



KETEPATAN KLASIFIKASI DENGAN MENGGUNAKAN METODE *MULTIVARIATE ADAPTIVE REGRESSION SPLINE* (MARS) PADA DATA KELOMPOK RUMAH TANGGA KABUPATEN CILACAP

Saroful Anam[✉], Sugiman, dan Sunarmi

Jurusan Matematika, FMIPA, Universitas Negeri Semarang, Indonesia
Gedung D7 Lt. 1, Kampus Sekaran Gunungpati, Semarang 50229

Info Artikel

Sejarah Artikel:
Diterima Maret 2017
Disetujui April 2017
Dipublikasikan Mei 2017

Keywords:

Classification, Group Household, Multivariate Adaptive Regression Spline (MARS).

Abstrak

Tujuan penelitian ini adalah mengklasifikasikan dan menghitung kesalahan klasifikasi kelompok rumah tangga di Kabupaten Cilacap berdasarkan faktor-faktor yang mempengaruhi kemiskinan menggunakan *Multivariate Adaptive Regression Spline* (MARS). Model terbaik yang dipilih adalah model dengan kombinasi $BF=45$, $MI=3$, $MO=1$ karena memiliki nilai GCV terkecil sebesar 0,030. Berdasarkan variabel-variabel yang mempengaruhi kelompok rumah tangga, diperoleh hasil klasifikasi dari 37 rumah tangga dengan kategori miskin, 34 rumah tangga tepat diklasifikasikan ke dalam kategori miskin, sedangkan 3 rumah tangga salah diklasifikasikan salah dan masuk ke dalam kategori miskin. Begitu juga dari 113 rumah tangga dengan kategori tidak miskin, 113 rumah tangga tepat diklasifikasikan ke dalam kategori tidak miskin, dan tidak ada rumah tangga yang salah diklasifikasikan ke dalam rumah tangga dengan kategori miskin. Diperoleh ketepatan klasifikasi sebesar 98,00% nilai *Apparent Error Rate* (APER) sebesar 2,00% dan nilai uji *Press's Q* menunjukkan bahwa secara statistik metode MARS sudah konsisten dalam mengklasifikasi data. Sehingga untuk penelitian mengenai klasifikasi selanjutnya disarankan menggunakan metode MARS karena terbukti menghasilkan ketepatan klasifikasi yang akurat.

Abstract

The purpose of this study is to classify and count the misclassification of households in Cilacap district based on factors that affect poverty using *Multivariate Adaptive Regression Spline* (MARS). The best model selected is a model with a combination of $BF = 45$, $MI = 3$, $MO = 1$ because it has the smallest value of 0,030 GCV. Based on the variables that affect households, the result of classification of 37 households with poor category, 34 households appropriately classified as poor, while 3 households incorrectly classified incorrectly and into the poor category. Likewise of the 113 households with non-poor category, 113 households appropriately classified into the category of not poor, and no housekeeping incorrectly classified into the category of poor households. Obtained classification accuracy of 98.00% value of *Apparent Error Rate* (APER) at 2.00% and the value of the *Press's Q* test showed that statistically MARS method has been consistent in classifying the data. So as to further research on the classification suggested using MARS method as shown to produce an accurate classification accuracy.

How to Cite

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.

PENDAHULUAN

Masalah kemiskinan merupakan salah satu permasalahan pemerintah yang diprioritaskan dalam menyusun strategi pembangunan setiap negara termasuk Indonesia. Meskipun pertumbuhan ekonomi negara Indonesia semakin meningkat, namun jumlah penduduk juga semakin bertambah sehingga beban ekonomi juga terus bertambah. Kemiskinan timbul apabila masyarakat tidak memiliki kemampuan-kemampuan utama, pengeluaran per kapita, pendidikan yang memadai, kondisi kesehatan yang baik, rasa aman, kepercayaan diri yang tinggi dan hak seperti kebebasan berbicara (Haughton dan Khandker, 2012). Menurut Badan Pusat Statistik (2014), kelompok rumah tangga berdasarkan pengeluaran per kapita yang sudah dikonversikan berdasarkan garis kemiskinan terdiri dari dua kelompok yaitu kelompok rumah tangga miskin dan kelompok rumah tangga tidak miskin.

Dalam ilmu statistika, banyak metode yang dapat digunakan untuk mengetahui pengaruh variabel prediktor terhadap variabel responnya yang bersifat kategori. Salah satunya adalah metode *Multivariate Adaptive Regression Spline* (MARS). Model MARS difokuskan untuk mengatasi permasalahan dimensi tinggi, memiliki variabel banyak, serta ukuran sampel besar sehingga diperlukan perhitungan yang rumit berdasarkan nilai *Generalized Cross Validation* (GCV) terkecil. Pada MARS, beberapa hal yang perlu diperhatikan dalam pemilihan model yang paling optimum (terbaik) adalah jika nilai GCV dari model tersebut mempunyai nilai yang paling rendah (minimum) diantara model-model yang lain (Friedman, 1991).

Metode klasifikasi merupakan bagian dari analisis statistika yang bertujuan memisahkan individu atau objek ke dalam suatu kelompok sehingga dapat diketahui suatu individu berada pada kelompok tertentu. Metode klasifikasi yang baik akan menghasilkan sedikit kesalahan klasifikasi atau akan menghasilkan peluang kesalahan alokasi yang kecil (Johnson dan Wichern, 1992).

