

**PEMODELAN *MULTIVARIATE ADAPTIVE REGRESSION SPLINES* (MARS)
PADA FAKTOR-FAKTOR RESIKO ANGKA KESAKITAN DIARE
(Studi Kasus : Angka Kesakitan Diare Di Jawa Tengah, Jawa Timur Dan Daerah
Istimewa Yogyakarta Tahun 2011)**

Wasis Wicaksono¹, Yuciana Wilandari², Suparti³

¹Mahasiswa Jurusan Statistika FSM UNDIP

^{2,3}Staff Pengajar Jurusan Statistika FSM UNDIP

ABSTRACT

Diarrhea morbidity is a number of diarrhea suffers in specific region in period of one year per 1000 populations. Diarrhea morbidity is the impact from some factors such as environment, education, socioeconomic, nutrition and foods. Environmental factors that can affect the morbidity of diarrhea include the percentage of families who have a healthy latrine and percentage of households using clean water. For educational factors include the average length of school and literacy rates. On socio-economic factors include the percentage of poor and average people per household. While the food and nutritional factors are the percentage TUPM (Public Places and Food Management) healthy. Diarrhea morbidity can be pressed by analyzing those factors so that the prevention can be devised. Regression curve is used to draw the relationship of response variable and predictor variable and mostly approached by parametric regression, where the curve design is known (such as linear, quadratic and cubic). If curve design is unknown, then regression curve can be derived by approaching using non parametric regression. *Multivariate Adaptive Regression Spline* (MARS) is one of nonparametric regression method that can be used on high dimension data. the best MARS model is derived by combination of Minimal Observation (MO), Maximum Basic Function (BF), and Maximal Interaction (MI) through *trial and error*. MARS model to predict diarrhea morbidity in Central Java, East Java and Yogyakarta is MARS (MO=2;BF=28;MI=3) and equation is $\hat{Y} = -0.526742 + 0.264444 * BF2 + 12.2382 * BF5 - 7.76719 * BF15 + 4.96445 * BF17$

Keywords: Diarrhea Morbidity, Nonparametric regression, MARS

1. PENDAHULUAN

Penyakit diare masih menjadi masalah kesehatan masyarakat di Indonesia karena angka kesakitan dan angka kematian yang ditimbulkan tergolong tinggi. Berdasarkan hasil Riset Kesehatan Dasar (Riskesdas) 2007, penyakit diare menempati urutan ke-3 penyakit menular penyebab kematian pada semua umur di Indonesia. Sedangkan angka kesakitan yang disebabkan oleh diare pada tahun 2010 mencapai 411 per 1000 penduduk. Tingginya angka kesakitan membuat penyakit diare sering menimbulkan Kejadian Luar Biasa (KLB). Jumlah kasus KLB Diare pada tahun 2010 sebanyak 2.580 dengan kematian sebesar 77 kasus sedangkan pada tahun 2009 KLB Diare sebanyak 3.037 kasus dengan kematian sebesar 21 kasus (Dinas Kesehatan, 2010).

Penyebaran penyakit diare dapat terjadi secara langsung maupun tak langsung. Diare dapat ditularkan dari orang satu ke orang lain secara langsung melalui *fecal – oral* dengan media penularan utama adalah makanan atau minuman yang terkontaminasi agen penyebab diare (Suharyono, 1991). Sedangkan penularan penyakit diare secara tidak langsung tak hanya disebabkan oleh kebersihan dan kesehatan lingkungan, tetapi juga dipengaruhi oleh faktor eksternal lainnya, seperti : faktor sosial, ekonomi dan juga

pendidikan. Untuk menekan angka penyebaran diare perlu dilakukan analisis pada faktor-faktor tersebut sehingga dapat dilakukan tindakan pencegahan terhadap penyakit diare.

Salah satu metode untuk menganalisis faktor-faktor penyebab diare adalah *Multivariate Adaptive Regression Spline* (MARS) yang pertama kali dipopulerkan oleh Friedman (1991). Model MARS berguna untuk mengatasi permasalahan data yang berdimensi tinggi, yaitu data yang memiliki jumlah variabel prediktor sebesar $3 \leq n \leq 20$. MARS merupakan pengembangan dari pendekatan *Recursive Partition Regression* (RPR) yang dikombinasikan dengan metode *spline* sehingga model yang dihasilkan kontinyu pada knot.

