

Penerapan Metode *Multivariate Adaptive Regression Spline* (MARS) untuk Mengetahui Faktor yang Mempengaruhi *Total Fertility Rate* (TFR) di Pulau Kalimantan

Aldi Dwi Putra*, Yayat Karyaana

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Islam Bandung, Indonesia.

*aldidp99@gmail.com, yayatkaryana@gmail.com

Abstract. Multivariate Adaptive Regression Spline (MARS) is a nonparametric regression method. The MARS model is useful for overcoming the problem of high-dimensional data, where data that has predictor variables as much as $3 \leq p \leq 20$ with a data sample of $50 \leq n \leq 1000$. The MARS model is obtained from a combination of Basis Function (BF), Maximum Interaction (MI), and Minimum Observation (MO) by trial and error, where the best MARS model is selected by selecting the minimum Generalized Cross Validation (GCV) value among other models. In this study, the researcher explains the use of MARS to analyze the factors that affect the Total Fertility Rate (TFR) in Kalimantan Island, because Kalimantan Island is the island with the highest average population growth rate in Indonesia. The response variable (Y) used is Total Fertility Rate (TFR) with several predictor variables, namely income (X_1), education (X_2), and use of contraceptives (X_3). The data used is secondary data from the 2015 Survei Penduduk Antar Sensus (SUPAS). Based on the results of the analysis, the best model obtained is a model with a combination of BF=6, MI=1, and MO=1 with a GCV value of 0,06611. These results indicate that there is one variable that is not included in the model, but only two variables that affect the Total Fertility Rate (TFR) in Kalimantan Island, namely in order of importance: use of contraceptive (X_3) and Education (X_2).

Keywords: Generalized Cross Validation (GCV), Multivariate Adaptive Regression Spline (MARS), SUPAS, TFR.

Abstrak. Multivariate Adaptive Regression Spline (MARS) adalah metode regresi nonparametrik. Model MARS berguna untuk mengatasi masalah data yang berdimensi tinggi, dimana data yang memiliki variabel prediktor sebanyak $3 \leq p \leq 20$ dengan sampel data yang berukuran $50 \leq n \leq 1000$. Model MARS diperoleh dari kombinasi nilai Basis Function (BF), Maximum Interaction (MI), dan Minimum Observation (MO) secara trial and error, dimana Model MARS terbaik dipilih dengan memilih nilai Generalized Cross Validation (GCV) minimum di antara model lainnya. Dalam penelitian ini, peneliti menjelaskan penggunaan MARS untuk menganalisis faktor-faktor yang mempengaruhi Total Fertility Rate (TFR) di Pulau Kalimantan, karena Pulau Kalimantan merupakan pulau dengan rata-rata laju pertumbuhan penduduk tertinggi di Indonesia. Variabel respon (Y) yang digunakan adalah Total Fertility Rate (TFR) dengan beberapa variabel prediktor yaitu pendapatan (X_1), pendidikan (X_2), dan penggunaan alat kontrasepsi (X_3). Data yang digunakan adalah data sekunder dari Survei Penduduk Antar Sensus (SUPAS) 2015. Berdasarkan hasil analisis, model terbaik yang didapatkan adalah model dengan kombinasi dari BF=6, MI=1, dan MO=1 dengan nilai GCV = 0,06611. Hasil ini menunjukkan bahwa ada satu variabel yang tidak masuk pada model, melainkan hanya dua variabel yang berpengaruh terhadap Total Fertility Rate (TFR) di Pulau Kalimantan yaitu menurut urutan kepentingan: penggunaan alat kontrasepsi (X_3) dan Pendidikan (X_2).

Kata Kunci: *Generalized Cross Validation* (GCV), *Multivariate Adaptive Regression Spline* (MARS), SUPAS, TFR.

1. Pendahuluan

Kondisi serta keadaan kependudukan amat mempengaruhi dinamika pembangunan yang dilakukan oleh pemerintah. Jumlah penduduk yang besar bila diiringi dengan kualitas penduduk yang mencukupi akan menjadi pendorong untuk perkembangan ekonomi. Kebalikannya, jumlah penduduk yang besar bila diiringi dengan tingkat kualitas yang rendah menjadikan penduduk tersebut menjadi beban untuk pembangunan (Suniarti *et al.*, 2015).