Metode statistika yang dapat digunakan untuk menganalisis dan mengklasifikasikan suatu objek dan variabel respon yang bersifat kualitatif atau kategori dengan sejumlah variabel prediktor kontinu atau kategori adalah metode MARS. MARS merupakan pendekatan untuk regresi multivariate nonparametrik yang dikembangkan oleh Friedman (1991). Teknik pemodelan MARS menjadi populer karena tidak mengasumsikan dan tidak menentukan tipe khusus seperti pada hubungan linier,

kuadratik, kubik di antara variabel prediktor dan respon (Budiantara, 2006). Penelitian mengenai klasifikasi dengan menggunakan MARS sebelumnya juga pernah dilakukan oleh Ghofar (2014) pada penelitian yang berjudul “*Klasifikasi Kelulusan Mahasiswa Fakultas Sains dan Matematika Universitas Diponegoro Menggunakan Multivariate Adaptive Regression Spline* (MARS)”. Metode MARS sangat baik untuk klasifikasi karena akan menghasilkan sedikit kesalahan klasifikasi. Berdasarkan uraian diatas, maka peneliti tertarik mengangkat topik yang berjudul “*Ketepatan Klasifikasi dengan Menggunakan Metode Multivariate Adaptive Regression Spline* (MARS) pada Data Kelompok Rumah Tangga Kabupaten Cilacap”.

Masalah yang dianalisis dalam penelitian ini adalah bagaimana hasil klasifikasi berdasarkan variabel yang mempengaruhi kelompok rumah tangga di Kabupaten Cilacap dengan menggunakan metode *Multivariate Adaptive Regression Spline* (MARS) dan bagaimana hasil perhitungan kesalahan klasifikasi kelompok rumah tangga di Kabupaten Cilacap dengan menggunakan metode *Multivariate Adaptive Regression Spline* (MARS). Tujuan dalam penelitian ini adalah mengklasifikasi kelompok rumah tangga di Kabupaten Cilacap berdasarkan faktor-faktor yang mempengaruhi tingkat kemiskinan di Kabupaten Cilacap dengan *Multivariate Adaptive Regression Spline* (MARS) dan menghitung besar kesalahan klasifikasi kelompok rumah tangga di Kabupaten Cilacap dengan metode MARS.

Metode MARS merupakan metode analisis regresi dengan pendekatan regresi nonparametrik. Model MARS berguna untuk mengatasi permasalahan data berdimensi tinggi, yaitu data yang memiliki jumlah variabel prediktor $3 \leq n \leq 20$, dimana n adalah banyak variabel dan sampel data yang berukuran $50 \leq n \leq 1000$ dimana N untuk ukuran sampel (Friedman, 1991).

Nisa' dan Budiantara (2012) menyatakan ada beberapa hal yang perlu diperhatikan dalam menggunakan model MARS yaitu *knot*, *Basis Function* (BF) dan interaksi. *Knot* yaitu akhir dari sebuah garis regresi (*region*) dan awal dari sebuah garis regresi (*region*) yang lain. Di setiap titik *knot*, diharapkan adanya kontinuitas dari fungsi basis antar satu *region* dengan *region* lainnya. *Basis Function* yaitu suatu fungsi yang digunakan untuk menjelaskan hubungan antara variabel respon dan variabel prediktor. Pada umumnya fungsi basis yang dipilih adalah berbentuk polinomial dengan turunan yang kontinu pada setiap titik *knot*. Friedman menyarankan jumlah maksimum fungsi basis

(BF) adalah 2-4 kali jumlah variabel prediktornya. Interaksi merupakan hasil perkalian silang antara variabel yang saling berkorelasi. Friedman membatasi jumlah maksimum interaksi (MI) yang diperbolehkan yaitu 1, 2, dan 3.

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

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

dengan,

- α_0 = konstanta regresi dari fungsi basis
- α_m = koefisien dari fungsi basis ke- m
- M = maksimum fungsi basis (*nonconstant* fungsi basis)
- K_m = derajat interaksi
- S_{km} = bernilai +1 jika *knot* terletak di kanan titik *knot*, dan bernilai -1 jika *knot* terletak di kiri titik *knot*
- $x_{v(k,m)}$ = variabel prediktor
- t_{km} = nilai *knot* dari variabel prediktor $x_{v(k,m)}$

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

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

dengan

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

Persamaan (1) dapat dituliskan kedalam bentuk sebagai berikut:

$$\hat{f}(x) = \alpha_0 + f_i(x_i) + f_{ij}(x_i, x_j) + f_{ijk}(x_i, x_j, x_k) + \dots \quad (3)$$

Persamaan (3) menunjukkan bahwa penjumlahan pertama meliputi semua fungsi basis untuk satu variabel, penjumlahan kedua meliputi semua fungsi basis untuk interaksi antara dua variabel, penjumlahan ketiga meliputi semua fungsi basis untuk interaksi antara tiga variabel dan seterusnya. Misalkan $V(m) = \{v(k, m)\}_1^{K_m}$ adalah himpunan dari variabel yang dihubungkan dengan fungsi basis B_m ke- m , maka setiap penjumlahan pertama pada persamaan (3) dapat dituliskan sebagai berikut,

$$f_i(x_i) = \sum_{\substack{k_m=1 \\ i \in V(m)}} \alpha_m B_m(x_i) \quad (4)$$

$f_i(x_i)$ merupakan penjumlahan semua fungsi basis untuk satu variabel x_i dan merupakan *spline* dengan derajat $q=1$ yang mempresentasikan fungsi *univariate*. Setiap fungsi *bivariate* pada persamaan (3) dapat dituliskan sebagai berikut,

$$f_{ij}(x_i, x_j) = \sum_{\substack{k_m=2 \\ (i,j) \in V(m)}} \alpha_m B_m(x_i, x_j) \quad (5)$$

dengan,

$$B_m(x_i, x_j) = [s_{2m}(x_{i(2,m)} - t_{2m})]_+ [s_{2m}(x_{j(2,m)} - t_{2m})]_+$$