Berdasarkan uraian di atas, pokok permasalahan yang dibahas oleh penulis adalah menganalisis faktor-faktor yang mempengaruhi banyaknya angka kesakitan diare dan menggunakan metode MARS untuk mendapatkan model angka kesakitan diare yang dibatasi pada faktor-faktor yang mempengaruhi banyaknya angka kesakitan diare pada 78 kabupaten / kota di Jawa Tengah, Jawa Timur dan Daerah Istimewa Yogyakarta tahun 2011.

2. TINJAUAN PUSTAKA

Diare adalah suatu penyakit dengan tanda-tanda adanya perubahan bentuk dan konsistensi dari tinja, yang melembek sampai mencair dan bertambahnya frekuensi buang air besar biasanya tiga kali atau lebih dalam sehari (Depkes RI, 2005). Diare bisa menyerang siapa saja. Menurut hasil Riskesdas 2007, diare merupakan penyakit menular penyebab kematian peringkat ke-3 pada semua umur. Sedangkan angka kesakitan diare pada tahun 2010 mencapai 411 per 1000 penduduk. Angka kesakitan adalah angka insidensi (q.v) yang dipakai untuk menyatakan jumlah keseluruhan orang yang menderita penyakit yang menimpa sekelompok penduduk pada periode waktu tertentu.

Diare disebabkan oleh beberapa organisme seperti bakteri, virus dan parasit. Beberapa organisme tersebut biasanya menginfeksi saluran pencernaan manusia melalui makanan dan minuman yang telah tercemar oleh organisme tersebut (Soenarto, 2011). Menurut Adisasmito (2007) ada beberapa faktor risiko penyebab diare, antara lain :

1. Faktor lingkungan

Diare disebabkan oleh bakteri, virus dan parasit yang cepat berkembang pada lingkungan yang tidak sehat. Penderita diare umumnya tinggal di daerah yang kotor dan tidak menerapkan Perilaku Hidup Bersih dan Sehat (PHBS). Contoh masyarakat yang menerapkan PHBS antara lain menggunakan air bersih untuk keperluan sehari-hari, seperti Mandi, Cuci dan Kakus (MCK) maupun untuk keperluan memasak. Selain itu keluarga yang ber-PHBS harus memiliki saluran pembuangan limbah / jamban yang sehat, agar bakteri penyebab penyakit diare tidak dapat menyebar luas.

2. Faktor pendidikan

Pendidikan merupakan faktor penting dalam upaya pencegahan dan penanggulangan penyakit diare. Tingkat pendidikan yang rendah membuat masyarakat sulit untuk menyerap berbagai informasi tentang penyakit diare, sehingga kurang tanggap dalam melakukan tindakan pencegahan maupun pengobatan diare. Tinggi rendahnya tingkat pendidikan masyarakat dapat dilihat dari berbagai indikator, antara lain rata-rata lama sekolah dan angka melek huruf di setiap kabupaten/ kota.

3. Faktor sosial ekonomi

Kondisi sosial ekonomi masyarakat sangat mempengaruhi banyaknya penderita

diare. Karena sebagian besar penderita diare berasal dari keluarga yang kondisi sosial ekonominya rendah. Indikator tinggi rendahnya kondisi sosial ekonomi di suatu kabupaten/ kota adalah persentase penduduk miskin dan rata-rata jiwa per rumah tangga.

4. Faktor gizi dan makanan

Salah satu media penyebaran bakteri, virus dan parasit penyebab diare adalah makanan. Makanan yang mengandung banyak zat-zat kimia dan kurang higienis dapat mempercepat penyebaran diare. Untuk itu perlu ditingkatkan pengawasan terhadap Tempat Umum dan Pengelolaan Makanan (TUPM) agar terjamin keamanan dan kesehatannya.

2.1 Analisis Regresi

Analisis regresi adalah suatu metode statistika yang umum digunakan untuk meneliti hubungan antara peubah bebas (variabel prediktor) dengan peubah tak bebas (variabel respon). Dalam analisis regresi terdapat dua pendekatan untuk mengestimasi kurva regresi yaitu regresi parametrik dan regresi nonparametrik. Jika kurva regresi merupakan model parametrik maka disebut sebagai regresi parametrik dan apabila model yang diasumsikan ini benar, maka pendugaan parametrik sangat efisien, tetapi jika tidak, menyebabkan interpretasi data yang menyesatkan (Hardle, 1990). Regresi parametrik memiliki asumsi-asumsi yang apabila salah satu asumsi tak terpenuhi, maka hasil regresi parametrik tersebut dianggap tidak valid.

