

**IDENTIFIKASI FAKTOR-FAKTOR PENCIRI TINGKAT
KESEJAHTERAAN RUMAH TANGGA
MENGUNAKAN *BOOTSTRAP AGREGATING
MULTIVARIATE ADAPTIVE REGRESSION SPLINES*
(BAGGING MARS)
(Kasus Rumah Tangga di Padang Pariaman)**

ARIE HARIADY ARIFIN, HAZMIRA YOZZA, IZZATI RAHMI H.G.

*Program Studi Matematika,
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas,
Kampus UNAND Limau Manis Padang, Indonesia,
ariehariady@gmail.com*

Abstrak. Tingkat kesejahteraan suatu rumah tangga dapat diduga berdasarkan faktor-faktor pencirinya. Pada penelitian ini akan diidentifikasi faktor-faktor tersebut dengan membuat suatu model klasifikasi berdasarkan faktor-faktor yang diduga menggunakan metode *Bagging MARS (Bootstrap Agregating Multivariate Adaptive Regression Splines)*. Tingkat kesejahteraan rumah tangga dikelompokkan menjadi dua kelompok berdasarkan garis kemiskinan, yaitu miskin dan tidak miskin. Kemudian dibuat model tingkat kesejahteraan berdasarkan 15 variabel yang mewakili aspek kependudukan, pendidikan, perumahan, ketenagakerjaan, sosial ekonomi rumah tangga, dan teknologi informasi dan komunikasi. Data yang digunakan adalah data Hasil Survei Sosial Ekonomi Nasional tahun 2011 di Padang Pariaman. Didapatkan 12 variabel yang berkontribusi pada model. Lima diantaranya yang memiliki tingkat kepentingan tertinggi berturut-turut adalah jumlah anggota rumah tangga (X_5), adanya anggota rumah tangga yang dapat menggunakan telepon genggam (X_{14}), umur kepala rumah tangga (X_2), ijazah tertinggi yang dimiliki kepala rumah tangga (X_4), pengalaman rumah tangga membeli beras raskin (X_{13}). Selanjutnya variabel-variabel yang berkontribusi pada model didefinisikan sebagai faktor penciri tingkat kesejahteraan rumah tangga.

Kata Kunci: Penciri Tingkat Kesejahteraan, Kemiskinan, bagging MARS

1. Pendahuluan

Kemiskinan sudah lama menjadi problematika di Indonesia. Pada tahun 2011 BPS mencatat 25,4 juta rumah tangga merupakan penduduk yang termasuk miskin maupun sudah keluar dari batas ambang kemiskinan, tapi masih termasuk kelompok yang rentan untuk jatuh kembali ke bawah batas kemiskinan. Banyak program dan strategi yang dilakukan pemerintah untuk mengentaskan masyarakat dari kemiskinan. Agar program-program tersebut tepat sasaran, harus ada suatu cara untuk mengidentifikasi apakah suatu rumah tangga tersebut miskin atau tidak miskin. Selama ini rumah tangga miskin dikelompokkan berdasarkan pendapatan atau pengeluaran. Namun pendapatan atau pengeluaran tersebut kadang belum menggambarkan kondisi sesungguhnya dari suatu rumah tangga. Untuk itu, perlu

digunakan cara lain untuk mengidentifikasi tingkat kesejahteraan. Salah satunya dengan melihat faktor-faktor yang dapat menjadi penciri tingkat kesejahteraan suatu rumah tangga.

Pada penelitian ini akan diidentifikasi faktor-faktor tersebut. Hasilnya dapat digunakan untuk perencanaan pembangunan, sehingga pembangunan lebih terarah pada peningkatan kesejahteraan rumah tangga. Untuk mendeskripsikan karakteristik kesejahteraan rumah tangga dan melihat faktor yang paling berpengaruh pada suatu kelompok maka perlu dibuat suatu klasifikasi berdasarkan faktor-faktor yang mempengaruhinya.

Metode klasifikasi yang digunakan dalam penelitian ini adalah MARS (*Multivariate Adaptive Regression Splines*) yang menghasilkan pemodelan regresi yang fleksibel untuk data dengan variabel prediktor $3 \leq k \leq 20$ dan ukuran contoh $50 \leq n \leq 1000$ [5]. Dalam penelitian ini, metode MARS dikombinasikan dengan salah satu pendekatan non parametrik *bootstrap aggregating (bagging)*. Konsep utama dari *bagging* adalah menggunakan *bootstrap resampling* untuk membangkitkan prediktor dalam banyak versi, dimana diharapkan ketika dikombinasikan hasilnya akan lebih baik dari pada prediktor tunggal yang dibangun untuk menyelesaikan permasalahan yang sama.