Persamaan (5) merupakan penjumlahan semua fungsi basis dua variabel x_i dan x_j . Untuk fungsi *trivariate* pada penjumlahan yang ketiga diperoleh dengan menjumlahkan semua fungsi basis untuk tiga variabel prediktor, yang dituliskan sebagai berikut,

$$f_{ijk}(x_i, x_j, x_k) = \sum_{\substack{k_m=3 \\ (i,j,k) \in V(m)}} \alpha_m B_m(x_i, x_j, x_k) \quad (6)$$

dengan,

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

Persamaan (3) dikenal dengan dekomposisi ANOVA dari model MARS. Interpretasi model MARS melalui dekomposisi ANOVA adalah mempresentasikan variabel prediktor yang masuk ke dalam model, baik untuk satu variabel prediktor maupun interaksi lebih dari satu variabel prediktor.

Penentuan *knot* pada metode MARS dapat ditentukan secara otomatis oleh data dan menghasilkan model yang kontinu pada *knot*. Penentuan *knot* pada MARS menggunakan *forward stepwise* dan *backward stepwise*. Untuk mendapatkan model dengan fungsi basis maksimum digunakan *forward stepwise*. Sedangkan *backward stepwise* digunakan untuk memenuhi prinsip parsimoni (model yang sederhana) dengan cara mengeliminasi fungsi basis yang berkontribusi kecil sampai tidak ada fungsi basis yang dapat dieliminasi. Pintowati dan Otok dalam Muslikah (2014) menyatakan bahwa, ukuran kontribusi pada *backward stepwise* ditentukan dengan kriteria *Generalized Cross Validation* (GCV). Bentuk fungsi GCV minimum menurut Friedman (1991) dapat didefinisikan pada persamaan berikut.

$$GCV(M) = \frac{\frac{1}{N} \sum_{i=1}^N [y_i - \hat{f}_M(x_i)]^2}{\left[1 - \frac{\tilde{C}(M)}{N}\right]^2} \quad (7)$$

dengan,
 N = ukuran sampel
 y_i = variabel respon
 x_i = variabel prediktor
 M = banyaknya fungsi basis (*nonconstant* fungsi basis)
 $\tilde{C}(M)$ = $C(M) + dM$
 $C(M)$ = $Trace[\mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T] + 1$
 d = nilai ketika setiap fungsi basis mencapai nilai optimal ($2 \leq d \leq 4$).

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:

$$\pi(\mathbf{x}) = \frac{e^{\hat{f}(\mathbf{x})}}{1 + e^{\hat{f}(\mathbf{x})}} \quad (8)$$

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

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

$$\begin{aligned} \text{logit } \pi(\mathbf{x}) &= \ln\left(\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right) \\ &= \alpha_0 + \sum_{m=1}^M \alpha_m \prod_{k=1}^{K_m} [s_{km}(x_{v(k,m)} - t_{km})]_+ \quad (10) \end{aligned}$$

Hosmer dan Lemeshow dalam Bisri (2014) menyatakan, 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.

Menurut Prasetyo (2012), sebuah sistem yang melakukan klasifikasi diharapkan dapat melakukan klasifikasi semua *data set* dengan benar, tetapi tidak dapat dipungkiri bahwa kinerja suatu sistem tidak bisa 100% benar sehingga sebuah sistem klasifikasi juga harus diukur kinerjanya. Pada umumnya, pengukuran klasifikasi dilakukan dengan matriks konfusi (*confusion matrix*). Jika y_0 dan y_1 merupakan subjek pengklasifikasian, maka bentuk matriks konfusi dapat dilihat pada tabel 1.

Tabel 1. Matriks Konfusi untuk Klasifikasi Dua Kelas

Kelas Observasi (Actual Class)	Kelas Prediksi (Predicted Class)	
	Kelas 0	Kelas 1
Kelas 0	f_{00}	f_{01}
Kelas 1	f_{10}	f_{11}

dengan,
 f_{00} = Jumlah subjek dari y_0 tepat diklasifikasikan sebagai y_0
 f_{01} = Jumlah subjek dari y_0 tidak tepat diklasifikasikan sebagai y_1
 f_{10} = Jumlah subjek dari y_1 tidak tepat diklasifikasikan sebagai y_0
 f_{11} = Jumlah subjek dari y_1 tepat diklasifikasikan sebagai y_1

Setiap sel f_{ij} dalam matriks menyatakan jumlah data dari kelas i yang hasil prediksinya masuk ke kelas j , f_{11} adalah jumlah data dalam kelas 1 yang secara benar dipetakan ke kelas 1, f_{10} adalah jumlah data dalam kelas 1 yang dipetakan secara salah ke kelas 0, f_{01} adalah jumlah data dalam kelas 0 yang dipetakan secara salah ke kelas 1, f_{00} adalah jumlah data dalam kelas 0 yang dipetakan secara benar ke kelas 0 (Prasetyo, 2012).

