

ANALISIS *MULTIVARIATE ADAPTIVE REGRESSION SPLINES* (MARS) PADA PREDIKSI KETERTINGGALAN KABUPATEN TAHUN 2014

Siskarossa Ika Oktora, SST, M.Stat
Sekolah Tinggi Ilmu Statistik

Abstrak

Penelitian ini bertujuan untuk membentuk model kabupaten tertinggal dan melakukan prediksi ketertinggalan kabupaten pada tahun 2014 berdasarkan kriteria perekonomian masyarakat, SDM, infrastruktur, kemampuan keuangan daerah, aksesibilitas, dan karakteristik daerah dengan metode MARS. MARS adalah salah satu metode pengklasifikasian yang mampu menangani data berdimensi tinggi dengan pola data yang tidak diketahui sebelumnya, serta dapat diterapkan untuk melihat interaksi diantara variabel yang digunakan. MARS digunakan untuk mengatasi beberapa kelemahan dari metode yang selama ini digunakan serta sebagai metode alternatif ketika data yang digunakan tidak memenuhi asumsi yang dibutuhkan pada statistika parametrik. Dari model MARS yang dibangun, terdapat tiga variabel utama yang berpengaruh terhadap ketertinggalan kabupaten diantaranya adalah pengeluaran konsumsi per kapita, angka harapan hidup, dan persentase rumah tangga pengguna listrik. Akurasi dari model MARS yang terbentuk sangat tinggi, yakni mencapai 97,83 persen dan dapat dipergunakan untuk melakukan prediksi ketertinggalan kabupaten. Berdasarkan model MARS, maka di akhir periode RPJM Nasional 2010-2014 diprediksikan terjadi transisi yang signifikan dari kabupaten dengan kondisi tertinggal menjadi tidak tertinggal serta terdapat beberapa kabupaten yang diindikasikan salah klasifikasi (yang sebelumnya dinyatakan tidak tertinggal namun seharusnya terkategorikan sebagai kabupaten tertinggal). Model ini juga dapat digunakan untuk memprediksi kondisi ketertinggalan daerah otonom baru berdasarkan data empiris yang ada, karena sebelumnya pengklasifikasian DOB hanya mengikuti status ketertinggalan daerah induknya saja.

Kata kunci : Multivariate Adaptive Regression Splines (MARS), Kabupaten tertinggal, Klasifikasi

ABSTRACT

The purposes of this research are to build underdeveloped regency model and make a prediction in 2014 based on economic categories, Human Resources (HR), infrastructures, fiscal capacity, accessibility, and regional characteristics with MARS method. MARS is a classification method which can handle high-dimensional data with unknown pattern in advance, and can be applied a model with interaction between variables. MARS is an alternative method when the data doesn't fulfil the parametric statistics assumptions. From MARS model, there are three variables that affect underdeveloped regency, they are consumption expenditure per capita, life expectancy, and percentage of household electricity users. The accuracy of MARS model is very high, 97.83 percent and can be used to make a prediction. Based on prediction of MARS model, at the end of the National Development Plan 2010-2014, there is a significant transition in regency's status. This model can also be used to predict the condition of new regency based on empirical data, because in the earlier classification, the status of regency just follows the status of parent region.

Keywords : Multivariate Adaptive Regression Splines (MARS), Underdeveloped regency, Classification

PENDAHULUAN

Di era reformasi dan otonomi daerah saat ini, ketimpangan antar wilayah di berbagai daerah di Indonesia masih sangat tinggi. Hal tersebut tercermin dari masih tingginya disparitas antar wilayah dari

segi pendidikan, perekonomian, infrastruktur, dan kualitas sumber daya manusia. Ketimpangan tersebut mengakibatkan beberapa daerah masuk ke dalam kategori kabupaten tertinggal. Kabupaten tertinggal merupakan kabupaten yang masyarakat serta wilayahnya relatif kurang berkembang dibandingkan daerah lain dalam skala nasional berdasarkan kategori perekonomian masyarakat, Sumber Daya Manusia (SDM), infrastruktur, kemampuan keuangan daerah, aksesibilitas, dan karakteristik daerah (berdasarkan RPJM 2010-2014 yang ditetapkan dengan Perpres No. 5 Tahun 2010).

Ketertinggalan suatu wilayah dapat terjadi akibat kondisi geografis yang menyebabkan daerah tersebut terisolir dan terpencil seperti daerah perbatasan negara, daerah pulau-pulau kecil, daerah pedalaman, serta daerah rawan bencana. Konflik sosial dan politik pun tidak luput menjadi salah satu penyebab ketertinggalan suatu wilayah. Untuk mengurangi tingkat kesenjangan tersebut, setiap tahunnya negara mengalokasikan Dana Alokasi Khusus untuk membantu kabupaten tertinggal dengan harapan agar pemerintah pusat dapat mengarahkan belanja daerah untuk percepatan pembangunan kabupaten tertinggal, dan sebagai implikasinya dapat meningkatkan kesejahteraan rakyat.

Kementerian Pembangunan Daerah Tertinggal (KPDT) bersama dengan Bappenas dan Kementerian Dalam Negeri melakukan evaluasi bersama pada tahun 2004 mengenai kabupaten tertinggal. Dari evaluasi tersebut ditetapkan 199 kabupaten yang tergolong kabupaten tertinggal, dimana 62 persen diantaranya berada di wilayah Indonesia Timur. Kabupaten-kabupaten tersebut menjadi target berbagai program percepatan pembangunan kabupaten tertinggal selama Rencana Pemerintah Jangka Menengah (RPJM) 2004-2009. Pada akhir periode tersebut, 50 kabupaten berhasil keluar dari daftar kabupaten tertinggal berdasarkan ukuran ketertinggalan.