2. Kesejahteraan

Berdasarkan Rancangan Undang-undang tentang Sistem Kesejahteraan Sosial Nasional, kesejahteraan sosial adalah kondisi sosial ekonomi yang memungkinkan bagi setiap warga negara untuk dapat memenuhi kebutuhan yang bersifat jasmani, rohani, dan sosial sesuai dengan harkat dan martabat manusia. Kesejahteraan merupakan kondisi seseorang dapat memenuhi kebutuhan dasar baik material maupun non material yang mencakup aspek gizi dan kesehatan, pengetahuan, dan kekayaan materi [3].

Penduduk miskin adalah penduduk yang memiliki rata-rata pengeluaran perkapita perbulan dibawah penjumlahan dari Garis Kemiskinan Makanan (GKM) dan Garis Kemiskinan Non Makanan (GKNM) [1]. Untuk tahun 2012, garis kemiskinan yang ditentukan oleh BPS adalah Rp 309.113,00. Penduduk yang memiliki rata-rata pengeluaran perkapita per bulan dibawah garis kemiskinan dikategorikan sebagai penduduk miskin.

Faktor-faktor yang diduga menjadi karakteristik kesejahteraan suatu rumah tangga dapat ditinjau dari beberapa aspek, antara lain: aspek kependudukan, pendidikan, perumahan, sosial ekonomi rumah tangga, dan teknologi informasi dan komunikasi [4].

3. *Multivariate Adaptive Regression Spline* (MARS)

Pada tahun 1990, Jerome H. Friedman memperkenalkan metode MARS sebagai suatu metode baru yang mengotomatiskan pembangunan model-model prediktif akurat untuk variabel-variabel respon kontinu dan biner. MARS merupakan pengembangan dari pendekatan *Recursive Partitioning* yang menghasilkan model yang tidak kontinu pada *knot* [5].

Beberapa hal yang perlu diperhatikan dalam membangun model MARS yaitu:

- *Knot*, merupakan nilai variabel prediktor ketika *slope* suatu garis regresi mengalami perubahan yang dapat didefinisikan sebagai akhir dari satu segmen sekaligus merupakan awal dari segmen yang lain. Di setiap titik *knot*, diharapkan adanya kontinuitas dari fungsi basis antar satu region dengan region lainnya. Minimum observasi antara *knot* (MO) yang biasa digunakan adalah 0, 1, 2, dan 3 observasi.
- fungsi basis (BF) yaitu selang antar *knot* yang berurutan. Pada umumnya fungsi basis yang dipilih berbentuk polinomial dengan turunan yang kontinu pada setiap titik *knot*. Maksimum fungsi basis yang diijinkan adalah 2-4 kali jumlah variabel prediktornya.
- *Interaction* (interaksi) yaitu hasil perkalian silang antar variabel yang saling berkorelasi. Jumlah maksimum interaksi (MI) yang diperbolehkan adalah 1, 2, atau 3. Jika $MI > 3$ akan dihasilkan model yang semakin kompleks dan model akan sulit untuk diinterpretasi.

Setelah modifikasi Friedman, model MARS menjadi:

$$\hat{f}(x) = \alpha_0 + \sum_{m=1}^m \alpha_m \prod_{k=1}^{k_m} (S_{km}(X_{v(k,m)} - t_{km})) \quad (3.1)$$

dengan:

α_0 = konstanta regresi dari fungsi basis

α_m = koefisien dari fungsi basis ke- m , $m = 1, \dots, M$

M = maksimum fungsi basis (*nonconstant* fungsi basis)

km = derajat interaksi

$S_{km} = \begin{cases} +1, & \text{jika } knot \text{ terletak dikanan } subregion; \\ -1, & \text{jika } knot \text{ terletak dikiri } subregion. \end{cases}$

$x_{v(k,m)}$ = variabel prediktor ke- v , pilahan ke- k dan *subregion* ke- m

t_{km} = nilai *knots* dari variabel prediktor $x_{v(k,m)}$

koefisien $\{\alpha_m\}_{m=1}^M$ ditentukan dengan metode kuadrat terkecil.