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

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

Hair et al dalam Oktiva (2014) menyatakan bahwa, uji *Press's Q* merupakan statistik uji untuk mengetahui kestabilan dalam pengklasifikasian atau sejauh mana kelompok-kelompok tersebut dapat dipisahkan dengan menggunakan variabel yang ada. Perhitungan nilai *Press's Q* dapat dihitung dengan menggunakan rumus:

$$\text{Press's } Q = \frac{[N - (rK)]^2}{N(K - 1)} \quad (12)$$

dengan,
 N = ukuran sampel
 r = banyaknya individu yang tepat diklasifikasikan
 K = banyaknya kelompok (kategori)

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

METODE

Penelitian ini menggunakan metode penelitian kajian pustaka. Kajian pustaka merupakan metode penelitian yang mengupas berbagai teori yang berhubungan dengan permasalahan dalam penelitian. Oleh karena itu, kajian pustaka digunakan sebagai dasar pemecahan masalah yang penulis angkat dalam penelitian ini. Langkah-langkah dalam metode ini adalah 1) Studi pustaka. 2) Perumusan masalah. 3) Pengumpulan data. 4) Pemecahan masalah. 5) Penarikan kesimpulan. Dalam tahap kajian pustaka dilakukan pengumpulan referensi, dan pengupasan teori yang dapat dijadikan sebagai suatu masalah.

Pemilihan dan perumusan masalah diperlukan untuk membatasi permasalahan sehingga diperoleh bahan kajian yang jelas. Sehingga akan lebih mudah untuk menentukan langkah dalam memecahkan masalah tersebut.

Dalam proses memperoleh jawaban dari masalah yang diangkat dalam penelitian ini dilakukan langkah-langkah analisis sebagai berikut.

- 1) Untuk mencapai tujuan pertama, dilakukan statistik deskriptif dengan bantuan *software* SPSS 16.0 untuk mengetahui karakteristik rumah tangga di Kabupaten Cilacap.
- 2) Untuk mencapai tujuan kedua, dilakukan analisis MARS dengan bantuan *software* MARS 2.0 untuk pembentukan model MARS. Langkah-langkah analisisnya adalah sebagai berikut.

- a) Pembentukan model MARS terbaik untuk data rumah tangga Kabupaten Cilacap dengan mengkombinasikan besarnya *Basis Function* (BF), *Maximum Interaction* (MI), dan *Minimum Observation* (MO). Dilakukan dengan cara *trial and error* dengan menentukan nilai maksimum BF, yaitu 2 sampai 4 kali banyaknya variabel prediktor yang digunakan, menentukan nilai MI yaitu 1,2 dan 3 dengan pertimbangan jika $MI > 3$ akan menghasilkan model yang semakin kompleks, dan menentukan minimal banyaknya pengamatan setiap *knot* (MO) yaitu 0,1,2 dan 3.

- b) Mendapatkan model MARS terbaik berdasarkan nilai GCV terkecil.
- c) Melakukan interpretasi model MARS terbaik dan interpretasi variabel-variabel yang berpengaruh di model tersebut.
- d) Mengklasifikasi ke dalam bentuk matriks konfusi (*confusion matrix*).
- e) Mendapatkan nilai kesalahan klasifikasi dari nilai APER. f) Menghitung nilai *Press's Q*

HASIL DAN PEMBAHASAN

Pada penelitian ini, statistik deskriptif bermanfaat untuk mengetahui karakteristik data kelompok rumah tangga di Kabupaten Cilacap. Data yang digunakan pada penelitian ini sebanyak 150 data hasil SUSENAS Kabupaten Cilacap tahun 2014 dengan prosentase 25% dikategorikan sebagai rumah tangga miskin dan 75% dikategorikan sebagai rumah tangga tidak miskin, ditunjukkan pada Gambar 1.



Gambar 1. Prosentase kelompok rumah tangga di Kabupaten Cilacap

Untuk mengklasifikasi kelompok rumah tangga di Kabupaten Cilacap, perlu dilakukan pembentukan model MARS terlebih dahulu. Pembentukan model dilakukan dengan cara *trial and error* dengan bantuan *software* MARS 2.0, dengan menzngkombinasikan beberapa kriteria model MARS yaitu fungsi basis (BF), maksimal interaksi antar variabel prediktor (MI) dan minimal pengamatan antar *knot* atau minimum observasi (MO) sampai diperoleh model MARS terbaik dengan nilai GCV minimum.

Pada penelitian ini, digunakan 15 variabel prediktor untuk pendugaan variabel respon sehingga penulis menentukan maksimum fungsi basis (BF) sebesar 30, 45 dan 60, sesuai dengan yang disarankan oleh Friedman (1991) yaitu dua sampai empat kali banyaknya variabel prediktor. Untuk interaksi maksimal (MI) sebesar 1, 2 dan 3 dengan pertimbangan apabila lebih dari 3, maka akan terbentuk model yang semakin kompleks sehingga akan sulit untuk diinterpretasikan. Sedangkan untuk observasi minimum (MO) sebesar 0, 1, 2, dan 3. Langkah selanjutnya adalah mengkombinasikan secara *trial and error*

dengan bantuan *software* MARS 2.0 terhadap nilai BF, MI dan MO. Tabel 2 merupakan hasil kombinasi BF, MI dan MO.

