

***Bootstrap Aggregating Multivariate Adaptive Regression Splines***  
**(Studi Kasus : Identifikasi Komponen Penciri Akreditasi Sekolah/Madrasah Pada Tingkat SD/MI di Provinsi Kalimantan Timur Tahun 2015)**

***Bootstrap Aggregating Multivariate Adaptive Regression Splines***  
**(Case Study : Identification of Identifier Components of Accreditation Schools at primary school in East Kalimantan Province 2015)**

**Marisa Nanda Rahmaniah<sup>1</sup>, Yuki Novia N<sup>2</sup>, Ika Purnamasari<sup>3</sup>**

<sup>1</sup>Mahasiswa Program Studi Statistika FMIPA Universitas Mulawarman

<sup>2,3</sup>Dosen Program Studi Statistika FMIPA Universitas Mulawarman

Email: syahidah\_green@yahoo.co.id

**Abstract**

*MARS is one of the classification methods that focus on the high dimension and discontinuity of the data. The level of accuracy in MARS can be improved by using Bagging method (Bootstrap Agregating). This method is used to improve stability, accuracy and strength's of prediction. This study discusses the MARS bagging applications in analyzing the issue of accreditation, which the accreditation level of a schools can be predicted based on the identifier components. Therefore, in this study will be identified these components to create a classification model. The data used is the accreditation data of the primary school in East Kalimantan Province 2015 issued by the Accreditation Board of the Provincial Schools (BAP-S/M) of East Kalimantan Province. This study obtained six components that affect the determination of the accreditation of schools at primary school level. The components are the variables that contribute to the classification. The variables are a standard component of content ( $X_1$ ), a standard component of the process ( $X_2$ ), a standard component of graduates ( $X_3$ ), standard components of teachers and staffs ( $X_4$ ), a standard component of infrastructure ( $X_5$ ) and standard component of financial ( $X_7$ ). Based on the result of the classification accuracy of MARS method (using Apparent Error Rate (APER), it is amounted to 78.87%, while the classification accuracy (using APER) with method of bagging of the best MARS models amounted to 89.44%. This means that the method of bagging MARS gives better classification accuracy of the classification than MARS.*

*Keywords: School accreditation, APER, bagging, MARS.*

**Pendahuluan**

Pendekatan regresi nonparametrik merupakan metode pendugaan model yang dilakukan berdasarkan pendekatan yang tidak terikat asumsi bentuk kurva regresi tertentu. Kurva regresi berdasarkan pendekatan nonparametrik ini, diwakili oleh model yang disebut model regresi nonparametrik. Karena sebelumnya tidak ada asumsi mengenai bentuk kurva regresi, model regresi nonparametrik dapat berbentuk fungsi apa saja, baik linier maupun nonlinier. Komputasi dan perhitungan dalam menduga model merupakan kendala utama dalam regresi nonparametrik. Namun dengan perkembangan media komputer yang sangat pesat, regresi nonparametrik turut berkembang pula (Hardle, 1990). Ada beberapa teknik pendugaan nilai variabel terikat dalam regresi nonparametrik, salah satunya adalah metode *Multivariat Adaptive Regression Splines (MARS)*.

*MARS* adalah salah satu metode klasifikasi yang inovatif dan relatif fleksibel untuk menyelidiki hubungan antara variabel terikat dan variabel bebas tanpa asumsi terhadap bentuk fungsionalnya. *MARS* difokuskan untuk mengatasi permasalahan data yang berdimensi

tinggi yaitu data yang memiliki jumlah variabel bebas sebesar  $3 \times n$  20. Pada metode *MARS* juga dapat ditambahkan atau melibatkan banyak interaksi antar variabel (Friedman, 1991).

Tingkat akurasi klasifikasi model *MARS* dapat ditingkatkan dengan menggunakan metode *resampling*, salah satu metode *resampling* adalah *Bagging (Bootstrap Agregating)*. Metode ini digunakan untuk memperbaiki stabilitas, meningkatkan akurasi dan kekuatan prediktif. *Bagging* dikenal sangat efektif bila pengklasifikasi tidak stabil, karena *bagging* meningkatkan kinerja generalisasi dengan cara mengurangi varians (*noise*) (Breiman, 1996).

Penerapan *MARS* telah banyak digunakan untuk menyelesaikan permasalahan klasifikasi dan peramalan, baik dalam bidang kesehatan maupun sosial ekonomi penduduk, misalnya dalam hal pembangunan, dimana pembangunan merupakan proses yang berkesinambungan yang mencakup seluruh aspek kehidupan masyarakat termasuk sosial, ekonomi, kultural dan politik, dengan tujuan utama meningkatkan kesejahteraan warga bangsa secara keseluruhan. Dalam proses pembangunan tersebut peranan pendidikan sangat strategis (BPS Kota Samarinda, 2009).

Pada beberapa penelitian sebelumnya, yaitu yang dilakukan oleh Budiarti (2014), metode *MARS* digunakan untuk memodelkan komponen yang berpengaruh terhadap peringkat akreditasi sekolah/madrasah pada tingkat SMA/SMK/MA di Provinsi Kalimantan Timur. Kemudian, penelitian yang dilakukan oleh Mandaku (2010), metode regresi logistik dan *MARS* digunakan untuk memodelkan kelulusan siswa masuk kelas akselerasi. Selanjutnya, penelitian yang dilakukan oleh Arleina dan Otok (2014), yang menggunakan metode *Bagging MARS* untuk pengklasifikasian rumah tangga miskin di Kabupaten Jombang.