Pembentukan model MARS diawali dengan terlebih dahulu menentukan *knot* dan fungsi basis setiap variabel prediktor dengan cara memplot setiap variabel prediktor dengan variabel respon. Pemilihan *knots* pada MARS dilakukan dalam dua tahap yaitu tahap *forward* dan tahap *backward*. *Forward* dilakukan untuk mendapatkan fungsi dengan jumlah basis maksimum dan nantinya akan dipilih fungsi basis dengan metode kuadrat terkecil. Pada tahap *backward* akan dipilih satu fungsi basis dan mengeluarkan basis tersebut jika kontribusi terhadap model kecil. Proses *backward* akan dilanjutkan hingga tidak ada fungsi basis yang dapat dikeluarkan. Ukuran kontribusi pada tahap *backward* ditentukan berdasarkan kriteria *Generalized Cross Validation* (GCV) yang diperkenalkan oleh Wahba pada tahun 1979. Fungsi GCV didefinisikan sebagai berikut:

$$GCV(M) = \frac{1/N \sum_{i=1}^N [y_i - \hat{f}_M(x_i)]^2}{[1 - (C(M))/N]^2} \quad (3.2)$$

dengan:

M = jumlah fungsi basis

x_i = variabel prediktor

y_i = variabel respon

N = banyaknya pengamatan

$$C(M) = \text{Trace}[B(B^T B)^{-1} B^T] + 1$$

$$B = \begin{bmatrix} \prod_{k=1}^{k_1} (S_{1M}(X_{1(1,M)} - t_{1M})) \cdots \prod_{k=1}^{k_m} (S_{k_m M}(X_{1(k_m,M)} - t_{k_m M})) \\ \vdots \qquad \qquad \qquad \ddots \qquad \qquad \qquad \vdots \\ \prod_{k=1}^{k_1} (S_{1M}(X_{n(1,M)} - t_{1M})) \cdots \prod_{k=1}^{k_m} (S_{k_m M}(X_{n(k_m,M)} - t_{k_m M})) \end{bmatrix}$$

Selanjutnya dilakukan uji signifikansi fungsi basis yang meliputi uji secara bersamaan dan uji individu. Uji signifikansi yang dilakukan secara bersamaan terhadap fungsi basis yang terdapat dalam model MARS ini bertujuan untuk mengetahui apakah secara umum model MARS terpilih merupakan model yang sesuai dan menunjukkan hubungan yang tepat antara variabel prediktor dengan variabel respon. Hipotesis yang digunakan adalah :

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_m = 0$$

$$H_1 : \text{ada } \alpha_j \neq 0, j = 1, 2, \dots, m$$

Statistik uji yang digunakan pada pengujian ini adalah statistik uji F yang diperoleh dari

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

Nilai F_{hitung} ini kemudian dibandingkan dengan $F_\alpha(v_1, v_2)$ yaitu nilai F dengan tingkat signifikansi α serta $v_1 = k$ dan $v_2 = n - k - 1$ dengan n adalah banyaknya sampel dan k adalah banyaknya fungsi basis yang berkontribusi terhadap model. Jika $F_{hitung} > F_\alpha(v_1, v_2)$ maka H_0 ditolak, artinya paling sedikit ada satu α_j yang tidak sama dengan nol.

Bila pada pengujian secara bersamaan disimpulkan untuk tolak H_0 , selanjutnya akan dilakukan pengujian untuk masing-masing fungsi basis yang. Hipotesisnya adalah sebagai berikut:

$$H_0 : \alpha_j = 0$$

$$H_1 : \alpha_j \neq 0, j = 1, 2, \dots, m$$

Nilai statistik hitung diperoleh dari $t_{hitung} = \frac{\alpha_j}{s_{\alpha_j}}$ dengan s_{α_j} merupakan standar error α_j yang diperoleh dari

$$s_{\alpha_j} = \sqrt{\left(\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - k - 1} \right) C_{jj}}$$

Nilai t_{hitung} ini dibandingkan dengan nilai tabel distribusi t dengan derajat bebas $v = n - k - 1$ dan tingkat signifikansi α . Jika daerah kritis $|t_{hitung}| > t(\frac{\alpha}{2}, v)$, maka H_0 ditolak, artinya ada pengaruh variabel prediktor pada fungsi basis tersebut terhadap variabel respon.

