ISBN:1858-1382

J M S K
Jurnal Matematika, Statistika, & Komputasi
http://journal.unhas.ac.id/index.php/jmsk
Vol. 7, No.1, 1-12, Juli 2010

Algoritma *Penalized Residual Sum of Square* pada Penentuan Model *Multivariate Adaptive Regression Spline* dengan Respon Kontinu

Raupong*

Abstrak

Analisis regresi digunakan untuk melihat pengaruh variabel independen terhadap variabel dependen dengan terlebih dahulu melihat pola hubungan variabel tersebut. Hal ini dapat dilakukan dengan dua pendekatan, yaitu pendekatan parametrik dan nonparametrik. Penerapan Algoritma *Penalized Residual Sum of Square* (PRSS) pada model MARS akan digunakan untuk menyelesaikan permasalahan dengan respon kontinu. Adapun tujuan dalam penulisan ini adalah mendapatkan model MARS terbaik pada penurunan tekanan darah terhadap pemberian ramuan buah mengkudu dan daun kumis kucing pada pasien hipertensi yang berobat di laboratorium P4OT Surabaya. Secara keseluruhan, model terbaik dipilih berdasarkan koefisien determinasi terbesar. Namun demikian untuk MARS, model terbaik dipilih berdasarkan pada *Generalized Cross Validation* (GCV), minimum *Mean Square Error (*MSE), dan koefisien determinasi terbesar.

Kata Kunci: *Regresi nonparametrik, spline, titik knot, MARS, koefisien determinasi.*

1. Pendahuluan

Analisis regresi digunakan untuk melihat pengaruh variabel prediktor terhadap variabel respon dengan terlebih dahulu melihat pola hubungan variabel tersebut. Hal ini dapat dilakukan dengan melakukan dua pendekatan. Pendekatan umum dan sering dilakukan adalah pendekatan parametrik yang mengasumsikan bentuk model sudah ditentukan. Apabila tidak ada informasi apapun tentang bentuk model, maka pendekatan yang dilakukan adalah pendekatan nonparametrik. Karena pendekatan tidak tergantung pada asumsi bentuk kurva tertentu, sehingga memberikan fleksibilitas yang lebih besar. Pendekatan regresi parametrik antara lain menggunakan regresi linear sederhana, kuadratik, dan kubik, sedangkan regresi nonparametrik antara lain menggunakan *B-Spline* dan *Multivariate Adaptive Regression Spline* (MARS). Secara keseluruhan, model terbaik dipilih berdasarkan koefisien determinasi terbesar, Mean Square Error (MSE) dan Cp Mallows. Namun demikian untuk MARS, model terbaik dipilih berdasarkan pada nilai *Generalized Cross Validation* (GCV), MSE dan koefisien determinasi terbesar (Budiantara, 2006).

MARS merupakan suatu prosedur adaptif dikarenakan seleksi fungsi basis adalah database dan terkhusus pada masalah yang ada. Algoritma ini merupakan prosedur regresi nonparametrik yang tidak membuat anggapan khusus mengenai hubungan fungsional penekanan antara variabel prediktor dan variabel respon (Taylan dkk., 2007).

* *Prodi Statistika, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin.*

Pada model MARS, selain penentuan knots yang dilakukan secara otomatis dari data, juga menghasilkan model yang kontinu pada knots. Pemilihan knots pada MARS menggunakan algoritma *forward stepwise* dan *backward stepwise* yang salah satunya didasarkan nilai GCV minimum (Budiantara, 2006).

Penerapan MARS pada umumnya digunakan untuk menyelesaikan dua permasalahan utama dalam statistika, yaitu respon kontinu dan respon kategorik. Pada respon kontinu, beberapa peneliti yang telah menerapkan MARS antara lain Dwinnell pada tahun 2000 dengan mengeksplorasi MARS sebagai alternatif Neural Network (NN), Park *et al.* pada tahun 2001 menggunakan MCMC pada model MARS untuk data korelasi temporal, dan Sephton di tahun yang sama memprediksi resesi menggunakan MARS dengan hasil yang lebih baik dibanding ARIMA. Sedangkan pada respon kategorik, Holmes dan Denisson tahun 2002 secara khusus membahas penerapan Bayesian statistika pada pembentukan model klasifikasi MARS (Otok, 2008).

Pada paper ini penelitian akan difokuskan untuk menentukan model MARS pada penurunan tekanan darah terhadap pemberian ramuan buah mengkudu dan daun kumis kucing pada pasien hipertensi yang berobat di laboratorium P4OT Surabaya.