Sekolah/madrasah tingkat SD/MI merupakan pendidikan tingkat pertama yang memberikan bekal kemampuan dasar kepada siswa untuk mengembangkan kehidupannya sebagai pribadi, anggota masyarakat, warga negara dan anggota umat manusia serta mempersiapkan siswa untuk mengikuti pendidikan menengah. Oleh karena itu penulis ingin mengidentifikasi dan memodelkan komponen penilaian apa saja yang menjadi penciri dalam penentuan akreditasi sekolah/madrasah di Provinsi Kalimantan Timur pada tingkat SD/MI dengan menggunakan metode *MARS* yang dikombinasikan dengan *Bagging*,

**Analisis Regresi**

Analisis regresi adalah suatu metode sederhana untuk melakukan investigasi tentang hubungan fungsional di antara beberapa variabel. Hubungan antara beberapa variabel tersebut diwujudkan dalam suatu model matematis (Nawari, 2010). Pada model regresi, variabel dibedakan menjadi dua bagian, yaitu variabel terikat dan variabel bebas. Misalkan  $x_1, x_2, \dots, x_n$  adalah variabel bebas dan  $y$  variabel terikat, maka dapat dirumuskan dalam persamaan sebagai berikut:

$$\hat{y} = f(x_1, x_2, \dots, x_n) \tag{1}$$

Nilai  $\hat{y}$  menyatakan dugaan terhadap variabel terikat  $y$  dan  $x$  menyatakan variabel bebas. Dalam kenyataannya nilai dugaan ( $\hat{y}$ ) yang diberikan oleh model regresi tidak selalu sama persis dengan nilai sebenarnya ( $y$ ), melainkan terdapat selisih. Selisih inilah yang kemudian disebut sebagai error ( $V$ ). Oleh karenanya, model persamaan regresi juga dapat dituliskan dengan rumus:

$$y_i = f(x_{1i}, x_{2i}, \dots, x_{ni}) + v_i \quad (i = 1, 2, \dots, n) \tag{2}$$

Berdasarkan rumus tersebut dapat diterangkan bahwa secara matematis model regresi yang baik adalah model yang mampu membangun nilai  $\hat{y}$  sedemikian rupa sehingga mendekati nilai  $y$  dan  $V$  mendekati nol (Nawari, 2010).

**Regresi Nonparametrik**

Regresi nonparametrik merupakan suatu metode statistika yang digunakan untuk mengetahui hubungan antara variabel terikat dan variabel bebas yang tidak diketahui bentuk kurva regresinya atau tidak terdapat informasi masa lalu yang lengkap tentang bentuk pola data (Eubank, 1999).

Model regresi nonparametrik secara umum adalah sebagai berikut (Eubank, 1999):

$$y_i = f(x_i) + v_i \quad (i = 1, 2, \dots, n) \tag{3}$$

dimana :

$y_i$  : variabel terikat

$x_i$  : variabel bebas

$f(x_i)$  : fungsi regresi

$v_i$  : error

dengan ( $i = 1, 2, \dots, n$ )

**Regresi Spline**

Regresi *spline* merupakan salah satu metode yang digunakan untuk menangani pola data yang mengikuti siklus nonlinier serta bentuk kurvanya tidak diketahui. Regresi *spline* terdiri dari beberapa penggal polinom berorde tertentu yang sering bersambung pada titik-titik ikat. Nilai absis dari titik ikat ini disebut *knot*. Regresi *spline* bersifat fleksibel sehingga model yang didapat akan cenderung sedekat mungkin menggambarkan kondisi sebenarnya (Eubank, 1999).

**Recursive Partitioning Regression (RPR)**

RPR adalah salah satu metode pemodelan regresi yang biasanya digunakan untuk data berdimensi tinggi karena penentuan *knot* tergantung dari data. Namun demikian, model RPR masih memiliki kelemahan, diantaranya yaitu model RPR menghasilkan himpunan bagian yang saling lepas dan diskontinyu pada basis himpunan bagian, serta model RPR tidak cukup mampu dalam menduga fungsi linier atau aditif.

Diberikan  $H[y]$ , yang merupakan suatu fungsi tangga (*step function*) yang berbentuk:

$$H[y] = \begin{cases} 1, & \text{untuk } v \geq 0 \\ 0, & \text{untuk lainnya} \end{cases}$$

maka fungsi basis yang dihasilkan pada prosedur RPR dapat dinyatakan sebagai berikut :

$$B_m(x) = \prod_{k=1}^K H \left[ S_{km} \cdot (x_{v(k,m)} - t_{km}) \right] \tag{4}$$

dimana :

$H[...]$  = fungsi tangga

- $K_m$  = jumlah penggolongan himpunan bagian ke- $m$
- $x_{v(k,m)}$  = variabel bebas ke- $v$ , pilahan ke- $k$  dan subregion ke- $m$
- $S_{km}$  = nilainya  $\pm 1$
- $t_{km}$  = nilai *knot* dari variabel bebas  $x_{v(k,m)}$

(Friedman, 1991)

**Multivariate Adaptive Regression Splines (MARS)**

MARS merupakan suatu pendekatan analisis untuk memodelkan regresi nonparametrik multivariat dengan kombinasi yang kompleks dari

$$\hat{f}(x) = s_0 + \sum_{m=1}^M s_m \prod_{k=1}^{K_m} (s_{km} (x_{v(k,m)} - t_{km}))$$