Selanjutnya digunakan APER (*Apparent Error Rate*) untuk mengukur peluang

kesalahan klasifikasi yang dihasilkan MARS. Nilai APER ini menunjukkan proporsi observasi yang salah diklasifikasikan oleh fungsi klasifikasi.

Nilai APER dihitung sebagai berikut,

$$APER(\%) = \frac{n_{12} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}} \times 100\%.$$

Keterangan :

n_{11} : Jumlah kelas aktual 1 yang tepat diklasifikasikan sebagai kelompok 1.

n_{12} : Jumlah kelas aktual 1 yang salah diklasifikasikan sebagai kelompok 2.

n_{21} : Jumlah kelas aktual 2 yang salah diklasifikasikan sebagai kelompok 1.

n_{22} : Jumlah kelas aktual 2 yang tepat diklasifikasikan sebagai kelompok 2.

4. *Bootstrap Agregating (Bagging)*

Metode *bagging* pertama kali digunakan oleh Breiman (1994) sebagai alat untuk membentuk pengklasifikasi yang lebih stabil. *Bagging predictor* adalah metode untuk membangkitkan *multiple version* dari prediktor dan menggunakannya untuk *aggregate* prediktor. *Multiple versions* dibentuk dengan replikasi *bootstrap* dari sebuah dataset.

Misalkan sebuah data set \mathcal{L} terdiri dari $\{(y_i, x_i), i = 1, 2, \dots, n\}$. Dari data set tersebut, dilakukan pengambilan sampel berukuran n dengan pengembalian sehingga didapatkan $\mathcal{L}_i^* = (y_i^*, x_i^*), i = 1, 2, \dots, n$. Pengembalian sampel semacam ini dilakukan sebanyak B kali, sehingga didapatkan $\mathcal{L}^{(B)} = \{\mathcal{L}_i^*, i = 1, 2, \dots, B\}$. Kemudian membuat model klasifikasi MARS dari masing-masing data set hasil sampel *bootstrap* $\mathcal{L}^{(B)}$ sehingga diperoleh klasifikasi *bagging* dari rata-rata ketepatan klasifikasi pada setiap pengambilan sampel B .

5. Metode Penelitian

Data yang digunakan dalam penelitian ini merupakan data hasil Survei Sosial Ekonomi Nasional tahun 2011 (SUSENAS 2011) Kabupaten Padang Pariaman. Variabel respon dalam penelitian ini adalah tingkat kesejahteraan rumah tangga, dimana rumah tangga dibagi menjadi dua kategori, yaitu:

- 1 = rumah tangga miskin ,
- 2 = rumah tangga tidak miskin.

Sedangkan variabel prediktor pada penelitian ini adalah:

X_1 = Jenis kelamin kepala rumah tangga

X_2 = Umur kepala rumah tangga

X_3 = Status perkawinan kepala rumah tangga

X_4 = Ijazah tertinggi kepala rumah tangga

X_5 = Jumlah anggota rumah tangga

X_6 = Kegiatan utama kepala rumah tangga

X_7 = Status pekerjaan utama kepala rumah tangga

X_8 = Status penguasaan tempat tinggal

X_9 = Cara memperoleh air minum

- X_{10} = Sumber penerangan
 X_{11} = Bahan bakar energi utama untuk memasak
 X_{12} = Pengalaman mendapatkan pelayanan kesehatan gratis selama enam bulan
 X_{13} = Pengalaman membeli beras raskin selama tiga bulan terakhir
 X_{14} = Ada anggota keluarga yang dapat menggunakan telepon seluler
 X_{15} = Ada anggota keluarga yang menguasai komputer/dekstop.

Langkah awal yang dilakukan dalam penelitian ini adalah melakukan analisis statistik deskriptif terhadap variabel-variabel prediktor. Selanjutnya baru dilakukan pemodelan *bagging* MARS dengan langkah-langkah sebagai berikut.

