

**MODEL KLASIFIKASI MENGGUNAKAN METODE REGRESI LOGISTIK
DAN MULTIVARIATE ADAPTIVE REGRESSION SPLINES (MARS)
(Studi Kasus: Data Survei Biaya Hidup (SBH) Kota Kediri Tahun 2012)**

Sumarno¹⁾, dan Bambang Widjanarko Otok²⁾

Institut Teknologi Sepuluh Nopember, Jurusan Statistika

Arief Rahman Hakim, Surabaya 60111, Indonesia

¹⁾ Mahasiswa S2 Jurusan Statistika ITS, Surabaya,

²⁾ Jurusan Statistika FMIPA ITS, Surabaya

Surel: kayla_marno@yahoo.com.

ABSTRACT

Good classification method should give the least misclassification rates. The Logistic Regression and Multivariate Adaptive Regression Spline (MARS) model was some of classification method which often used when there were any categorical variable of response, wanted to be predicted with categorical or continuous type of predictor variable. This research aim to get the best model among the two on classification, taking the case of study of household preference on clothing shopping place in Kediri, and its influencing variables, based on 2012's Cost of Living Surveys (SBH). Method performance is measured by its accuracy rate, Noise Signal Ratio (NSR) and G-Mean from classification table. The classification using logistic regression shows that response variable were only influenced by household income, numbers of household member, the education level of household leader, car ownership, the source of household income, and food expenditure percentage. Meanwhile, the result given by MARS added two more influenced variables named the age of household leader and housing ownership statuses. The accuracy rate, NSR, and G-mean performance gained by MARS method reached consecutively 71.2, 26.2, and 60.6 percent, Meanwhile the same results by Logistic Regression consecutively shows only 70.6, 30.3, and 36.3 percent. It was lead to the conclusion that in this case of study, classification using MARS performs better than Logistic Regression.

Keywords: classification, logistic regression, MARS, SBH

PENDAHULUAN

Pasar sebagai salah satu fasilitas perbelanjaan selama ini sudah menyatu dan memiliki tempat penting dalam kehidupan masyarakat. Pasar atau tempat berbelanja barang dapat dibedakan menjadi dua, yaitu pasar modern dan bukan pasar modern. Pasar modern yang dimaksud disini adalah yang terdiri dari *Supermarket*, *Hypermarket*, *Minimarket*, *Swalayan* dan *Department Store*, sedangkan yang dikategorikan bukan pasar modern yaitu pasar tradisional, toko/warung, pedagang keliling dan lainnya.

Seiring dengan kemunculan pusat-pusat pembelajaran pasar modern di Kota Kediri belakangan ini merupakan cermin terjadinya perubahan gaya hidup masyarakat di Kota Kediri. Berdasarkan data dari Dinas Perindustrian, Perdagangan, Pertambangan dan Energi Pemerintah Kota Kediri, pada tahun 2012 Kota Kediri yang hanya terdiri dari 3 kecamatan memiliki 23 pasar tradisional yang sebarannya merata hampir ke seluruh wilayah Kota Kediri, dan memiliki 41 pusat perbelanjaan modern yang terdiri dari: 31 minimarket, 1 supermarket, 3 departemen store dan 6 hypermarket, di mana sebaran untuk minimarket juga hampir merata ke seluruh wilayah Kota Kediri, sedangkan untuk supermarket, departemen store dan hypermarket cenderung mengelompok pada daerah-daerah tertentu, seperti di Jl. Panglima Sudirman, Jl. Hasanudin, Jl. Hayam Muruk, dan Jl. Brigjen Katamso. Ke empat lokasi ini merupakan pusat-pusat keramaian di Kota Kediri.

Menurut Rahardja (2010) konsep preferensi berkaitan dengan kemampuan konsumen menyusun prioritas pilihan agar dapat mengambil keputusan. Sehingga dapat dikatakan seorang konsumen akan mengatur pembeliannya sesuai dengan pendapatan yang dimilikinya dengan memilih berbelanja di pasar tradisional atau di ritel modern. Dalam hal pengambilan keputusan untuk memilih tempat berbelanja di pasar modern atau bukan pasar modern, tentu saja akan lebih menarik untuk dilakukan penelitian lebih lanjut dengan mengklasifikasikan sehingga dapat diketahui faktor-faktor apa saja yang mempengaruhi rumah tangga dalam menentukan pilihannya berbelanja di pasar modern atau bukan pasar modern di Kota Kediri.