Tabel 2. *Trial and Error* kombinasi BF, MI, dan MO

No. Model	BF	MI	MO	GCV	KETEPATAN KLASIFIKASI
1	30	1	0	0.080	96.00%
2	30	1	1	0.082	94.00%
3	30	1	2	0.081	96.00%
4	30	1	3	0.083	94.67%
5	30	2	0	0.052	97.33%
6	30	2	1	0.038	97.33%
7	30	2	2	0.042	100.00%
8	30	2	3	0.045	98.67%
9	30	3	0	0.052	100.00%
10	30	3	1	0.031	98.00%
11	30	3	2	0.036	99.33%
12	30	3	3	0.040	97.33%
13	45	1	0	0.080	96.00%
14	45	1	1	0.085	95.33%
15	45	1	2	0.085	94.67%
16	45	1	3	0.085	94.67%
17	45	2	0	0.049	99.33%
18	45	2	1	0.044	99.33%
19	45	2	2	0.044	100.00%
20	45	2	3	0.046	96.00%
21	45	3	0	0.053	100.00%
*22	45	3	1	0.030	98.00%
23	45	3	2	0.034	100.00%
24	45	3	3	0.032	100.00%
25	60	1	0	0.083	96.67%
26	60	1	1	0.086	95.33%
27	60	1	2	0.086	94.67%
28	60	1	3	0.087	94.67%
29	60	2	0	0.045	99.33%
30	60	2	1	0.041	98.67%
31	60	2	2	0.051	98.67%
32	60	2	3	0.041	96.67%
33	60	3	0	0.052	99.33%
34	60	3	1	0.037	97.33%
35	60	3	2	0.039	90.33%
36	60	3	3	0.031	99.33%

Model terbaik merupakan kombinasi BF, MI dan MO secara *trial and error*. Model terbaik yang diperoleh yaitu dengan mengkombinasikan BF = 45, MI = 3 dan MI = 1. Model tersebut memiliki nilai GCV minimum sebesar 0,030. Diperoleh persamaan model sebagai berikut.

$$\hat{y} = 1,322 - (0,522586) \cdot BF3 + (0,197) \cdot BF5 - (0,189886) \cdot BF11 + (0,625259) \cdot BF13 + (0,843470) \cdot BF14 - (0,215113) \cdot BF15 - (0,283164) \cdot BF18 + (0,282436) \cdot BF20 + (0,647352) \cdot BF21 - (0,781267) \cdot BF22 - (0,114) \cdot BF25 - (0,176323) \cdot BF27 - (0,258350) \cdot BF28 - (0,136) \cdot BF29 - (0,578586) \cdot BF33 + (0,115312) \cdot BF34 - (0,313) \cdot BF39 - (0,163) \cdot BF41 + (0,143)BF43. \tag{13}$$

dengan,

$$\begin{aligned} BF2 &= \max(0; 2300071 - X5); \\ BF3 &= \max(0; X8 - 4).BF2; \\ BF4 &= \max(0; 4 - X8).BF2; \\ BF5 &= \max(0; X8 - 4); \\ BF8 &= \max(0; 1812940 - X5).BF5; \\ BF10 &= \max(0; 1470257 - X5); \\ BF11 &= (X13 = 1).BF10; \\ BF13 &= \max(0; X8 - 4).BF11; \\ BF14 &= \max(0; 4 - X8).BF11; \\ BF15 &= (X11 = 2) * BF4; \\ BF18 &= \max(0; 2800000 - X4).BF5; \\ BF20 &= \max(0; 26 - X2).BF11; \\ BF21 &= \max(0; X2 - 25).BF3; \\ BF22 &= \max(0; 25 - X2).BF3; \\ BF25 &= (X12 = 1 atau X12 = 4 atau X1= 5).BF5; \\ BF26 &= (X12 = 2 atau X12 = 3).BF5; \\ BF27 &= \max(0; X4 - 1100000).BF26; \\ BF28 &= \max(0; 1100000 - X4).BF26; \\ BF29 &= (X11 = 3).BF26; \\ BF33 &= \max(0; X2 - 67).BF8; \\ BF34 &= \max(0; 67 - X2).BF8; \\ BF39 &= (X6 = 1); \\ BF41 &= \max(0; X8 - 6); \\ BF43 &= (X15 = 1 atau X15 = 2).BF41; \end{aligned}$$

Interpretasi dari koefisien-koefisien fungsi basis pada persamaan model MARS sebagai berikut.

1. $BF3 = \max(0; X8 - 4).BF2$
 $BF2 = \max(0; 2300071 - X5)$

Artinya, kontribusi BF3 akan bermakna jika jumlah anggota rumah tangga lebih dari 4 anggota atau ditulis $\text{Max}(0; X8 - 4)$ dan jumlah pengeluaran perkapita rumah tangga kurang dari Rp2.300.071,- atau ditulis $\text{Max}(0, 2300071 - X5)$. Tetapi jika kriteria

tersebut tidak terdapat pada kepala rumah tangga tersebut, maka koefisien BF3 tidak bermakna. Bermaknanya kontribusi BF3 mengurangi nilai Y sebesar 0,522586.

$$2. \quad BF5 = (0; X8 - 4)$$

Artinya, kontribusi BF5 akan bermakna jika jumlah anggota rumah tangga lebih dari 4 anggota atau ditulis $\text{Max}(0; X8 - 4)$. Tetapi jika kriteria tersebut tidak terdapat pada kepala rumah tangga tersebut, maka koefisien BF5 tidak bermakna. Bermaknanya kontribusi BF5 menambah nilai Y sebesar 0,197.

$$3. \quad BF11 = (X13 = 1).BF10$$

$$BF10 = \text{max}(0; 1470257 - X5)$$

Artinya, kontribusi BF11 akan bermakna jika jenis tempat pembuangan air besar milik sendiri atau ditulis $(X13 = 1)$ dan jumlah pengeluaran perkapita rumah tangga kurang dari Rp1.470.257 atau ditulis $\text{Max}(0; 1470257 - X5)$. Tetapi jika kriteria tersebut tidak terdapat pada kepala rumah tangga tersebut, maka koefisien BF11 tidak bermakna. Bermaknanya kontribusi BF11 mengurangi nilai Y sebesar 0,189886.