Perubahan sistem pemerintahan Indonesia dari sentralisasi menjadi desentralisasi melahirkan cukup banyak provinsi dan kabupaten baru, atau yang biasa disebut Daerah Otonom Baru (DOB). Untuk pengkategorian ketertinggalan bagi DOB didasarkan pada kabupaten induknya, jika kabupaten induknya bukan merupakan kabupaten tertinggal, maka DOB tersebut otomatis tidak masuk dalam kategori kabupaten tertinggal, dan berlaku kondisi sebaliknya. Dengan semakin banyaknya daerah otonom baru, maka bertambah pula daerah-daerah yang masuk ke dalam kategori kabupaten tertinggal.

Pada RPJM Nasional 2010-2014 terdapat 183 kabupaten yang masuk dalam kategori kabupaten tertinggal dan menjadi fokus kinerja pemerintah dalam penanganan kabupaten tertinggal, yang terdiri dari 149 kabupaten lama dan 34 kabupaten baru hasil pemekaran. Berdasarkan Tabel 1 berikut dapat diketahui bahwa jika dilihat per provinsi, maka dari 33 provinsi yang ada di Indonesia, 26 provinsi diantaranya memiliki kabupaten tertinggal. Provinsi yang memiliki jumlah kabupaten tertinggal terbanyak adalah Provinsi Papua yaitu 27 kabupaten, diikuti Provinsi Nusa Tenggara Timur sebanyak

20 kabupaten, dan Provinsi Aceh sebanyak 12 kabupaten. Jika dibandingkan dengan jumlah kabupaten/kota yang ada pada provinsi bersangkutan, Provinsi Sulawesi Barat merupakan provinsi yang paling tertinggal karena seluruh kabupatennya (100%) termasuk dalam kategori kabupaten tertinggal.

Tabel 1. Daerah Tertinggal per Provinsi di Indonesia Periode 2010-2014

No	Provinsi	Jumlah Kabupaten/Kota	Jumlah Daerah Tertinggal	% Daerah Tertinggal
(1)	(2)	(3)	(4)	(5)
1	Aceh	23	12	52,17
2	Sumatera Utara	33	6	18,18
3	Sumatera Barat	19	8	42,11
4	Kepulauan Riau	7	2	28,57
5	Sumatera Selatan	15	7	46,67
6	Bangka Belitung	7	1	14,29
7	Bengkulu	10	6	60,00
8	Lampung	14	4	28,57
9	Jawa Barat	26	2	7,69
10	Banten	8	2	25,00
11	Jawa Timur	38	5	13,16
12	Kalimantan Barat	14	10	71,43
13	Kalimantan Tengah	14	3	21,43
14	Kalimantan Timur	14	3	21,43
15	Sulawesi Utara	15	3	20,00
16	Gorontalo	6	3	50,00
17	Sulawesi Tengah	11	10	90,91
18	Sulawesi Selatan	24	4	16,67
19	Sulawesi Barat	5	5	100,00
20	Sulawesi Tenggara	12	9	75,00
21	NTB	10	8	80,00
22	NTT	21	20	95,24
23	Maluku	11	8	72,73
24	Maluku Utara	9	7	77,78
25	Papua	29	27	93,10
26	Papua Barat	11	8	72,73
Total		406	183	45,07

Sumber : Kementerian Keuangan

Untuk melakukan evaluasi kabupaten tertinggal periode 2010-2014 Badan Pusat Statistik dilibatkan untuk melakukan penghitungan dengan menggunakan metode yang sudah dikembangkan. Jika sebelumnya hanya dilakukan perbandingan dengan rata-rata hitung pada masing-masing variabel, maka untuk evaluasi saat ini, ke-27 variabel tersebut diberikan bobot berdasarkan hasil analisis faktor dan dilakukan penghitungan *Z-score* untuk masing-masing variabel. Nilai tersebut kemudian

dikelompokkan ke dalam lima kelas interval dengan kategori Potensi Maju, Agak Tertinggal, Tertinggal, Sangat Tertinggal, dan Sangat Parah. Dari kelima kategori tersebut, hanya kabupaten yang berada pada status Potensi Maju saja yang dikeluarkan dari kategori ketertinggalan.

Berdasarkan kondisi tersebut di atas, penulis tertarik untuk melakukan kajian yang lebih mendalam tentang fenomena kabupaten tertinggal di Indonesia melalui variabel-variabel yang telah digunakan sebelumnya dan melihat ketepatan klasifikasi dengan sebuah alat statistik sehingga menghasilkan suatu keterbandingan dengan metode yang selama ini sudah digunakan. Penulis juga melihat beberapa kelemahan dari metode yang selama ini digunakan, diantaranya kementerian terkait belum pernah melakukan pemodelan dan melihat sejauh mana variabel-variabel yang digunakan tersebut memberikan kontribusi terhadap ketertinggalan suatu wilayah. Selain itu dengan mengklasifikasikan DOB berdasarkan kabupaten induknya tanpa dilakukan evaluasi ulang terhadap kabupaten tersebut akan menimbulkan misklasifikasi, dimana kondisi DOB bisa saja berbeda dengan daerah induknya seiring dengan berjalannya pembangunan di daerah tersebut. Dengan demikian diperlukan adanya penilaian ketepatan klasifikasi dari masing-masing wilayah. Selain itu, RPJM 2010-2014 sudah menginjak periode akhir, sehingga diperlukan suatu kajian analisis untuk mengevaluasi sejauh mana pencapaian yang terjadi dalam hal penanganan kabupaten tertinggal.

Penentuan kabupaten tertinggal dan tidak tertinggal pada dasarnya adalah bagaimana cara mengelompokkan atau mengklasifikasikan sejumlah observasi ke dalam kelompok tersebut dengan memperhatikan indikator yang ada. Friedman (1991) memperkenalkan metode klasifikasi yang relatif fleksibel untuk menyelidiki pola hubungan antara variabel respon dan variabel prediktor tanpa asumsi awal terhadap bentuk hubungan fungsionalnya yang dikenal dengan *Multivariate Adaptive Regression Splines* (MARS). Metode ini merupakan kombinasi yang kompleks dari spline dan *recursive partitioning* serta melibatkan dimensi data yang besar yakni dengan jumlah observasi dan jumlah variabel yang cukup banyak. Selain itu MARS dapat secara efektif mengeksplorasi hubungan non linier yang tersembunyi diantara variabel respon dan variabel prediktor serta efek interaksi pada struktur data yang kompleks (Li-Yen Chang, 2014).