Asumsi-asumsi tersebut adalah :

1. Normalitas
2. Kesamaan Varian (Homoskedastisitas)
3. Autokorelasi
4. Multikolinieritas

2.2 Regresi Spline

Regresi spline merupakan salah satu metode regresi nonparametrik yang bertujuan untuk memperkecil keragaman dan mengestimasi perilaku data yang cenderung berbeda. Pendekatan spline memiliki kemampuan untuk mengatasi pola data yang menunjukkan naik atau turun yang tajam dengan bantuan titik-titik knot, serta kurva yang dihasilkan relatif mulus. Titik-titik knot ialah titik perpaduan bersama yang menunjukkan terjadinya perubahan pola perilaku data (Budiantara, 2009). Spline merupakan potongan (*piecewise*) polinomial orde q dan memiliki turunan yang kontinyu dengan knot sampai orde (q-1) (Friedman, 1991). Dalam spline univariat dengan K knot memiliki fungsi basis sebagai berikut :

$$1, \quad \{x^j\}_1^q, \quad \{(x - t_k)_+\}_1^K$$

Sehingga model spline dapat dituliskan menjadi :

$$f(x) = \beta_0 + \beta_1 x + \dots + \beta_q x^q + \sum_{k=1}^K \gamma_k (x - t_k)_+^q$$

dimana $q \geq 1$ dan (t_1, t_2, \dots, t_K) merupakan titik titik knot. Sedangkan fungsi dari *truncated power* $(x - t_k)_+^q$ adalah:

$$(x - t_k)_+^q = \begin{cases} (x - t_k)^q & ; x - t_k > 0 \\ 0 & ; x - t_k \leq 0 \end{cases}$$

2.3 Recursive Partitioning Regression (RPR)

Penghitungan komputasi sangat diperlukan dalam pengolahan data, karena akan selalu memperoleh hasil yang sesuai dengan algoritmanya. *Recursive Partitioning Regression* (RPR) merupakan salah satu dari program komputasi yang memiliki keunggulan dalam mengolah data yang berdimensi tinggi. Tujuan dari RPR adalah menggunakan data untuk mengestimasi *subregion* dan parameter yang berasosiasi pada setiap *subregion* (Friedmann, 1991).

Fungsi basis pada RPR memiliki bentuk

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

- dengan K_m = derajat interaksi
 s_{km} = tanda pada titik knot (nilainya ± 1)
 $x_{v(k,m)}$ = variabel prediktor
 t_{km} = nilai *knots* dari variabel prediktor $x_{v(k,m)}$

2.4 Multivariate Adaptive Regression Spline (MARS)

Multivariate Adaptive Regression Spline (MARS) merupakan salah satu metode regresi nonparametrik yang pertama kali dipopulerkan oleh Friedman (1991).

Model MARS berguna untuk mengatasi permasalahan data yang berdimensi tinggi. Yang dimaksud dengan data berdimensi tinggi adalah data yang memiliki jumlah variabel prediktor sebesar $3 \leq n \leq 20$. Menurut Friedman (1991), jumlah sampel yang diharuskan untuk pendekatan MARS adalah $50 \leq N \leq 1000$. MARS memiliki estimator model sebagai berikut :

$$\hat{f}(x) = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} [s_{km} \cdot (x_{v(k,m)} - t_{km})]_{\pm}$$

- dengan a_0 = parameter fungsi basis induk
 a_m = parameter fungsi basis ke-m (vektor koefisien) $m=1,2,\dots,M$
 M = maksimum basis fungsi (*nonconstant* basis fungsi)

2.5 Pemilihan Model MARS Terbaik

Pada MARS, model terbaik diperoleh berdasarkan nilai *generalized cross-validation* (GCV) dari model tersebut mempunyai nilai yang paling rendah (minimum) diantara model-model yang lain (Otok *et al.*, 2012).