- (1) Membentuk model MARS terbaik untuk dataset awal dengan mengkombinasikan banyaknya $BF = 30, 45, 60$, $MI = 1, 2, 3$, dan $MO = 0, 1, 2, 3$.
- (2) Mendapatkan model MARS terbaik untuk dataset awal berdasarkan nilai GCV terkecil.
- (3) Mendapatkan variabel-variabel yang signifikan berpengaruh terhadap model MARS terbaik untuk dataset awal.
- (4) Melakukan *bagging* dari pasangan variabel respon dan variabel prediktor yang signifikan dari model MARS terbaik untuk dataset awal dengan 15, 20, 25, 30, dan 35 replikasi *bootstrap*.
- (5) Melakukan pemodelan MARS pada setiap pengambilan sampel B replikasi *bootstrap* dengan jumlah BF , MI , dan MO di antara *knot* yang sama dengan jumlah BF , MI , dan MO diantara *knot* pada model MARS terbaik untuk dataset awal.
- (6) Mendapatkan nilai ketepatan klasifikasi pada setiap pengambilan sampel B replikasi *bootstrap*.
- (7) Mendapatkan nilai ketepatan klasifikasi *bagging* dari rata-rata ketepatan klasifikasi pada setiap pengambilan sampel sampai B .
- (8) Model *bagging* MARS yang didapatkan adalah model MARS terbaik untuk data set awal. Hal ini dikarenakan nilai *knot* yang berubah-ubah untuk setiap replikasi sehingga estimasi parameternya tidak bisa dirata-rata.
- (9) Menginterpretasikan tingkat kontribusi variabel prediktor yang mempunyai kepentingan dalam pengelompokan variabel respon.

6. Analisis dan Pembahasan

Pembentukan model MARS dilakukan dengan *trial and error*, dengan mengombinasikan beberapa kriteria model MARS, yaitu jumlah fungsi basis maksimum (BF), *maximum interaction* (MI), dan *minimum observation* (MO), sampai diperoleh model terbaik dengan GCV minimum. Didapatkan model MARS terbaik dengan kombinasi $BF=60$, $MI=3$, $MO=1$ dengan GCV minimum 0,062, yaitu:

$$\begin{aligned}
 Y = & 1,00014 + 0,78624 * BF_5 + 0,572332 * BF_7 - 0,276479 * BF_{10} \\
 & - 0,644501 * BF_{11} - 0,611178 * BF_{17} - 0,0236544 * BF_{20} \\
 & - 0,20185 * BF_{21} + 0,150439 * BF_{23} + 0,650067 * BF_{25} \\
 & + 1,13016 * BF_{29} - 0,0412475 * BF_{35} - 0,690529 * BF_{37}
 \end{aligned}$$

$$\begin{aligned}
&+0,0411356 * BF_{39} + 0,11673 * BF_{42} - 0,022368 * BF_{43} \\
&-0,0459884 * BF_{45} - 0,0330924 * BF_{49} + 0,0631606 * BF_{51} \\
&-0,0219992 * BF_{53} - 0,0767398 * BF_{57};
\end{aligned}$$

dengan

$$\begin{aligned}
BF_1 &= \max(0, X_5 - 6); & BF_{29} &= \max(0, X_5 - 6) * BF_{27} \\
BF_3 &= (X_{14} \text{ in } (1)); & BF_{31} &= \max(0, X_2 - 61) \\
BF_4 &= (X_{14} \text{ in } (2)) & BF_{34} &= (X_{13} \text{ in } (1)) * BF_{31} \\
BF_5 &= \max(0, X_5 - 5) * BF_4 & BF_{35} &= (X_{12} \text{ in } (2)) * BF_{34} \\
BF_7 &= (X_{10} \text{ in } (1)) * BF_5 & BF_{37} &= \max(0, X_5 - 4) * BF_{27} \\
BF_{10} &= \max(0, 36 - X_2) * BF_5 & BF_{39} &= (X_4 \text{ in } (3, 2, 5)) * BF_{34} \\
BF_{11} &= \max(0, X_5 - 4) * BF_4 & BF_{42} &= \max(0, 47 - X_2) * BF_{37} \\
BF_{15} &= (X_{11} \text{ in } (4, 5)) * BF_3 & BF_{43} &= (X_{14} \text{ in } (1)) * BF_{34} \\
BF_{17} &= (X_4 \text{ in } (2)) * BF_1 & BF_{45} &= \max(0, X_2 - 66) * BF_{17} \\
BF_{20} &= \max(0, 55 - X_2) * BF_{17} & BF_{47} &= \max(0, X_5 - 3) * BF_4 \\
BF_{21} &= (X_8 \text{ in } (3)) * BF_{11} & BF_{49} &= \max(0, X_2 - 51) * BF_{47} \\
BF_{23} &= \max(0, X_2 - 72) * BF_{17} & BF_{51} &= \max(0, X_2 - 56) * BF_{11} \\
BF_{25} &= (X_6 \text{ in } (1)) * BF_{17} & BF_{53} &= (X_3 \text{ in } (4)) * BF_{34} \\
BF_{27} &= (X_4 \text{ in } (1)) & BF_{57} &= (X_1 \text{ in } (1)) * BF_{15}
\end{aligned}$$