Penelitian ini akan menggunakan dua teknik *modelling* yaitu regresi logistik dan MARS tujuannya untuk mendapatkan model yang ideal mengenai faktor-faktor yang mempengaruhi tempat berbelanja barang kebutuhan sandang rumah tangga di Kota Kediri tahun 2012 yaitu di pasar modern atau bukan pasar modern kemudian dari model yang sudah didapatkan akan ditentukan metode mana yang lebih tepat untuk dipergunakan pada klasifikasi tempat berbelanja barang kebutuhan sandang rumah tangga di Kota Kediri tahun 2012. Pengukuran performa klasifikasi menggunakan nilai ketepatan klasifikasi (*Accuracy*), *Noise Signal Ratio* (NSR) dan *Geometric Mean* (*G-Mean*).

Regresi Logistik

Analisis regresi logistik adalah analisis yang digunakan untuk melihat hubungan antara variabel respon kategorik dengan variabel-variabel prediktor kategorik maupun kontinyu (Agresti, 2002). Regresi logistik dapat digunakan untuk pengklasifikasian sejumlah objek ke dalam dua kelompok, karena variabel respon (Y) hanya terdiri dari dua kategori (misal: 1 dan 0).

Bentuk umum fungsi regresi logistik dengan k variabel prediktor diformulasikan sebagai berikut: $\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}$, $\beta_i =$ Nilai parameter ke- i dan $i = 1, 2, \dots, k$

Transformasi logit dari $\pi(x)$ menghasilkan persamaan sebagai berikut (Hosmer dan Lameshow, 2000): $\hat{g}(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$

Pendugaan Parameter

Model regresi logistik dengan variabel respon bernilai 0 atau 1 dimana antar pengamatan diasumsikan saling bebas maka penduga parameter β diperoleh dengan metode *Maximum Likelihood Estimation* (MLE). Pada dasarnya prinsip dari pendugaan maksimum *likelihood* adalah dengan mencari nilai dari β dengan memaksimalkan nilai dari fungsi *likelihood* $L(\beta_0, \beta_1, \dots, \beta_k)$. Karena setiap pengamatan diasumsikan saling bebas, maka fungsi *likelihood* merupakan fungsi kepadatan gabungan yaitu:

$L(\beta) = \prod_{i=1}^n f(\beta, y_i)$. Berhubung variabel respon regresi logistik mengikuti sebaran distribusi bernoulli, sehingga fungsi *likelihood* menjadi $L(\beta) = \prod_{i=1}^n [\pi(x_i)]^{y_i} [1 - \pi(x_i)]^{1-y_i}$, dimana $\pi(x_i)$ menunjukkan nilai $\pi(x)$ ke- x_i dapat ditulis menjadi:

$$\ln[L(\beta)] = \sum_{i=1}^n \left[y_i (\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}) + \ln(1 + \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}))^{-1} \right].$$

Untuk mendapatkan nilai β yang memaksimalkan $L(\beta)$ maka perlu mendifferensialkan $L(\beta)$ terhadap $(\beta_0, \beta_1, \beta_2, \dots, \beta_k)$ dan membuat persamaan yang

diperoleh sama dengan nol, dengan demikian diperoleh: $\sum_{i=1}^n [y_i - \pi(x_i)] = 0$ dan

$\sum_{i=1}^n [y_i x_{ij} - x_{ij} \pi(x_i)] = 0$ dengan $j = 1, 2, \dots, k$. Hasil dari turunan pertama biasanya

dimasukkan dalam sebuah vektor, yang disebut *vector gradient* (\vec{g}). Sedangkan turunan kedua hasilnya dimasukkan dalam matriks yang disebut matriks Hessian (\mathbf{H}), dengan

bentuk umum turunannya adalah:
$$\frac{\partial^2 L(\beta_0, \beta_1, \dots, \beta_k)}{\partial \beta_j \partial \beta_w} = -\sum_{i=1}^n x_{ij} x_{iw} \pi(x_i) (1 - \pi(x_i)),$$

dimana $j, w = 0, 1, 2, \dots, k$.