*spline* dan partisi rekursif yang diperkenalkan oleh Friedman (1991). Metode ini merupakan pengembangan dari pendekatan *Recursive Partitioning Regression (RPR)* yang masih memiliki kelemahan di mana model yang dihasilkan tidak kontinu pada *knot*. Metode MARS mampu mengatasi semua kelemahan yang dimiliki oleh metode RPR dengan menggunakan algoritma RPR yang dimodifikasi. Model MARS sebagai berikut :

(5)

dengan :

- $s_0$  = konstanta regresi dari fungsi basis
- $s_m$  = koefisien dari fungsi basis ke- $m$
- $M$  = maksimum fungsi basis (fungsi basis yang tidak konstan)
- $K_m$  = banyaknya interaksi pada fungsi basis ke- $m$
- $S_{km}$  = nilainya  $\pm 1$
- $x_{v(k,m)}$  = variabel bebas ke- $v$ , pilahan ke- $k$  dan subregion ke- $m$
- $t_{km}$  = nilai *knot* dari variabel bebas  $x_{v(k,m)}$
- $v$  = banyaknya variabel bebas
- $k$  = banyaknya interaksi

Beberapa hal yang perlu diperhatikan dalam membangun model MARS yaitu (Friedman, 1991) :

1. *Knot*, merupakan variabel bebas ketika *slope* suatu garis regresi mengalami perubahan yang dapat didefinisikan sebagai akhir dari segmen sekaligus merupakan awal dari segmen yang lain. Di setiap titik *knot*, diharapkan ada kontinuitas dari fungsi basis antar satu region dengan region lainnya. Minimum observasi (MO) antara *knot* yang

biasa digunakan adalah 0, 1, 2 dan 3 observasi.

2. Fungsi basis (BF) yaitu selang antara *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 sampai 4 kali jumlah variabel bebasnya.
3. *Interaction* (interaksi) yaitu hasil perkalian silang antara 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 diinterpretasikan.

Pembentukan model MARS diawali dengan terlebih dahulu menentukan *knot* dan fungsi basis setiap variabel bebas. Pemilihan *knot* pada MARS dilakukan dalam dua tahap yaitu tahap *forward* dan tahap *backward*.

1. Tahap *Forward*

Pada tahap ini, digunakan kriteria pemilihan fungsi basis yaitu dengan meminimumkan *Average Sum of Square Residual (ASR)* atau rata-rata jumlah kuadrat, untuk mendapatkan jumlah fungsi basis maksimum.

$$ASR = \left( \frac{1}{N} \right) \sum_{i=1}^N [y_i - \hat{f}_M(x_i)]^2 \tag{6}$$

dimana :

- $N$  = jumlah pengamatan
- $y_i$  = variabel terikat
- $x_i$  = variabel bebas
- $\hat{f}_M(x_i)$  = taksiran parameter
- $M$  = banyaknya fungsi basis (fungsi basis yang tidak konstan)

2. Tahap *Backward*

Pada tahap *backward* akan dipilih satu fungsi basis dan mengeluarkan basis tersebut jika kontribusi terhadap model terkecil. Proses *backward* akan dilanjutkan hingga tidak ada fungsi basis yang 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 :

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

dimana :

$D$  = nilai ketika setiap fungsi basis mencapai optimasi ( $2 \leq d \leq 4$ )

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

$$C(\hat{M}) = C(M) + dM$$

dilakukan tahap *backward*, bertujuan untuk memilih fungsi basis yang dihasilkan dari tahap *forward* dengan meminimumkan nilai GCV (Friedman dan Silverman, 1989).

**Klasifikasi pada MARS**

Jika melakukan klasifikasi pada variabel respon yang terdiri dari dua nilai, maka dapat digunakan model probabilitas dengan persamaan sebagai berikut (Cox dan Snell, 1989) :

$$f(x) = \frac{e^{\hat{f}(x)}}{1 + e^{\hat{f}(x)}}$$

$$\text{dan } \{1 - f(x)\} = \frac{1}{1 + e^{\hat{f}(x)}} \tag{8}$$

dengan,

$$P(Y = 1 / X = x) = f(x) = \frac{e^{\hat{f}(x)}}{1 + e^{\hat{f}(x)}}$$

$$P(Y = 0 / X = x) =$$

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

Untuk menghitung ketepatan klasifikasi pada hasil pengelompokan digunakan *APER* (*Apparent Error Rate*) dengan perhitungan pada tabel klasifikasi sebagai berikut :

Tabel 1. Tabel Klasifikasi

Hasil Observasi	Taksiran Observasi	
	$y_1$	$y_2$
$y_1$	$n_{11}$	$n_{12}$
$y_2$	$n_{21}$	$n_{22}$

dengan,

$y_1$  = variabel terikat kategori 1

$y_2$  = variabel terikat kategori 2

$n_{11}$  = jumlah observasi dari  $y_1$  yang tepat diklasifikasikan sebagai  $y_1$

$n_{22}$  = jumlah observasi dari  $y_2$  yang tepat diklasifikasikan sebagai  $y_2$

$n_{21}$  = jumlah observasi dari  $y_2$  yang salah diklasifikasikan sebagai  $y_1$

$n_{12}$  = jumlah observasi dari  $y_1$  yang salah diklasifikasikan sebagai  $y_2$