Tujuan yang ingin dicapai dari penelitian ini diantaranya membentuk model kabupaten tertinggal berdasarkan kriteria perekonomian masyarakat, SDM, infrastruktur, kemampuan keuangan daerah, aksesibilitas, dan karakteristik daerah dengan metode MARS dimana pemodelan dilakukan secara simultan (*multivariate*) dengan kondisi data yang tidak diketahui polanya; menentukan ketepatan klasifikasi kabupaten tertinggal dengan menggunakan metode MARS, membuat prediksi

ketertinggalan kabupaten pada akhir periode RPJM Nasional 2010-2014 berdasarkan model MARS yang terbentuk.

METODOLOGI

Tinjauan referensi

Proses penghitungan dan penentuan daerah tertinggal mengalami perubahan dari waktu ke waktu guna penyempurnaan. Penentuan kabupaten tertinggal pada tahun 2004 menggunakan metode rata-rata hitung, dimana dari seluruh kabupaten yang ada pada saat itu diperoleh rata-rata hitung untuk 27 variabel dari 6 kriteria yang digunakan, yaitu perekonomian masyarakat, Sumber Daya Manusia (SDM), infrastruktur, kemampuan keuangan daerah, aksesibilitas, dan karakteristik daerah. Kabupaten-kabupaten yang memiliki nilai variabel di bawah rata-rata hitung akan dikategorikan sebagai daerah tertinggal. Berdasarkan kajian statistik, metode ini memiliki kelemahan, yaitu jika hanya digunakan rata-rata hitung, maka penentuan kabupaten tertinggal akan menjadi bias akibat adanya outlier. Selain itu, metode ini hanya bersifat multiindikator, yakni melibatkan banyak sekali indikator namun tidak melakukan penghitungan secara simultan (*multivariate*), dan tidak melihat efek interaksi diantara variabel-variabel yang digunakan.

Selanjutnya berdasarkan panduan Penjelasan Penetapan Daerah Tertinggal dijelaskan bahwa teknis penghitungan daerah tertinggal menggunakan data hasil standarisasi karena masing-masing data memiliki variasi satuan. Dari hasil standarisasi tersebut selanjutnya dikalikan dengan bobot untuk masing-masing variabel dan dilakukan penjumlahan. Namun, sebelum dilakukan penjumlahan, hasil perkalian tersebut harus dengan arah yang sama. Indikator yang bersifat mengukur tingkat keburukan seperti jumlah penduduk miskin maka arahnya positif, dan sebaliknya. Hasil total indeks inilah yang dijadikan patokan penetapan kabupaten tertinggal, dimana kabupaten-kabupaten yang memiliki total indeks di atas 0 merupakan kabupaten tertinggal. Secara statistika kondisi penghitungan tersebut baik, karena melakukan proses standarisasi yang disebabkan satuan yang berbeda dari masing-masing variabel. Namun akibat proses standarisasi ini, nilai yang semula positif dan bisa dikalikan ± 1 guna membedakan mana variabel yang mengukur tingkat keburukan dan yang tidak, justru akan diperoleh hasil yang kurang representatif, karena standarisasi akan menghasilkan nilai yang tidak hanya positif, melainkan juga negatif. Sehingga jika kemudian dikalikan dengan ± 1 akan memberikan peluang untuk menghasilkan kesimpulan yang salah.

Selain itu asumsi bagi Daerah Otonom Baru (DOB) yang dimekarkan dari daerah induk dengan status daerah tertinggal dan kemudian langsung ditetapkan sebagai daerah tertinggal membutuhkan

kajian yang lebih mendalam, karena bisa saja terjadi bahwa DOB tersebut justru merupakan daerah tidak tertinggal. Atau sebaliknya, DOB yang dimekarkan dari non daerah tertinggal justru kondisinya lebih tertinggal yang tentunya menjadi lebih berhak untuk mendapatkan perhatian.

Untuk melakukan pemodelan dengan respon biner (dalam hal ini kabupaten tertinggal dan kabupaten tidak tertinggal, biasanya dilakukan dengan analisis regresi. Analisis regresi digunakan untuk memperlihatkan hubungan dan pengaruh variabel prediktor terhadap variabel respon dengan terlebih dahulu melihat pola hubungan dari variabel tersebut. Hal ini dapat dilakukan dengan dua pendekatan. Pendekatan yang paling umum dan seringkali digunakan adalah pendekatan parametrik, yang mengasumsikan bentuk model sudah ditentukan sebelumnya. Namun apabila tidak ada informasi apapun tentang bentuk dari fungsi regresi, maka pendekatan yang digunakan adalah pendekatan nonparametrik. Karena pendekatan ini tidak tergantung pada asumsi bentuk kurva tertentu, maka akan memberikan fleksibilitas yang lebih besar (Budiantara, dkk., 2006).

Metode klasifikasi merupakan bagian dari analisis statistika. Metode yang paling sering digunakan untuk masalah klasifikasi adalah analisis diskriminan. Penggunaan analisis ini membutuhkan sejumlah asumsi diantaranya populasi berdistribusi normal dengan varians-kovarians sama. Tetapi pada penerapannya analisis diskriminan sering melibatkan variabel-variabel kategorik yang tidak mengikuti pola distribusi normal sehingga akibatnya diperoleh hasil yang tidak optimal. Metode lain yang juga sering digunakan untuk masalah klasifikasi adalah regresi logistik. Analisis regresi logistik digunakan untuk analisis data respon kategorik dengan variabel-variabel bebas dapat berupa kategorik maupun kontinu (Otok, 2003). Namun analisis ini mensyaratkan adanya asumsi tidak terjadinya multikolinieritas pada variabel prediktornya (Nash dan Bradford, 2001). Padahal dalam penelitian di bidang sosial, masalah multikolinieritas seringkali tidak bisa dihindari. Selain itu, jika asumsi independensi tidak terpenuhi maka akan memberikan pendugaan yang tidak tepat.

Metode Analisis

Multivariate Adaptive Regression Splines (MARS)

MARS adalah salah satu pendekatan regresi non parametrik serta merupakan metode yang relatif baru dan dikembangkan oleh Jerome H. Friedman pada tahun 1991 untuk mengatasi kelemahan *recursive partitioning*. MARS difokuskan untuk mengatasi permasalahan data berdimensi tinggi dengan jumlah variabel dan observasi yang cukup banyak dan menghasilkan model yang kontinu pada knots. Prinsip dasar MARS adalah memberikan fleksibilitas tinggi untuk mengeksplorasi hubungan non linier yang terjadi diantara variabel respon dan variabel prediktor melalui fungsi yang berbeda

untuk setiap interval yang berbeda. Selain itu, melalui metode ini juga dapat diketahui interaksi yang terjadi diantara variabel prediktor.

Pemilihan model pada MARS dilakukan dengan menggunakan metode *stepwise* yang terdiri dari *forward* dan *backward*. *Forward stepwise* dilakukan untuk mendapatkan jumlah fungsi basis maksimum dengan kriteria pemilihan fungsi basis adalah dengan meminimumkan *Average Square Residual* (ASR). Sedangkan untuk memenuhi konsep parsemoni (model sederhana) dilakukan *backward stepwise* yaitu memilih fungsi basis yang dihasilkan dari *forward stepwise* dengan meminimumkan *Generalized Cross Validation* (GCV) (Friedman, 1991, Budiantara, dkk, 2006).

Secara umum model MARS dapat ditulis sebagai berikut :

$$\hat{f}(x) = \alpha_0 + \sum_{m=1}^M \alpha_m \prod_{k=1}^{K_m} [s_{km} \cdot (x_{v(k,m)} - t_{km})] \quad (1)$$

dimana :

α_0 = koefisien konstan fungsi basis

α_m = koefisien dari fungsi basis ke-m

M = jumlah maksimum fungsi basis (*non constant* fungsi basis)

K_m = derajat interaksi

s_{km} = nilainya ± 1

$x_{v(k,m)}$ = variabel prediktor

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

Menurut Friedman (1991), klasifikasi pada model MARS dapat didasarkan pada pendekatan analisis regresi. Regresi logistik linier sering digunakan ketika variabel respon diasumsikan memiliki dua nilai atau yang biasa disebut *binary response*.

Berdasarkan B. W. Otok (2009) jika $f(x) = y$ dan $y \sim Ber(1, \pi(x))$ dengan $y \in (0,1)$ dan $x \in \mathfrak{R}^p$ maka :

$$P(Y_i = 1) = \pi(x) \text{ dan } P(Y_i = 0) = 1 - \pi(x)$$

$x \in \mathfrak{R}^p$ adalah vektor dari p variabel prediktor dan $P(Y = 1|x) = \pi(x)$

Lemma 1 : Jika hubungan dengan model logistik, $\sigma_L : R \rightarrow (0,1)$, dimana

$$\sigma_L = \pi(x) = \left[\frac{e^z}{1 + e^z} \right]$$

maka invers dari σ_L dapat dikatakan sebagai transformasi logit, yakni :

$$\text{logit } \pi(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = z$$

$$z = \hat{f}(x) = \alpha_0 + \sum_{m=1}^M \alpha_m \prod_{k=1}^{K_m} [s_{km} \cdot (x_{v(k,m)} - t_{km})]$$

maka dapat ditulis ke dalam model :

$$\text{logit } \pi(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \alpha_0 + \sum_{m=1}^M \alpha_m \prod_{k=1}^{K_m} [s_{km} \cdot (x_{v(k,m)} - t_{km})] \quad (2)$$

Dan dalam bentuk matriks dapat ditulis sebagai berikut :

$$\text{logit } \pi(x) = \mathbf{B}\alpha \quad (3)$$

Karena variabel respon memiliki 2 kategori (biner), maka digunakan titik potong (*cut off*) sebesar 0,5 dengan ketentuan apabila $\pi(x) \geq 0,5$ maka hasil prediksi adalah 1. Dan jika $\pi(x) < 0,5$ maka hasil prediksi adalah 0.

Data dan Variabel yang Digunakan

Data yang digunakan dalam penelitian ini diperoleh dari Kementerian Pembangunan Daerah Tertinggal yang bersumber dari Pendataan PODES 2008, PODES 2011, SUSENAS 2009, SUSENAS 2012-2013 (yang bersumber dari Badan Pusat Statistik), serta realisasi Kemampuan Keuangan Daerah (KKD) 2009 dan realisasi KKD 2012 (yang bersumber dari Kementerian Keuangan). Sementara variabel yang digunakan pada penelitian ini dapat dilihat pada Tabel 2 berikut :

Tabel 2. Variabel yang Digunakan dalam Pemodelan Kabupaten Tertinggal

Variabel	Keterangan
Y	Kabupaten tertinggal (1) Kabupaten tidak tertinggal (0)
X1	Persentase penduduk miskin
X2	Pengeluaran konsumsi per kapita
X3	Angka harapan hidup
X4	Rata-rata lama sekolah
X5	Angka melek huruf
X6	Jumlah desa dengan jenis permukaan jalan terluas aspal/beton
X7	Jumlah desa dengan jenis permukaan jalan terluas diperkeras
X8	Jumlah desa dengan jenis permukaan jalan terluas tanah
X9	Jumlah desa dengan jenis permukaan jalan terluas lainnya
X10	Persentase rumah tangga pengguna listrik
X11	Persentase rumah tangga pengguna telepon
X12	Persentase rumah tangga pengguna air bersih
X13	Jumlah desa yang memiliki pasar tanpa bangunan permanen

X14	Jumlah prasarana kesehatan per 1000 penduduk
X15	Jumlah dokter per 1000 penduduk
X16	Jumlah SD dan SMP per 1000 penduduk
X17	Kemampuan keuangan daerah
X18	Rata-rata jarak dari kantor desa/kelurahan ke kantor kabupaten yang membawahi
X19	Jumlah desa dengan akses ke pelayanan kesehatan >5 km
X20	Jarak desa ke pelayanan pendidikan dasar
X21	Persentase desa gempa bumi
X22	Persentase desa tanah longsor
X23	Persentase desa banjir
X24	Persentase desa bencana lainnya
X25	Persentase desa di kawasan hutan lindung
X26	Persentase desa berlahan kritis
X27	Persentase desa konflik satu tahun terakhir

HASIL DAN PEMBAHASAN

Pembentukan Model Kabupaten Tertinggal

Pembentukan model kabupaten tertinggal dengan metode MARS dilakukan dengan cara *trial and error* terhadap kombinasi antara jumlah maksimum *basis function* (BF) yang nilainya adalah 2 sampai dengan 4 kali jumlah variabel prediktor, dalam kasus ini BF yang digunakan adalah 54, 81, dan 108; jumlah maksimum interaksi (MI) yaitu 1, 2, dan 3, dengan asumsi model yang melibatkan lebih dari 3 interaksi akan menghasilkan model yang terlalu kompleks; serta nilai minimum observasi (MO) yaitu sebesar 12, yang didasarkan atas teori dari Friedman (1991) bahwa jumlah minimum observasi terbaik dihasilkan dari rumus berikut ini:

$$\begin{aligned}
 L(\alpha) &= 3 - \log_2(\alpha/n) \\
 &= 3 - \log_2\left(0,05/27\right) \\
 &= 12,077 \approx 12
 \end{aligned}$$

dimana α = tingkat kesalahan, dan n = jumlah variabel prediktor. Penentuan model yang terbaik didasarkan pada nilai GCV minimum.

Setelah dilakukan proses pengolahan dengan kombinasi BF, MI dan MO, maka diperoleh hasil kombinasi seperti terlihat pada Tabel 3.

Tabel 3. Hasil Kombinasi dari BF, MI, dan MO untuk Model Kabupaten Tertinggal

Model MARS	BF	MI	MO	GCV	R²	Jumlah Variabel Sig.	Jumlah BF Sig.	Var. X yang Berkontribusi
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	54	1	12	0.0934	0.686	13	17	X2
2	54	2	12	0.0778	0.784	17	26	X2
3	54	3	12	0.0727	0.808	14	29	X2
4	81	1	12	0.0949	0.703	14	23	X2
5	81	2	12	0.0746	0.809	19	31	X2
6	81	3	12	0.0669	0.846	16	37	X2
7	108	1	12	0.0940	0.702	13	22	X2
8	108	2	12	0.0777	0.855	21	48	X2
9	108	3	12	0.0662	0.850	20	38	X2

Keterangan : BF = *Basis Function* MI = Maksimum Interaksi MO = Minimum Observasi

Berdasarkan kombinasi tersebut dan kriteria dalam pemilihan model, maka model terbaik yang dihasilkan adalah model 9 (sembilan) yang merupakan kombinasi antara BF = 108, MI = 3, dan MO = 12. Model tersebut merupakan model dengan nilai GCV minimum yaitu 0,0662. Dari kombinasi tersebut dihasilkan 20 variabel yang berpengaruh terhadap model dengan jumlah fungsi basis sebanyak 38.

Dari tabel tersebut juga dapat disimpulkan bahwa dari sisi jumlah interaksi antar variabel, model-model dengan maksimum interaksi sebanyak 3 memiliki nilai GCV yang paling minimum dibandingkan dengan model untuk 2 interaksi maupun tanpa interaksi. Kondisi tersebut mengindikasikan bahwa diantara variabel-variabel yang mempengaruhi ketertinggalan suatu wilayah saling berinteraksi satu dengan lainnya. Selain itu juga terlihat bahwa untuk semua kombinasi, variabel X2 (pengeluaran konsumsi per kapita) adalah variabel yang paling berpengaruh terhadap kondisi ketertinggalan wilayah dibandingkan dengan variabel lainnya di dalam model.

3.2 Model Kabupaten Tertinggal

Berdasarkan hasil kombinasi BF, MI, dan MO, maka model yang dihasilkan adalah sebagai berikut :

$$\hat{Y} = 1,168 - 0,0278 * BF1 - 0,2188 * BF2 + 0,1162 * BF3 - 0,01775 * BF4 + 0,004347 * BF5 + 4,49x10^{-5} * BF6 + 0,000232 * BF7 - 0,00677 * BF8 + 0,01119 * BF9 - 0,00507 * BF10 - 0,02704 * BF11 + 0,000929 * BF12 - 9,1x10^{-5} * BF13 + 0,004137 * BF14 + 0,007318 * BF15 - 0,00272 * BF16 + 0,000212 * BF17 + 1,68x10^{-5} * BF18 + 4,39x10^{-5} * BF19 + 0,004126 * BF20 - 0,00068 * BF21 + 0,000224 * BF22 + 0,006894 * BF23 + 0,197 * BF24 - 2,4x10^{-5} * BF25 - 0,00506 * BF26 - 0,00248 * BF27 - 0,01275 * BF28 - 0,01113 * BF29 + 0,000859 * BF30 + 0,001872 * BF31 + 0,000261 * BF32 - 0,1033 * BF33 + 0,000207 * BF34 - 0,00614 * BF35 + 0,007584 * BF36 + 0,000239 * BF37$$

dimana :

BF1 = h(X2-600.79)	BF20 = h(X1-24.96)*h(X10-73.58)
BF2 = h(X3-69.62)	BF21 = h(24.96-X1)*h(X10-73.58)
BF3 = h(69.62-X3)	BF22 = h(24.96-X1)*h(X3-67.82)*h(X10-73.58)
BF4 = h(X10-73.58)	BF23 = h(X2-600.79)*h(X12-53.03)*h(X16-1.74)
BF5 = h(X2-600.79)*h(X3-69.38)	BF24 = h(69.62-X3)*h(X5-98.88)*h(X26-38.87)
BF6 = h(X2-600.79)*h(69.38-X3)*h(X18-25.57)	BF25 = h(X2-600.79)*h(69.38-X3)*h(X6-95)
BF7 = h(X2-600.79)*h(69.38-X3)*h(25.57-X18)	BF26 = h(610.39-X2)*h(69.62-X3)
BF8 = h(600.79-X2)*h(X11-1.9)	BF27 = h(X2-610.39)*h(69.62-X3)*h(X5-96.31)
BF9 = h(600.79-X2)*h(1.9-X11)	BF28 = h(610.39-X2)*h(69.62-X3)*h(X26-45.7)
BF10 = h(40.68-X21)	BF29 = h(69.62-X3)*h(4.69-X11)
BF11 = h(600.79-X2)*h(X10-90.98)*h(X11-1.9)	BF30 = h(611.66-X2)*h(X10-73.58)*h(X20-1.76)
BF12 = h(X2-611.66)*h(X10-73.58)	BF31 = h(611.66-X2)*h(X10-73.58)*h(1.76-X20)
BF13 = h(X2-600.79)*h(69.38-X3)*h(30.22-X22)	BF32 = h(X1-14.73)*h(X2-600.79)*h(X24-48.61)
BF14 = h(611.66-X2)*h(X10-73.58)*h(X12-85.76)	BF33 = h(1-X9)
BF15 = h(X2-600.79)*h(3.36-X11)	BF34 = h(69.62-X3)*h(X11-4.69)*h(147-X13)
BF16 = h(X2-600.79)*h(3.36-X11)*h(3-X20)	BF35 = h(X1-14.73)*h(X2-600.79)*h(X4-8.38)
BF17 = h(X2-600.79)*h(X3-69.38)*h(X8-58)	BF36 = h(X6-111)*h(X8-106)*h(1-X9)
BF18 = h(X2-600.79)*h(69.38-X3)*h(X13-100)	BF37 = h(X6-111)*h(1-X9)*h(X25-38.13)
BF19 = h(X2-600.79)*h(69.38-X3)*h(100-X13)	

Interpretasi pada model MARS tidak hanya melibatkan 1 variabel saja, melainkan juga terdapat interaksi antar variabel. Karena model yang digunakan adalah MARS untuk respon biner, maka interpretasi model didekati dengan model logistik, yakni menggunakan *odds ratio*.

Variabel-variabel yang Berpengaruh Signifikan Terhadap Pembentukan Kabupaten Tertinggal

Dari pemodelan dengan metode MARS, dapat diketahui variabel-variabel mana saja yang berpengaruh secara signifikan terhadap model MARS yang dibangun. Selain itu, variabel tersebut dapat diurutkan berdasarkan tingkat kepentingan variabel tersebut di dalam model (*variable importance*). Kriteria yang digunakan untuk mengestimasi tingkat kepentingan variabel pada model MARS adalah *nsubsets*, GCV, dan RSS (*Residual Sum of Squares*). Kriteria *nsubsets* digunakan dengan cara menghitung jumlah *subsets* model yang memasukkan suatu variabel di dalamnya. Variabel yang dilibatkan dalam *subset* yang lebih banyak dianggap sebagai variabel yang lebih penting. Pada kriteria RSS, yang dilakukan adalah dengan menghitung penurunan RSS untuk setiap *subsets*. Setiap penambahan variabel akan dihitung penurunan RSS-nya. Variabel yang menyebabkan penurunan RSS yang lebih besar dianggap sebagai variabel yang lebih penting. Hal yang serupa juga dilakukan jika menggunakan kriteria GCV. Penambahan variabel yang mengakibatkan peningkatan nilai GCV, dianggap memiliki pengaruh yang tidak baik pada model. Selanjutnya untuk kemudahan interpretasi, penurunan nilai RSS atau GCV dibuat skala sehingga penurunan terbesar memiliki skala 100.

Dari hasil pemodelan, dapat diketahui bahwa dari 27 variabel yang digunakan dalam penentuan kabupaten tertinggal, hanya 20 variabel yang signifikan dalam pembentukan model. Variabel-variabel tersebut telah diurutkan berdasarkan jumlah kemunculannya pada *subsets* dan pengaruhnya terhadap penurunan GCV dan RSS. Variabel X2 (pengeluaran konsumsi per kapita) adalah variabel yang paling berpengaruh terhadap status ketertinggalan suatu kabupaten. Hal tersebut dapat dijelaskan bahwa dengan mengeluarkan variabel X2 dari model maka akan sangat berpengaruh terhadap model secara keseluruhan karena secara otomatis akan mengakibatkan peningkatan nilai RSS dan GCV, sehingga model yang dihasilkan menjadi kurang baik. Selanjutnya untuk melihat 10 variabel utama yang berkontribusi dan seberapa besar peranannya terhadap pembentukan model, dapat dilihat pada Tabel 4 berikut ini.

Tabel 4. Sepuluh Variabel Utama yang Berpengaruh terhadap Pembentukan Kabupaten Tertinggal di Indonesia

Variabel	Jumlah Subset	GCV	RSS
(1)	(2)	(3)	(4)
X2 (Pengeluaran konsumsi per kapita)	37	100,0	100
X3 (Angka harapan hidup)	36	68,1	72,7
X10 (Persentase rumah tangga pengguna listrik)	35	54,8	61,8
X18 (Rata-rata jarak dari kantor desa/kelurahan ke kantor kabupaten yang membawahi)	32	45,6	53,3
X13 (Jumlah desa yang memiliki pasar tanpa bangunan permanen)	31	45,9	52,9
X11 (Persentase rumah tangga pengguna telepon)	30	44,0	51,0
X5 (Angka melek huruf)	30	43,5	50,9
X26 (Persentase desa berlahan kritis)	29	42,4	49,5
X1 (Persentase penduduk miskin)	27	38,7	45,7
X24 (Persentase desa bencana lainnya)	27	38,7	45,7