Uji Signifikansi Parameter

a. Uji Serentak

Adapun hipotesis yang digunakan dalam pengujian ini adalah:

$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$ $H_1 : \text{Paling sedikit ada satu } \beta_j \neq 0$, dengan $j = 1, 2, \dots, k$.

Statistik Uji-G dirumuskan sebagai:
$$G = -2 \ln \left[\frac{L_0}{L_1} \right]$$

dimana L_0 = Penduga *likelihood* tanpa variabel prediktor

L_1 = Penduga *likelihood* dengan variabel prediktor (model penuh).

Hipotesis nol (H_0) akan ditolak jika nilai $G_{hitung} > \chi^2_{(k, \alpha)}$ atau $P_{value} < \alpha$.

b. Uji Parsial

Hipotesis yang digunakan dalam pengujian ini adalah:

$H_0 : \beta_j = 0$ $H_1 : \beta_j \neq 0$, dengan $j = 1, 2, \dots, k$.

Statistik uji Wald dirumuskan sebagai berikut:
$$W = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} ; j = 0, 1, 2, \dots, k$$

dimana $\hat{\beta}_j$ adalah nilai penduga untuk parameter β_j dengan $SE(\hat{\beta}_j)$ merupakan dugaan galat baku untuk parameter β_j . Jika nilai $P_{value} < \alpha$ atau $W_j^2 > \chi^2_{(\alpha, 1)}$ dimana α adalah taraf signifikansi atau tingkat kesalahan yang ditentukan, maka variabel prediktor mempengaruhi variabel respon.

c. Interpretasi Koefisien Parameter

Interpretasi model regresi logistik dilakukan dengan melihat nilai *odds ratio* (*OR*) pada masing-masing variabel dengan menganggap variabel lain konstan, *OR* didefinisikan sebagai perbandingan antara dua nilai logit pada $X = 1$ dan $X = 0$, maka:

$$OR = \frac{\pi(1)/[1-\pi(1)]}{\pi(0)/[1-\pi(0)]} = \frac{\left[\frac{\exp(\beta_0 + \beta_j)}{1 + \exp(\beta_0 + \beta_j)} \right] \left[\frac{1}{1 + \exp(\beta_0)} \right]}{\left[\frac{\exp(\beta_0)}{1 + \exp(\beta_0)} \right] \left[\frac{1}{1 + \exp(\beta_0 + \beta_j)} \right]} = \frac{\exp(\beta_0 + \beta_j)}{\exp(\beta_0)} = \exp(\beta_j)$$

Sehingga $OR = \exp(\hat{\beta}_j)$ dan $\ln OR = (\hat{\beta}_j)$

Pemodelan *Multivariate Adaptive Regression Splines (MARS)*

Pada dasarnya MARS merupakan kombinasi yang kompleks dari *Spline* dan *Recursive Partitioning Regression (RPR)*. Menurut Dillon (1978) dan Sharma (1996) metode MARS merupakan salah satu metode untuk klasifikasi statistik modern yang sudah memanfaatkan fleksibilitas model dan menduga suatu distribusi di dalam masing-masing kelas yang pada akhirnya menyediakan suatu aturan pengelompokan. Untuk penentuan *knot* yaitu menggunakan algoritma *forward stepwise* dan *backward stepwise*. Pemilihan model yang paling optimum (terbaik) dalam model MARS yaitu jika nilai GCV dari model tersebut mempunyai nilai GCV yang paling rendah (minimum) diantara model-model lain.

Menurut Friedman (1991) setelah dilakukan modifikasi model RPR, diperoleh model MARS adalah sebagai berikut: $f(x) = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} [s_{km} \cdot (x_{v(k,m)} - t_{km})] + \varepsilon_i$

Model MARS di atas dapat disajikan sebagai berikut:

$$y_i = a_0 + \sum_{m=1}^M a_m B_m(x) + \varepsilon_i \quad \text{dimana,} \quad B_m(x) = \prod_{k=1}^{K_m} [s_{km} (x_{v(k,m)} - t_{km})]_+$$

METODE

Data yang akan digunakan dalam penelitian ini adalah data skunder dari Badan Pusat Statistik (BPS) mengenai data rumah tangga hasil Survei Biaya Hidup tahun 2012 (SBH 2012) di Kota Kediri, dengan jumlah sampel rumah tangga SBH 2012 sebanyak 1.600 rumah tangga. Unit observasi/analisis dari penelitian ini adalah rumah tangga. Dari data ini kemudian diolah menggunakan program SPSS 20 dan paket MARS 2.0.