Nilai *APER* didapatkan dengan perhitungan sebagai berikut :

$$APER(\%) = \frac{n_{21} + n_{12}}{n} \times 100\% \tag{9}$$

Sebagai statistik uji untuk mengetahui kestabilan dalam ketetapan klasifikasi digunakan uji *Press's Q*, dengan persamaan sebagai berikut :

$$Press's Q = \frac{[N - (n_i K)]^2}{N(K - 1)} \tag{10}$$

dimana,

$N$  = jumlah pengamatan

$n_i$  = jumlah individu yang tepat untuk diklasifikasikan ( $n_{11} + n_{22}$ )

$K$  = jumlah kelompok

Jika *Press's Q* melebihi nilai kritis, maka klasifikasi dapat dianggap sudah stabil dan konsisten secara statistik (Hair dkk, 2006). Dengan hipotesis sebagai berikut :

$H_0$  : Klasifikasi dapat dianggap belum stabil dan konsisten secara statistik

$H_1$  : Klasifikasi dapat dianggap sudah stabil dan konsisten secara statistik

**Metode Bootstrap**

Metode *bootstrap* pertama kali diperkenalkan oleh Bradley Efron pada tahun 1979. *Bootstrap* merupakan salah satu metode *resampling* (pengambilan sampel ulang dari data yang ada). Efron (1979) menyatukan dan menggabungkan metode-metode *resampling* menjadi suatu metode *resampling* baru yaitu *bootstrap* nonparametrik sederhana. Pada metode ini *resampling* pada data dilakukan dengan pengembalian dari sampel hasil observasi dengan replikasi  $B$  kali.

Prinsip metode *bootstrap* ialah untuk memperkirakan parameter masing-masing sampel *bootstrap* sebanyak  $B$  yang merupakan sampel acak berukuran  $n$  yang diambil dengan pengembalian dari populasi  $n$  pengamatan. Pengamatan ke- $k$  ( $k = 1, 2, 3, \dots, n$ ) dari sampel awal mungkin akan muncul beberapa kali pada sampel *bootstrap* replikasi ke- $r$  ( $r = 1, 2, 3, \dots, B$ ), sedangkan pengamatan lain mungkin tidak akan muncul sama sekali (Gujarati, 1999). Maksudnya, dalam sampel *bootstrap*, beberapa nilai sampel asli akan menjadi berulang, dan beberapa di antaranya mungkin tidak akan terjadi sama sekali (Spren, 1991). Kedudukan sampel asli dalam *bootstrap* dipandang sebagai populasi. *Bootstrap* merupakan suatu metode pendekatan nonparametrik untuk menaksir berbagai kuantitas statistik seperti rata-rata (*mean*), standar deviasi *error* dan bias suatu estimator atau untuk

membentuk interval konfidensi dengan memanfaatkan kecanggihan teknologi komputer.

Berdasarkan hasil penelitian Efron dan Tibshirani (1993), berikut dua poin penting dalam menentukan besaran  $B$  :

1. Menggunakan replikasi *bootstrap* yang sedikit, katakanlah  $B = 25$ , biasanya sudah cukup informatif.  $B = 50$  terkadang sudah cukup untuk memberikan estimasi simpangan baku untuk dengan baik.
2. Penggunaan replikasi  $B = 200$  ke atas jarang diperlukan untuk estimasi simpangan baku. Akan tetapi, penggunaan  $B$  yang lebih besar diperlukan untuk membuat interval kepercayaan *bootstrap*.

**Bootstrap Aggregating (Bagging) pada MARS**

*Bagging* pertama kali digunakan oleh Breiman (1994) sebagai alat untuk membentuk pengklasifikasi yang lebih stabil. Metode ini digunakan untuk mereduksi variansi estimator pada metode klasifikasi dan regresi, dan penggunaannya tidak dibatasi hanya untuk memperbaiki estimator. Teknik ini dapat juga digunakan untuk memperbaiki stabilitas, meningkatkan akurasi dan kekuatan prediktif (Breiman, 1996). Dengan kata lain, misalkan sebuah data set  $L_1^*$  terdiri dari  $\{(y_i, x_i), i = 1, 2, \dots, n\}$ . Dari himpunan data set tersebut, dilakukan pengambilan sampel berukuran  $n$  dengan pengembalian sehingga didapatkan  $L_1^* = (y_i^*, x_i^*), i = 1, 2, \dots, n$ . Pengembalian sampel seperti ini dilakukan sebanyak  $B$  kali, sehingga didapatkan  $L^{(B)} = \{L_i^*, i = 1, 2, \dots, B\}$ . Kemudian membuat model klasifikasi MARS dari masing-masing data set hasil sampel *bootstrap*  $\{L^{(B)}\}$  sehingga diperoleh ketetapan klasifikasi pada setiap pengambilan sampel  $B$  (Breiman, 1996).

**Akreditasi Sekolah/Madrasah**

Akreditasi sekolah/madrasah adalah suatu kegiatan penilaian kelayakan suatu sekolah/madrasah berdasarkan kriteria yang telah ditetapkan dan dilakukan oleh BAN-S/M yang hasilnya diwujudkan dalam bentuk pengakuan peringkat kelayakan. BAN-S/M atau Badan Akreditasi Nasional Sekolah/Madrasah adalah badan evaluasi mandiri yang menetapkan kelayakan program dan/atau satuan pendidikan jenjang pendidikan dasar dan menengah jalur formal dengan mengacu pada standar nasional pendidikan. Dalam melaksanakan akreditasi sekolah/madrasah, BAN-S/M dibantu oleh Badan Akreditasi Provinsi Sekolah/Madrasah (BAP-S/M).

