

# MEMODELKAN FAKTOR-FAKTOR YANG MEMPENGARUHI GIZI BURUK BALITA DENGAN METODE *MULTIVARIATE ADAPTIVE REGRESSION SPLINE (MARS)*

Irwan<sup>i</sup>, Mawar Wardani<sup>ii</sup>

<sup>i</sup> Prodi Matematika FST, UINAM, irwan.msi@uin-alauddin.ac.id

<sup>ii</sup> Program Studi Matematika-FST, UINAM

**ABSTRAK**, *Multivariate Adaptive Regression Spline (MARS)* merupakan suatu kombinasi antara *Recursive Partitioning Regression (RPR)* dan metode *Spline* yang mampu mengolah data berdimensi tinggi dan berukuran besar serta mampu mengolah data dengan variabel respon kontinu ataupun biner. MARS akan membangun suatu model terbaik sebagai model klasifikasi yang melibatkan beberapa fungsi basis yang memuat variabel prediktor yang berpengaruh. Oleh karena itu penelitian ini dilakukan untuk mengetahui klasifikasi dan model dari faktor-faktor yang mempengaruhi terjadinya gizi buruk balita di Provinsi Sulawesi Selatan. Hasil yang diperoleh yaitu variabel prediktor yang paling berpengaruh terhadap gizi buruk balita adalah pemberian kapsul vitamin A, berat badan, inisialisasi menyusui dini dan umur dengan hasil klasifikasi bersifat stabil secara statistik berdasarkan statistik uji *Press's Q*. Model klasifikasi MARS terbaik untuk data gizi buruk balita dihasilkan dari kombinasi BF=18, MI=2 dan MO=1.

**Kata Kunci:** *Recursive Partitioning Regression (RPR)*, *Spline*, *Multivariate Adaptive Regression Spline (MARS)*, *Gizi Buruk*

## 1. PENDAHULUAN

Sumber daya manusia yang sehat dan berkualitas merupakan modal utama atau investasi dalam pembangunan kesehatan. Gizi merupakan penentu kualitas sumber daya manusia. Gizi buruk dapat dipandang sebagai suatu kondisi dimana kurangnya asupan nutrisi yang diperlukan oleh tubuh. Adanya kekurangan asupan nutrisi dapat menimbulkan berbagai macam efek timbulnya penyakit [1]. Berbagai penelitian telah membuktikan bahwa ada hubungan yang sangat erat antara kematian bayi dengan kekurangan gizi. Keadaan gizi yang buruk bisa menurunkan daya tahan anak sehingga anak rentan terserang penyakit dan bisa berujung pada kematian. Badan Kesehatan Dunia (WHO) memperkirakan bahwa 54% kematian bayi dan anak dilatar belakangi keadaan gizi yang buruk [2]. Dari berbagai dampak negatif yang telah dipaparkan di atas maka pengetahuan

mengenai faktor-faktor yang mempengaruhi gizi buruk balita sangatlah diperlukan. Salah satu metode dalam statistik yang dapat digunakan untuk memodelkan faktor-faktor yang mempengaruhi gizi buruk balita adalah dengan menggunakan metode *Multivariate Adaptive Regression Spline (MARS)*.

*Multivariate Adaptive Regression Spline (MARS)* merupakan salah satu pemodelan regresi *nonparametric* yang dapat digunakan pada data yang memiliki sebaran tidak normal. MARS juga dapat digunakan untuk klasifikasi dan untuk mengatasi permasalahan data berdimensi tinggi. MARS dapat digunakan untuk mencari model terbaik dari faktor-faktor yang mempengaruhi gizi buruk serta mengklasifikasikannya berdasarkan faktor-faktor yang telah diperoleh yang mengandung variabel-variabel yang jenisnya berbeda (kontinu dan diskrit), yakni dengan menggunakan beberapa fungsi basis *spline*.

Beberapa penelitian sebelumnya oleh Azzikra dan kawan-kawan mengkaji metode *Multivariate Adaptive Regression Spline (MARS)* untuk mengidentifikasi komponen apa saja yang mempengaruhi pengelompokan akreditasi sekolah [3] dan penelitian lain oleh Nurul dengan metode yang sama meramalkan indeks enso dan hujan bulanan [4].