Dalam penelitian ini variabel-variabel yang diduga berhubungan atau berpengaruh terhadap tempat berbelanja barang kebutuhan sandang rumah tangga (Y) di Kota Kediri tahun 2012 adalah variabel pendapatan rumah tangga (X_1), umur kepala rumah tangga (X_2), banyaknya anggota rumah tangga (X_3), jenis kelamin kepala rumah tangga (X_4), pendidikan terakhir kepala rumah tangga (X_5), status penguasaan bangunan tempat tinggal (X_6), kepemilikan mobil (X_7), sumber pendapatan utama rumah tangga (X_8) dan persentase pengeluaran untuk kelompok non makanan (X_9).

Untuk mencapai tujuan dari penelitian ini maka dilakukan langkah-langkah sebagai berikut:

1. Pemisahan data penelitian menjadi 2 bagian, yaitu data *training* (sebanyak 1.440 rumah tangga) dan *testing* (sebanyak 160 rumah tangga) untuk validasi.
2. Untuk mendapatkan model dengan regresi logistik tahapannya sebagai berikut:
 - a. Mendefinisikan variabel respon dan prediktor dalam pembentukan model.
 - b. Menduga parameter model dengan menggunakan MLE.
 - c. Melakukan uji simultan dengan menggunakan statistik uji-G.
 - d. Melakukan uji parsial dengan menggunakan statistik uji wald.
 - e. Dari langkah (a) sampai dengan (d) dapat diketahui variabel-variabel prediktor (X) apa saja yang mempengaruhi variabel respon (Y).
 - f. Menginterpretasikan model.
3. Untuk mendapatkan model dengan MARS, dilakukan tahapan sebagai berikut:
 - a. Mendefinisikan variabel respon (Y) dan variabel prediktor (X).
 - b. Menentukan model terbaik, diperoleh dengan cara *trial and error* sampai didapatkan model dengan nilai GCV minimum. Tahapannya sebagai berikut:
 - Menentukan maksimum fungsi basis (BF) yaitu 2 sampai 4 kali banyaknya variabel prediktor.
 - Menentukan jumlah maksimum interaksi (MI) yaitu 1, 2, dan 3.

- Menentukan minimal jumlah pengamatan setiap *knot* atau observasi minimum (MO) antar *knot* sebanyak 0, 1, 2, atau 3 (Friedman, 1991).
 - c. Mengulangi langkah (b) sampai didapat model dengan nilai GCV minimum dan tingkat ketepatan klasifikasi terbesar.
 - d. Dari urutan hasil tersebut diatas akan diperoleh variabel-variabel yang berpengaruh secara signifikan dari pembentukan model MARS.
 - e. Menginterpretasikan model MARS terbaik.
4. Menentukan metode mana yang mempunyai performa terbaik antara metode regresi logistik dan metode MARS dengan membandingkan nilai *Accuracy*, *Noise Signal Ratio* (NSR) dan *G-Mean*.

HASIL DAN PEMBAHASAN

Pemodelan dengan Regresi Logistik

Dari hasil pengolahan dengan menggunakan SPSS 20 diperoleh model regresi logistik, dimana dari 9 variabel prediktor yang diduga mempengaruhi variabel respon, ternyata hanya ada 6 variabel saja yang secara signifikan mempengaruhi tempat berbelanja barang kebutuhan sandang rumah tangga di Kota Kediri (nilai signifikasinya kurang dari $\alpha = 0,05$) yaitu: pendapatan rumah tangga (X_1), banyaknya anggota rumah tangga (X_3), pendidikan terakhir kepala rumah tangga (X_5), kepemilikan mobil (X_7), sumber pendapatan utama rumah tangga (X_8) dan persentase pengeluaran untuk kelompok non makanan (X_9). Dengan tingkat *accuracy* sebesar 70,0 persen untuk data *training* dan setelah di validasi dengan data *testing* naik menjadi 70,6 persen, NSR dan *G-Mean* masing-masing 30,3 persen dan 36,3 persen.