2. Landasan Teori

2.1. Regresi Nonparametrik

Diberikan data (t_j, y_j) dengan hubungan antara t_j dan y_j diasumsikan mengikuti model regresi nonparametrik

$$y_j = f(t_j) + \varepsilon_j,\ j = 1,2,\ldots,n \tag{1}$$

dengan $f(t_j)$ merupakan kurva regresi yang diasumsikan halus (*smooth*) bagian demi bagian, dalam arti merupakan anggota ruang fungsi tertentu, khususnya $f \in W_2^m[a,b]$, dengan $W_2^m[a,b] = \left\{ f : \int_a^b \left[f^m(t) \right]^2 dt < \infty \right\}$ adalah ruang fungsi *sobolev* orde dua untuk suatu *m* bilangan bulat positif. Galat acak ε_j berdistribusi normal dengan mean nol dan variansi σ^2 (Budiantara, 2001).

2.2 Spline

Menurut Mahler (1995), analisis regresi nonparametrik spline dikembangkan untuk mendapatkan fleksibilitas dalam persoalan regresi untuk analisis data. Di samping itu, menurut Budiantara (2003) spline juga dikembangkan untuk menangani data yang bersifat lokal dan mempunyai perilaku yang berbeda pada tiap-tiap sub interval. Terdapat beberapa model spline yang telah dikembangkan oleh banyak peneliti, masing-masing mempunyai peran yang berbeda sebagai estimasi kurva regresi nonparametrik maupun semiparametrik (Budiantara, 2004).

Spline adalah salah satu jenis piecewise polinomial, yaitu polinomial yang memiliki sifat tersegmen. Sifat tersegmen ini memberikan fleksibilitas lebih dari polinomial biasa, sehingga memungkinkan untuk menyesuaikan diri secara lebih efektif terhadap karakteristik suatu fungsi atau data. Spline mempunyai kelemahan pada saat orde Spline tinggi, knots yang banyak, dan knots yang terlalu dekat yang akan membentuk matriks dalam perhitungan yang hampir singular, sehingga persamaan normal tidak dapat diselesaikan. Secara umum, fungsi Spline berorde k adalah sembarang fungsi yang dinyatakan sebagai berikut

$$S(t)=\sum_{i-0}^{k-1}\alpha_i t^i+\sum_{j=1}^{h}\delta_j\left(t-u_j\right)^{k-1}, \tag{2}$$

dengan

$$\left(t-u_j\right)_+^{k=1}=\begin{cases}\left(t-u_j\right)^{k-1}, t\geq u_j\\ 0, t<u_j\end{cases}, \tag{3}$$

dimana α dan δ adalah konstanta, dan $u_1,...,u_h$ adalah titik-titik knot. Menurut Budiantara (2001, 2004), spline pada dasarnya merupakan suatu fungsi yang digeneralisasikan dari polinomial. Spline $S(x)$ merupakan potongan polinomial yang dapat dinyatakan sebagai :

$$S(x)=P(x)+T(x),\ x\epsilon R, \tag{4}$$

dengan P merupakan polinomial dan T merupakan fungsi *truncated*.

2.3. Multivariate Adaptive Regression Spline (MARS)

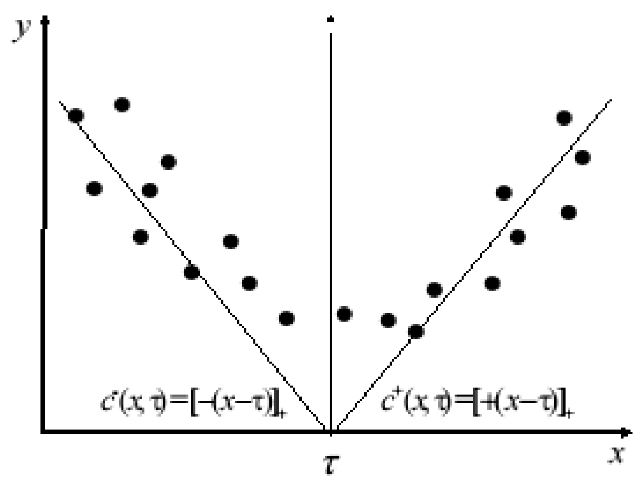
Metode statistik yang digunakan dalam analisis diskriminan multivariat telah banyak dikembangkan mulai dari metode analisis klasik hingga metode yang berbasis komputasi. Analisis diskriminan linear yang merupakan bagian dari analisis multivariat, digunakan dengan mengasumsikan kenormalan sisaan dan memiliki kovarian sisaan yang sama. Dalam hal terbatasnya informasi tentang pola data sehingga sulit untuk membuat asumsi terhadap bentuk kurva atau dalam bentuk pola data nonlinear dan berdimensi tinggi, diskriminan linear sering menimbulkan masalah dan sulit diinterpretasikan. Salah satu pendekatan yang dapat dilakukan adalah dengan menggunakan metode *Multivariate Adaptive Regression Spline* (MARS). MARS merupakan pendekatan regresi multivariat nonparametrik yang diharapkan dapat meningkatkan tingkat ketepatan hasil. Pendekatan ini digunakan untuk model regresi nonlinear yang merupakan pengembangan dari dari prosedur *Recursive Partitioning Regression* (RPR) dengan menggunakan spline untuk menduga model. Dalam hal pengelompokan, pendekatan regresi logistik digunakan dalam MARS (Abraham dkk., 2003).