## 2. TINJAUAN PUSTAKA

Analisis regresi adalah suatu alat statistik yang dapat digunakan untuk melihat hubungan sebab akibat. Dalam analisis regresi terdapat peubah bebas dan peubah tak bebas. Berdasarkan banyak peubah bebas yang digunakan, analisis regresi linier dibagi menjadi dua yaitu analisis regresi linier sederhana dan analisis regresi linear berganda [5].

**REGRESI NONPARAMETRIC**

Regresi *nonparametric* adalah salah satu metode yang digunakan untuk mengestimasi pola hubungan antara variabel respon dan variabel prediktor dimana bentuk kurva regresinya tidak diketahui. Secara umum, regresi *nonparametric* dapat dimodelkan sebagai berikut:

$$Y_i = f(X_i) + \varepsilon_i, \quad i = 1, 2, \dots, n.$$

dengan  $Y_i$  merupakan variabel respon pada pengamatan ke- $i$ ,  $X_i$  variabel prediktor pada pengamatan ke- $i$ ,  $\varepsilon_i$  merupakan *error* random pada pengamatan ke- $i$  dengan mean nol dan varians  $\sigma^2$ . Dan  $f$  merupakan kurva regresi yang tidak diketahui bentuknya, hanya diasumsikan *smooth* (mulus) dalam arti termuat dalam suatu ruang fungsi tertentu sehingga model regresi *nonparametric* lebih fleksibel [6].

**MULTIVARIATE ADAPTIVE REGRESSION SPLINE (MARS)**

Friedman menyatakan bahwa model *MARS* adalah kombinasi yang kompleks antara metode *spline* dengan rekursif partisi untuk menghasilkan estimasi fungsi regresi yang kontinu. Model *MARS* hasil dari kombinasi kompleks antara RPR dan pendekatan *spline* diperoleh sebagai berikut:

$$y_i = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{k_m} (S_{km}(X_{v(k,m)} - t_{km})) + \varepsilon_i$$

dengan,

- $a_0$  : konstanta regresi dari fungsi basis
- $a_m$  : koefisien dari fungsi basis ke- $m$ ,  $m = 1, \dots, M$
- $M$  : maksimum fungsi basis (nonconstant fungsi basis)
- $k_m$  : derajat interaksi
- $S_{km}$  : +1; jika knot terletak dikanan subregion  
-1; jika knot terletak dikiri subregion
- $X_{v(km)}$ : variabel prediktor
- $t_{km}$  : nilai knots dari variabel prediktor  $X_{v(km)}$
- $\varepsilon_i$  : *error* acak yang bersifat independen

Pemilihan knots pada *MARS* dilakukan dalam dua tahap yaitu tahap *forward* dan tahap *backward*. Ukuran kontribusi pada tahap *backward* ditentukan berdasarkan kriteria *Generalized Cross Validation* (GCV). Fungsi GCV didefinisikan sebagai berikut :

$$GCV(M) = \frac{MSE}{[1 - (C(M))/N]^2}$$

Dengan,

MSE : *Mean square Error*

- $M$  : jumlah fungsi basis
- $x_i$  : variabel prediktor ke- $i$
- $y_i$  : variabel respon ke- $i$
- $N$  : banyaknya pengamatan
- $C(M)$  : Trace  $[B(B^T B)^{-1} B^T] + 1$ .

**KLASIFIKASI MARS**

Hosmer dan Lemeshow menyatakan bahwa langkah awal klasifikasi dari variabel respon biner adalah menentukan titik potong. Variabel respon yang memiliki dua kategori (biner) dapat digunakan titik potong sebesar 0,50 dengan ketentuan jika  $\pi(x) \geq 0,50$ , maka hasil prediksi adalah 1 dan jika  $\pi(x) \leq 0,50$  maka hasil prediksinya adalah 0. Model *MARS* untuk respon biner adalah sebagai berikut [7]:

$$\begin{aligned} \text{logit } \pi(x) &= \ln \left( \frac{\pi(x)}{1 - \pi(x)} \right) \\ &= a_0 \\ &+ \sum_{m=1}^M a_m \prod_{k=1}^{k_m} (S_{km}(X_{v(k,m)} - t_{km})) \end{aligned}$$