Sehingga model regresi logistik yang didapatkan adalah sebagai berikut:

$$\pi(x) = \frac{\exp(0,989 + 0,000X_1 - 0,089X_3 - 0,764X_5(1) - 0,476X_5(2) - 0,811X_7(1) - 0,270X_8(1) + 0,000X_9)}{1 + \exp(0,989 + 0,000X_1 - 0,089X_3 - 0,764X_5(1) - 0,476X_5(2) - 0,811X_7(1) - 0,270X_8(1) + 0,000X_9)}$$

atau apabila dilakukan transformasi logit maka persamaan di atas menjadi:

$$\hat{g}(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = 0,989 + 0,000X_1 - 0,089X_3 - 0,764X_5(1) - 0,476X_5(2) - 0,811X_7(1) - 0,270X_8(1) + 0,000X_9$$

Dari output SPSS diperoleh nilai $\text{Exp}(B)$ atau nilai *Odds Ratio* (OR) untuk (X_1), (X_3), ($X_5(1)$), ($X_5(2)$), (X_8) dan (X_9) masing-masing sebesar: 1,000, 0,914, 0,466, 0,621, 0,445, 0763 dan 1,000 dapat diambil kesimpulan sebagai berikut:

- Variabel X_1 , dengan nilai OR=1,000: Jika pendapatan rumah tangga bertambah 1 rupiah, maka kecenderungan rumah tangga tersebut berbelanja di pasar modern berlipat sebanyak 1,000 kali dengan asumsi variabel lainnya konstan.
- Variabel X_3 , dengan nilai OR=0,914: Jika banyaknya anggota rumah tangga bertambah 1 orang, maka kecenderungan rumah tangga tersebut berbelanja di pasar modern berlipat sebanyak 0,914 kali atau menurun sebesar $(1/0,914)$ 1,09 kali dengan asumsi variabel lainnya konstan.
- Variabel X_9 , dengan nilai OR=1,000: Jika rumah tangga persentase pengeluaran untuk kelompok non makanan bertambah 1 persen, maka kecenderungan rumah tangga tersebut berbelanja di pasar modern berlipat sebanyak 1,000 kali dengan asumsi variabel lainnya konstan.

Pemodelan dengan MARS

Pembentukan model MARS dilakukan dengan cara *trial and error* terhadap maksimum Basis Fungsi (BF), Maksimum Interaksi (MI), dan minimum jumlah pengamatan diantara *knot* atau Minimum Observasi (MO) sampai diperoleh model optimal dengan nilai GCV dan MSE minimum. Dari cara *trial and error* terhadap kombinasi antara BF, MI dan MO, Maka diperoleh model yang terbaik adalah **model 35**. Model MARS yang terbentuk memiliki nilai GCV dan MSE terkecil sebesar 0,202 dan 0,194 dengan 8 variabel prediktor yang berpengaruh terhadap tempat berbelanja barang kebutuhan sandang rumah tangga di Kota Kediri, tingkat ketepatan klasifikasi data *training* sebesar 70,56 persen, dan setelah divalidasi dengan data testing naik menjadi 71,2 persen, nilai NSR dan *G-Mean* sebesar 26,2 persen dan 60,6 persen. **Model 35** ini merupakan kombinasi antara BF=36, MI=3, dan MO=2. Sehingga persamaan model MARS adalah sebagai berikut:

$$Y = 0,764 - 0,868E-07*BF2 + 0,427E-08* BF6 + 0,424E-07*BF7 - 0,162E-07*BF9 \\ - 0,241E-07*BF13 + 0,017*BF22 - 0,448E-07*BF26 + 0,217E-07*BF29 \\ - 0,116E-07*BF30 - 0,008*BF36$$