RPR merupakan pendekatan dari fungsi $f(t)$ yang tidak diketahui dengan

$$\hat{f}(t)=\sum_{j=1}^{S}c_j(t)B_j(t) \tag{5}$$

dengan $B_j(t) = I\lfloor t \in R_j \rfloor, I[.]$ menunjukkan fungsi indikator yang mempunyai nilai 1 jika pernyataan benar $(t \in R_j)$, dan 0 (nol) jika salah, $c_j(t)$ merupakan koefisien (konstanta) yang ditentukan dalam sub region (Budiantara, 2006).

MARS merupakan suatu prosedur *adaptif* dikarenakan seleksi fungsi basis adalah database dan dikhususkan pada masalah yang ada. Model ini merupakan prosedur regresi nonparametrik yang tidak membuat anggapan khusus mengenai penekanan hubungan fungsional antara variabel tidak bebas dengan variabel bebas. MARS sangat berguna untuk data dimensional tinggi serta dapat menjelaskan kesesuaian fungsi multivariabel nonlinear. Manfaat khusus MARS terdapat pada kemampuannya untuk memperkirakan kontribusi fungsi basis sehingga efek aditif dan interaktif prediktor dapat menentukan variabel respon. MARS menggunakan perluasan pada bagian fungsi basis linear pada bentuk $c^{+}(x,\tau) = [+(x-\tau)]_{+}$, $c^{-}(x-\tau) = [-(x-\tau)]_{+}$, dimana $[q]_{+} = \max\{0, q\}$ dan τ merupakan knot variasi tunggal. Tiap fungsi merupakan bagian linear dengan nilai knot τ, dan disebut sebagai *reflected pair* (cerminan), sebagaimana yang diperlihatkan pada Gambar 1.



Gambar 1. Elemen Basis pada Regresi dengan MARS.

Titik-titik pada gambar ini mengilustrasikan data $(\bar{x}_i, \bar{y}_i)$, $i = 1,2,\ldots,N$, tersusun oleh spesifikasi input dimensional-*p* pada variabel *x*, dan yang berhubungan dengan respon dimensional-l yang ditentukan oleh variabel y (Taylan dkk., 2007).

Pada model MARS, pemilihan model menggunakan metode *stepwise* yang terdiri dari *forward* dan *backward*. *Forward stepwise* dilakukan untuk mendapatkan jumlah basis fungsi maksimum dengan kriteria pemilihan basis fungsi adalah meminimumkan *Average Sum Of Square Residual* (ASR). Menurut Friedman dan Silverman (1989), serta Friedman (1990, 1991), untuk memenuhi konsep parsemoni (model sederhana) dilakukan *backward stepwise* yaitu memilih basis fungsi yang dihasilkan dari *forward stepwise* dengan meminimumkan nilai *Generalized Cross-Validation* (GCV). Model MARS dapat ditulis sebagai berikut

$$\hat{f}(t) = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{K_m} \left[s_{km}.\left(t_{v(k,m)} - u_{km}\right)\right] \qquad (6)$$

dimana

a_0 = basis fungsi induk,
a_m = koefisien dari basis fungsi ke-*m*,
M = maksimum basis fungsi (*non constant basis fungsi*),
K_m = derajat interaksi,
S_{km} = nilainya ± 1,
$t_{v(k,m)}$ = variabel independen,
u_{km} = nilai knots dari variabel independen $t_{v(k,m)}$.