**Metodologi Penelitian**

Data yang digunakan dalam penelitian ini merupakan data akreditasi sekolah/madrasah tingkat SD/MI yang telah divitasi pada tahun 2015 di Provinsi Kalimantan Timur (sebanyak 142 sekolah/madrasah), yang diperoleh dari BAP-S/M Provinsi Kalimantan Timur. Variabel dalam penelitian ini meliputi :

$$Y = \text{Akreditasi Sekolah} = \begin{cases} 0 = \text{Akreditasi B} \\ 1 = \text{Akreditasi A} \end{cases}$$

- $X_1$  = Komponen standar isi
  - $X_2$  = Komponen standar proses
  - $X_3$  = Komponen standar lulusan
  - $X_4$  = Komponen standar pendidik dan tenaga didik
  - $X_5$  = Komponen standar sarana dan prasarana
  - $X_6$  = Komponen standar pengelolaan
  - $X_7$  = Komponen standar pembiayaan
  - $X_8$  = Status penguasaan tempat tinggal
  - $X_9$  = Komponen standar penilaian pendidikan
- Dimana variabel  $X_1$  hingga  $X_8$  memiliki kisaran nilai antara 0 – 100.

Adapun teknik analisis data dalam penelitian ini adalah :

1. Analisis deskriptif
2. *Bagging MARS*  
 Analisis *bagging MARS* yang dilakukan pada penelitian ini, dengan langkah-langkah sebagai berikut :
  - a. Membentuk model MARS terbaik untuk dataset awal dengan mengkombinasikan banyaknya BF (*Basis Function*) = 16, 24, 32, MI (Maksimum Interaksi) = 1,2,3 dan minimal jumlah pengamatan setiap *knot* (MO) = 0,1,2,3.
  - b. Mendapatkan model MARS terbaik untuk dataset awal berdasarkan nilai GCV terkecil.
  - c. Mendapatkan variabel-variabel yang signifikan berpengaruh terhadap model MARS terbaik untuk dataset awal.
  - d. Menginterpretasikan tingkat kontribusi dan pengurangan GCV variabel yang mempunyai kepentingan dalam pengelompokkan variabel terikat.
  - e. Menginterpretasikan model MARS terbaik untuk dataset awal.
  - f. Menguji keakurasian prediksi model MARS (ketepatan klasifikasi) yang terbentuk dari data dengan menggunakan *APER* serta menghitung kestabilan klasifikasi dengan statistik uji *Press's Q*.
  - g. Melakukan *bagging* dari pasangan variabel terikat dan variabel bebas yang signifikan dari model MARS terbaik untuk dataset awal dengan 50 replikasi *bootstrap*.
  - h. Melakukan pemodelan MARS pada setiap pengambilan sampel  $B$  replikasi

bootstrap dengan jumlah BF, MI, MO diantara knot pada model MARS terbaik untuk dataset awal.

- i. Mendapatkan model bagging MARS terbaik berdasarkan nilai GCV terkecil.
- j. Mendapatkan nilai keakuratan ketepatan klasifikasi dari model bagging MARS.

**Hasil dan Pembahasan**

**Analisis Deskriptif**

- 1. Kelompok Akreditasi Sekolah/Madrasah Tingkat SD/MI dengan Peringkat A

Dari data pada lampiran 1 dilakukan analisis deskriptif terhadap sekolah/madrasah tingkat SD/MI dengan peringkat A untuk mengetahui rata-rata, nilai minimum dan maksimum pada tiap-tiap variabel, diperoleh sebagai berikut :

Tabel 2. Statistika Deskriptif Kelompok Akreditasi Sekolah/Madrasah Tingkat SD/MI di Provinsi Kalimantan Timur dengan Peringkat A

Variabel Bebas	Nilai		
	Rata-rata	Minimum	Maksimum
X <sub>1</sub>	91,78	83	100
X <sub>2</sub>	90,95	80	100
X <sub>3</sub>	85,50	63	98
X <sub>4</sub>	88,18	76	100
X <sub>5</sub>	89,88	76	100
X <sub>6</sub>	89,55	64	100
X <sub>7</sub>	93,58	77	100
X <sub>8</sub>	91,62	80	100

Berdasarkan Tabel 2, diketahui bahwa komponen akreditasi sekolah/madrasah tingkat SD/MI dengan peringkat A yang memiliki rata-rata tertinggi terdapat pada komponen standar pembiayaan (X<sub>7</sub>) yaitu sebesar 93,58. Kemudian untuk komponen yang memiliki nilai minimum paling besar adalah komponen standar isi (X<sub>1</sub>) yaitu sebesar 83. Selanjutnya untuk komponen yang memiliki nilai maksimum terkecil adalah komponen standar lulusan (X<sub>3</sub>) yaitu sebesar 98.

- 2. Kelompok Akreditasi Sekolah/Madrasah Tingkat SD/MI dengan Peringkat B

Dari data dilakukan analisis deskriptif terhadap sekolah/madrasah tingkat SD/MI dengan peringkat B untuk mengetahui rata-rata, nilai minimum dan maksimum pada tiap-tiap variabel, diperoleh pada Tabel 3.