Berikut adalah interpretasi dari BF_{10} :

$$\begin{aligned}
BF_{10} &= \max(0, 36 - X_2) * BF_5; \\
BF_5 &= \max(0, X_5 - 5) * BF_4; \\
BF_4 &= (X_{14} \text{ in } (2));
\end{aligned}$$

dengan koefisien -0,276479, artinya setiap kenaikan BF_{10} sebesar satu satuan, akan mengurangi kecendrungan suatu rumah tangga masuk ke dalam kelompok rumah tangga tidak miskin sebesar 0,276479. Berkurangnya kecendrungan rumah tangga ini terjadi jika kepala dari rumah tangga tersebut berumur kurang dari 36 tahun ($0,36 - X_2$) dan jumlah anggota rumah tangga tersebut lebih dari lima orang ($0, X_5 - 5$) dan tidak ada anggota keluarga tersebut yang menggunakan telepon genggam ($X_{14} \text{ in } (2)$). Untuk BF lainnya dapat dilakukan interpretasi yang serupa.

Selanjutnya dilakukan uji signifikansi fungsi basis secara keseluruhan (secara serempak/bersama) dengan hipotesis sebagai berikut:

$$H_0 : \alpha_j = 0$$

$$H_1 : \text{ada } \alpha_j \neq 0$$

dengan $j = 5, 7, 10, 11, 15, 17, 20, 21, 23, 25, 27, 29, 31, 34, 35, 37, 39, 42, 43, 45, 49, 51, 53, \text{ dan } 57$

Berdasarkan hasil pengolahan MARS diperoleh nilai $F_{hitung} = 24,363$. Dengan $\alpha = 0,05$, nilai $v_1 = 20$ dan $v_2 = 560$ diperoleh nilai $F_{0,05(20,560)} = 1,589$. Karena $F_{hitung} > 1,589$, maka keputusan yang diambil adalah menolak H_0 yang berarti terdapat paling tidak satu α_j yang tidak sama dengan nol atau dengan kata lain terdapat paling tidak satu fungsi basis yang memuat variabel prediktor yang berpengaruh terhadap variabel respon.

Kemudian dilakukan uji secara individual terhadap setiap fungsi basis menggunakan hipotesis sebagai berikut :

$$H(0) : \alpha_j = 0$$

$H(1) : \alpha_j \neq 0$

dengan $j = 5, 7, 10, 11, 15, 17, 20, 21, 23, 25, 27, 29, 31, 34, 35, 37, 39, 42, 43, 45, 49, 51, 53,$ dan 57 .

Dalam pengujian ini digunakan $\alpha = 0,05$ sehingga $t_{(0,025;560)} = 2,247$. Dari perhitungan MARS diperoleh semua fungsi basis menolak H_0 yang berarti semua fungsi basis yang terpilih dalam model MARS memiliki pengaruh terhadap variabel respon.

Tingkat kepentingan masing-masing variabel prediktor ditunjukkan pada tabel berikut:

Tabel 1. Tingkat Kepentingan Variabel Prediktor

Variabel	Importance	-GCV	Variabel	Importance	-GCV
X_5	100,00	0,09353	X_8	18,09	0,06339
X_{14}	84,20	0,08446	X_1	16,65	0,06323
X_2	68,48	0,07698	X_{11}	16,65	0,06323
X_4	62,94	0,07472	X_3	12,08	0,06283
X_{13}	44,68	0,06859	X_7	0	0,06283
X_{10}	37,94	0,06686	X_9	0	0,06283
X_{12}	36,88	0,06661	X_{15}	0	0,06283
X_6	29,83	0,06514			

Variabel-variabel yang berkontribusi pada model inilah yang didefinisikan sebagai faktor-faktor penciri tingkat kesejahteraan.

Dari klasifikasi tingkat kesejahteraan menggunakan model MARS diperoleh nilai APER=6,2%. Sehingga dapat disimpulkan bahwa ketepatan klasifikasi dari pendekatan MARS adalah sebesar 93,8% dengan tingkat kesalahan 6,2%.

