

Pemodelan IPM Di Kawasan Timur Indonesia Menggunakan Multivariate Adaptive Regression Spline (MARS)

(HDI modelling in eastern of Indonesia using Multivariate Adaptive Regression Spline (MARS))

Annisa Nur Insany^{1*}, Nur'eni¹, dan Mohammad Fajri¹

Program Studi Statistika, Jurusan Matematika, Universitas Tadulako Jl Soekarno-Hatta Km 9 Tondo Palu 94118

Keywords: GCV, HDI, Eastern of Indonesia, MARS

Keywords: GCV, IPM, Kawasan Timur Indonesia, MARS

* Corresponding Author :
annisanurinsany20@gmail.com

Abstract

Human Development Index (HDI) is an important issue in designing and strategizing of sustainable development. Multivariate Adaptive Regression Spline (MARS) is a regression approach that produces models with continuous character on knots. MARS models are determined based on trial and error for a combination of basis function (BF), maximum interaction (MI), and minimum observation (MO). The determination of knots is based on the minimum Generalized Cross Validation (GCV) value. The results of this study are the combination value of BF = 52, MI = 3, and MO = 2 with a minimum GCV of 0,00049. The factors that influence HDI are average school length (X2) per capita expenditure (X4), life expectancy (X3), percentage of poor woman aged 15-49 who use the birth control tool (X5).

Abstrak

Indeks Pembangunan Manusia (IPM) menjadi isu penting dalam perancangan dan strategi pembangunan berkelanjutan. Multivariate Adaptive Regression Spline (MARS) merupakan pendekatan regresi yang menghasilkan model dengan sifat kontinu pada knot. Pemodelan MARS ditentukan berdasarkan trial and error untuk kombinasi basis function (BF), maksimum interaksi (MI), dan minimum observasi (MO). Penentuan knot didasarkan pada nilai Generalized Cross Validation (GCV) minimum. Hasil dari penelitian ini, diperoleh nilai kombinasi BF = 20, MI = 3, dan MO = 2 dengan GCV minimum sebesar 0,01651. Faktor-faktor yang mempengaruhi IPM adalah rata-rata lama sekolah (X2), pengeluaran per kapita (X4), angka harapan hidup (X3), persentase perempuan berstatus miskin usia 15-49 tahun yang menggunakan alat KB (X5).

Latar Belakang

Indeks Pembangunan Manusia (IPM) adalah merupakan tolak ukur dalam menilai keberhasilan upaya membangun kualitas hidup manusia (UNDP, 1990). United Nations (UN) memperkenalkan agenda pembangunan yang disebut dengan Sustainable Development Goals (SDGs). Konsep yang digunakan pada agenda tersebut berkaitan dengan pembangunan manusia. Terdapat beberapa target SDGs yang berkenaan dengan pembangunan manusia, yaitu

tujuan ketiga (menjamin kehidupan yang sehat dan meningkatkan kesejahteraan penduduk disegala usia), tujuan keempat (menjamin kualitas pendidikan yang adil dan menyeluruh serta meningkatkan kesempatan belajar seumur hidup untuk seluruh penduduk disegala usia), dan tujuan kedelapan (meningkatkan pertumbuhan ekonomi yang komprehensif dan berkelanjutan, kesempatan kerja penuh dan produktif, serta pekerjaan yang layak untuk semua) (Nugroho, 2016).

Keadaan IPM di Indonesia mengalami kenaikan, namun kenaikan IPM tersebut tidak merata. Hal ini dibuktikan dengan nilai IPM di kawasan timur Indonesia, khususnya Papua yang tergolong pada kategori sangat rendah yaitu 59,09 (Kompas, 2018). Oleh karena ketimpangan kenaikan IPM tersebut, maka dilakukan penelitian mengenai faktor-faktor apa saja yang mempengaruhi nilai IPM di kawasan timur Indonesia.

Bahan dan Metode

Data yang digunakan dalam penelitian ini adalah data sekunder di tiap kabupaten dan kota di Indonesia Timur yang berasal dari Badan Pusat Statistik yang terangkum dalam buku Data dan Informasi Kemiskinan Kabupaten Kota Tahun 2017 dan Provinsi Dalam Angka 2018, dimana provinsi tersebut adalah Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Papua, Papua Barat, Maluku, Maluku Utara, Nusa Tenggara Barat, Provinsi Nusa Tenggara Timur, Gorontalo, Sulawesi Barat, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Tenggara, dan Sulawesi Utara. Variabel respon yang digunakan pada penelitian ini adalah indeks pembangunan manusia (Y), sedangkan variabel prediktor yang diduga mempengaruhi variabel respon adalah persentase penduduk miskin (X1), rata-rata lama sekolah (X2), angka harapan hidup (X3), pengeluaran per kapita (X4), persentase perempuan berstatus miskin usia 15-49 tahun yang menggunakan alat KB (X5).