Berdasarkan Tabel 3, diketahui bahwa komponen akreditasi sekolah/madrasah tingkat SD/MI dengan peringkat B yang memiliki rata-rata tertinggi dan nilai maksimum terbesar terdapat pada komponen standar pembiayaan (X<sub>7</sub>) yaitu sebesar 83,90 dan 100. Kemudian untuk komponen yang memiliki nilai minimum paling besar adalah komponen standar isi (X<sub>1</sub>) yaitu sebesar 69.

Tabel 3. Statistika Deskriptif Kelompok Akreditasi Sekolah/Madrasah Tingkat SD/MI di Provinsi Kalimantan Timur dengan Peringkat B

Variabel Bebas	Nilai		
	Rata-rata	Minimum	Maksimum
X <sub>1</sub>	80,22	69	98
X <sub>2</sub>	80,30	55	98
X <sub>3</sub>	72,68	45	92
X <sub>4</sub>	75,75	54	90
X <sub>5</sub>	73,85	47	90
X <sub>6</sub>	77,99	54	97
X <sub>7</sub>	83,90	62	100
X <sub>8</sub>	80,66	58	96

**Bagging MARS**

**Konstruksi Model MARS**

Pembentukan model dengan menggunakan metode MARS dilakukan dengan trial and error BF, MI dan MO sampai diperoleh model terbaik dengan nilai GCV minimum. Nilai BF yang digunakan dalam pembentukan model adalah 16, 24, dan 32. Kemudian nilai MI yang digunakan adalah satu sampai dengan tiga, sedangkan nilai MO yang digunakan adalah nol sampai dengan tiga.

Dari hasil analisis diperoleh kombinasi BF=32, MI=3 dan MO=2 yang memiliki nilai GCV minimum yaitu sebesar 0,02143, dengan 6 variabel yang masuk ke dalam model, yaitu X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub>, X<sub>4</sub>, X<sub>5</sub>, dan X<sub>7</sub>. Berikut merupakan model MARS terbaik:

$$\hat{f}(x) = -0,0185072 + 0,0312087 BF_3(x) - 0,00397983 BF_6(x) - 0,00712428 BF_7(x) + 0,0165526 BF_9(x) + 0,00451169 BF_{11}(x) - 0,00131878 BF_{16}(x) - 0,022762 BF_{19}(x) + 0,018023 BF_{21}(x) - 0,0010072 BF_2(x) + 0,000215826 BF_{28}(x) - 0,00317385 BF_{29}(x) - 0,0133896 BF_{31}(x)$$

dengan,

- BF<sub>1</sub>(x) = max (0, X<sub>5</sub> - 78)
- BF<sub>3</sub>(x) = max (0, X<sub>1</sub> - 81)
- BF<sub>6</sub>(x) = max (0, 85 - X<sub>3</sub>) BF<sub>3</sub>(x)
- BF<sub>7</sub>(x) = max (0, X<sub>1</sub> - 87) BF<sub>1</sub>(x)
- BF<sub>9</sub>(x) = max (0, X<sub>7</sub> - 91) BF<sub>6</sub>(x)
- BF<sub>11</sub>(x) = max (0, X<sub>5</sub> - 70) BF<sub>3</sub>(x)
- BF<sub>16</sub>(x) = max (0, 80 - X<sub>2</sub>) BF<sub>11</sub>(x)
- BF<sub>17</sub>(x) = max (0, X<sub>4</sub> - 78)
- BF<sub>19</sub>(x) = max (0, X<sub>1</sub> - 84) BF<sub>17</sub>(x)
- BF<sub>21</sub>(x) = max (0, X<sub>4</sub> - 78) BF<sub>3</sub>(x)
- BF<sub>24</sub>(x) = max (0, 77 - X<sub>5</sub>) BF<sub>21</sub>(x)
- BF<sub>28</sub>(x) = max (0, 78 - X<sub>5</sub>) BF<sub>6</sub>(x)
- BF<sub>29</sub>(x) = max (0, X<sub>7</sub> - 88) BF<sub>6</sub>(x)

$BF_{31}(x) = \max(0, X_7 - 92) BF_6(x)$   
 $B_m$  yang masuk ke dalam model adalah  $BF_3(x)$ ,  $BF_6(x)$ ,  $BF_7(x)$ ,  $BF_9(x)$ ,  $BF_{11}(x)$ ,  $BF_{16}(x)$ ,  $BF_{19}(x)$ ,  $BF_{21}(x)$ ,  $BF_{24}(x)$ ,  $BF_{28}(x)$ ,  $BF_{29}(x)$  dan  $BF_{31}(x)$ .

**Ketepatan Klasifikasi pada Model MARS**

Evaluasi prosedur klasifikasi adalah suatu evaluasi yang melihat peluang kesalahan klasifikasi yang dilakukan oleh suatu fungsi klasifikasi (Johnson dan Wichern, 1991).

Tabel 4. Tingkat Akurasi Klasifikasi MARS

Hasil Observasi	Taksiran		Total
	Akreditasi A	Akreditasi B	
Akreditasi A	40	0	40
Akreditasi B	30	72	102
Total	70	72	142

$$APER(\%) = \frac{30+0}{142} \times 100\% = 21,13\%$$

Berdasarkan Tabel 4 dan perhitungan *APER*, terlihat bahwa kesalahan klasifikasi secara keseluruhan adalah 21,13%. Sehingga dari dugaan dengan model *MARS*, hanya menghasilkan 78,87% sekolah/madrasah pada tingkat SD/MI dengan akreditasi A ataupun B yang terklasifikasi sesuai dengan data asli.