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

dengan

- MSE = Kuadrat tengah eror
 y_i = variabel respon
 x_i = variabel prediktor
 N = banyaknya pengamatan
 $\hat{f}_M(x_i)$ = nilai taksiran variabel respon pada M fungsi basis di x_i
 M = maksimal jumlah fungsi basis
 $\tilde{C}(M)$ = $C(M) + dM$
 $C(M)$ = $Trace [B (B^T B)^{-1} B^T] + 1$; B adalah matriks dari M fungsi basis
 d = nilai ketika setiap fungsi basis mencapai nilai optimal ($2 \leq d \leq 4$)

Menurut Friedman (1991) ada beberapa hal yang harus diperhatikan dalam pemodelan menggunakan metode MARS, yaitu :

1. Knot, nilai variabel prediktor ketika *slope* suatu garis regresi mengalami perubahan disebut dengan *knot* (Ryan & Porth, 2007). Setiap garis regresi mendefinisikan satu *region* sehingga *knot* dapat didefinisikan sebagai akhir dari satu *region* dan awal dari *region* yang lain. Minimum jarak antara knot atau minimum observasi antara knot (MO) ditentukan dengan cara *trial and error* sampai diperoleh GCV minimum.
2. Fungsi basis, yaitu kumpulan dari fungsi yang digunakan untuk mewakili informasi. Fungsi basis terdiri dari satu atau lebih variabel. Fungsi basis ini merupakan fungsi parametrik yang didefinisikan pada *region*. Pada umumnya fungsi basis yang dipilih berbentuk polinomial dengan derivatif yang kontinu pada setiap *knot*. Maksimum fungsi basis (BF) yang dibolehkan adalah sebanyak dua sampai empat kali dari banyaknya variabel prediktor yang digunakan.
3. Interaksi, merupakan *cross product* antar variabel yang saling berkorelasi. Friedman (1991) membatasi jumlah maksimum interaksi (MI) yang diperbolehkan yaitu 1, 2, dan 3. Karena apabila terdapat lebih dari 3 interaksi, maka akan menimbulkan interpretasi model yang sangat kompleks.

2.6 Pengujian Signifikansi Model MARS

Apabila residual dalam model MARS memenuhi asumsi pada regresi parametrik, maka dilakukan pengujian untuk mengecek signifikansi parameter dan mengevaluasi kecocokan model. Pengujian dilakukan dengan menguji koefisien regresi secara simultan maupun secara parsial.

1. Pengujian koefisien regresi simultan

- a. Rumusan hipotesis :

$H_0 : a_1 = a_2 = \dots = a_M = 0$ (model tidak signifikan)

$H_1 : \text{minimal terdapat satu } a_m \neq 0 ; m = 1, 2, \dots, M$ (model signifikan)

- b. Taraf signifikansi : α

- c. Statistik uji :

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

- d. Daerah Kritis:

Tolak H_0 jika nilai $F > F_{\alpha(M; N-M-1)}$ atau $p_value < \alpha$

2. Pengujian koefisien regresi parsial

- a. Rumusan hipotesis :

$H_0 : a_m = 0$ (koefisien a_m tidak berpengaruh terhadap model)

$H_1 : a_m \neq 0 ; \text{ untuk setiap } m, \text{ dimana } m = 1, 2, \dots, M$ (koefisien a_m berpengaruh terhadap model)

- b. Taraf signifikansi : α

- c. Statistik Uji :

$$t_{hitung} = \frac{\hat{a}_m}{se(\hat{a}_m)}$$

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

- d. Daerah Kritis:

Tolak H_0 jika $t > t_{(\frac{\alpha}{2}; N-M)}$ atau $p_value < \alpha$

3. METODOLOGI PENELITIAN

Data yang digunakan dalam penelitian tugas akhir ini berupa data sekunder yang diambil dari *website* resmi BPS dan Dinas Kesehatan. Data yang diambil berupa data hasil Susenas 2011 yang terangkum dalam “Jawa Tengah dalam Angka 2011”, “Yogyakarta dalam Angka 2011”, “Jawa Timur dalam Angka 2011” dan data dari dinas kesehatan yang terangkum dalam “Profil Kesehatan Jawa Tengah 2011” “Profil Kesehatan Yogyakarta 2011” “Profil Kesehatan Jawa Timur 2011”.

Adapun variabel yang digunakan dalam penelitian ini terbagi menjadi 2, yaitu variabel respon dan variabel prediktor.