Selanjutnya dilakukan pendekatan *bagging* untuk mengetahui akurasi klasifikasi dari model MARS. Pasangan variabel respon dan variabel prediktor yang memberikan kontribusi pada model MARS terbaik direplikasi sebanyak 15, 20, 25, 30, dan 35 kali dengan metode *bootstrap*. Pada setiap dataset baru $\{\mathcal{L}^{(B)}\}$ hasil replikasi dilakukan pemodelan MARS dengan kriteria *BF*, *MI*, dan *MO* yang sama dengan kriteria pembentuk model terbaik, yaitu $BF=60$, $MI=3$, dan $MO=20$. Didapatkan nilai tingkat kesalahan klasifikasi sebanyak B yang berasal dari pemodelan MARS untuk $\{\mathcal{L}^{(B)}\}$. Selanjutnya tingkat kesalahan klasifikasi baru tersebut dirata-rata sehingga menghasilkan nilai agregat kesalahan klasifikasi $_B$.

Tabel 2. Kesalahan Klasifikasi $_B$

$\{\mathcal{L}^{(B)}\}$	Kesalahan Klasifikasi $_B$	Kesalahan Klasifikasi model MARS dataset tunggal	Penurunan kesalahan klasifikasi
15	0,0489	0,062	0,0131
20	0,0496		0,0124
25	0,0505		0,0115
30	0,0505		0,0115
35	0,0513		0,0107

Dapat dilihat pada tabel $\{\mathcal{L}^{(B)}\}$ berhasil memberikan akurasi klasifikasi yang lebih baik daripada klasifikasi model MARS dataset tunggal. Penurunan tingkat klasifikasi terbesar diberikan oleh $\{\mathcal{L}^{(15)}\}$.

7. Kesimpulan

Dari 15 variabel yang diteliti dalam penelitian ini didapatkan 12 faktor penciri tingkat kesejahteraan rumah tangga, lima diantaranya adalah jumlah anggota rumah tangga (X_5), adanya anggota rumah tangga yang menggunakan telepon seluler (X_{14}), umur kepala rumah tangga (X_2), ijazah tertinggi yang dimiliki kepala rumah tangga (X_4), pengalaman membeli beras raskin selama tiga bulan terakhir (X_{13}), dengan tingkat kepentingan masing-masing yaitu 100%, 84,20%, 68,48%, 62,94%, dan 44,68%. Dalam penelitian ini, pendekatan *bagging* dengan 15, 20, 25, 30 dan 35 replikasi terbukti dapat menunjukkan akurasi klasifikasi dari model MARS. Penurunan tingkat klasifikasi terbesar diberikan oleh model MARS *bagging* dengan 15 replikasi.

8. Ucapan Terima kasih

Penulis mengucapkan terima kasih kepada Ibu Hazmira Yozza, M.Si, Ibu Izzati Rahmi H.G, M.Si, Bapak Dr. Dodi Devianto, Ibu Riri Lestari, M.Si, Ibu Dr. Maiyasatri, Bapak Yudiantri Asdi, M.Sc dan Ibu Dr. Lyra Yulianti yang telah memberikan masukan dan saran sehingga paper ini dapat diselesaikan dengan baik.

Daftar Pustaka

- [1] Badan Pusat Statistik. 2012. **Berita Resmi Statistik: Profil Kemiskinan di Indonesia Maret 2012**. BPS. Jakarta
- [2] Buhlman, P., Yu, B. 2002. Analyzing *bagging*, *The Annals of Statistics* **30** (4) : 927 – 961
- [3] Cahyat, A., Gonner, C., dan Haug, M.. 2007. **Mengkaji Kemiskinan dan Kesejahteraan Rumah Tangga : Sebuah Panduan dengan Contoh dari Kutai Barat**. Indonesia. CIFOR. Bogor.
- [4] Dimas, Arvian P. P. 2011. *Klasifikasi Kesejahteraan Rumah Tangga di Jawa Timur dengan Pendekatan Multivariate Adaptive Regression Spline - Bootstrap Aggregating (MARS bagging)*. Tugas Akhir, tidak dipublikasikan, Jurusan Statistika Institut Teknologi Sepuluh Nopember, Surabaya.
- [5] Friedman, J.H. 1991. *Multivariate Adaptive Regression Splines*. *The Annals of Statistics*, Vol. 19 No. 1.