Penjabaran persamaan (6) dapat disajikan sebagai berikut $\hat{f}(t) = a_0 + \sum_{m=1}^{M} a_m \left[s_{1m}.\left(t_{v(1,m)} - u_{1m}\right)\right]$

$$+ \sum_{m=1}^{M} a_m \left[s_{1m}.\left(t_{v(1,m)} - u_{1m}\right)\right]\left[s_{2m}.\left(t_{v(2,m)} - u_{2m}\right)\right]$$

$$+ \sum_{m=1}^{M} a_m \left[s_{1m}.\left(t_{v(1,m)} - u_{1m}\right)\right]\left[s_{2m}.\left(t_{v(2,m)} - u_{2m}\right)\right]\left[s_{3m}.\left(t_{v(3,m)} - u_{3m}\right)\right] + \cdots \quad (7)$$

Secara umum persamaan (6) dapat dituliskan sebagai berikut

$$\hat{f}(t) = a_0 + \sum_{k_m=1} f_i(t_i) + \sum_{k_m=2} f_{ij}\left(t_i, t_j\right) + \sum_{k_m=3} f_{ijk}\left(t_i, t_j, t_k\right) + \ldots \quad (8)$$

Persamaan (8) menunjukkan bahwa penjumlahan pertama meliputi semua basis fungsi untuk satu variabel, penjumlahan kedua meliputi semua basis fungsi untuk interaksi antara dua variabel, penjumlahan ketiga meliputi semua basis fungsi untuk interaksi antara tiga variabel dan seterusnya.

Misalkan $V(m) = \{v(k,m)\}_1^{K_m}$ adalah himpunan dari variabel yang dihubungkan dengan basis fungsi B_m ke-*m*, maka setiap penjumlahan pertama pada Persamaan (8) dapat dinyatakan sebagai

$$f_i\left(t_i\right) = \sum_{\substack{K_m=1 \\ i \in V(m)}} a_m B_m\left(t_i\right), \quad (9)$$

$f_1(t_i)$ merupakan penjumlahan semua basis fungsi untuk satu variabel x_i, dan merupakan Spline dengan derajat q=1 yang mempresentasikan fungsi univariat. Setiap fungsi bivariat pada Persamaan (8) dapat ditulis sebagai

$$f_{ij}\left(t_i, t_j\right) = \sum_{\substack{K_m=2 \\ (i,j) \in V(m)}} a_m B_m\left(t_i, t_j\right), \quad (10)$$

yang mempresentasikan penjumlahan semua basis fungsi dua variabel t_i dan t_j. Penambahan ini untuk menghubungkan kontribusi univariat, yang dituliskan sebagai berikut :

$$f_{ij}^{*}\left(t_i, t_j\right) = f_i\left(t_i\right) + f_j\left(t_j\right) + f_{ij}\left(t_i, t_j\right). \quad (11)$$

Untuk fungsi trivariat pada penjumlahan yang ketiga diperoleh dengan menjumlahkan semua basis fungsi untuk tiga variabel, yang dituliskan menjadi

$$f_{ijk}(t_i, t_j, t_k) = \sum_{\substack{K_m=3 \\ (i,j,k)\in V(m)}} a_m B_m(t_i, t_j, t_k) \ . \tag{12}$$

Penambahan fungsi univariat dan bivariat mempunyai kontribusi dalam bentuk

$$\begin{aligned} f^*_{ijk}(t_i, t_j, t_k) = {} & f_i(t_i) + f_j(t_j) + f_k(t_k) + f_{ij}(t_i, t_j) + f_{ik}(t_i, t_k), \\ & + f_{jk}(t_j, t_k) + f_{ijk}(t_i, t_j, t_k) \ . \end{aligned} \tag{13}$$

Persamaan (8) merupakan dekomposisi dari analisis varians untuk tabel kontingensi, yang dikenal dengan dekomposisi ANOVA dari model MARS. Interpretasi model MARS melalui dekomposisi ANOVA adalah merepresentasikan variabel yang masuk dalam model, baik untuk satu variabel maupun interaksi antara variabel, selanjutnya merepresentasikan secara grafik. Penambahan aditif persamaan (9) dapat ditunjukkan dengan membuat plot antara $f_i(t_i)$ dengan t_i sebagai salah satu model aditif. Kontribusi interaksi antara dua variabel dapat divisualisasikan dengan membuat plot antara $f^*_{ij}(t_i, t_j)$ dengan t_i dan t_j menggunakan kontur plot. Model dengan interaksi yang lebih tinggi dalam visualisasi dapat dibuat dengan menggunakan plot dalam beberapa variabel fixed dengan variabel komplemen (Budiantara, 2006).