Selanjutnya untuk melihat kestabilan klasifikasi, maka digunakan uji *statistic Press's Q*, ialah sebagai berikut :

Hipotesis

$H_0$ : Klasifikasi dapat dianggap belum stabil dan konsisten secara statistik

$H_1$ : Klasifikasi dapat dianggap sudah stabil dan konsisten secara statistik

Kriteria Pengujian

Jika *Press's Q*  $\geq$  nilai kritis (tabel *Chi-Square* dengan derajat bebas 1), maka  $H_0$  ditolak.

Statistik Uji

$$Press's Q = \frac{[N - (n_i K)]^2}{N(K-1)} = \frac{[142 - (112 \times 2)]^2}{142(2-1)} = 47,35$$

Keputusan dan kesimpulan

Nilai (*Press's Q*=47,35)  $>$  (3,841 =  $t_{(1;0,05)}^2$ ), maka  $H_0$  ditolak, sehingga keakuratan pengklasifikasian dengan pendekatan *MARS* dapat dikatakan sudah konsisten secara statistik.

**Pendekatan Bagging MARS**

Pasangan variabel terikat dan variabel bebas yang memberikan kontribusi pada model *MARS* terbaik direplikasi sebanyak 50 kali dengan metode *bootstrap*. Pada dataset baru  $\{L^{(50)}\}$  hasil replikasi dilakukan pemodelan *MARS* dengan

kriteria BF, MI, MO yang sama dengan kriteria pembentuk model terbaik. Setelah dilakukan *bagging MARS*, diperoleh model terbaik pada pengulangan ke-45 dengan nilai GCV sebesar 0,0006 dengan model yang diperoleh sebagai berikut :

$$\hat{f}(x) = 0,00108073 - 0,0264876 BF_3(x) - 0,00352463 BF_7(x) - 0,00903669 BF_8(x) + 0,00604773 BF_9(x) - 0,10431 BF_{10}(x) + 0,00980986 BF_{11}(x) + 0,0132783 BF_{12}(x) - 0,0446058 BF_{13}(x) - 0,00723745 BF_{15}(x) - 0,0017768 BF_{16}(x) + 0,303644 BF_{17}(x) - 0,0267962 BF_{19}(x) - 0,0151973 BF_{20}(x) + 0,00100123 BF_{21}(x) + 0,000898042 BF_{22}(x) + 0,00242552 BF_{23}(x) - 0,000166413 BF_{24}(x) + 0,00170871 BF_{25}(x) + 0,00339154 BF_{26}(x) + 0,192372 BF_{27}(x) + 0,00661582 BF_{29}(x) + 0,00282233 BF_{31}(x) + 0,000679721 BF_{32}(x)$$

dengan,

- $BF_1(x) = \max(0, X_5 - 80)$
- $BF_3(x) = \max(0, X_3 - 88) BF_1(x)$
- $BF_4(x) = \max(0, 88 - X_3) BF_1(x)$
- $BF_7(x) = \max(0, X_1 - 87) BF_1(x)$
- $BF_8(x) = \max(0, 87 - X_1) BF_1(x)$
- $BF_9(x) = \max(0, X_7 - 92) BF_1(x)$
- $BF_{10}(x) = \max(0, 92 - X_7) BF_1(x)$
- $BF_{11}(x) = \max(0, X_4 - 80) BF_{10}(x)$
- $BF_{12}(x) = \max(0, 80 - X_4) BF_{10}(x)$
- $BF_{13}(x) = \max(0, X_5 - 91)$
- $BF_{15}(x) = \max(0, X_2 - 94) BF_{13}(x)$
- $BF_{16}(x) = \max(0, 94 - X_2) BF_{13}(x)$
- $BF_{17}(x) = \max(0, X_3 - 88)$
- $BF_{19}(x) = \max(0, X_7 - 91) BF_{17}(x)$
- $BF_{20}(x) = \max(0, 91 - X_7) BF_{17}(x)$
- $BF_{21}(x) = \max(0, X_1 - 82) BF_4(x)$
- $BF_{22}(x) = \max(0, 82 - X_1) BF_4(x)$
- $BF_{23}(x) = \max(0, X_3 - 91) BF_9(x)$
- $BF_{24}(x) = \max(0, 91 - X_3) BF_9(x)$
- $BF_{25}(x) = \max(0, X_1 - 89) BF_{10}(x)$
- $BF_{26}(x) = \max(0, 89 - X_1) BF_{10}(x)$
- $BF_{27}(x) = \max(0, X_7 - 98) BF_8(x)$
- $BF_{29}(x) = \max(0, X_3 - 79) BF_1(x)$
- $BF_{31}(x) = \max(0, X_2 - 77) BF_1(x)$
- $BF_{32}(x) = \max(0, 77 - X_2) BF_1(x)$

$B_m$  yang masuk ke dalam model adalah  $BF_3(x)$ ,  $BF_7(x)$ ,  $BF_8(x)$ ,  $BF_9(x)$ ,  $BF_{10}(x)$ ,  $BF_{11}(x)$ ,

$BF_{12}(x)$ ,  $BF_{13}(x)$ ,  $BF_{15}(x)$ ,  $BF_{16}(x)$ ,  $BF_{17}(x)$ ,  $BF_{19}(x)$ ,  $BF_{20}(x)$ ,  $BF_{21}(x)$ ,  $BF_{22}(x)$ ,  $BF_{23}(x)$ ,  $BF_{24}(x)$ ,  $BF_{25}(x)$ ,  $BF_{26}(x)$ ,  $BF_{27}(x)$ ,  $BF_{29}(x)$ ,  $BF_{31}(x)$  dan  $BF_{32}(x)$ .