$$4. \quad BF13 = \text{max}(0; X8 - 4).BF11$$

$$BF11 = (X13 = 1).BF10$$

$$BF10 = \text{max}(0; 1470257 - X5)$$

Artinya, kontribusi BF13 akan bermakna jika jumlah anggota rumah tangga lebih dari 4 anggota atau ditulis $\text{Max}(0; X8 - 4)$ dan jenis tempat pembuangan air besar milik sendiri atau ditulis $(X13 = 1)$ dan jumlah pengeluaran perkapita rumah tangga kurang dari Rp1.470.257,- atau ditulis $\text{Max}(0; 1470257 - X5)$. Tetapi jika kriteria tersebut tidak terdapat pada kepala rumah tangga tersebut, maka koefisien BF13 tidak bermakna. Bermaknanya kontribusi BF13 menambah nilai Y sebesar 0,625259.

$$5. \quad BF14 = \text{max}(0; 4 - X8).BF11$$

$$BF11 = (X13 = 1).BF10$$

$$BF10 = \text{max}(0; 1470257 - X5)$$

Artinya, kontribusi BF14 akan bermakna jika jumlah anggota rumah tangga kurang dari 4 anggota atau ditulis $\text{Max}(0; 4 - X8)$ dan jenis tempat pembuangan air besar milik sendiri atau ditulis $(X13 = 1)$ dan jumlah pengeluaran perkapita rumah tangga kurang dari Rp1.470.257,- atau ditulis $\text{Max}(0; 1470257 - X5)$. Tetapi jika kriteria tersebut tidak terdapat pada kepala rumah tangga tersebut, maka koefisien BF14 tidak bermakna. Bermaknanya kontribusi BF14 mengurangi nilai Y sebesar 0,843470.

$$6. \quad BF15 = (X11 = 2).BF4$$

$$BF4 = \text{max}(0; 4 - X8).BF2$$

$$BF2 = \text{max}(0; 2300071 - X5)$$

Artinya, kontribusi BF15 akan bermakna jika jenis lantai terluas dalam rumah tangga

tersebut berupa tegel/teraso atau ditulis $(X11 = 2)$ dan jumlah anggota rumah tangga kurang dari 4 anggota atau ditulis $\text{Max}(0; 4 - X8)$ dan jumlah pengeluaran perkapita rumah tangga kurang dari Rp2.300.071,- atau ditulis $\text{Max}(0; 2300071 - X5)$. Tetapi jika kriteria tersebut tidak terdapat pada kepala rumah tangga tersebut, maka koefisien BF15 tidak bermakna. Bermaknanya kontribusi BF15 menambah nilai Y sebesar 0,215113.

$$7. \quad BF18 = \text{max}(0; 2800000 - X4).BF5$$

$$BF5 = \text{max}(0; X8 - 4)$$

Artinya, kontribusi BF18 akan bermakna jika pendapatan kepala rumah tangga kurang dari Rp2.800.000,- atau ditulis $\text{Max}(0; 2800000 - X4)$ dan jumlah anggota rumah tangga lebih dari 4 anggota $\text{Max}(0; 4 - X8)$. Tetapi jika kriteria tersebut tidak terdapat pada kepala rumah tangga tersebut, maka koefisien BF18 tidak bermakna. Bermaknanya kontribusi BF18 mengurangi nilai Y sebesar 0,283164.

$$8. \quad BF20 = \text{max}(0; 26 - X2).BF11$$

$$BF11 = (X13 = 1).BF10$$

$$BF10 = \text{max}(0; 1470257 - X5)$$

Artinya, kontribusi BF20 akan bermakna jika umur kepala rumah tangga tersebut lebih dari 26 tahun atau ditulis $\text{Max}(0; 26 - X2)$ dan jenis tempat pembuangan air besar milik sendiri atau ditulis $(X13 = 1)$ dan jumlah pengeluaran perkapita rumah tangga kurang dari Rp1.470.257,- atau ditulis $\text{Max}(0; 1470257 - X5)$. Tetapi jika kriteria tersebut tidak terdapat pada kepala rumah tangga tersebut, maka koefisien BF20 tidak bermakna. Bermaknanya kontribusi BF20 menambah nilai Y sebesar 0,031.

$$9. \quad BF21 = \text{max}(0; X2 - 25).BF3$$

$$BF3 = \text{max}(0; X8 - 4).BF2$$

$$BF2 = \text{max}(0; 2300071 - X5)$$

Artinya, kontribusi BF21 akan bermakna jika umur kepala rumah tangga lebih dari 25 tahun, atau ditulis $\text{Max}(0; X2 - 25)$ dan jumlah anggota rumah tangga kurang dari 4 anggota atau ditulis $\text{Max}(0; X8 - 4)$ dan jumlah pengeluaran perkapita rumah tangga lebih dari Rp2.300.071,- atau ditulis $\text{Max}(0; 2300071 - X5)$. Tetapi jika kriteria tersebut tidak terdapat pada kepala rumah tangga tersebut, maka koefisien BF21 tidak bermakna. Bermaknanya kontribusi BF21 menambah nilai Y sebesar 0,647352.