dan dalam bentuk matriks dapat ditulis:

$$\text{logit } \pi(x) = \mathbf{B} \mathbf{a}$$

dengan

$$\mathbf{a} = (a_0, a_1, \dots, a_M)^T$$

$$\mathbf{B} = \begin{pmatrix} \prod_{k=1}^{k_m} (S_{1m}(X_{1(1,m)} - t_{1m})) & \dots & \prod_{k=1}^{k_M} (S_{Mm}(X_{1(M,m)} - t_{Mm})) \\ 1 & \dots & 1 \\ \prod_{k=1}^{k_m} (S_{1m}(X_{2(1,m)} - t_{1m})) & \dots & \prod_{k=1}^{k_M} (S_{Mm}(X_{2(M,m)} - t_{Mm})) \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \\ \prod_{k=1}^{k_m} (S_{1m}(X_{n(1,m)} - t_{1m})) & \dots & \prod_{k=1}^{k_M} (S_{Mm}(X_{n(M,m)} - t_{Mm})) \end{pmatrix}$$

**KETEPATAN KLASIFIKASI MARS**

Untuk menghitung ketepatan klasifikasi pada hasil pengelompokan digunakan *apparent error rate* (APER). Penentuan kesalahan klasifikasi untuk data respon biner dapat dihitung dengan matriks konfusi (*confussion matrix*). Nilai APER dihitung sebagai berikut:

$$APER(\%) = \frac{n_{11} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}} \times 100\%$$

Sebagai statistik uji untuk mengetahui sejauh mana kelompok-kelompok ini dapat dipisahkan dengan menggunakan variabel yang ada mempunyai kestabilan dalam ketepatan

klasifikasi digunakan *Press's Q*, yang diformulasikan sebagai berikut:

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

dimana:

N = jumlah total sampel

n = jumlah individu yang tepat diklasifikasikan

K = jumlah kelompok

### GIZI BURUK

Gizi buruk dapat dipandang sebagai suatu kondisi dimana kurangnya asupan nutrisi yang diperlukan oleh tubuh. Adanya kekurangan asupan nutrisi dapat menimbulkan berbagai macam efek timbulnya penyakit. Untuk mewujudkannya maka memberikan nutrisi yang cukup dan baik dimulai dari masa kanak-kanak sehingga bisa tumbuh kembang dengan sempurna, sehat dan cerdas yang membuat mereka mudah untuk mewujudkan cita-citanya. Ketidacacuhan orang tua terhadap nutrisi anak akan membuat keadaan gizi mereka menjadi buruk.

### 3. METODOLOGI

Jenis data yang digunakan adalah jenis data sekunder berupa data gizi buruk balita dari Januari 2015 sampai Desember 2015 yang diambil di Dinas Kesehatan Provinsi Sulawesi Selatan. Data sekunder merupakan sumber data penelitian yang diperoleh oleh peneliti secara tidak langsung melalui media perantara. Data sekunder umumnya berupa bukti, catatan atau laporan historis yang telah tersusun dalam arsip yang dipublikasikan dan tidak dipublikasikan.

### VARIABEL PENELITIAN

Variabel yang digunakan dalam penelitian ini yaitu:

Y = Balita yang berstatus gizi buruk (0);

Balita yang tidak berstatus gizi buruk (1)

X1= Jenis Kelamin

X2= Penyakit Penyerta

X3= Umur

X4= Berat Badan

X5= Inisiasi Menyusui Dini (IMD)

X6= Pemberian Kapsul Vitamin A

### Prosedur Analisis

Tahapan penelitian pada penelitian ini adalah:

1. Mendeskripsikan data dan membuat diagram pencar
2. Pembentukan model *MARS* dengan langkah-langkah sebagai berikut:
  - Menentukan kemungkinan maksimum banyaknya basis function (BF) yaitu 12, 18 dan 24.
  - Menentukan jumlah maksimum interaksi (MI), yaitu 1, 2 dan 3.
  - Menentukan minimal jumlah pengamatan setiap knots (MO), yaitu 0 dan 1
3. Menetapkan dan mendapatkan model *MARS* terbaik dengan melihat nilai GCV terkecil.
4. Melakukan uji signifikansi fungsi basis model *MARS* yaitu:
  - Pengujian parameter model secara simultan
  - Pengujian parameter model secara parsial
5. Menginterpretasikan model gizi buruk dan variabel-variabel yang berpengaruh di dalam model tersebut serta tingkat kepentingan masing-masing variable prediktor yang berpengaruh.
6. Mengklasifikasi ke dalam bentuk matriks konfusi (*confussion matrix*).
7. Menguji keakurasian prediksi model *MARS* (ketepatan klasifikasi) yang terbentuk dari data dengan menggunakan *APER*. Menghitung kestabilan klasifikasi dengan statistik uji *Press's Q*.

### 4. PEMBAHASAN

Tahapan pembentukan model *MARS* dilakukan dengan mencobakan semua nilai BF, MI dan MO. Dalam hal ini, nilai BF yang digunakan adalah 2-4 kali variable predictor yaitu 12, 18, 24. Nilai MI yang digunakan adalah 1, 2 dan 3 karena menurut Friedman jika  $MI > 3$  maka hasil akan semakin kompleks dan sulit untuk diinterpretasikan. Sedangkan nilai MO yang digunakan adalah 0 dan 1. Hasil *trial and error* dalam pembentukan model *MARS* disajikan dalam tabel berikut.

**Tabel 1.** *Trial and Error* Pembentukan Model MARS

BF	MI	MO	GCV	Variabel yang Masuk Model
12	1	0	0.13117	X <sub>2</sub> , X <sub>4</sub> , X <sub>5</sub> , X <sub>6</sub>
12	2	0	0.13261	X <sub>3</sub> , X <sub>4</sub> , X <sub>5</sub> , X <sub>6</sub>
12	3	0	0.13766	X <sub>4</sub> , X <sub>5</sub> , X <sub>6</sub>
12	1	1	0.13117	X <sub>2</sub> , X <sub>4</sub> , X <sub>5</sub> , X <sub>6</sub>
12	2	1	0.13257	X <sub>3</sub> , X <sub>4</sub> , X <sub>5</sub> , X <sub>6</sub>
12	3	1	0.13753	X <sub>4</sub> , X <sub>5</sub> , X <sub>6</sub>
18	1	0	0.14027	X <sub>4</sub> , X <sub>5</sub> , X <sub>6</sub>
18	2	0	0.12997	X <sub>3</sub> , X <sub>4</sub> , X <sub>5</sub> , X <sub>6</sub>
18	3	0	0.13152	X <sub>4</sub> , X <sub>5</sub> , X <sub>6</sub>
18	1	1	0.13140	X <sub>4</sub> , X <sub>5</sub> , X <sub>6</sub>
<b>18</b>	<b>2</b>	<b>1</b>	<b>0.12991</b>	<b>X<sub>3</sub>, X<sub>4</sub>, X<sub>5</sub>, X<sub>6</sub></b>
18	3	1	0.13786	X <sub>4</sub> , X <sub>5</sub> , X <sub>6</sub>
24	1	0	0.13161	X <sub>4</sub> , X <sub>5</sub> , X <sub>6</sub>
24	2	0	0.13379	X <sub>4</sub> , X <sub>5</sub> , X <sub>6</sub>
24	3	0	0.13149	X <sub>4</sub> , X <sub>5</sub> , X <sub>6</sub>
24	1	1	0.14927	X <sub>5</sub> , X <sub>6</sub>
24	2	1	0.13214	X <sub>3</sub> , X <sub>4</sub> , X <sub>5</sub> , X <sub>6</sub>
24	3	1	0.13505	X <sub>4</sub> , X <sub>5</sub> , X <sub>6</sub>

Model MARS terbaik adalah model MARS dengan nilai GCV yang terkecil. Model MARS terbaik yang diperoleh dari hasil *trial and error* berdasarkan Tabel 1 adalah model MARS dengan kombinasi BF=8, MI=2 dan MO=1 sebagai berikut.

$$\hat{f}(x) = 0.222659 + 0.543709BF1 - 0.292086BF4 - 0.326626BF5 - 0.0130215BF7 + 0.0450897BF9$$

Dengan,

- BF1 = (X<sub>6</sub> = 0)
- BF4 = max (0, 4 - X<sub>4</sub>) BF1
- BF5 = (X<sub>5</sub> = 1)
- BF6 = (X<sub>5</sub> = 0)
- BF7 = max (0, X<sub>3</sub> - 17) BF6
- BF9 = max (0, X<sub>4</sub> - 1.8)

Pada model MARS yang telah diperoleh dilakukan uji signifikansi yang meliputi uji fungsi basis secara keseluruhan (secara simultan/bersama) dan uji masing-masing fungsi basis (secara parsial).

```

=====
MARS Regression: Training Data
=====
W: 270.00                                R-SQUARED: 0.53547
MEAN DEP VAR: 0.42593                    ADJ R-SQUARED: 0.52667
                                           UNCENTERED R-SQUARED = R-0 SQUARED: 0.73333

Parameter      Estimate      S.E.      T-Ratio      P-Value
-----
Constant       | 0.22266     0.06605    3.37129     0.00086
Basis Function 1 | 0.54371     0.04706   11.55425    0.00000
Basis Function 4 | -0.29209    0.06979   -4.18541    0.00004
Basis Function 5 | -0.32663    0.05021   -6.50511    0.00000
Basis Function 7 | -0.01302    0.00370   -3.51946    0.00051
Basis Function 9 | 0.04509     0.01280    3.52313     0.00050
-----
F-STATISTIC = 60.86320                    S.E. OF REGRESSION = 0.34083
P-VALUE = 0.00000                        RESIDUAL SUM OF SQUARES = 30.66760
[MDF,NDF] = [ 5, 264 ]                   REGRESSION SUM OF SQUARES = 35.35092
=====
    
```

**Gambar 1.** *Output* Pemodelan Data Menggunakan Metode MARS

Uji signifikansi yang dilakukan secara simultan menggunakan hipotesis berikut:

$$H_0 : \alpha_0 = \alpha_1 = \alpha_4 = \alpha_5 = \alpha_7 = \alpha_9 = 0$$

$$H_1 : \text{minimal ada satu } \alpha_j \neq 0, \quad j = 0, 1, 4, 5, 7 \text{ dan } 9.$$

Berdasarkan Gambar 1, hasil pengolahan MARS dengan menggunakan *software* SPM v8.0 diperoleh F hitung sebesar 60.86320 dengan *p-value* sebesar 0.00000. Karena *p-value* < 0.05 maka *H<sub>0</sub>* ditolak pada taraf signifikansi 0.05. Dengan begitu dapat disimpulkan bahwa model MARS yang dimiliki adalah model yang signifikan.

Uji yang dilakukan secara parsial terhadap setiap fungsi basis menggunakan hipotesis sebagai berikut:

$$H_0 : \alpha_j = 0$$

$$H_1 : \alpha_j \neq 0, \quad j = 0, 1, 4, 5, 7 \text{ dan } 9.$$

Karena seluruh estimator parameter yang ditunjukkan pada Gambar 1, yaitu  $\alpha_0, \alpha_1, \alpha_4, \alpha_5, \alpha_7$  dan  $\alpha_9$  mempunyai nilai *p-value* < 0.05 maka untuk seluruh estimator parameter, *H<sub>0</sub>* ditolak pada taraf signifikansi 0.05. Dengan begitu dapat disimpulkan bahwa konstanta dan koefisien fungsi basis 1, 4, 5, 7, dan 9 mempunyai pengaruh yang signifikan terhadap model.

Adapun kontribusi untuk masing-masing variabel terhadap model yang diperoleh adalah sebagai berikut.

**Tabel 2.** Besarnya Kontribusi Masing-masing Variabel

Variabel	Besarnya Kontribusi
X <sub>3</sub>	21,56 %
X <sub>4</sub>	58,18 %
X <sub>5</sub>	49,32 %
X <sub>6</sub>	100,00 %

Pengklasifikasian MARS dengan menggunakan matriks konfusi untuk status gizi buruk dengan prediksi yang digunakan sebanyak 52,02 % dari banyaknya data dan terdapat 270 data yang diprediksi. Adapun pengklasifikasiannya sebagai berikut.

**Tabel 3.** Matriks Konfusi untuk Status Gizi Balita

Actual Class	Predicted Class		Total Actual Class
	0	1	
0	<b>130</b>	25	155
1	17	<b>98</b>	115
Total Predicted	147	123	<b>270</b>

Untuk menguji tingkat kesalahan klasifikasi pada tabel matriks konfusi, digunakan *Apparent Error Rate* (APER) dan diperoleh hasil yaitu tingkat kesalahan klasifikasi sebesar 15,55 % dan tingkat ketepatan klasifikasi sebesar 84,44 %. Jika nilai uji statistic APER masih di bawah 50% maka ketepatan hasil klasifikasi masih dapat diterima dan digunakan untuk mengklasifikasikan status gizi balita.

Evaluasi selanjutnya adalah menguji kestabilan model dengan statistic uji *Press'sQ* dan diperoleh hasil uji *Press'sQ* sebesar 128,13. Nilai uji *Press'sQ* selanjutnya dibandingkan dengan nilai *chi-square* dan diperoleh nilai 3,841. Dari perhitungan hasil uji *Press'sQ* jauh lebih besar dibandingkan dengan nilai *chi-square* sehingga dapat dikatakan bahwa pengklasifikasian status gizi buruk menggunakan model MARS sudah baik dan konsisten secara statistik

### 5. KESIMPULAN

Dari hasil yang diperoleh dapat disimpulkan bahwa faktor-faktor yang paling berpengaruh terhadap gizi buruk balita di provinsi Sulawesi

Selatan melalui pendekatan *Multivariate Adaptive Regression Spline* adalah pemberian kapsul vitamin A, berat badan, inisiasi menyusui dini dan umur dengan hasil klasifikasi bersifat stabil secara statistik berdasarkan statistik uji *Press'sQ*. Model klasifikasi MARS terbaik untuk data gizi buruk balita dihasilkan dari kombinasi BF=18, MI=2 dan MO=1 dengan persamaan berikut:

$$\hat{f}(x) = 0.222659 + 0.543709BF1 - 0.292086BF4 - 0.326626BF5 - 0.0130215BF7 + 0.0450897BF9$$

Dengan BF1 = (X<sub>6</sub> = 0), BF4 = max (0, 4 - X<sub>4</sub>) BF1, BF5 = (X<sub>5</sub> = 1), BF6 = (X<sub>5</sub> = 0), BF7 = max (0, X<sub>3</sub> - 17) BF6 dan BF9 = max (0, X<sub>4</sub> - 1.8).

### 6. DAFTAR PUSTAKA

- [1] Hasdiana H.R, dkk., *Gizi Pemanfaatan Gizi, Diet dan Obesitas*, Yogyakarta: Nuha Medika, 2013
- [2] Lubis, Petti dan Irina Damayanti. "54% Kematian Bayi di Dunia Akibat Kurang Gizi," *Viva.Co.Id*, 12 Maret 2010. <http://life.viva.co.id/news/read/13611254> (17 November 2015).
- [3] Febriyanti, Azzikra. Vol. 2 No. 2. "Penerapan Metode *Multivariate Adaptive Regression Spline* (Mars) Untuk Mengidentifikasi Komponen Yang Berpengaruh Terhadap Peringkat Akreditasi Sekolah", Jurusan Matematika FMIPA UNAND, 2012.
- [4] Astuty, Nurul Yensy, Vol.5, No.2. "Penggunaan Regresi *Splines Adaptif Berganda* Untuk Peramalan Indeks Enso Dan Hujan Bulana', Program Studi Matematika FKIP Universitas Bengkulu, 2007.
- [5] Ayu, Ida Prasetya Utami, "Regresi Kuantil Median Untuk Mengatasi Heteroskedastisitas Pada Analisis.
- [6] Haslinda, Andi. *Multivariate Adaptive Regression Spline*. Cet. I; Makassar: Alauddin University Press, 2013.

- [7] Binadari, Ratih, "*Perbandingan Metode Regresi Logistik Biner dan Multivariate Adaptive Regression Spline (MARS) pada Peminatan Jurusan SMA*", Jurusan Statistik Universitas Diponegoro, 2015