1. Variabel respon :

Y = Angka kesakitan diare

2. Variabel-variabel prediktor :

X_1 = persentase keluarga yang memiliki jamban sehat

X_2 = persentase keluarga yang menggunakan air bersih

X_3 = persentase TUPM (Tempat Umum dan Pengelolaan Makanan) sehat

X_4 = rata-rata lama sekolah

X_5 = persentase melek huruf penduduk usia di atas 10 tahun

X_6 = rata-rata jiwa per rumah tangga

X_7 = persentase penduduk miskin

3.1 Metode Analisis

3.3.1 Analisis Deskriptif

Analisis deskriptif merupakan analisis yang menggambarkan keadaan suatu hal secara umum dan bertujuan untuk mempermudah penafsiran dan penjelasan dengan menganalisis tabel, grafik, atau diagram. Analisis deskriptif digunakan sebagai pendukung untuk menambah dan mempertajam analisis yang dilakukan. Dalam hal ini, analisis deskriptif digunakan untuk mengetahui bagaimana gambaran umum variabel-variabel penentu angka kesakitan diare setiap kabupaten/ kota di Jawa Tengah, Daerah Istimewa Yogyakarta, dan Jawa Timur.

3.3.2 Analisis MARS

Dalam penelitian ini data dianalisis dengan metode MARS. Untuk mendapatkan model terbaik dilakukan langkah-langkah sebagai berikut :

1. Melakukan standardisasi terhadap setiap variabel yang terlibat (variabel respon dan variabel prediktor) agar memiliki skala nilai yang sama.
2. Membuat grafiks plot antara variabel respon angka kesakitan diare kabupaten/ kota dengan masing-masing variabel prediktor sebagai langkah awal pendekteksian pola hubungan antar variabel tersebut
3. Menentukan maksimum jumlah fungsi basis (BF), maksimum fungsi basis yang dibolehkan adalah sebanyak dua sampai empat kali dari banyaknya variabel prediktor yang digunakan. Dalam penelitian ini digunakan 7 variabel prediktor sehingga maksimum jumlah BF adalah 14, 21 dan 28.
4. Menentukan jumlah maksimum interaksi, dalam penelitian ini jumlah maksimum interaksi (MI) yaitu 1, 2, dan 3. Karena apabila terdapat lebih dari 3 interaksi, maka akan menimbulkan interpretasi model yang sangat kompleks.
5. Menentukan minimum observasi antara knot dengan cara *trial and error* karena belum ada landasan atau batasan yang tetap untuk penentuan minimum observasi antara knot.

6. Menentukan model terbaik dari kombinasi nilai BF, MI, dan MO yang mungkin dengan kriteria nilai GCV minimum serta melakukan penaksiran parameter.
7. Menguji signifikansi model MARS dengan uji koefisien regresi secara simultan (Uji F) maupun secara parsial (Uji t).
8. Menginterpretasikan model MARS dan variabel - variabel yang berpengaruh di dalam model tersebut.

4. ANALISIS DAN PEMBAHASAN

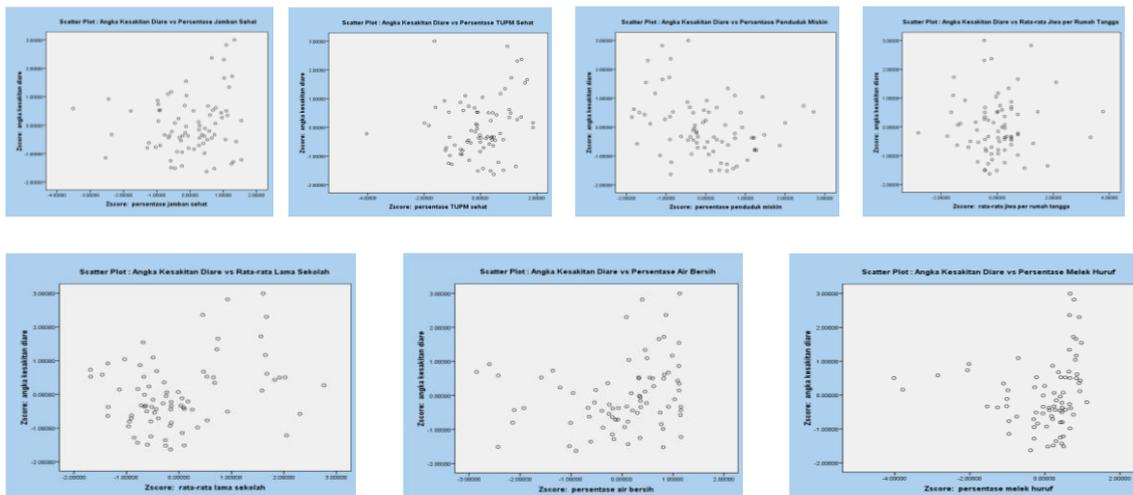
4.1 Deskripsi Data

Angka kesakitan diare di provinsi Jawa Tengah, Jawa Timur dan Daerah Istimewa Yogyakarta di tahun 2011 masih tergolong tinggi. Angka kesakitan terendah mencapai 68 per 1000 penduduk, angka kesakitan tertinggi berada di kabupaten Mojokerto yaitu sebesar 696 dan rata-rata angka kesakitan pada ketiga provinsi tersebut mencapai 289.47 per 1000 penduduk.