Metode yang digunakan pada penelitian ini adalah Multivariate Adaptive Regression Spline (MARS). MARS memiliki model yang fleksibel untuk pemodelan regresi dengan data berdimensi tinggi dimana Friedman (1991) menyarankan jumlah variabel prediktor (X_p) adalah $3 \leq p \leq 20$ dan ukuran sampel $50 \leq n \leq 1000$. Menurut Friedman (1991), pada masalah yang besar (jumlah $p > 20$ dan $n > 1000$) dibutuhkan waktu pengerjaan dan perhitungan yang lebih lama serta rumit. Pada penelitian ini, variabel prediktor berjumlah 5 dan sampel yang digunakan berjumlah 232 kota dan kabupaten di kawasan timur Indonesia. Penelitian ini dilakukan dengan prosedur sebagai berikut.

- a. Melakukan analisis deskriptif setiap variabel prediktor (X) dan respon (Y).
- b. Membuat scatterplot antara variabel prediktor (Y) dengan masing-masing variabel respon (X).
- c. Melakukan standarisasi terhadap variabel yang digunakan agar memiliki skala yang sama.

- d. Mengkombinasikan nilai jumlah maksimum jumlah basis function (BF), maksimum interaksi (MI), dan minimum observasi (MO).
- e. Melakukan estimasi model.
- f. Memilih model MARS dengan nilai GCV minimum.
- g. Menguji signifikansi parameter MARS secara simultan dan parsial.
- h. Menginterpretasi model MARS yang terpilih dan faktor-faktor yang berpengaruh.

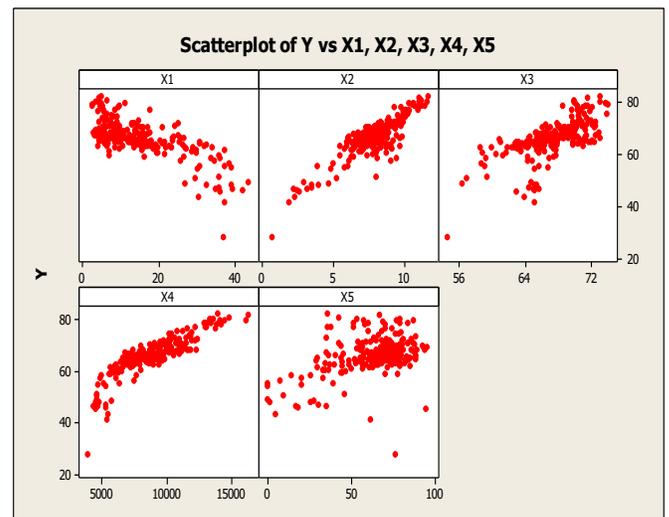
Hasil dan Pembahasan

Statistik Deskriptif

Statistik deskriptif digunakan untuk menggambarkan karakteristik data. Statistik deskriptif masing-masing variabel dapat dilihat pada Tabel 1.

Tabel 1. Statistik Deskriptif

Var	Min	Max	Mean	Stdev
Y	27,87	81,83	65,68	7,430
X ₁	2,73	43,63	15,24	9,966
X ₂	0,71	11,68	7,76	1,767
X ₃	54,60	73,97	67,40	3,502
X ₄	3972,00	16367,00	8932,00	2438,377
X ₅	0,00	95,56	63,05	19,999



Gambar 1. Scatterplot Masing-Masing Variabel Prdiktor Terhadap Respon

Scatterplot Masing-Masing Variabel Prediktor terhadap Variabel Respon

Scatterplot digunakan untuk mengidentifikasi pola data. Berdasarkan Gambar 1, dapat diketahui bahwa terdapat beberapa pola data variabel prediktor

terhadap variabel respon yang tidak mengikuti pola tertentu. Pada saat pola data yang digunakan tidak dapat diidentifikasi atau tidak mengikuti pola distribusi tertentu, maka digunakan analisis regresi nonparametrik (Eubank, 1988).

