

KLASIFIKASI STATUS RUMAH TANGGA DI PROVINSI BENGKULU MENGUNAKAN *MULTIVARIATE ADAPTIVE REGRESSION SPLINE* (MARS)

Esther Damayanti Sihombing¹, Idhia Sriliana^{2*}, Dyah Setyo Rini³,

^{1,2,3} Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Bengkulu

*e-mail : idhiasriliana@unib.ac.id

DOI: 10.14710/j.gauss.13.1.145-155

Article Info:

Received: 2023-12-24

Accepted: 2024-10-07

Available Online: 2024-10-09

Keywords:

Poverty; Classification; MARS

Abstract: Poverty is a global issue that captures the attention of governments in any country because it is a complex population-related problem. Poverty is a high-dimensional case, involving numerous predictor variables that interact with each other. This study was conducted to obtain a model that is capable of classifying household in Bengkulu Province. Multivariate Adaptive Regression Spline (MARS) is one of the methods used for classification of high-dimensional data. The MARS model is performed with combining Maximum Base Function (BF), Minimal Observation (MO), and Maximum Interaction (MI) with a small Generalized Cross Validation (GCV) by trial and error. The data used in this study is data from the 2022 National Socioeconomic Survey sourced from the Central Statistics Agency of Bengkulu Province. The variables used is the poverty status of households classified as poor and not poor households as a response variable as well as several predictor variables. The results of this study indicate that it produces a MARS model with a combination of Basis Function (BF) = 48, Maximum Interaction (MI) = 3, and Minimum Observation (MO) = 0 which has the minimum GCV criteria of 0.06799. The results of the accuracy evaluation of the classification obtained an accuracy of 91.65%.

1. PENDAHULUAN

Analisis regresi merupakan suatu metode statistika yang digunakan untuk menyelidiki dan memodelkan hubungan antar variabel. Analisis regresi memberikan penjelasan tentang pola hubungan antara variabel respon dan variabel prediktor (Montgomery, Peck, dan Vining, 2012). Ada beberapa pendekatan analisis regresi di antaranya yaitu pendekatan parametrik dan nonparametrik. Regresi parametrik digunakan apabila pola hubungan antara variabel respon dan prediktor diketahui secara jelas, sedangkan regresi nonparametrik digunakan apabila pola data tidak diketahui apakah memiliki pola linier, kuadrat, kubik, atau lainnya (Eubank, 1999).

Multivariate Adaptive Regression Spline (MARS) merupakan salah satu metode regresi nonparametrik yang dapat menyelesaikan masalah regresi dan klasifikasi. MARS merupakan kombinasi metode *spline truncated* dengan *recursive partitioning regression* (Friedman, 1991). Model MARS mengatasi kelemahan kedua metode tersebut sehingga menghasilkan model yang kontinu pada *knot* dan mampu mengidentifikasi adanya fungsi linear dan aditif. MARS dapat menghasilkan estimasi variabel respon yang lebih akurat dengan permasalahan data berdimensi tinggi yakni cukup banyaknya jumlah amatan dan variabel prediktor sebesar $3 \leq n \leq 20$ (Hastie, Tibshirani, dan Friedman, 2009). Penelitian dengan menggunakan MARS telah banyak dilakukan, beberapa di antaranya yaitu Tamonob (2020) menggunakan MARS untuk mengklasifikasikan status desa di Provinsi Nusa Tenggara Timur dengan kelompok data yang tidak seimbang menghasilkan akurasi sebesar 99.40%, sensitivitas sebesar 99.84%, dan spesifisitas sebesar 92.80%.

Berdasarkan hasil tersebut diketahui bahwa metode MARS dapat mengklasifikasikan dengan baik meskipun dalam kondisi data yang tidak seimbang. Penelitian oleh Surya (2016) membandingkan ketepatan klasifikasi metode regresi logistik biner dengan metode MARS pada status stroke pasien memberikan kesimpulan bahwa metode klasifikasi MARS memiliki performa dan akurasi pendugaan yang lebih baik.

Kemiskinan merupakan isu global yang menjadi perhatian pemerintah karena merupakan masalah kependudukan yang kompleks. Salah satu provinsi dengan tingkat kemiskinan yang cukup tinggi adalah Provinsi Bengkulu yakni sebesar 14.62% pada tahun 2022. Angka ini menunjukkan bahwa kemiskinan di Provinsi Bengkulu masih lebih tinggi dibandingkan dengan angka kemiskinan nasional. Berbagai program dan kebijakan pemerintah di berbagai bidang seperti bidang kesehatan, sosial, maupun bidang lainnya telah diupayakan untuk mengurangi rumah tangga berstatus miskin. Mengidentifikasi karakteristik rumah tangga miskin penting untuk dikaji, hal ini dapat menjadi acuan agar program pemerintah terlaksana sesuai target. Salah satu metode untuk menentukan karakteristik dari rumah tangga yang tergolong miskin dan tidak miskin dapat dilakukan dengan metode klasifikasi. Kemiskinan merupakan kasus berdimensi tinggi yaitu terdiri dari banyak variabel prediktor dan saling berinteraksi (Hickel, 2016), sehingga metode yang tepat yaitu dengan menggunakan MARS. Metode MARS merupakan penyempurnaan dari metode *recursive partitioning regression* dan *spline truncated*. Proses pembentukan model pada metode MARS tidak memerlukan asumsi dan menghasilkan model yang kontinu dalam titik *knot* berdasarkan *Generalized Cross Validation* (GCV).