Berdasarkan data SUPAS Pulau Kalimantan memiliki rata-rata laju pertumbuhan penduduk tertinggi di Indonesia dengan laju pertumbuhan penduduk sebesar 2,5. Sementara itu di Pulau Jawa memiliki rata-rata laju pertumbuhan penduduk terendah di Indonesia sebesar 1,31. Oleh sebab itu penelitian ini akan membahas permasalahan laju pertumbuhan penduduk di Pulau Kalimantan.

Salah satu masalahnya karena masih tingginya *Total Fertility Rate* (TFR), yaitu angka yang menunjukkan rata-rata banyak anak dari seorang perempuan pada akhir masa reproduksi (Hatmadji, 1981).

Pada penelitian terdahulu oleh Mahendra (2016), terdapat banyak faktor yang bisa mempengaruhi tingkat fertilitas, beberapa faktor yang mempengaruhinya merupakan pendapatan, pendidikan, dan penggunaan alat kontrasepsi. Demikian juga dilakukan oleh Karyana *et al* (2020), meneliti pengaruh pendapatan, pendidikan, dan penggunaan alat kontrasepsi terhadap TFR namun untuk kasus di Indonesia

Pendekatan analisis yang digunakan untuk mengetahui faktor yang mempengaruhi TFR, salah satunya ialah analisis regresi. Dalam penelitian ini, *Total Fertility Rate* (TFR) merupakan variabel respon dan variabel prediktornya yaitu faktor yang mempengaruhi TFR yaitu pendapatan, pendidikan, dan penggunaan alat kontrasepsi.

Metode regresi nonparametrik salah satunya ialah *Multivariate Adaptive Regression Spline* (MARS) yang dipopulerkan oleh (Friedman, 1991). Model MARS berguna untuk mengatasi masalah data yang berdimensi tinggi, ialah data yang memiliki variabel prediktor sebesar $3 \leq p \leq 20$ dengan sampel data yang berukuran $50 \leq n \leq 1000$. MARS merupakan pengembangan dari pendekatan *Recursive Partitioning Regression* (RPR) yang dikombinasikan dengan metode Spline maka model yang didapatkan kontinu pada knot yakni garis regresi selalu menyambung, dimana setiap knot selalu menyambung dengan fungsi basisnya, dan Model MARS terbaik dipilih dengan memilih nilai *Generalized Cross Validation* (GCV) minimum di antara model lainnya.

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah dalam penelitian ini sebagai berikut: “Bagaimana pemodelan dan pengaruh variabel-variabel prediktor terhadap model yang diperoleh”. Selanjutnya, tujuan dalam penelitian ini diuraikan dalam pokok-pokok sbb.

1. Mengetahui hasil pemodelan *Multivariate Adaptive Regression Spline* (MARS) terbaik terhadap *Total Fertility Rate* (TFR) di Pulau Kalimantan menggunakan kriteria *Generalized Cross Validation* (GCV).
2. Mengetahui pengaruh dan tingkat pentingnya variabel-variabel prediktor terhadap model terbaik yang diperoleh.

2. Metodologi

Bahan

Data yang digunakan dalam penelitian ini adalah data sekunder. Yaitu data hasil dari Survei Penduduk Antar Sensus (SUPAS) 2015, yang didapat dari Badan Pusat Statistik (BPS). Fokus penelitian pada data SUPAS ini yaitu 56 kabupaten/kota yang terdapat di 5 provinsi di Pulau Kalimantan.

Variabel Penelitian

Variabel-variabel yang digunakan dalam penelitian dapat dilihat pada Tabel 1 di bawah ini:

Tabel 1. Variabel Penelitian

Variabel	Keterangan
Y	<i>Total Fertility Rate</i> (TFR)
X ₁	Pendapatan
X ₂	Pendidikan
X ₃	Penggunaan Alat Kontrasepsi