Evaluasi Model Kabupaten Tertinggal Tahun 2009

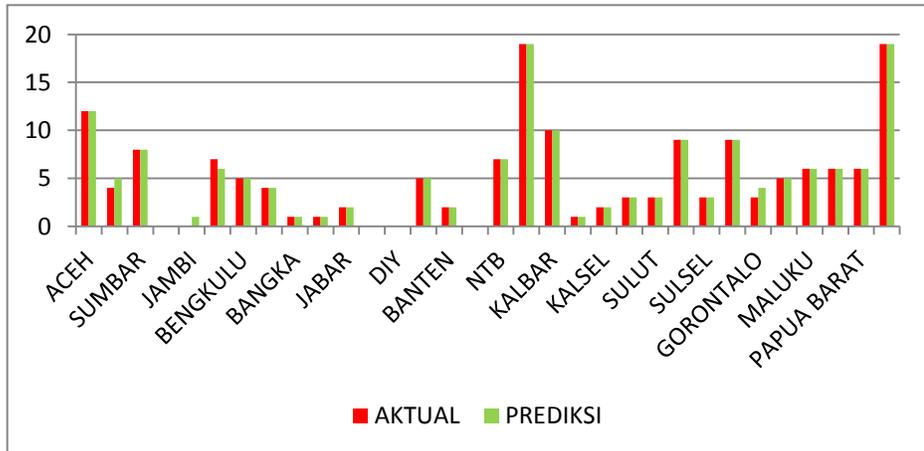
Berdasarkan hasil pemodelan kabupaten tertinggal tahun 2009, maka diperoleh tabel klasifikasi sebagai berikut :

Tabel 5. Klasifikasi Kabupaten Tertinggal di Indonesia Tahun 2009

Kelompok Aktual	Kelompok Prediksi		Total
	0	1	
(1)	(2)	(3)	(4)
0	202	5	207
1	3	159	162
Total	205	164	369

Dengan menggunakan metode MARS kombinasi BF = 108, MI = 3, dan MO = 12, maka ketepatan klasifikasi secara keseluruhan yang dihasilkan mencapai 97,83 persen. *Sensitivity* atau dengan kata lain kabupaten yang tepat diklasifikasikan sebagai kabupaten tidak tertinggal mencapai 0,976. Sementara *Specificity* atau kabupaten yang tepat diklasifikasikan sebagai kabupaten tertinggal mencapai 0,981. Nilai Press's Q yang dihasilkan untuk menilai ketepatan dalam pengelompokkan

adalah sebesar 337,69. Nilai ini lebih besar dari nilai kritis yakni sebesar 3,841. Dengan demikian dapat dikatakan bahwa klasifikasi ini konsisten secara statistik. Selain itu model juga dapat dikatakan baik karena selain memiliki GCV minimum juga memiliki tingkat akurasi yang tinggi dan tingkat kesalahan yang sangat kecil.



Gambar 1 Perbandingan Kabupaten Tertinggal Tahun 2009 antara Kondisi Aktual dan Prediksi per Provinsi di Indonesia

Jika dikaji lebih jauh, maka akan dapat dilihat perbedaan antara kondisi yang sebenarnya dengan hasil prediksi yang digambarkan oleh Gambar 1 di atas. Berdasarkan gambar tersebut dapat disimpulkan bahwa ketepatan klasifikasi pada model MARS cukup baik karena dari 33 provinsi yang ada di Indonesia, hanya 4 provinsi yang memiliki perbedaan antara kondisi aktual dengan prediksinya, yaitu Provinsi Sumatera Utara, Provinsi Jambi, Provinsi Sumatera Selatan dan Provinsi Gorontalo. Rasio misklasifikasi terhadap total kabupaten hanya sebesar 0,02. Berdasarkan indikator-indikator tersebut, maka selanjutnya model dapat digunakan untuk memprediksi ketertinggalan suatu kabupaten di akhir periode RPJM Nasional 2010-2014.

Estimasi Pengklasifikasian Kabupaten Tertinggal pada Akhir Periode RPJM Nasional 2010-2014

Dengan menggunakan pemodelan dari data untuk penentuan kabupaten tertinggal tahun 2009, selanjutnya akan dilakukan evaluasi terhadap pencapaian yang terjadi dalam hal penanganan kabupaten tertinggal di Indonesia di akhir periode RPJM Nasional 2010-2014. Tujuan yang ingin dicapai adalah memprediksi kabupaten-kabupaten mana yang masih berstatus kabupaten tertinggal dan kabupaten tidak tertinggal. Selain itu, melakukan prediksi kabupaten mana saja yang berpotensi lepas dari

ketertinggalan dan kabupaten yang mengalami misklasifikasi, yakni kabupaten yang seharusnya berstatus kabupaten tertinggal namun diklasifikasikan sebagai kabupaten tidak tertinggal.

**Tabel 6. Transisi Kabupaten Tertinggal di Indonesia
pada akhir RPJM Nasional 2010- 2014**

Kelompok Aktual	Kelompok Prediksi		Total
	0	1	
(1)	(2)	(3)	(4)
0	195	20	215
1	68	115	183
Total	263	135	398

Dari tabel tersebut dapat disimpulkan bahwa setelah kurun waktu 5 tahun proses pembangunan yang terfokus pada kabupaten tertinggal, diprediksikan bahwa di akhir periode (tahun 2014) terdapat 68 kabupaten yang berpotensi lepas dari ketertinggalan. Sehingga saat ini terdapat 263 kabupaten dengan status kabupaten tidak tertinggal. Dari 183 kabupaten tertinggal pada periode sebelumnya, 62,84 persen diantaranya tetap berada pada status yang sama. Selain itu, dari hasil pemodelan menggunakan metode MARS, diperkirakan terdapat 20 kabupaten yang salah klasifikasi (misklasifikasi), salah satu penyebabnya adalah penentuan status ketertinggalan DOB yang hanya didasarkan oleh status ketertinggalan kabupaten induknya. Dengan demikian sekitar 33,92 persen kabupaten menjadi target pembangunan kabupaten tertinggal untuk periode selanjutnya.