2. TINJAUAN PUSTAKA

Multivariate Adaptive Regression Spline (MARS) merupakan salah satu metode regresi nonparametrik yang dapat menyelesaikan masalah regresi dan klasifikasi. MARS merupakan kombinasi metode *spline truncated* dengan *recursive partitioning regression*. MARS dapat digunakan pada permasalahan data berdimensi tinggi yakni cukup banyaknya jumlah amatan dan variabel prediktor sebesar $3 \leq n \leq 20$. Ada beberapa hal yang perlu dipertimbangkan saat menggunakan model MARS yaitu (Friedman, 1991):

- Knot*, adalah titik belakang dari satu garis regresi dan titik depan dari garis regresi lainnya. Diharapkan pada setiap titik *knot* antara satu *region* dengan *region* lainnya berkesinambungan atau kontinu. Minimum Observasi (MO) atau jarak minimum antar *knot* yang disarankan adalah 0, 1, 2, dan 3. MO yang lebih besar dari jumlah yang disarankan akan semakin mengurangi fleksibilitas model dan akurasinya.
- Basis function*, adalah fungsi yang menjabarkan hubungan antara variabel respon dan variabel prediktor. Basis fungsi terdapat di setiap *region*. Batas maksimal basis fungsi yaitu 2-4 kali dari total variabel prediktornya.
- Interaksi, adalah hubungan korelasi antar variabel. Maksimum Interaksi (MI) yang diperbolehkan adalah 1, 2, dan 3. Jika MI lebih dari 3, interpretasi model mungkin lebih rumit serta sulit diartikan.

Model umum MARS dituliskan sebagai berikut (Friedman, 1991):

$$f(x_i) = \beta_0 + \sum_{m=1}^M \beta_m \prod_{k=1}^{K_m} [S_{km}(x_{v(k,m)} - t_{km})]_+ \quad (1)$$

dengan:

- β_0 = koefisien basis fungsi induk
- β_m = koefisien basis fungsi ke- m
- M = maksimum basis fungsi
- K_m = maksimum derajat interaksi

s_{km} = bernilai ± 1 jika data berada di sebelah kanan atau kiri titik *knot*
 $x_{v(k,m)}$ = variabel prediktor ke- v
 t_{km} = nilai titik *knot* dari variabel prediktor

Model MARS berdasarkan fungsi regresi nonparametrik dinyatakan dalam persamaan berikut:

$$y_i = f(x_i) + \varepsilon_i = \beta_0 + \sum_{m=1}^M \beta_m \prod_{k=1}^{K_m} [s_{km}(x_{v(k,m)} - t_{km})]_+ + \varepsilon_i \quad (2)$$

Jika dijabarkan, Persamaan (2) dapat ditulis sebagai berikut:

$$\begin{aligned}
 y_i = & \beta_0 + \sum_{m=1}^M \beta_m [s_{km}(x_{v(k,m)} - t_{km})]_+ \\
 & + \sum_{m=1}^M \beta_m [s_{km}(x_{v(k,m)} - t_{km})]_+ [s_{km}(x_{v(k,m)} - t_{km})]_+ \\
 & + \sum_{m=1}^M \beta_m [s_{km}(x_{v(k,m)} - t_{km})]_+ [s_{km}(x_{v(k,m)} - t_{km})]_+ [s_{km}(x_{v(k,m)} - t_{km})]_+ \\
 & + \dots + \varepsilon_i
 \end{aligned} \quad (3)$$

Persamaan (3) pada basis fungsi di penjumlahan pertama merupakan basis fungsi yang melibatkan satu variabel. Penjumlahan kedua meliputi semua basis fungsi untuk interaksi yang melibatkan dua variabel. Penjumlahan ketiga meliputi semua basis fungsi untuk interaksi yang melibatkan tiga variabel dan seterusnya.