Tabel 1. Statistik deskriptif data angka kesakitan diare

	Minimum	Maksimum	Mean	Std Deviasi
Angka kesakitan diare	68.00	696.00	289.47	135.69
Persentase jamban sehat	10.66	100.00	72.75	17.67
Persentase air bersih	11.03	100.00	74.32	22.20
Persentase TUPM sehat	8.67	100.00	71.21	15.48
Rata-rata lama sekolah	5.22	11.52	7.61	1.42
Melek huruf	41.16	99.46	86.78	11.35
Rata-rata jiwa per rumah tangga	2.39	5.19	3.59	0.43
Persentase Penduduk miskin	4.81	31.94	15.86	5.92

Untuk mengetahui pola hubungan antara variabel respon dan variabel prediktor dapat dilihat melalui *scatter plot*. Namun sebelumnya terlebih dahulu dilakukan standardisasi terhadap nilai pada setiap variabel menjadi nilai normal baku. Standardisasi dilakukan karena berbedanya skala nilai variabel respon dan beberapa variabel prediktor. Proses Standardisasi dilakukan dengan rumus : $Z_{score} = \frac{(X - \mu)}{\sigma}$



Gambar 1. Scatter plot variabel respon dengan variabel prediktor

Berdasarkan *scatter plot* pada Gambar 1, pola hubungan antara angka kesakitan diare dan ketujuh variabel prediktor tidak mengikuti model pola tertentu. Sehingga memenuhi persyaratan untuk memodelkan pola tersebut dengan pendekatan regresi nonparametrik.

4.2 Pemodelan Angka Kesakitan Diare Menggunakan MARS

Dalam penelitian ini, penulis menentukan maksimal jumlah fungsi basis (BF) 14, 21 dan 28 sesuai dengan Friedman (1991), yang menganjurkan pemilihan maksimal jumlah fungsi basis sebesar dua sampai empat kali banyaknya jumlah variabel prediktor. Dalam penelitian ini menggunakan 7 variabel prediktor untuk pendugaan variabel respon. Untuk maksimal jumlah interaksi variabel (MI) yang ditentukan sebesar 1, 2, dan 3. Karena apabila terdapat lebih dari 3 interaksi, maka akan menimbulkan interpretasi model yang sangat kompleks. Sedangkan untuk minimum observasi (MO) dipilih 0, 1, 2, 3, 4 dan 5. Nilai MO telah dicobakan dengan berbagai nilai kemungkinan (0 sampai 30), namun pada titik-titik tersebut diperoleh nilai GCV minimum. Terdapat 54 model dari hasil kombinasi dengan nilai BF dan nilai MI.

Model MARS terbaik dipilih berdasarkan nilai GCV yang terkecil. Dari hasil kombinasi nilai MO, BF dan MI diperoleh model terbaik pada MO=2, BF=28 dan MI=3 dengan nilai GCV sebesar 0.56959 dan bentuk modelnya sebagai berikut :

$$\hat{Y} = -0.526742 + 0.264444 * BF2 + 12.2382 * BF5 - 7.76719 * BF15 + 4.96445 * BF17;$$

dengan

$$\begin{aligned} BF1 &= \max(0, X5 - 0.302193); & BF11 &= \max(0, X3 + 1.08918) * BF1; \\ BF2 &= \max(0, 0.302193 - X5); & BF15 &= \max(0, X2 - 0.356956) * BF11; \\ BF4 &= \max(0, 0.354093 - X3) * BF1; & BF17 &= \max(0, X2 + 0.248386) * BF11; \\ BF5 &= \max(0, X2 - 0.749258) * BF4; \end{aligned}$$

4.3 Pengujian Signifikansi Model MARS

Setelah residual pada model MARS memenuhi asumsi regresi parametrik, maka dapat dilakukan pengujian untuk mengecek signifikansi parameter dan mengevaluasi kecocokan model.