$$10. \quad BF22 = \text{max}(0; 25 - X2).BF3$$

$$BF3 = \text{max}(0; X8 - 4).BF2$$

$$BF2 = \text{max}(0; 2300071 - X5)$$

Artinya, kontribusi BF22 akan bermakna jika umur kepala rumah tangga tersebut lebih dari 25 tahun atau ditulis $\text{Max}(0; 25 - X2)$ dan

jumlah anggota rumah tangga kurang dari 4 anggota $\text{Max}(0; X_8 - 4)$ dan dan jumlah pengeluaran perkapita rumah tangga lebih dari Rp2.300.071,- atau ditulis $\text{Max}(0; 2300071 - X_5)$. Tetapi jika kriteria tersebut tidak terdapat pada kepala rumah tangga tersebut, maka koefisien BF22 tidak bermakna. Bermaknanya kontribusi BF22 mengurangi nilai Y sebesar 0,781267.

Model MARS terbaik pada persamaan (13) dipengaruhi oleh 9 variabel prediktor yang secara signifikan memberikan kontribusi pada model. Tabel 3 menunjukkan kontribusi dari masing-masing variabel prediktor terhadap model.

Tabel 3. Tingkat kepentingan variabel

Variabel	Tingkat Kepentingan	Pengurangan GCV
X_5	100,00%	0,171%
X_8	90,48%	0,146%
X_{13}	61,93%	0,084%
X_{12}	30,56%	0,043%
X_2	29,72%	0,043%
X_4	25,97%	0,040%
X_{11}	20,78%	0,036%
X_{15}	15,86%	0,034%
X_6	13,22%	0,033%
X_1	0,00%	0,030%
X_3	0,00%	0,030%
X_7	0,00%	0,030%
X_9	0,00%	0,030%
X_{10}	0,00%	0,030%
X_{14}	0,00%	0,030%

Tingkat kepentingan dari setiap variabel ditaksir oleh kenaikan nilai GCV. Nilai "pengurangan GCV" menunjukkan besarnya GCV yang berkurang apabila variabel tersebut dimasukkan ke dalam model. Berdasarkan Tabel 3 dapat diketahui bahwa variabel pengeluaran perkapita rumah tangga (X_5) merupakan variabel yang paling berpengaruh terhadap model dengan tingkat kepentingan 100,00%. Variabel kedua adalah variabel jumlah anggota rumah tangga (X_8) dengan tingkat kepentingan 90,48%. Variabel ketiga adalah variabel penggunaan tempat buang air besar (X_{13}) dengan tingkat kepentingan 61,93%.

Variabel keempat adalah variabel sumber air minum utama (X_{12}) dengan tingkat kepentingan 30,56%. Variabel kelima adalah variabel umur kepala rumah tangga (X_2) dengan tingkat kepentingan 29,72%. Variabel keenam adalah variabel pendapatan kepala rumah tangga (X_4) dengan tingkat kepentingan 25,97%. Variabel ketujuh adalah variabel jenis lantai terluas (X_{11}) dengan tingkat kepentingan 20,78%. Variabel kedelapan adalah variabel bahan atau energi untuk memasak (X_{15}) dengan tingkat kepentingan 15,86%. Variabel kesembilan adalah variabel status penguasaan bangunan tempat tinggal (X_6) dengan tingkat kepentingan 13,22%. Sedangkan variabel jenis kelamin kepala rumah tangga (X_1), variabel Ijazah/STTB tertinggi yang dimiliki (X_3), variabel pengalaman membeli raskin (X_7), variabel jenis atap terluas (X_9), jenis dinding terluas (X_{10}), variabel tempat pembuangan akhir tinja (X_{14}), tidak memberikan kontribusi pada model.

Hasil kombinasi antara BF, MI, dan MO yang telah dilakukan secara *trial and error* berdasarkan model persamaan terbaik, diketahui terdapat variabel-variabel prediktor yang mempengaruhi variabel respon yaitu variabel pengeluaran perkapita rumah tangga (X_5), variabel jumlah anggota rumah tangga (X_8), variabel penggunaan tempat buang air besar (X_{13}), variabel sumber air minum utama (X_{12}), variabel umur kepala rumah tangga (X_2), variabel pendapatan kepala rumah tangga (X_4), variabel jenis lantai terluas (X_{11}), variabel bahan atau energi untuk memasak (X_{15}), dan variabel status penguasaan bangunan tempat tinggal (X_6), Sedangkan variabel jenis kelamin kepala rumah tangga (X_1), variabel Ijazah/STTB tertinggi yang dimiliki (X_3), variabel pengalaman membeli raskin (X_7), variabel jenis atap terluas (X_9), jenis dinding terluas (X_{10}), variabel tempat pembuangan akhir tinja (X_{14}), tidak memberikan kontribusi pada model.

Tabel 4. Matriks konfusi untuk klasifikasi kelompok rumah tangga

Kelas Observasi	Kelas Prediksi		Total Aktual
	Kelas 0	Kelas 1	
Kelas 0	34	3	37
Kelas 1	0	113	113
Total Prediksi	34	116	150

Tabel 4 merupakan matriks konfusi yang menampilkan frekuensi sampel yang tepat diklasifikasikan maupun yang salah diklasifikasikan. Dari 37 rumah tangga dengan

kategori miskin, 34 rumah tangga tepat diklasifikasikan ke dalam kategori miskin, sedangkan 3 rumah tangga salah diklasifikasikan dan masuk ke dalam kategori tidak miskin. Begitu juga dari 113 rumah tangga dengan kategori tidak miskin, 113 rumah tangga tepat diklasifikasikan ke dalam kategori tidak miskin, dan tidak ada rumah tangga yang salah diklasifikasikan ke dalam rumah tangga dengan kategori miskin.

$$APER = \frac{3+0}{34+3+0+113} \times 100\% = 2,00\%$$

ketepatan klasifikasi = $100\% - 2,00\% = 98,00\%$

Sehingga dapat disimpulkan bahwa nilai APER dari klasifikasi kelompok rumah tangga di Kabupaten Cilacap dengan menggunakan metode MARS sebesar 2,00% dan nilai ketepatan klasifikasinya sebesar 98,00%.