Persamaan (2) jika ditulis dalam bentuk matriks adalah sebagai berikut:

$$\mathbf{y} = \mathbf{B}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (4)$$

dengan $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_m)^T$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$

$$\mathbf{B} = \begin{bmatrix}
 1 & \prod_{k=1}^{K_1} [s_{km}(x_{v(k,m)} - t_{km})] & \dots & \prod_{k=1}^{K_M} [s_{km}(x_{v(k,m)} - t_{km})] \\
 1 & \prod_{k=1}^{K_1} [s_{km}(x_{v(k,m)} - t_{km})] & \dots & \prod_{k=1}^{K_M} [s_{km}(x_{v(k,m)} - t_{km})] \\
 \vdots & \vdots & \vdots & \vdots \\
 1 & \prod_{k=1}^{K_1} [s_{km}(x_{v(k,m)} - t_{km})] & \dots & \prod_{k=1}^{K_M} [s_{km}(x_{v(k,m)} - t_{km})]
 \end{bmatrix}$$

Persamaan (4) dalam melakukan estimasi parameter menggunakan metode kuadrat terkecil sehingga diperoleh estimator:

$$\hat{\boldsymbol{\beta}} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{Y} \quad (5)$$

Pemilihan model MARS terbaik dilakukan dengan dua pendekatan yaitu *forward stepwise* dan *backward stepwise*. Tahapan *forward stepwise* berfungsi untuk mendapatkan fungsi dengan jumlah basis maksimum (Zhang, 1997). Dalam membatasi model, maka digunakan maksimum banyaknya fungsi (Kriner, 2007). Meskipun telah dibatasi, tahap *forward* tetap memperoleh model basis fungsi yang sangat banyak. Dalam hal ini perlu dilakukan penghapusan sebagian basis fungsi, sehingga dapat menghasilkan model yang lebih sederhana. Menurut Zhang (1997), *backward stepwise* bertujuan untuk mendapatkan model yang lebih sederhana (prinsip parsemoni) dengan cara menghilangkan basis fungsi yang memiliki kontribusi kecil terhadap respon pada tahap *forward stepwise* dengan meminimumkan *Generalized Cross Validation* (Sheskin, 2011).

Menurut Friedman (1991) formula GCV adalah sebagai berikut:

$$GCV = \frac{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{f}_M(x_i))^2}{\left[1 - \frac{C(M)}{N}\right]^2} \quad (6)$$

di mana $C(M) = \text{trace}[\mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T] + 1$ dengan \mathbf{B} merupakan matriks dari M fungsi-fungsi *nonconstant* dengan ukuran $M \times N$. N merupakan banyaknya amatan.

Klasifikasi adalah pengelompokan ke dalam suatu kelas ataupun kategori yang sudah ditetapkan (Johnson dan Wichern, 1992). Klasifikasi merupakan proses penemuan model yang dapat menggambarkan dan membedakan kelas data. klasifikasi bertujuan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Han dan Kamber, 2006).

Menurut Cox dan Snell (1989), jika pada variabel respon terdiri dari dua nilai (biner), maka dikatakan sebagai regresi binary response. Salah satu metode nya adalah dengan regresi logistik biner. Klasifikasi pada MARS berdasarkan pendekatan analisis regresi logistik biner (Agresti, 2007), sehingga dapat digunakan model probabilitas dengan persamaan sebagai berikut:

$$\pi(x) = \frac{e^{\hat{y}}}{1 + e^{\hat{y}}} \quad (7)$$

dengan $\hat{y} = \beta_0 + \sum_{m=1}^M \beta_m \prod_{k=1}^{K_m} [s_{km}(x_{v(k,m)} - t_{km})]_+$

Menurut Hosmer dan Lemeshow (2000) untuk menetapkan klasifikasi variabel respon biner (1 dan 0) dapat dilakukan dengan menggunakan titik potong (*cut point*) sebesar 0.5 dengan ketentuan jika estimasi peluang > 0.5 maka hasil prediksi akan masuk ke kelompok 1, namun jika estimasi peluang kurang dari atau sama dengan 0.5 maka hasil prediksi akan masuk ke kelompok 0 (Cox & Snell, 2018).

Metode klasifikasi yang baik akan menghasilkan sedikit kesalahan klasifikasi atau akan menghasilkan peluang kesalahan klasifikasi yang kecil (Johnson dan Wichern 1992). Evaluasi kinerja (*performance*) klasifikasi dapat dilakukan dengan beberapa cara, di antaranya dengan menggunakan tabulasi silang (*confusion matrix*). *Confusion matrix* menyajikan data aktual dan data hasil prediksi dari sebuah model klasifikasi (Han dan Kamber, 2006). Bentuk dari *confusion matrix* disajikan pada Tabel 1.