dimana,

$$\begin{aligned}
 \text{BF2} &= \max(0, 6788333.000 - \text{PENDAPAT}) & \text{BF3} &= (\text{MOBIL} = 0) * \text{BF2} \\
 \text{BF6} &= \max(0, 34.000 - \text{UMUR}) * \text{BF3} & \text{BF7} &= (\text{PENDIDIK} = 3) * \text{BF3} \\
 \text{BF9} &= (\text{SUMBER} = 0) * \text{BF3} & \text{BF13} &= \max(0, \text{JMLART} - \\
 & & & 1.000) * \text{BF3} \\
 \text{BF17} &= (\text{PENDIDIK} = 2 \text{ OR } \text{PENDIDIK} = 3) & \text{BF19} &= (\text{MOBIL} = 0) \\
 \text{BF22} &= \max(0, 48.420 - \text{NONMAKAN}) * \text{BF17} & \text{BF30} &= (\text{RUMAH} = 0) * \text{BF29} \\
 \text{BF24} &= \max(0, 30.470 - \text{NONMAKAN}) * \text{BF17} & \text{BF34} &= (\text{RUMAH} = 1) * \text{BF19} \\
 \text{BF26} &= \max(0, 4878000.000 - \text{PENDAPAT}) * \text{BF24} & \text{BF29} &= \max(0, \text{JMLART} - \\
 & & & 1.000) * \text{BF2} \\
 \text{BF36} &= \max(0, 52.000 - \text{UMUR}) * \text{BF34}
 \end{aligned}$$

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

- Koefisien $\text{BF2} = \max(0, 6788333.000 - \text{PENDAPAT})$ dengan koefisien $-0,869\text{E}-07$: Artinya, koefisien BF2 akan bermakna jika nilai pendapatan rumah tangga (X_1) lebih kecil dari 6788333 rupiah maka setiap kenaikan fungsi basis (BF2) dapat mengurangi resiko rumah tangga berbelanja barang kebutuhan sandang di pasar modern sebesar $0,869\text{E}-07$ dengan rata-rata pendapatan rumah tangga kurang dari 6788333 rupiah.
- $\text{BF6} = \max(0, 34.000 - \text{UMUR}) * \text{BF3}$ dengan koefisien $0,427\text{E}-08$: Artinya, koefisien BF6 akan bermakna jika umur kepala rumah tangga (X_2) lebih kecil dari 34 tahun dan kepemilikan mobil (X_7) adalah tidak memiliki mobil dan pendapatan rumah tangga (X_1) lebih kecil dari 6788333 rupiah maka setiap kenaikan satu fungsi basis (BF6) dapat menambah resiko rumah tangga berbelanja barang kebutuhan sandang di pasar modern sebesar $0,427\text{E}-08$ dengan rata-rata umur kepala rumah tangga (X_2) kurang dari 34 tahun dan rata-rata pendapatan rumah tangga (X_1) kurang dari 6788333 rupiah.
- Koefisien $\text{BF36} = \max(0, 52.000 - \text{UMUR}) * \text{BF34}$ dengan koefisien $-0,008$: Artinya, koefisien BF36 akan bermakna jika umur kepala rumah tangga (X_2) lebih besar dari 52 tahun dan status penguasaan bangunan tempat tinggal (X_6) adalah milik sendiri dan kepemilikan mobil (X_7) adalah tidak memiliki mobil maka setiap kenaikan satu fungsi basis (BF36) dapat mengurangi resiko rumah tangga

berbelanja barang kebutuhan sandang di pasar modern sebesar 0,008 dengan rata-rata umur kepala rumah tangga (X_2) lebih dari dari 52 tahun.

Ketepatan Klasifikasi pada Model Regresi Logistik

Dengan menggunakan regresi logistik, Ketepatan klasifikasi yang dihasilkan sebesar 70,0 persen untuk data *training* dan setelah divalidasi dengan data *testing* naik sebesar 70,6 persen, nilai NSR sebesar 30,3 persen dan nilai *G-Mean* sebesar 36,3 persen.

Ketepatan Klasifikasi pada Model MARS

Dengan menggunakan metode MARS, ternyata Ketepatan klasifikasi yang dihasilkan sebesar 70,6 persen untuk data *training* dan divalidasi dengan data *testing* naik menjadi 71,2 persen, nilai NSR sebesar 26,2 persen dan nilai *G-Mean* sebesar 60,6 persen.