KESIMPULAN DAN SARAN

Kesimpulan

Dari hasil penelitian tentang kabupaten tertinggal di Indonesia dengan model MARS maka dapat disimpulkan beberapa hal sebagai berikut :

1. Metode MARS merupakan pendekatan yang sesuai untuk penentuan kabupaten tertinggal jika dilihat dari kondisi data dan keterbatasan beberapa metode statistika yang ada dan metode yang digunakan oleh KPDT selama ini. Dari hasil pemodelan diperoleh tingkat akurasi yang sangat tinggi, yakni mencapai 97,83 persen dan dapat dipergunakan untuk melakukan prediksi ketertinggalan kabupaten.

2. Dari model MARS yang dibangun dapat diketahui bahwa terdapat keterkaitan/interaksi antara variabel prediktor yang digunakan dalam penentuan kabupaten tertinggal.
3. Lima prediktor utama yang berpengaruh terhadap ketertinggalan kabupaten diantaranya adalah pengeluaran konsumsi per kapita, angka harapan hidup, persentase rumah tangga pengguna listrik, rata-rata jarak dari kantor desa/kelurahan ke kantor kabupaten yang membawahi, serta jumlah desa yang memiliki pasar tanpa bangunan permanen.
4. Berdasarkan model MARS, maka di akhir periode RPJM Nasional 2010-2014 diprediksikan terjadi transisi yang signifikan dari kabupaten dengan kondisi tertinggal menjadi tidak tertinggal serta terdapat beberapa kabupaten yang diindikasikan salah klasifikasi (yang sebelumnya dinyatakan tidak tertinggal namun seharusnya terkategori sebagai kabupaten tertinggal).

Saran

Untuk penelitian dan pengembangan lebih lanjut, maka berdasarkan hasil penelitian ini dapat disarankan beberapa hal, diantaranya :

1. Pengklasifikasian yang dilakukan pada saat ini adalah dengan respon biner, sehingga untuk penyempurnaan selanjutnya diharapkan dapat dilakukan pengklasifikasian dengan lebih dari dua kategori agar dapat dibedakan kabupaten mana saja yang berada pada kondisi maju, berkembang, tertinggal, dan sangat tertinggal.
2. Penelitian yang dilakukan saat ini adalah pemodelan kabupaten tertinggal dengan skala nasional agar hasilnya dapat diperbandingkan dengan metode sebelumnya. Namun demikian diperlukan pemodelan MARS dan prediksi ketertinggalan kabupaten untuk masing-masing wilayah Indonesia Barat, Indonesia Tengah, dan Indonesia Timur yang diperkirakan memiliki karakteristik yang berbeda.
3. Model MARS yang digunakan berbasis pada spline, untuk pengembangan metode selanjutnya dapat dilakukan modifikasi dengan fungsi basis yang setara dengan spline seperti *wavelet* untuk kemudian dapat dilihat perbedaan tingkat akurasi yang diperoleh.

DAFTAR PUSTAKA

Agresti, A. 2002. *Categorical Data Analysis*. New Jersey : John Wiley & Sons, Inc

Budiantara, I.N., Suryadi, F., Otok, B.W., Guritno, S. 2006. Pemodelan B-Spline dan MARS pada Nilai Ujian Masuk terhadap IPK Mahasiswa Jurusan Disain Komunikasi Visual UK. Petra Surabaya. *Jurnal Teknik Industri*. Vol 8 (1),1-13.

- Chang, Li-Yen. 2014. Analysis of Bilateral Air Passenger Flows: A Non-Parametric Multivariate Adaptive Regression Spline Approach. *Journal of Air Transport Management*, Vol 34, 123-130
- Direktoral Jenderal Perimbangan Keuangan Kementerian Keuangan. 2013. *Affirmative Policy Dalam Percepatan Pembangunan Daerah Untuk Peningkatan Kesejahteraan Rakyat*. Jakarta : Kementerian Keuangan.
- Fernandez, J. R. A., Nieto, P. J. G., Muniz, C. D., Anton, J. C. A. 2014. Modelling Eutrophication and Risk Prevention in a Reservoir in the Northwest of Spain by Using Multivariate Adaptive Regression Splines Analysis. *Journal of Ecological Engineering*, Vol 68, 80-89
- Friedman, J. H. 1991. *Multivariate Adaptive Regression Splines*. The Annals of Statistics, Vol. 19 (1), 1-141.
- Hair, J.F, Rolph E.A, Ronald L.T, William C.B. 2006. *Multivariate Data Analysis, Sixth Edition*. Pearson Education Prentice Hall, Inc.
- Kementerian Pembangunan Daerah Tertinggal. 2010. *Rencana Strategis Tahun 2010-2014*. Jakarta: KPDT.
- Nash, M. S. dan David F.B. 2001. *Parametric and Non Parametric Logistic Regression for Prediction of Precense/ Absence of an Amphibian*. Las Vegas, Nevada : US Environmental Protection Agency Office of Research and Development National Exposure Research Laboratory Environmental Sciences Division
- Otok, B. W. 2003. Perbandingan MARS dengan Regresi Logistik pada Respon Biner. *Prosiding Seminar Nasional Matematika dan Statistika VI*. ITS, Surabaya.
- Otok, B. W., Akbar, M. S., Guritno, S., Subanar. 2007. Pendekatan Bootstrap pada Klasifikasi Pemodelan Respon Ordinal. *Jurnal Ilmu Dasar*, Vol. 8 (1), 54-67.
- Otok, B. W. 2009. Konsistensi dan Asimtotik Normalitas Model Multivariate Adaptive Regression Splines (MARS) pada Respon Biner. *Jurnal Ilmu Dasar*, Vol. 10 (2), 133-140.
- Quiros, E., Felicimo, A. M., Cuartero, A. 2009. *Testing Multivariate Adaptive Regression Splines (MARS) as a Method of Land Cover Classification of TERRA-ASTER Satellite Images*. Sensors 2009, 9.