4.3.1 Pengujian koefisien regresi simultan

- Rumusan hipotesis :
 $H_0 : a_2 = a_5 = a_{15} = a_{17} = 0$ (model tidak signifikan)
 $H_1 : \text{minimal terdapat satu } a_m \neq 0 ; m = 2, 5, 15, \text{ dan } 17$ (model signifikan)
 Taraf signifikansi : $\alpha = 0.05$
- Statistik uji : $F_{hitung} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 / M}{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / N - M - 1}$
 Daerah Kritis : Tolak H_0 jika nilai $F > F_{0.05(4;73)}$ atau $p_value < \alpha$
- Keputusan :
 Berdasarkan hasil pengolahan data, diperoleh p-value sebesar 0.00000, dan nilai $F = 25.55471$ serta dalam tabel F didapat nilai $F_{0.05(4;73)} = 2.50$. Sehingga $p_value < \alpha$ atau $F > F_{0.05(4;73)}$ maka H_0 ditolak yang berarti bahwa model signifikan.
- Kesimpulan :
 Pada taraf signifikansi 5% dapat disimpulkan bahwa model signifikan sehingga dapat digunakan untuk memprediksi angka kesakitan diare.

4.3.2 Pengujian Koefisien Regresi Parsial

- Rumusan hipotesis :
 $H_0 : a_m = 0$ (koefisien a_m tidak berpengaruh terhadap model)
 $H_1 : a_m \neq 0$; $m = 2,5,15$, dan 17 (koefisien a_m berpengaruh terhadap model)
Taraf signifikansi : $\alpha = 0.05$
- Statistik Uji : $t_{hitung} = \frac{\hat{a}_m}{se(\hat{a}_m)}$ dengan $Se(\hat{a}_m) = \sqrt{var(\hat{a}_m)}$
- Daerah Kritis: Tolak H_0 jika $t > t_{(\frac{\alpha}{2}; 74)}$ atau $p_value < \alpha$
- Keputusan :
Berdasarkan hasil pengolahan data, diperoleh p-value dari masing-masing fungsi basis :
Fungsi Basis 2 ($m = 2$) : 0.00398 Fungsi Basis 15 ($m = 15$) : 0.00000
Fungsi Basis 5 ($m = 5$) : 0.00000 Fungsi Basis 17 ($m = 17$) : 0.00000
Dapat dilihat nilai p-value pada setiap $m < \alpha$. Sehingga H_0 ditolak, yang berarti bahwa koefisien a_2, a_5, a_{15} , dan a_{17} berpengaruh signifikan terhadap model.
- Kesimpulan :
Pada taraf signifikansi 5% dapat disimpulkan bahwa model signifikan sehingga dapat digunakan untuk memprediksi angka kesakitan diare.

4.4 Interpretasi Model MARS Terbaik

Interpretasi dari beberapa koefisien-koefisien fungsi basis pada model MARS di atas adalah sebagai berikut:

- $BF2 = \max(0, 0.302193 - X5)$ dengan koefisien 0.264444; artinya bahwa setiap kenaikan $BF2$ sebesar satu satuan akan meningkatkan angka kesakitan diare sebesar 0.264444 pada daerah dengan nilai baku melek huruf kurang dari 0.302193
- $BF5 = \max(0, X2 - 0.749258) * BF4$
dengan koefisien 12.2382
 $BF4 = \max(0, 0.354093 - X3) * BF1$
 $BF1 = \max(0, X5 - 0.302193)$
Jadi $BF5 = \max(0, X2 - 0.749258) * \max(0, 0.354093 - X3) * \max(0, X5 - 0.302193)$
Artinya bahwa setiap kenaikan $BF5$ sebesar satu satuan akan meningkatkan angka kesakitan diare sebesar 12.2382 pada daerah dengan nilai baku persentase penggunaan air bersih lebih dari 0.749258, nilai baku persentase TUPM sehat kurang dari 0.354093 dan nilai baku melek huruf lebih dari 0.302193.
- $BF15 = \max(0, X2 - 0.356956) * BF11$
dengan koefisien -7.76719
 $BF11 = \max(0, X3 + 1.08918) * BF1$
 $BF1 = \max(0, X5 - 0.302193)$
Jadi $BF15 = \max(0, X2 - 0.356956) * \max(0, X3 + 1.08918) * \max(0, X5 - 0.302193)$
Artinya bahwa setiap kenaikan $BF15$ sebesar satu satuan akan menurunkan angka kesakitan diare sebesar 7.76719 pada daerah dengan nilai baku persentase penggunaan air bersih lebih dari 0.356956, nilai baku persentase TUPM sehat lebih dari -1.08918 dan nilai baku melek huruf lebih dari 0.302193
- $BF17 = \max(0, X2 + 0.248386) * BF11$
dengan koefisien 4.96445
 $BF11 = \max(0, X3 + 1.08918) * BF1$
 $BF1 = \max(0, X5 - 0.302193)$

Jadi $BF_{17} = \max(0, X_2 + 0.248386) * \max(0, X_3 + 1.08918) * \max(0, X_5 - 0.302193)$
 Artinya bahwa setiap kenaikan BF_{17} sebesar satu satuan akan meningkatkan angka kesakitan diare sebesar 4.96445 pada daerah dengan nilai baku persentase penggunaan air bersih lebih dari -0.248386, nilai baku persentase TUPM sehat lebih dari -1.08918 dan nilai baku melek huruf lebih dari 0.302193

5. KESIMPULAN

Berdasarkan analisis yang dilakukan, didapat suatu kesimpulan :

1. MARS merupakan suatu metode yang dapat digunakan untuk mencari model hubungan antara variabel respon dan banyak variabel prediktor ($3 \leq n \leq 20$) yang mana bentuk kurva antara variabel respon dan variabel prediktor tidak memiliki pola tertentu. Kelebihan metode MARS adalah model yang dihasilkan dapat berupa interaksi antara variabel prediktor yang diinterpretasikan dalam suatu fungsi basis. MARS menggunakan kriteria GCV minimum untuk memperoleh model terbaik.
2. Model MARS yang digunakan untuk memprediksi angka kesakitan diare di Provinsi Jawa Tengah, Jawa Timur dan Daerah Istimewa Yogyakarta adalah model $MARS(MO=2;BF=28;MI=3)$ karena memiliki nilai GCV paling kecil yaitu sebesar 0.56959 dengan bentuk persamaan :

$$\hat{Y} = -0.52674 + 0.26444 * BF_2 + 12.238 * BF_5 - 7.76719 * BF_{15} + 4.96445 * BF_{17};$$

dengan

$$\begin{aligned} BF_1 &= \max(0, X_5 - 0.302193); & BF_{11} &= \max(0, X_3 + 1.08918) * BF_1; \\ BF_2 &= \max(0, 0.302193 - X_5); & BF_{15} &= \max(0, X_2 - 0.356956) * BF_{11}; \\ BF_4 &= \max(0, 0.354093 - X_3) * BF_1; & BF_{17} &= \max(0, X_2 + 0.248386) * BF_{11}; \\ BF_5 &= \max(0, X_2 - 0.749258) * BF_4; \end{aligned}$$

DAFTAR PUSTAKA

- Adisasmito, W., 2007. *Faktor Risiko Diare Pada Bayi Dan Balita Di Indonesia: Systematic Review Penelitian Akademik Bidang Kesehatan Masyarakat*. Makara Of Health Series, Vol 11, No 1. FKM UI.
- Badan penelitian dan pengembangan Kesehatan, Departemen Kesehatan RI. Riset Kesehatan Dasar (Riskesdas) . 2007.
- Budiantara, I.N. 2009, *Spline dalam Regresi Nonparametrik dan Semiparametrik: Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang*, Institut Teknologi Sepuluh November, Surabaya.
- Depkes RI , 2005. *Pedoman Pemberantasan Penyakit Diare*. Edisi 4. Ditjen PPM dan Friedman, J.H. 1991. *Multivariate Adaptive Regression Splines. The Annals of Statistics*, Vol. 19, No. 1.
- Hardle, W., 1990, *Applied Nonparametric Regression*, Cambridge University.
- Otok, B. W. dan Pintowati, W., 2012, *Pemodelan Kemiskinan di Jawa Timur dengan Pendekatan Multivariate Adaptive*. Jurnal Sains dan Seni, FMIPA ITS.
- Ryan, S.E dan Porth, L.S. 2007. *A Tutorial on The Piecewise Regression Approach Applied to Bedload Transport Data*. Rocky Research Station. USA
- Suharyono, 1991. *Diare Akut Klinik dan Labotarik*, Cetakan Pertama. PT Rineka Cipta, Jakarta.