Tabel 1. *Confusion Matrix* pada Klasifikasi Dua Kelas

Prediksi	Aktual	
	Kelompok 1	Kelompok 2
Kelompok 1	n_{11}	n_{12}
Kelompok 2	n_{21}	n_{22}

n_{11} : total amatan dari prediksi kelompok 1 yang diklasifikasikan dengan tepat sebagai kelompok 1

n_{12} : total amatan dari prediksi kelompok 1 yang diklasifikasikan secara keliru sebagai kelompok 2

n_{21} : total amatan dari prediksi kelompok 2 yang diklasifikasikan secara keliru sebagai kelompok 1

n_{22} : total amatan dari prediksi kelompok 2 yang diklasifikasikan dengan tepat sebagai kelompok 2

Menurut Johnson dan Wichern (1992), *Apparent Error Rate* (APER) adalah prosedur evaluasi yang digunakan untuk melihat kesalahan klasifikasi yang dilakukan oleh suatu fungsi klasifikasi. *Total Accuracy Rate* (TAR) digunakan untuk menghitung ketepatan klasifikasi pada hasil pengelompokan amatan. Nilai *Total Accuracy Rate* (TAR) dapat menyatakan representasi proporsi amatan yang tepat diklasifikasikan. Nilai *Apparent Error Rate* (APER) dan *Total Accuracy Rate* (TAR) didapatkan dengan perhitungan sebagai berikut:

$$APER (\%) = \frac{\text{jumlah prediksi salah}}{\text{jumlah total prediksi}} = \frac{n_{12} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}} \times 100\%$$

$$TAR (\%) = \frac{\text{jumlah prediksi tepat}}{\text{jumlah total prediksi}} = \frac{n_{11} + n_{22}}{n_{11} + n_{12} + n_{21} + n_{22}} \times 100\%$$

Pada tabel klasifikasi juga perlu dipertimbangkan nilai *sensitivity* yang menggambarkan akurasi amatan pada kelompok positif dan *specificity* yang menggambarkan akurasi amatan pada kelompok negatif (Agresti, 2007). Kemampuan mengukur *sensitivity* dan *specificity*, dengan baik menandakan bahwa metode klasifikasi tersebut sudah baik dalam memprediksi amatan tiap kelompok (Morton, Hebel, dan McCarter, 2008).

$$sensitivity (\%) = \frac{n_{11}}{n_{11} + n_{12}} \times 100\%$$

$$specificity (\%) = \frac{n_{22}}{n_{21} + n_{22}} \times 100\%$$

3. METODE PENELITIAN

Data yang digunakan dalam penelitian ini berupa data sekunder hasil Survei Sosial Ekonomi Nasional (SUSENAS) bulan Maret tahun 2022 bersumber dari Badan Pusat Statistik Provinsi Bengkulu. Objek pengamatan dalam penelitian ini adalah rumah tangga di Provinsi Bengkulu. Data tersebut terdiri dari variabel prediktor sebanyak 13 variabel dan 1 variabel respon. Variabel respon dalam penelitian ini berskala biner yaitu rumah tangga dengan status miskin dan rumah tangga dengan status tidak miskin. Variabel penelitian sesuai dengan Tabel 2 berikut:

Tabel 2. Variabel Penelitian

Variabel	Keterangan	Skala
Status Rumah Tangga (Y)	0 : Tidak Miskin 1 : Miskin	Nominal
Wilayah (X_1)	0 : Perkotaan 1 : Pedesaan	Nominal
Status Kepemilikan Rumah (X_2)	0 : Milik sendiri 1 : Kontrak/sewa 2 : Bebas sewa 3 : Dinas	Nominal
Bahan Bangunan Atap (X_3)	0 : Genteng 1 : Seng 2 : Asbes 3 : Bambu	Nominal
Bahan Utama Dinding Rumah (X_4)	0 : Tembok 1 : Plesteran anyaman 2 : Kayu/papan	Nominal

	3 : Anyaman bambu	
	4 : Batang kayu	
	5 : Lainnya	
Bahan Utama Lantai Rumah (X_5)	0 : Marmer/granit	
	1 : Keramik	
	2 : Teraso	Nominal
	3 : Kayu/papan	
	4 : Semen/bata merah	
	5 : Bambu	
	6 : Tanah	
Penggunaan Fasilitas BAB (X_6)	0 : Ada, digunakan hanya ART sendiri	
	1 : Ada, digunakan bersama rumah	Nominal
	2 : Ada, di MCK umum	
	3 : Ada, ART tidak menggunakan	
	4 : Tidak ada fasilitas	
Sumber Penerangan (X_7)	0 : Listrik PLN	
	1 : Listrik PLN tanpa meteran	Nominal
	2 : Listrik non PLN	
	3 : Bukan listrik	
Bahan Bakar Utama Sumber Memasak (X_8)	0 : Listrik	
	1 : Elpiji 5,5 kg/bluegaz	
	2 : Elpiji 12 kg	Nominal
	3 : Elpiji 3 kg	
	4 : Minyak tanah	
	5 : Kayu bakar	
Status Kepemilikan Tanah (X_9)	0 : Ya	
	1 : Tidak	Nominal
Luas Lantai (X_{10})	m^2	Rasio
Banyak ART (X_{11})	Cacah	Rasio
Jumlah keluarga (X_{12})	Cacah	Rasio
Banyaknya Konsumsi Kalori Perkapita Sehari (X_{13})	Kkal	Rasio