Menurut Friedman (1991), algoritma MARS untuk memperkirakan model fungsi *f(x)* terdiri dari dua algoritma, yaitu:

1. Algoritma *forward stepwise* (seselangkah maju): *forward stepwise* mencari fungsi basis berlangsung dengan fungsi basis yang konstan, satu-satunya yang ada pada awalnya. Pada tiap langkah, dari tiap basis fungsi terpilih pembagian yang meminimalkan beberapa *"lack of fit"* atau kriteria kekurangcocokan dari seluruh kemungkinan pembagian yang ada. Prosesnya berhenti ketika nilai M_{max} dicapai.
2. Algoritma *backward stepwise* (seselangkah mundur): tujuan dari algoritma ini adalah untuk mencegah *over-fitting* dengan mengurangi kerumitan model tanpa menurunkan kecocokan dengan data. Karena itu, algoritma *backward stepwise* melibatkan penjauhan dari fungsi basis model yang memberikan kontribusi peningkatan terkecil pada kesalahan sisa hasil perkalian di tiap tempat serta menghasilkan model estimasi optimal $\hat{f}_\alpha$ dengan mempertimbangkan setiap jumlah suku, yang dinamakan α. Untuk memperkirakan nilai optimal α, GCV dapat digunakan, yang menunjukkan kekurangcocokan ketika menggunakan MARS. Kriteria tersebut ditetapkan dengan
3.

$$GCV := \frac{\sum_{i=1}^{N}\left(y_i - \hat{f}_\alpha(x_i)\right)^2}{\left(1 - M(\alpha)/N\right)^2} , \tag{14}$$

dimana *M(α)* adalah jumlah knot yang terpilih pada proses ke depannya, $M(\alpha) = u + dM$, *N* adalah jumlah sampel observasi, *M* adalah jumlah fungsi basis yang secara linear tidak terikat, dan *d* adalah nilai untuk optimisasi fungsi basis juga sebagai parameter pelancar dalam prosedur. Namun, algoritma *backward stepwise* dapat dihilangkan (tidak digunakan) dalam model MARS. Sebagai gantinya digunakan terminologi penalti di samping perkiraan hasil perkalian terkecil untuk mengontrol kekurangcocokan dari sudut pandang kompleksitas dari perkiraan. Menurut Taylan dkk. (2007), metode tersebut dinamakan *Penalized Residual Sum of Square* (PRSS).

Algoritma PRSS yang digunakan dengan fungsi basis M_{max} telah terakumulasi pada *algoritma forward stepwise*. Untuk model MARS, PRSS memiliki bentuk sebagai berikut

$$PRSS := \sum_{i=1}^{N}\left(y_i - f\left(\bar{x}_i\right)\right)^2 + \sum_{m=1}^{M_{\max}}\lambda_m \sum_{\substack{|\alpha|=1 \\ \alpha=(\alpha_1,\alpha_2)}}^{2} \sum_{\substack{r<s \\ r,s\in V(m)}} \int \theta_m^2 \left[D_{r,s}^{\alpha}\psi_m\left(t^m\right)\right]^2 dt^m, \qquad (15)$$

dimana $V(m) = \left\{K_j^m \mid j = 1,2,...,K_m\right\}$ adalah perangkat variabel ke-*m* yang berkenaan dengan fungsi basis ψ_m, $t^m = \left(t_{m1}, t_{m2}, ..., t_{mK_m}\right)^T$ menggambarkan vektor variabel yang menambah/memperbesar fungsi basis ke-*m*. Selanjutnya,

$$D_{r,s}^{\alpha}\psi_m\left(t^m\right) := \frac{\partial^{\alpha}\psi_m}{\partial^{\alpha_1} t_r^m \partial^{\alpha_2} t_s^m}\left(t^m\right). \qquad (16)$$

Untuk $\alpha = (\alpha_1, \alpha_2), |\alpha| = \alpha_1 + \alpha_2$, dimana $\alpha_1, \alpha_2 \in \{0,1\}$, α_i =2, turunan $D_{r,s}^{\alpha}\psi_m\left(t^m\right)$ tak bersisa, dan dengan menunjukan indeks $r < s$ digunakan aplikasi teorema *Schwarz* (Taylan dkk., 2007).