Metode analisis penelitian

1. Analisis Deskriptif
Melakukan statistik deskriptif dan grafik plot antara variabel respon TFR dengan masing-masing variabel prediktor sebagai langkah awal mengetahui pola hubungan antar variabel tersebut.
2. Membuat model MARS
 - a. Menentukan kemungkinan maksimum banyaknya *basis function* (BF) yaitu 2-4 kali jumlah variabel prediktor yang akan digunakan. Karena pada penelitian ini akan menggunakan tiga variabel prediktor maka kemungkinan banyaknya fungsi basis pada penelitian ini adalah 6, 9, dan 12.
 - b. Menentukan jumlah maksimum interaksi, dalam penelitian ini jumlah maksimum interaksi (MI) yaitu 1, 2, dan 3. Apabila terdapat lebih dari 3 interaksi, maka akan menimbulkan interpretasi model yang sangat kompleks.
 - c. Menentukan minimal jumlah pengamatan setiap knots (MO), yaitu 0, 1, 2, dan 3.
 - d. Melakukan estimasi model MARS dan menentukan model MARS terbaik. Ditentukan dengan trial and error sampai diperoleh model optimal dengan nilai *Generalized Cross Validation* (GCV) yang terkecil yang diperoleh dari kombinasi antara BF, MI, dan MO.
 - e. Melakukan interpretasi model MARS terbaik dan interpretasi variabel-variabel yang berpengaruh di dalam model tersebut, serta tingkat kepentingan masing-masing variabel prediktor yang berpengaruh

3. Pembahasan dan Diskusi

Deskripsi Data

Deskripsi data yaitu tahap awal eksplorasi data yang dilakukan untuk mendapatkan gambaran umum dari data yang digunakan pada penelitian. Data penelitian ini merupakan data Survei Penduduk Antar Sensus (SUPAS) 2015, dalam penelitian ini yaitu 56 kabupaten/kota di Pulau Kalimantan.

Tabel 2. Statistik Deskriptif Data Penelitian

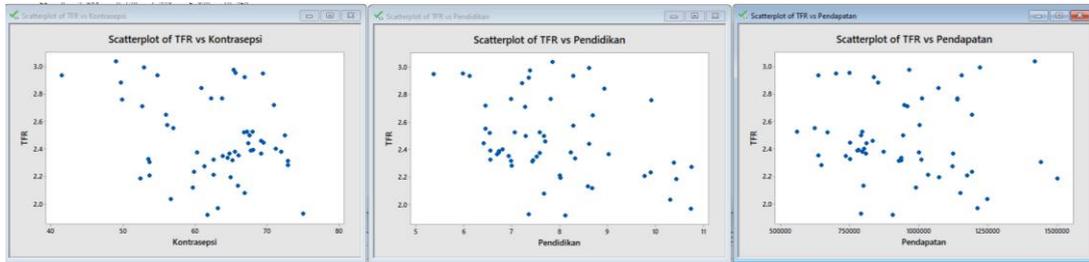
Variabel	N	Minimum	Maksimum	Rata-rata
TFR (Y)	56	1,92	3,04	2,47
Pendapatan (X ₁)	56	560.200	1.501.600	946.183,93
Pendidikan (X ₂)	56	5,4	10,8	7,8
Alat Kontrasepsi (X ₃)	56	41,49	75,03	62,70

Sumber: Data Penelitian yang Sudah Diolah, 2021.

Berdasarkan Tabel 2 menunjukkan secara rata-rata *Total Fertility Rate* (TFR) di Pulau Kalimantan yaitu sebesar 2,47. Nilai tersebut menunjukkan bahwa secara rata-rata bayi yang lahir hidup di Pulau Kalimantan yaitu terbesar dari seluruh pulau di Indonesia. Dengan TFR terendah yaitu di Kabupaten Paser sebesar 1,92 dan tertinggi di Kabupaten Tana Tidung sebesar 3,04.

Pola hubungan variabel respon dengan variabel prediktor dapat dilihat dari Scatterplot

berikut:



Gambar 1. Scatterplot antara Variabel Respons dengan Variabel Prediktor

Dapat dilihat bahwa variabel respon TFR dengan masing-masing variabel prediktor yaitu pendapatan (X_1), pendidikan (X_2), dan penggunaan alat kontrasepsi (X_3) dengan menggunakan *scatterplot* menunjukkan pola hubungan yang tidak mengikuti suatu pola tertentu. Karena tidak diketahui bentuk pola datanya maka estimasi model dilakukan dengan analisis regresi nonparametrik.

Dalam penelitian ini metode MARS akan diterapkan dalam memodelkan Total Fertility Rate (TFR) berdasarkan faktor-faktor yang diduga mempengaruhinya yaitu Pendapatan (X_1), Pendidikan (X_2), dan Penggunaan Alat Kontrasepsi (X_3).

Langkah selanjutnya yaitu menentukan maksimum jumlah *basis function* (BF), *maximum interaction* (MI), dan *minimum observation* (MO). Fungsi basis merupakan fungsi yang didefinisikan dari subregion. Fungsi basis yang digunakan adalah dua sampai empat kali jumlah variabel prediktor. