klasifikasi adalah data *testing* sebanyak 10% atau 10 data. Berikut ini adalah hasil klasifikasi yang disajikan pada tabel 4.

Tabel 4 – Hasil Klasifikasi

Kelas Aktual	**Kelas Prediksi** Kelas 1 (Lulus tepat waktu)	Kelas 0 (Lulus tidak tepat waktu)	**Total Aktual**
Kelas 1 (Lulus tepat waktu)	1	3	4
Kelas 0 (Lulus tidak tepat waktu)	1	5	6
Total Prediksi	2	8	10

Berdasarkan tabel 4 maka dapat dihitung nilai kesalahan dan ketepatan klasifikasi lama studi mahasiswa Program Studi Matematika FMIPA UNRAM angkatan 2014-2017 yaitu:

a. Nilai APER (kesalahan klasifikasi)
Nilai APER yang diperoleh adalah:

$$APER = \frac{n_{12} + n_{21}}{N} \times 100\%$$
$$APER = \frac{3+1}{10} \times 100\%$$
$$= 40\ \%$$

Nilai APER sebesar 40% menunjukkan persentase objek yang salah diklasifikasikan (*missclassification*). Karena nilai APER masih dibawah 50%, maka ketepatan hasil klasifikasi lama studi mahasiswa masih dapat diterima dan digunakan untuk mengklasifikasikan lama studi mahasiswa Program Studi Matematika FMIPA UNRAM (Annur, dkk., 2015).

b. Ketepatan Klasifikasi
Ketepatan klasifikasi yang diperoleh adalah:

$$100\% - 40\% = 60\%$$

Hal ini menunjukkan bahwa ketepatan klasifikasi lama studi mahasiswa Program Studi Matematika FMIPA UNRAM angkatan 2014-2017 berdasarkan lulus tepat waktu dan lulus tidak tepat waktu adalah sebesar 60%.

5. Penutup

Berdasarkan analisis yang dilakukan, dapat disimpulkan bahwa variabel prediktor yang berpengaruh terhadap lama studi mahasiswa yaitu Nilai Mata Kuliah Wajib Teori Peluang (X_7) dan Nilai Mata Kuliah Wajib Analisis Real I (X_8) dan ketepatan klasifikasi lama studi mahasiswa Program Studi Matematika berdasarkan status lulus tepat waktu atau lulus tidak tepat waktu adalah sebesar 60%.

DAFTAR PUSTAKA

Amin, M. F. (2014). https://budosen.id, *diakses* pada pukul 14.00 WITA, tanggal 20/12/2021.

Anam, S., Sugiman, & Sunarmi. (2017). Ketepatan Klasifikasi Dengan Menggunakan Metode *Multivariate Adaptive Regression Spline* (MARS) Pada Data Kelompok Rumah Tangga Kabupaten Cilacap. *UNNES Journal of Mathematics*, 6(1), 92-101.

Annur, M., Dahlan, J. A., & Agustina, F. (2015). Penerapan Metode *Multivariate Adaptive Regression Spline* (MARS) Untuk Menentukan Faktor Yang Mempengaruhi Masa Studi Mahasiswa FPMIPA UPI. *Eureka Matika*, 3(1), 135-155.

Cox, D. R., & Snell, E. J. (1989). *Analysis of Binary Data*. London: Chapman and Hall.

Eubank, R. (1999). *Nonparametric Regression and Spline Smoothing*. New York: Marcel Dekker.

Febriyanti, A., Yozza, H., & HG, I. R. (2013). Penerapan Metode *Multivariate Adaptive Regression Spline* (MARS) Untuk Mengidentifikasi Komponen Yang Berpengaruh Terhadap Peningkatan Akreditas Sekolah. *Jurnal Matematika UNPAD*, 2(2), 44-53.

Friedman, J. H. (1991). *Multivariate Adaptive Regression Spline*. The Annals of Statistics, Vol. 19 No. 1.

Ghofar, R. Y., Safitri, D., & Rusgiyono, A. (2014). Klasifikasi Kelulusan Mahasiswa Fakultas dan Matematika Universitas Diponegoro Menggunakan *Multivariate Adaptive Regression Spine* (MARS). *Jurnal Gaussian*, 3(4), 839-848.

Hardle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied Logistic Regression*. New York: Wiley.

Johnson, R. A., & Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis*. Prentice Hall, New Jersey.

Nisa' & Budiantara, N. (2012). Analisis Survival dengan Pendekatan *Multivariate Adaptive Regression Spline* pada Kasus Demam Berdarah Dengue (DBD). *Jurnal Sains dan Seni*, 1(1), 318-328.

Oktora, S. I. (2016). Analisis *Multivariate Adaptive Regression Splines* (MARS) Pada Prediksi Ketertinggalan Kabupaten Tahun 2014. *Jurnal Aplikasi Statistika & Komputasi Statistik*, 7(2), 115-128.

Parmadani, T. S., & Latifah, L. (2016). Pengaruh Minat Baca, Sumber Belajar dan Lingkungan Teman Sebaya Terhadap Prestasi Belajar Ekonomi. *Economics Educations Analysis Journal*, 4(2), 496-508.

Rofiq, A., Wuryandari, T., & Rahmawati, R. (2016). Perbandingan Analisis Diskriminan Fisher dan Naïve Bayes Untuk Klasifikasi Risiko Kredit. *Jurnal Gaussian*, 5(1), 1-10.

Wicaksono, W., Wilandari, Y., & Suparti. (2014). Pemodelan *Multivariate Adaptive Regression Spline* (MARS) Pada Faktor-Faktor Resiko Angka Kesakita Diare. *Jurnal Gaussian*, 3(2), 253-262.

Wulandari, I. D. A. M. I., & Budiantara, I. N. (2014). Analisis Faktor-Faktor yang Mempengeruhi Persentase Miskin dan Pengeluaran Perkapita Makanan di Jawa Timur Memggunakan Regresi Nonparametrik Birespon Spline. *Jurnal Sains dan Seni POMITS*, 3(1), 30-35.

Zhang, H., & Singer, B. H. (2010). *Recursive Partitioning and Application 2nd Edition*. New York: Springer.