3. Metodologi Penelitian

Sumber data yang akan dipergunakan sebagai aplikasi penggunaan algoritma PRSS pada penentuan model MARS adalah data sekunder dimana jumlah pengamatan sebanyak sembilan puluh sembilan dengan dua variabel respon yaitu Y_1 dan Y_2, merupakan tekanan darah pasien yang diukur sebelum dan sesudah diberikan ramuan buah mengkudu dan daun kumis kucing, dan tiga variabel bebas yaitu umur (X_1), jenis kelamin (X_2) dan jenis pekerjaan (X_3). Untuk variabel X_2 dikategorikan dalam dua indikator yakni 1 untuk laki-laki dan 2 untuk perempuan, sedangkan untuk variabel X_3 dikategorikan dalam lima indikator yaitu 1 untuk pegawai negeri, 2 untuk swasta, 3 untuk pensiunan, 4 untuk ibu rumah tangga, dan 5 untuk lainnya. Data tersebut diperoleh dari Tesis karya Stang**,** Program Pascasarjana, Universitas Airlangga. Adapun langkah-langkah yang dilakukan adalah sebagai berikut

1. Membuat basis fungsi dengan menggunakan metode *forward stepwise* untuk mendapatkan jumlah basis fungsi maksimum dengan kriteria pemilihan basis fungsi adalah meminimumkan ASR.
2. Memilih basis fungsi dengan menggunakan algoritma PRSS yang digunakan dengan fungsi basis M_{max} yang telah terakumulasi pada algoritma *forward stepwise* dengan menggunakan rumus (15).
3. Menetukan model MARS dengan variasi pada basis fungsi, interaksi antar variabel independen, dan minimum observasi sub region dengan menggunakan *software* MARS.
4. Memilih model MARS terbaik dengan menggunakan kriteria yang berdasarkan pada nilai GCV, nilai minimum MSE dan koefisien determinasi terbesar (R^2).

4. Hasil dan Pembahasan

Model MARS pada penurunan tekanan darah terhadap pemberian ramuan buah mengkudu dan daun kumis kucing pada pasien hipertensi diolah dengan menggunakan software MARS 2.0 yang dikembangkan oleh Salford-Systems. Untuk mendapatkan jumlah basis fungsi maksimum yang menggunakan metode *forward stepwise* digunakan program MARS dengan hasil sebagai berikut.

a. Jumlah basis fungsi dengan variabel respon kontinu Y_1 dan variabel tak bebas (*predictor*) X_1, X_2, dan X_3 sebagai berikut.

Tabel 1. Jumlah Basis Fungsi dengan Variabel Y_1 terhadap $\boldsymbol{X}$.

Basis Fungsi	**Nilai GCV**	**Variabel**	**Knot**
0	98.217		
2 1	97.388	X_3	
3	98.767	X_1	23
5 4	100.459	X_1	48
7 6	100.899	X_1	60
9 8	102.505	X_3	
11 10	104.193	X_3	
12	104.204	X_1	23
14 13	103.089	X_1	65
15	104.298	X_3	

Berdasarkan Tabel 1, jumlah basis fungsi maksimum adalah sembilan dengan nilai GCV terbesar dihasilkan pada variabel X_3 sebesar 104.298. Terdapat beberapa basis fungsi yang tidak memiliki titik knot yang menunjukkan bahwa basis fungsi tersebut dihasilkan oleh variabel bebas kategorik.

b. Jumlah basis fungsi dengan variabel respon kontinu Y_2 dan variabel bebas (*predictor*) X_1, X_2, dan X_3 adalah sebagai berikut.

Tabel 2. Jumlah Basis Fungsi dengan Variabel Y_2 terhadap $\boldsymbol{X}$.

Basis Fungsi	**Nilai GCV**	**Variabel**	**Knot**
0	189.176		
1	186.453	X_1	23
3 2	189.315	X_3	
4	193.91	X_1	23
6 5	199.178	X_1	60
8 7	205.46	X_2	
10 9	212.033	X_2	
11	214.073	X_1	23
12	216.143	X_1	23
13	218.243	X_1	23
14	220.373	X_1	23
15	222.536	X_1	23

Berdasarkan Tabel 2 di atas diperoleh jumlah basis fungsi maksimum sebanyak 11 dengan nilai GVC terbesar terdapat pada variabel X_1 sebesar 222.536 dan terletak pada titik knot 23, yang menunjukkan bahwa titik knot tersebut dihasilkan oleh variabel kontinu.