KESIMPULAN

1. Dari hasil klasifikasi dengan model regresi logistik didapatkan hanya 6 variabel yang mempengaruhi tempat berbelanja barang kebutuhan sandang rumah tangga di Kota Kediri yaitu: (X_1), (X_3), (X_5), (X_7), (X_8) dan (X_9). Model regresi logistik yang diperoleh adalah sebagai berikut:

$$\hat{g}(x) = 0,989 + 0,000X_1 - 0,089X_3 - 0,764X_5(1) - 0,476X_5(2) - 0,811X_7(1) - 0,270X_8(1) + 0,000X_9$$

Sedangkan dari hasil klasifikasi dengan model MARS variabel prediktor yang berpengaruh bertambah menjadi 8 variabel yaitu: (X_1), (X_2), (X_3), (X_5), (X_6), (X_7), (X_8) dan (X_9). Model MARS yang diperoleh sebagai berikut:

$$Y = 0,764 - 0,868E-07*BF2 + 0,427E-08* BF6 + 0,424E-07*BF7 - 0,162E-07*BF9 \\ - 0,241E-07*BF13 + 0,017*BF22 - 0,448E-07*BF26 + 0,217E-07*BF29 \\ - 0,116E-07*BF30 - 0,008*BF36$$

Dengan:

$$BF2 = \max(0, 6788333.000 - \text{PENDAPAT})$$

$$BF3 = (\text{MOBIL} = 0)*BF2$$

$$\begin{aligned}
 \text{BF6} &= \max(0, 34.000 - \text{UMUR}) * \text{BF3} & \text{BF7} &= (\text{PENDIDIK} = \\
 &3) * \text{BF3} & & \\
 \text{BF9} &= (\text{SUMBER} = 0) * \text{BF3} & \text{BF13} &= \max(0, \text{JMLART} - \\
 &1.000) * \text{BF3} & & \\
 \text{BF17} &= (\text{PENDIDIK} = 2 \text{ OR } \text{PENDIDIK} = 3) & \text{BF19} &= (\text{MOBIL} = 0) \\
 \text{BF22} &= \max(0, 48.420 - \text{NONMAKAN}) * \text{BF17} & \text{BF30} &= (\text{RUMAH} = 0) * \\
 &\text{BF29} & & \\
 \text{BF24} &= \max(0, 30.470 - \text{NONMAKAN}) * \text{BF17} & \text{BF34} &= (\text{RUMAH} = \\
 &1) * \text{BF19} & & \\
 \text{BF26} &= \max(0, 4878000.000 - \text{PENDAPAT}) * \text{BF24} & \text{BF29} &= \max(0, \text{JMLART} - \\
 &1.000) * \text{BF2} & & \\
 \text{BF36} &= \max(0, 52.000 - \text{UMUR}) * \text{BF34} & &
 \end{aligned}$$

2. Berdasarkan penghitungan ketepatan klasifikasi dalam model maka metode MARS mempunyai ketepatan klasifikasi yang lebih baik jika dibandingkan dengan regresi logistik yaitu sebesar 71,2 persen untuk data *testing*, nilai NSR yaitu sebesar 26,2 persen dan nilai *G-Mean* yaitu sebesar 60,6 persen. sehingga dalam kasus ini metode yang tepat untuk mengklasifikasikan terhadap tempat berbelanja barang kebutuhan sandang rumah tangga di Kota Kediri lebih tepat menggunakan metode MARS.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Bapak Dr. Bambang Widjanarko Otok, M.Si atas bimbingan dan saran-saran yang diberikan selama ini.

DAFTAR PUSTAKA

- Agresti, A, (2002), *Categorical Data Analysis*, John Wiley & Sons, Hoboken, New Jersey.
- BPS, (2012), *Pedoman Survei Biaya Hidup 2012*, Badan Pusat Statistik, Jakarta.
- Dillon, W.R, (1978), "On The Performance of Some Multinomial Classification Rules", *Journal of American Statistical Association*, vol 73:305-313
- Friedman, J.H, (1991), Multivariate Adaptive Regression Spline, *The Annals of Statistics*, Vol.19, No.1.(Mar., 1991).Hlm 1-67



Hosmer, D.W., dan Lemeshow, S., (2000), *Applied Logistic Regression*, Second Edition, John Wiley and Sons.

Kubat, M., dan Matwin, S., (1997), Addressing the Curse of Imbalanced Training Sets: One-Sided Selection, in *Proc.of the 14th Intl. Conf.on Machine Learning*, Hlm. 179-186.

Rahardja, P, (2010), *Teori Mikroekonomi*, Lembaga Penerbit FE UI, Jakarta.

Sharma, S, (1996), *Applied Multivariate Techniques*, Jhon Wiley and Sons, inc, New York.