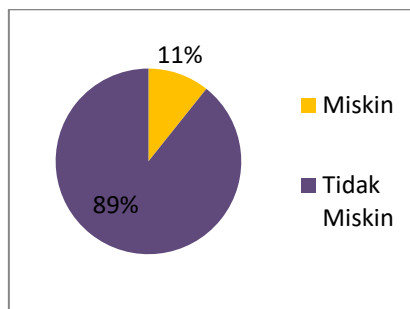
Analisis data dilakukan dengan langkah-langkah sebagai berikut:

1. Melakukan eksplorasi data untuk melihat karakteristik data secara umum
2. Identifikasi hubungan variabel respon dan prediktor
3. Membagi data menjadi data *training* dan data *testing*
4. Pembentukan dan estimasi model MARS dengan tahapan sebagai berikut:
 - a. Menetapkan jumlah basis fungsi (BF), basis fungsi yang diperbolehkan adalah 2-4 kali dari total variabel prediktor yang dimiliki
 - b. Menetapkan maksimum jumlah interaksi yaitu 1, 2, 3
 - c. Menetapkan minimum observasi (MO) yaitu 0, 1, 2, dan 3

- d. Melakukan estimasi model MARS berdasarkan Persamaan (5)
- e. Menentukan MARS terbaik menurut nilai *Generalized Cross Validation* (GCV) terkecil yang didapat dari kombinasi antara BF, MI, dan MO berdasarkan Persamaan (6)
- f. Menginterpretasi basis fungsi dari model MARS yang terpilih
- g. Mengalokasikan rumah tangga berdasarkan hasil klasifikasi terhadap kelompok status rumah tangga ke dalam bentuk tabel klasifikasi
- h. Menguji dan mengevaluasi hasil klasifikasi

4. HASIL DAN PEMBAHASAN

Survei Sosial Ekonomi Nasional (SUSENAS) Provinsi Bengkulu pada bulan Maret tahun 2022 ada sebanyak 5731 rumah tangga yang disurvei. Sebanyak 616 rumah tangga berstatus miskin dan 5115 rumah tangga berstatus tidak miskin. Rumah tangga berstatus miskin apabila pengeluaran konsumsi per kapita per bulan berada di bawah garis kemiskinan. Adapun garis kemiskinan di Provinsi Bengkulu pada bulan Maret tahun 2022 adalah Rp590.754,00. Gambar 4.1 menunjukkan bahwa terdapat perbedaan jumlah yang cukup besar di antara rumah tangga berstatus sangat miskin dan rumah tangga berstatus miskin. Persentase rumah tangga miskin sebesar 11% sedangkan untuk rumah tangga tidak miskin sebesar 89%.



Gambar 1. Persentase Kategori pada Peubah Respon

Sebelum data tersebut digunakan untuk melakukan analisis lebih lanjut dengan MARS, perlu diketahui terlebih dahulu hubungan antara variabel respon yang bersifat kategorik dengan variabel prediktor yang bersifat numerik. Hubungan antara variabel respon yang bersifat kategorik dengan variabel prediktor yang bersifat kategorik diuji menggunakan *chi-square* (χ^2). Sedangkan uji korelasi *point biserial* digunakan untuk mengidentifikasi hubungan variabel prediktor bersifat numerik dengan variabel respon kategorik. Hipotesis yang digunakan yaitu:

H_0 : Tidak terdapat hubungan antara dua variabel

H_1 : Terdapat hubungan antara dua variabel

Berikut hasil pengujian antara variabel respon dan prediktor.

Tabel 3. Uji Asosiasi Variabel Prediktor dengan Variabel respon

No	Variabel	<i>Chi-square</i> (χ^2)	r_{pb}	p_{value}	Kesimpulan
1	X_1	8,724	-	0,0031	Tolak H_0
2	X_2	19,347	-	0,0002	Tolak H_0
3	X_3	4,894	-	0,2984	Gagal Tolak H_0
4	X_4	43,649	-	$8,76 \times 10^{-9}$	Tolak H_0
5	X_5	195,31	-	$< 2,2 \times 10^{-16}$	Tolak H_0