Selanjutnya untuk menghasilkan model *MARS*, jumlah basis fungsi maksimum yang telah diperoleh akan dipilih dengan menggunakan metode *backward stepwise* atau metode PRSS dengan hasil sebagai berikut.

- Model akhir dengan variabel respon kontinu Y_1 dan variabel bebas X_1, X_2, dan X_3 adalah sebagai berikut.

Tabel 3. Model Variabel Respon Y_1 terhadap $\boldsymbol{X}$.

Basis Fungsi	Koefisien	Variabel	Nilai GCV	MSE	R^2
0	80.832				
1	18.448	X_3			
4	-1.911	X_1	96.917	90.946	0.117
10	1.841	X_3			
12	1.16	X_1			
13	1.489	X_1			

Berdasarkan Tabel 3 dapat dilihat bahwa model akhir dari MARS terdiri dari lima basis fungsi dengan model sebagai berikut.

$$\hat{Y}_1 = 80.832 + 18.448 * BF1 - 1.911 * BF4 + 1.841 * BF10$$
$$+ 1.160 * BF12 + 1.489 * BF13$$

dimana

$BF1 = (X_3 = 4 \text{ atau } X_3 = 5)$,
$BF2 = (X_3 = 1 \text{ atau } X_3 = 2 \text{ atau } X_3 = 3)$,
$BF4 = \max(0, X_1 - 48)$,
$BF10 = (X_3 = 4 \text{ atau } X_3 = 5) * BF4$,
$BF12 = \max(0, X_1 - 23) * BF2$,
$BF13 = \max(0, X_1 - 65) * BF2$.

Model ini menunjukkan bahwa pada variabel respon Y_1 dipengaruhi oleh basis fungsi (variabel bebas kategorik dan kontinu). Model di atas menunjukkan pula bahwa tekanan darah pada pasien hipertensi sebelum diberi ramuan buah mengkudu dan daun kumis kucing dipengaruhi variabel-variabel bebasnya yaitu umur dan jenis pekerjaan.

- Model akhir dengan variabel respon kontinu Y_2 dan variabel prediktor X_1, X_2, dan X_3 sebagai berikut.

Tabel 4. Model Variabel Respon Y_2 terhadap $\boldsymbol{X}$.

Basis Fungsi	Koefisien	Variabel	Nilai GCV	MSE	R2
0	95.8		188.087	181.323	0.042
5	-0.521	X_1			

Berdasarkan Tabel 4, model MARS untuk respon kontinu Y_2 hanya menghasilkan satu basis fungsi dengan model $\hat{Y}_2 = 95.8 - 0.521 * BF5$, dengan $BF5$ = max (0, X_1 – 60). Model di atas menunjukkan bahwa variabel respon Y_2 dipengaruhi oleh basis fungsi (variabel bebas kontinu), ini menunjukkan pula bahwa tekanan darah pada pasien hipertensi setelah diberi ramuan buah mengkudu dan daun kumis kucing dipengaruhi variabel bebas yaitu umur.

Untuk melihat variabel-variabel bebas yang paling berpengaruh terhadap variabel respon ditunjukkan dalam Tabel 5 dan 6 berikut.

Tabel 5. Pengaruh ***X*** terhadap Y_1.

Variabel	**Nilai Kepentingan**	**GCV**
X_3	100	99.775
X_1	40.616	97.388
X_2	0	96.917

Tabel 6. Pengaruh ***X*** terhadap Y_2.

Variabel	**Nilai Kepentingan**	**GCV**
X_1	100	189.176
X_2	0	188.087
X_3	0	188.087

Tabel 5 menunjukkan pengaruh variabel bebas terhadap variabel respon kontinu Y_1, dimana variabel X_3 sangat berpengaruh yang dapat dilihat dari nilai kepentingan dan nilai GCV dengan nilai masing-masing 100 dan 99.775. Sementara variabel X_1 memiliki nilai kepentingan sebesar 40.616 dan nilai GCV 97.388. Sedangkan variabel X_2 tidak memiliki pengaruh yang berarti terhadap variabel respon Y_1, dapat dilihat dari nilai kepentingan yang besarnya 0 dan nilai GCV yang paling kecil yaitu 96.917. Tabel 6 menunjukkan pengaruh variabel bebas X_1, X_2, X_3 terhadap variabel respon kontinu Y_2. Dari tabel terlihat bahwa variabel X_1 merupakan satu-satunya variabel yang sangat berpengaruh terhadap variabel respon Y_2 dengan nilai kepentingan sebesar 100 dan nilai GCV sebesar 189.176. Sementara untuk variabel X_2 dan X_3 tidak memiliki pengaruh yang berarti terhadap variabel respon yang dapat dilihat dari nilai kepentingan dan nilai GCV yang sama yaitu masing-masing 0 dan 188.087.