Kemudian dilakukan uji *Press's Q* untuk menguji kestabilan pengklasifikasian model. Nilai *Press's Q* dapat dihitung dengan rumus.

$$Press's Q = \frac{[150 - (147 * 2)]^2}{150(2 - 1)} = 138,24$$

Selanjutnya dibandingkan dengan nilai chi-square dengan derajat bebas 1 ($\chi^2_{(1;0,05)} = 3,841$). Nilai *Press's Q* yang didapatkan jauh lebih besar dibandingkan dengan $\chi^2_{(1;0,05)}$, sehingga dapat dikatakan keakuratan klasifikasi kelompok rumah tangga di kabupaten Cilacap menggunakan model MARS sudah konsisten secara statistik.

PENUTUP

Berdasarkan analisis yang telah dilakukan dapat disimpulkan bahwa, berdasarkan variabel-variabel yang mempengaruhi kelompok rumah tangga di Kabupaten Cilacap, diperoleh hasil klasifikasi dari 37 rumah tangga dengan kategori miskin, 34 rumah tangga tepat diklasifikasikan ke dalam kategori miskin, sedangkan 3 rumah tangga salah diklasifikasikan dan masuk ke dalam kategori tidak miskin. Begitu juga dari 113 rumah tangga dengan kategori tidak miskin, 113 rumah tangga tepat diklasifikasikan ke dalam kategori tidak miskin, dan tidak ada rumah tangga yang salah diklasifikasikan ke dalam rumah tangga dengan kategori miskin. Diperoleh ketepatan hasil klasifikasinya sebesar 98,00%, nilai *Apparent Error Rate* (APER) sebesar 2,00% dan nilai uji *Press's Q* menunjukkan bahwa secara statistik metode MARS sudah konsisten dalam mengklasifikasi

data kelompok rumah tangga Kabupaten Cilacap.

Saran yang dapat diberikan untuk penelitian berikutnya adalah (1) Penelitian mengenai klasifikasi selanjutnya dapat menggunakan metode MARS karena telah terbukti menghasilkan ketepatan klasifikasi yang akurat untuk klasifikasi kelompok rumah tangga. (2) Penelitian selanjutnya dapat menambahkan variabel-variabel prediktor yang lain yang dapat mempengaruhi kelompok rumah tangga di Kabupaten Cilacap.

DAFTAR PUSTAKA

- Badan Pusat Statistik 2014a (2014). *Indikator Kesejahteraan Rakyat 2014*. Jakarta: Badan Pusat Statistik.
- Budiantara, N, S. Guritno, Otok, B.W. & Suryadi, F. 2012. "Pemodelan B-Spline dan MARS pada Nilai Ujian Masuk Terhadap IPK Mahasiswa Jurusan Disain Komunikasi Visual UK Petra Surabaya," *Jurnal Teknik Industri*, vol 8, Surabaya (2006).
- Cox, D.R. & Snell, E.J. 1989. Cox, D.R. & Snell, E.J. 1989. *Analysis of Binary Data*. Second Edition. Chapman and Hall, London.
- Friedman, J.H. 1991. *Multivariate Adaptive Regression Splines*. *The Annals of Statistics*, Vol. 19 No. 1.
- Ghofar, Y.R. Safitri, D. & Rusgiyono, A. 2014. Klasifikasi Kelulusan Mahasiswa Fakultas SAINS dan Matematika Universitas Diponegoro Menggunakan *Multivariate Adaptive Regression Spline* (MARS). *Jurnal Gaussian*. 3(4): 839-848.
- Houghton, J. & Khandker, S. R. 2012. *Pedoman Tentang Kemiskinan dan Ketimpangan*. Jakarta: Salemba Empat.
- Bisri, M. 2014. Perbandingan Analisis Klasifikasi menggunakan Metode *K-Nearest Neighbour* (K-NN) dan *Multivariate Adaptive Regression Splines* (MARS) pada Data Akreditasi Sekolah Dasar Negeri di Kota Semarang. *Jurnal Gaussia*, 3(3): 313 – 322.
- 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 Splines* pada Kasus Demam Berdarah *Dengue* (DBD). *Jurnal Sains dan Seni*, 1(1): 318-323.
- Muslikhah, R. 2014. *Multivariate Adaptive Regression Spline* (MARS) untuk

- Klasifikasi Kejadian Konstipasi terhadap Pemberian Air Susu Ibu dan Pemberian Air Susu Formula. *Jurnal Statistika*, 3(2).
- Oktiva D. A. (2014). *Bootstrap Aggregating Multivariate Adaptive Regression Splines (Bagging MARS)* untuk Mengklasifikasikan Rumah Tangga Miskin di Kabupaten Jombang. *Jurnal Sains dan Seni Pomits*. 3(2): 2337-3520.
- Prasetyo, E. 2012. *Data Mining Konsep dan Aplikasi menggunakan MATLAB*. Yogyakarta: Penerbit Andi Yogyakarta.