6	X_6	111,57	-	$< 2,2 \times 10^{-16}$	Tolak H_0
7	X_7	76,201	-	$< 2,2 \times 10^{-16}$	Tolak H_0
8	X_8	72,335	-	$1,36 \times 10^{-13}$	Tolak H_0
9	X_9	13,307	-	0,0002	Tolak H_0
10	X_{10}	-	-0,114	$< 2,2 \times 10^{-16}$	Tolak H_0
11	X_{11}	-	0,204	$< 2,2 \times 10^{-16}$	Tolak H_0
12	X_{12}	-	0,092	$3,87 \times 10^{-12}$	Tolak H_0
13	X_{13}	-	-0,315	$< 2,2 \times 10^{-16}$	Tolak H_0

Setelah menguji hubungan antara masing-masing variabel prediktor dengan variabel respon, diperoleh bahwa ada sebanyak 12 variabel prediktor yang secara signifikan pada taraf 5% memiliki hubungan dengan variabel respon. Variabel prediktor yang berpengaruh secara signifikan diantaranya adalah $X_1, X_2, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}$, dan X_{13} . Data penelitian dibagi menjadi data *training* 75% dan data *testing* 25%. Selanjutnya dilakukan estimasi parameter dalam analisis dengan klasifikasi dengan menggunakan metode MARS.

Pembentukan model MARS diperoleh berdasarkan kombinasi Basis Fungsi (BF), Maksimum Interaksi (MI), dan Minimum Observasi (MO). Jumlah variabel prediktor yang signifikan adalah 12 variabel, sehingga nilai BF yang digunakan adalah 24, 36, dan 48 yaitu kelipatan 2-4 dari jumlah variabel prediktor yang signifikan (Friedman, 1991). Langkah selanjutnya yaitu menentukan model MARS secara *trial and error* berdasarkan nilai GCV terkecil.

Tabel 4. Kombinasi BF, MI, dan MO

BF	MI	MO	GCV	BF	MI	MO	GCV
24	1	0	0,0747218	36	2	2	0,06973
24	1	1	0,07472188	36	2	3	0,06973
24	1	2	0,0747218	36	3	0	0,06902
24	1	3	0,07472188	36	3	1	0,06889
24	2	0	0,07066869	36	3	2	0,06902
24	2	1	0,0706679	36	3	3	0,06902
24	2	2	0,07066869	48	1	0	0,07438
24	2	3	0,0706679	48	1	1	0,07439
24	3	0	0,07064533	48	1	2	0,07439
24	3	1	0,07064483	48	1	3	0,07439
24	3	2	0,07064533	48	2	0	0,06914
24	3	3	0,07064483	48	2	1	0,06928
36	1	0	0,07438757	48	2	2	0,06931
36	1	1	0,07438849	48	2	3	0,06931
36	1	2	0,07438849	48	3	0	0,06799
36	1	3	0,07438799	48	3	1	0,06807
36	2	0	0,06975823	48	3	2	0,06829
36	2	1	0,06973764	48	3	3	0,06822

Diperoleh model terbaik yaitu pada BF=48, MI=3, dan MO=0 yang memiliki nilai GCV terkecil yaitu 0.0679963. Maksimum basis fungsi yang telah ditetapkan sebanyak 48.

Setelah memperoleh basis fungsi optimum pada tahapan *backward stepwise*, diperolehlah GCV terkecil adalah pada subset yang berjumlah 26 dengan model sebagai berikut:

$$\begin{aligned} \hat{f}(x) = & (-9,56 \times 10^{-2}) + (-3,47 \times 10^{-5})BF2 + (6,54 \times 10^{-4})BF3 \\ & + (-2,49 \times 10^{-4})BF4 + (1,13 \times 10^{-5})BF6 + (-2,53 \times 10^{-4})BF7 \\ & + (-8,08 \times 10^{-6})BF9 + (-6,43 \times 10^{-2})BF11 + (5,08 \times 10^{-2})BF12 \\ & + (8,71 \times 10^{-2})BF14 + (-4,58 \times 10^{-4})BF15 + (-5,46 \times 10^{-2})BF17 \\ & + (-7,92 \times 10^{-2})BF18 + (-6,02 \times 10^{-3})BF19 + (1,14 \times 10^{-3})BF20 \\ & + (2,69 \times 10^{-1})BF22 + (7,59 \times 10^{-5})BF23 + (-1,18 \times 10^{-4})BF24 \\ & + (4,07 \times 10^{-2})BF25 + (3,53 \times 10^{-1})BF26 + (1,46 \times 10^{-4})BF27 \\ & + (1,13 \times 10^{-3})BF30 + (5,88 \times 10^{-3})BF29 + (-5,89 \times 10^{-2})BF31 \\ & + (8,26 \times 10^{-5})BF33 + (-3,99 \times 10^{-4})BF31 + (9,49 \times 10^{-4})BF32 \\ & + (-2,29 \times 10^{-4})BF33 + (7,11 \times 10^{-4})BF34 \end{aligned}$$