5. Kesimpulan dan Saran

5.1. Kesimpulan

Dari hasil simulasi diperoleh dua persamaan model MARS yaitu

1. Persamaan dengan variabel respon Y_1 sebagai berikut

$$\hat{Y}_1 = 80.832 + 18.448 * BF1 - 1.911 * BF4 + 1.841 * BF10$$
$$+ 1.160 * BF12 + 1.489 * BF13$$

dengan

$BF1$ = (X_3 = 4 atau X_3 = 5),
$BF2$ = (X_3 = 1 atau X_3 = 2 atau X_3 = 3),
$BF4$ = max (0, X_1 – 48),
$BF10$ = (X_3 = 4 atau X_3 = 5) * $BF4$,
$BF12$ = max (0, X_1 – 23) * $BF2$,
$BF13$ = max (0, X_1 – 65) * $BF2$.

2. Persamaan dengan variabel respon Y_2 sebagai berikut

$$\hat{Y}_2 = 95.8 - 0.521 * BF5$$

dengan $BF5$ = max (0, X_1 – 60).

Hasil analisis model MARS, menunjukkan bahwa terjadi penurunan tekanan darah setelah pemberian ramuan buah mengkudu dan daun kumis kucing terhadap pasien hipertensi yang berobat di laboratorium P4OT Surabaya. Hal ini dapat dilihat dari model yang dihasilkan pada tekanan darah setelah pemberian ramuan (Y_2).

Ketiga variabel bebas yaitu umur (X_1), jenis kelamin (X_2) dan jenis pekerjaan (X_3) yang paling berpengaruh terhadap tekanan darah sebelum diberikan ramuan (variabel respon kontinu Y_1) adalah jenis pekerjaan (X_3) dengan nilai kepentingan dan nilai GCV masing-masing sebesar 100 dan 99.775. Sedangkan yang paling berpengaruh terhadap tekanan darah setelah diberikan ramuan (Y_2) adalah umur (X_1) dengan nilai kepentingan dan nilai GCV masing-masing sebesar 100 dan 189.176.

5.2. Saran

Dalam tulisan penulis hanya membahas model MARS dengan respon kontinu, bagi pembaca yang ingin mendalami masalah ini lebih jauh lagi maka perlu dikembangkan suatu model MARS dengan respon kategorik.

Daftar Pustaka

Abraham, A. dkk., 2003. *Instruction Detection Systems Using Adaptive Regression Splines.* Natural Computation Lab. Dep. Computer Science, Oklahama State University, Tulsa, U.S.A.

Otok, B.W., 2008. *Materi Pelatihan Nonparametric Regression Multivariate Adaptive Regression Spline* (*MARS*). Laboratorium Statistika Sosial dan Bisnis, Institut Teknologi Sepuluh November, Surabaya.

Budiantara, I.N., 2001. Aplikasi spline estimator terboboti. *Jurnal Teknik Industri Universitas Petra Surabaya*, Vol. 3, 57-62.

_______________, 2004. Spline: historis, motivasi dan perannya dalam regresi nonparametrik. *Makalah Pembicara Utama pada Konferensi nasional Matematika XII,* Jurusan

Matematika, Fakultas matematika dan Ilmu Pengetahuan Alam Universitas Udayana, Denpasar- Bali.

Budiantara, I.N. dkk., 2006. Pemodelan B-Spline dan MARS pada nilai ujian masuk terhadap IPK mahasiswa Jurusan Disain Komunikasi Visual UK. Petra Surabaya. *Jurnal UK Petra*, Surabaya.

Hidayat, U., 2003. *Analisis Pengelompokan dengan Metode Multivariate Adaptive Regression Spline* (*MARS*). *Studi Kasus: Pengelompokan Desa/Kelurahan di Jawa Timur,* Surabaya.

Taylan dkk., 2007. *Multivariate Adaptive Regression Spline and Continuous Optimization for Modern Application in Science, Economy and Technology*. Department of Mathematics, Dicle University, Turkey.

Wahba, G., 1990. Spline models for observational data. *SIAM, CBMS-NSF Regional Conferense Series in Applied Mathematics*, Philadelphia.