Standarisasi Data

Standarisasi data digunakan untuk menyempitkan perbedaan satuan serta penyebaran nilai yang tinggi (Santoso, 2010). Standarisasi data dilakukan menggunakan persamaan sebagai berikut:

$$z = \frac{x_i - \bar{x}}{s}$$

Kombinasi Nilai Basis Function, Minimum Interaction, dan Maximum Observation

Kombinasi nilai basis function (BF), nilai maximum interaction (MI), dan nilai minimum observation (MO) digunakan untuk memperoleh nilai GCV minimum. Friedman (1991) menyarankan nilai BF sebesar 2 sampai 4 kali jumlah variabel prediktor (X), nilai MI sebesar 1, 2, dan 3, nilai MO sebesar 0, 1, 2, dan 3. Pada penelitian ini, jumlah variabel prediktor (X) yang digunakan adalah 13, maka diperoleh nilai BF menjadi 10, 15, dan 20. Nilai GCV untuk masing-masing kombinasi nilai BF, MI, dan MO diperoleh menggunakan persamaan sebagai berikut.

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

Kombinasi nilai BF, MI, dan MO yang menghasilkan nilai GCV ditunjukkan pada Tabel 2. Kombinasi nilai yang menghasilkan GCV minimum dipilih dengan tujuan untuk digunakan dalam pengujian signifikansi model MARS. Nilai GCV yang minimum dapat menjadi model MARS yang terbaik (Friedman, 1991). Berdasarkan kombinasi tersebut, dapat diketahui bahwa model MARS yang ke 35 memiliki nilai GCV yang minimum.

Tabel 2. Nilai GCV Untuk Seluruh Model

Model	Nilai BF	Nilai MI	Nilai MO	Nilai GCV
1	10	1	0	0,02237
2	10	1	1	0,02379
3	10	1	2	0,02379
4	10	1	3	0,02355
5	10	2	0	0,01970

6	10	2	1	0,01965
7	10	2	2	0,01966
8	10	2	3	0,02041
9	10	3	0	0,01970
10	10	3	1	0,01965
11	10	3	2	0,01966
12	10	3	3	0,01966
13	15	1	0	0,02156
14	15	1	1	0,02228
15	15	1	2	0,02229
16	15	1	3	0,02277
17	15	2	0	0,01770
18	15	2	1	0,01782
19	15	2	2	0,01814
20	15	2	3	0,01925
21	15	3	0	0,01770
22	15	3	1	0,01782
23	15	3	2	0,01849
24	15	3	3	0,01892
25	20	1	0	0,02073
26	20	1	1	0,02192
27	20	1	2	0,02193
28	20	1	3	0,02198
29	20	2	0	0,01734
30	20	2	1	0,01664
31	20	2	2	0,01660
32	20	2	3	0,01731
33	20	3	0	0,01738
34	20	3	1	0,01703
35	20	3	2	0,01651
36	20	3	3	0,01738

Estimasi Model MARS Berdasarkan GCV Minimum

Model MARS yang ke 35 pada Tabel 2 merupakan model dengan nilai GCV minimum. Persamaan model tersebut adalah sebagai berikut:

$$Y = -0.654 + 0.297*BF_1 - 0.727*BF_2 - 1.268*BF_3 - 0.673*BF_4 + 0.076*BF_5 - 0.242*BF_6 - 0.075*BF_{10} + 1.494*BF_{11} - 0.026*BF_{13} - 0.035*BF_{14} + 4.623*BF_{15} - 1.288*BF_{16} - 2.176*BF_{17} - 2.360*BF_{19};$$

Pengujian Signifikansi Model MARS Secara Simultan

Pengujian ini bertujuan untuk mengetahui signifikansi koefisien regresi secara simultan yang dilakukan secara bersamaan terhadap model. Rumusan hipotesis yang digunakan yaitu :

$$H_0 : a_m = 0 \text{ (model MARS yang ke 35 signifikan);}$$

H_1 : minimal terdapat satu $a_m \neq 0$ (model MARS yang ke 35 tidak signifikan);
 dengan m adalah nilai basis function (BF) yang terdapat pada model 35.

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

Apabila $F_{hitung} > F_{0,05((5);(218))}$ atau $P\text{-value} < \alpha$ maka H_0 ditolak. Pada Tabel 3 diperoleh nilai $P\text{-Value}$ sebesar 0,00000 dan nilai hitung dari $F\text{-Statistic}$ sebesar 1367,127; sehingga diperoleh keputusan bahwa H_0 di tolak karena nilai $P\text{-value} < \alpha$ ($0,00000 < 0,05$). Oleh karena itu, maka dapat disimpulkan bahwa model MARS yang ke 35 signifikan terhadap pengujian signifikansi model secara simultan, sehingga model MARS tersebut dapat digunakan untuk menganalisis kasus indeks pembangunan manusia (IPM).