dengan

$$\begin{aligned} BF2 &= \text{Max}(0, X_{13} - 1878,52) \\ BF3 &= \text{Max}(0, 1878,52 - X_{13}) \\ BF4 &= (X_5 = 1)\text{Max}(0, 1878,52 - X_{13}) \\ BF6 &= \text{Max}(0, 96 - X_{10})\text{Max}(0, 1878,52 - X_{13}) \\ BF7 &= \text{Max}(0, X_{11} - 7)\text{Max}(0, 1878,52 - X_{13}) \\ BF9 &= (X_1 = 0)\text{Max}(0, 96 - X_{10})\text{Max}(0, 1878,52 - X_{13}) \\ BF11 &= \text{Max}(0, 1 - X_6) \\ BF12 &= \text{Max}(0, X_{11} - 4) \\ BF15 &= \text{Max}(0, X_4 - 1)\text{Max}(0, 1878,52 - X_{13}) \\ BF17 &= (X_5 = 1)\text{Max}(0, X_{11} - 4) \\ BF18 &= \text{Max}(0, X_6 - 1)\text{Max}(0, X_{11} - 5) \\ BF19 &= \text{Max}(0, X_6 - 1)\text{Max}(0, 5 - X_{11}) \\ BF20 &= (X_2 = 1)\text{Max}(0, 1 - X_4)\text{Max}(0, 1878,52 - X_{13}) \\ BF22 &= (X_5 = 6)\text{Max}(0, 4 - X_{11}) \\ BF23 &= \text{Max}(0, X_4 - 1)\text{Max}(0, 7 - X_{11})\text{Max}(0, 1878,52 - X_{13}) \\ BF24 &= \text{Max}(0, 1 - X_4)\text{Max}(0, 7 - X_{11})\text{Max}(0, 1878,52 - X_{13}) \\ BF25 &= (X_4 = 2)\text{Max}(0, X_6 - 1)\text{Max}(0, X_{11} - 5) \\ BF26 &= \text{Max}(0, 1 - X_6)(X_7 = 1)(X_8 = 6) \\ BF27 &= (X_5 = 4)\text{Max}(0, X_{13} - 2886,87) \\ BF30 &= \text{Max}(0, 84 - X_{10})\text{Max}(0, X_{11} - 4) \\ BF31 &= \text{Max}(0, 1 - X_4)(X_9 = 1)\text{Max}(0, 1878,52 - X_{13}) \\ BF32 &= (X_7 = 1)\text{Max}(0, 1878,52 - X_{13}) \\ BF33 &= (X_7 = 1)\text{Max}(0, 7 - X_{11})\text{Max}(0, 1878,52 - X_{13}) \\ BF34 &= \text{Max}(0, X_6 - 1)\text{Max}(0, X_{10} - 24) \end{aligned}$$

Berdasarkan model MARS yang terbentuk, interpretasi dari beberapa koefisien basis fungsi adalah sebagai berikut:

1. $BF2 = \text{Max}(0, X_{13} - 1987,41)$
Artinya koefisien BF2 akan bermakna jika nilai $X_{13} > 1987,41$, tetapi jika nilai $X_{13} \leq 1987,41$ maka koefisien BF2 tidak bermakna sehingga nilainya 0.
2. $BF4 = \text{Max}(0, X_{10} - 48)\text{Max}(0, 1987,41 - X_{13})$
Artinya koefisien BF4 akan bermakna jika nilai $X_{10} > 48$ dan $X_{13} < 1987,41$, tetapi jika nilai $X_{10} \leq 48$ dan $X_{13} \geq 1987,41$ maka koefisien BF2 tidak bermakna sehingga nilainya 0.
3. $BF20 = (X_6 = 5)\text{Max}(0, X_{10} - 35)\text{Max}(0, X_{11} - 5)$

Artinya koefisien BF20 akan bermakna jika $X_6 = 5$, $X_{10} > 1987,41$ dan $X_{11} > 5$ tetapi jika $X_6 \neq 5$, $X_{10} \leq 1987,41$ dan $X_{11} \leq 5$ maka koefisien BF2 tidak bermakna sehingga nilainya 0.