**Ketepatan Klasifikasi pada Model Bagging MARS**

Tabel 5. Ketepatan Klasifikasi bagging MARS

Hasil Observasi	Taksiran		Total
	Akreditasi A	Akreditasi B	
Akreditasi A	35	5	40
Akreditasi B	10	92	102
Total	46	96	142

$$APER(\%) = \frac{10 + 5}{142} \times 100\% = 10,56\%$$

Berdasarkan Tabel 5, dapat diketahui bahwa ketepatan klasifikasi dengan menggunakan 50 replikasi  $\{L^{(50)}\}$  adalah sebesar 89,44% dengan tingkat kesalahan sebesar 10,56%. Dengan kata lain, dari dugaan dengan model bagging MARS, dapat menghasilkan 89,44% sekolah/madrasah pada tingkat SD/MI di Provinsi Kalimantan Timur dengan akreditasi A ataupun B yang terklasifikasi sesuai dengan data asli.

**Kesimpulan**

Berdasarkan hasil penelitian yang diperoleh, maka penulis dapat menarik kesimpulan sebagai berikut :

1. Setelah dilakukan kombinasi antara BF, MI, dan MO sampai diperoleh nilai GCV minimum, sehingga didapatkan model terbaik MARS sebagai berikut :

$$\hat{f}(x) = -0,0185072 + 0,0312087 BF_3(x) - 0,00397983 BF_6(x) - 0,00712428 BF_7(x) + 0,0165526 BF_9(x) + 0,00451169 BF_{11}(x) - 0,00131878 BF_{16}(x) - 0,022762 BF_{19}(x) + 0,018023 BF_{21}(x) - 0,0010072 BF_{24}(x) + 0,000215826 BF_{28}(x) - 0,00317385 BF_{29}(x) - 0,0133896 BF_{31}(x)$$

dengan kombinasi BF=32, MI=3 dan MO=2.

2. Ketepatan klasifikasi akreditasi sekolah/madrasah pada tingkat SD/MI di Provinsi Kalimantan Timur menggunakan metode MARS adalah sebesar 78,87%, dimana secara keseluruhan didapatkan hasil bahwa model MARS terbaik yang telah diperoleh dapat mengklasifikasikan sebanyak sebanyak 112 sekolah di antara 142 sekolah/madrasah.

3. Tingkat akurasi klasifikasi akreditasi sekolah/madrasah pada tingkat SD/MI di Provinsi Kalimantan Timur melalui metode Bagging dari model MARS terbaik sebesar 89,44%, hal ini berarti bahwa metode bagging MARS memberikan akurasi klasifikasi yang lebih baik dibandingkan metode MARS.

**Daftar Pustaka**

Arleina, Octiva D. dan Bambang W Otok. 2014. *Bootstrap Aggregating Multivariate Adaptive Regression Spline (Bagging MARS)* untuk Mengklasifikasikan Rumah Tangga Miskin di Kabupaten Jombang. 3(2). ITS. Jurnal Sains dan Seni Pomits.

Badan Pusat Statistika. 2009. *Indikator Sosial Budaya Kota Samarinda 2009*. Samarinda: BPS dan BAPPEDA Kota Samarinda.

Breiman, L. 1996. *Bagging Predictors, Machine Learning*. 24:123-140.

Budiarti, Frediana. 2014. Penerapan Metode *Multivariate Adaptive Regression Spline (MARS)* untuk Memodelkan Komponen yang Berpengaruh Terhadap Peringkat Akreditasi Sekolah. UNMUL.

Efron, Bradley. 1979. *Bootstrap Methods: Another Look at The Jackknife. The Annals of Statistics* 7(1):1-26.

Efron, Bradley dan Robert J, Tibshirani. 1993. *An Intoduction to The Bootstrap*. New York: Chapman & Hall.

Eubank, Randall. 1999. *Spline Smoothing and Nonparametrics Regression*. New York: a Series of Textbooks and Monographs.

Friedman, J.H. 1991. *Multivariate Adaptive Regression Spline. The Annals of Statistics*. 19(1):1-65.

Gujarati, D. 1999. *Ekonometrika Dasar*. Jakarta: Erlangga.

Hair, J.F., W.C Black., L.T Ronald., dan R.E Anderson. 2006. *Multivariate Data Analysisis*. Sixth Edition. Pearson Education Prentice Hall.Inc.

Hardle, W. 1990. *Applied Nonparametric Regression*. New York: Cambridge University Press.

Mandaku, Fentje. dan Hanok Mandaku. 2010. *Pemodelan Terhadap Kelulusan Siswa Masuk Kelas Akselerasi Menggunakan Analisis Regresi Logistik dan Multivariate Adaptive Regression Spline (MARS)*. 4(2):159-166. Universitas Patimura Ambon. ARIKA.

Nawari. 2010. *Analisis Regresi dengan MS Excel 2007 dan SPSS 17*. Jakarta: Elex Media Kompotindo.

Sprent, P. 1991. *Metode Statistika Nonparametrik Terapan*. Jakarta: UI- Press.