Tabel 3. Pengujian Signifikansi Model MARS Secara Simultan

$F\text{-Statistic} = 1367,127$	$S. E. \text{ of Regression} = 0,109$
$P\text{-Value} = 0,00000$	$\text{Residual Sum of Square} = 1,590$

Pengujian Signifikansi Model MARS Secara Parsial

Hipotesis yang digunakan pada pengujian ini adalah sebagai berikut:

H_0 : $a_m = 0$ (BF tidak memiliki pengaruh yang signifikan);

H_1 : $a_m \neq 0$ (BF memiliki pengaruh yang signifikan) untuk setiap m ;

dengan m adalah nilai *basis function* (BF) yang terdapat di model MARS yang ke 35. Persamaan yang digunakan adalah sebagai berikut:

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

Menurut Budiantara (2012), apabila nilai dari $|t_{hitung}| > t_{(\alpha/2);n-(p+r)-1}$ atau $P\text{-value} < \text{taraf signifikansi } (\alpha)$, maka H_0 ditolak. Berdasarkan Tabel 4, dapat diketahui bahwa seluruh $P\text{-value} < \alpha$ sehingga dapat disimpulkan bahwa seluruh *basis function* yang terdapat dalam model adalah signifikan terhadap pengujian signifikansi model secara parsial. Sehingga variabel-variabel yang terdapat dalam model MARS yang ke 35 memiliki pengaruh yang signifikan pula.

Tabel 4. Pengujian Signifikansi Model MARS Secara Parsial

Parameter	$P\text{-Value}$
Constant	0,00000
Basis Function 1	0,00000
Basis Function 2	0,00000
Basis Function 3	0,00003
Basis Function 4	0,00003
Basis Function 5	0,00000
Basis Function 6	0,00000
Basis Function 10	0,00000
Basis Function 11	0,00000
Basis Function 13	0,00001
Basis Function 14	0,00038
Basis Function 15	0,00000
Basis Function 16	0,00000
Basis Function 17	0,00000
Basis Function 19	0,00067

Interpretasi Model MARS yang Terbaik

Berdasarkan hasil pengujian signifikansi model MARS secara simultan dan parsial, maka diperoleh kesimpulan bahwa model MARS yang ke 35 merupakan model MARS yang terbaik.

- $BF_1 = \max(0, ZX_4 + 1.01461)$;
- $BF_2 = \max(0, -1.01461 - ZX_4)$;
- $BF_3 = \max(0, ZX_2 + 2.70323)$;
- $BF_4 = \max(0, -2.70323 - ZX_2)$;
- $BF_5 = \max(0, ZX_3 + 1.83791) * BF_3$;
- $BF_6 = \max(0, -1.83791 - ZX_3) * BF_3$;
- $BF_8 = \max(0, -1.00565 - ZX_2)$;
- $BF_{10} = \max(0, 0.0688965 - ZX_4) * BF_3$;
- $BF_{11} = \max(0, ZX_2 + 2.40898)$;
- $BF_{13} = \max(0, ZX_5 + 0.447908) * BF_{11}$;
- $BF_{14} = \max(0, -0.447908 - ZX_5) * BF_{11}$;
- $BF_{15} = \max(0, ZX_4 + 1.29636) * BF_8$;
- $BF_{16} = \max(0, -1.29636 - ZX_4) * BF_8$;
- $BF_{17} = \max(0, ZX_4 + 1.74255) * BF_8$;
- $BF_{19} = \max(0, ZX_4 + 1.05685) * BF_8$;

Berdasarkan 15 *basis function* (BF) di atas, dapat diketahui bahwa variabel prediktor yang paling banyak

terdapat dalam *basis function* (BF) adalah rata-rata lama sekolah (X_2), pengeluaran per kapita (X_4), angka harapan hidup (X_3), persentase perempuan berstatus miskin usia 15-49 tahun yang menggunakan alat KB (X_5).

Daftar Pustaka

Eubank, R.L., 1988, Nonparametric Regression and Spline Smoothing. New York : Marcel Dekker.
Friedman, J.H. 1991. Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19: 1-67.
Santoso, A., 2010, Studi Deskriptif Effect Size Penelitian-Penelitian di Fakultas Psikologi Universitas Sanata Darma. *Jurnal Penelitian*, 14:1-17.

[UNDP] United Nations Development Programme. 1990. Human Development Report 1990. New York : Oxford university Press.

[Kompas] Harian Kompas. 2018. IPM Indonesia Timur Sangat Rendah. Diperoleh dari website Harian Kompas:

<https://kompas.id/baca/video/2018/04/18/ipm-indonesia-timur-sangat-rendah/> Diakses 19 November 2018.

Budiantara, I. N., dan Nisa', S. F., 2012, Analisis Survival dengan Pendekatan Multivariate Adaptive Regression Splines pada Kasus Demam Berdarah Dengue (DBD). *Jurnal Sains dan Seni ITS*: 1, 318-323.