Berikut disajikan *confusion matrix* klasifikasi MARS:

Tabel 5. *Confusion Matrix* Klasifikasi MARS

Prediksi	Aktual	
	Tidak Miskin	Miskin
Tidak Miskin	3767	290
Miskin	69	172

Berdasarkan tabel di atas dapat diketahui evaluasi model sebagai berikut:

$$APER (\%) = \frac{290 + 69}{3767 + 290 + 69 + 172} \times 100\% = 8,35\%$$

$$TAR (\%) = \frac{3767 + 172}{3767 + 290 + 69 + 172} \times 100\% = 91,65\%$$

$$Specificity (\%) = \frac{172}{69 + 172} \times 100\% = 71,37\%$$

$$Sensitivity (\%) = \frac{3767}{3767 + 290} \times 100\% = 92,85\%$$

Nilai APER menghasilkan kesalahan klasifikasi sebesar 8,35% dan menghasilkan ketepatan klasifikasi yaitu nilai TAR sebesar 91,65%. Nilai TAR sebesar 8,35% menunjukkan bahwa model mampu memprediksi dengan benar kelas rumah tangga miskin atau tidak miskin untuk sebagian besar data yang diuji. Nilai *specifity* menggambarkan akurasi amatan pada kelompok negatif yaitu rumah tangga yang berstatus tidak miskin, diperoleh nilai *specifity* sebesar 71,37%. Nilai *sensitivity* menggambarkan akurasi amatan pada kelompok positif yaitu rumah tangga yang berstatus miskin, diperoleh nilai *sensitivity* sebesar 92,85% yang mengindikasikan bahwa model memiliki kemampuan yang baik dalam mendeteksi kelas positif atau kelas minoritas dalam data.

5. KESIMPULAN

Model MARS terbaik untuk mengklasifikasikan status rumah tangga sebagai variabel respon (Y) di Provinsi Bengkulu berdasarkan nilai GCV terkecil yaitu model MARS dengan kombinasi BF=48, MI=3, dan MO=0 karena menghasilkan nilai GCV paling kecil yaitu sebesar 0.0679963. Hasil evaluasi klasifikasi dihasilkan nilai APER sebesar 8,35% dan ketepatan klasifikasi yaitu nilai TAR sebesar 91,65%. Nilai *sensitivity* sebesar 92,85% dan nilai *specifity* sebesar 71,37%. Penelitian selanjutnya dapat dilakukan dengan menggabungkan metode MARS dengan metode yang dapat menangani ketidakseimbangan kelas data. Saran kepada pemerintah Provinsi Bengkulu agar lebih memperhatikan kesejahteraan rumah tangga di wilayah Provinsi Bengkulu.

DAFTAR PUSTAKA

- Agresti, A. 2007. *An Introduction to Categorical Data Analysis*. New Jersey: John Wiley & Sons.
 Cox, D.R., & Snell, E.J. 2018. *Analysis of Binary Data*. Boca Raton (US) : Chapman & Hall/CRC.
 Eubank, R.L. 1999. *Spline Smoothing and Nonparametric Regression*. New York: Marcel Dekker.

- Friedman, J.H. 1991. Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1), 1-67.
- Gujarati, D.N. 2022. *Basic Econometrics*. Prentice Hall.
- Han, J., & Kamber, M. 2006. *Data Mining Concepts and Techniques Second Edition*. San Francisco: Morgan Kaufmann.
- Hastie, T., Tibshirani, R., & Friedman, J.H. 2009. *The Elements of Statistical Learning: Data Mining, Inference, And Prediction (Vol. 2)*. New York: Springer.
- Hickel, J. 2016. The True Extent of Global Poverty and Hunger: Questioning The Good News Narrative Of The Millennium Development Goals. *Third World Quarterly*, 37(5), 749-767.
- Hosmer, D.W. & Lemeshow, S. 2000. *Applied Logistic Regression*. USA: John Willey and Sons.
- Johnson, R.A., & Wichern, D.W. 1992. *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.
- Kriner, M. 2007. *Survival Analysis with Multivariate Adaptive Regression Splines Disertation*. Germany (GR): Munchen University.
- Morton, R., Hebel, J., & McCarter, R. 2008. *A Study Guide to Epidemiology and Biostatistics*. Sudbury: Jones and Bartlett Publishers, Inc.
- Surya, D.W.T. 2016. Perbandingan Ketepatan Klasifikasi Metode Regresi Logistik Biner dan Multivariate Adaptive Regression Splines Pada Status Stroke Pasien (Studi Kasus: Rsud Dr. H. Slamet Martodirdjo Pamekasan Tahun 2015). *Disertasi*. Surabaya: Institut Teknologi Sepuluh Nopember.
- Sheskin, D.J. 2011. *Handbook of Parametric and Nonparametric Statistical Procedures Fifth Edition*. London: Chapman & Hall.
- Tamonob, O. 2020. Analisis Multivariate Adaptive Regression Splines (MARS) Untuk Mengklasifikasikan Status Desa Di Provinsi Nusa Tenggara Timur. *Skripsi*. Bogor: Institut Bogor Pertanian.
- Zhang, H. 1997. Multivariate Adaptive Splines for Analysis Of Longitudinal Data. *Journal of Computational and Graphical Statistics*, 6(1), 74-91.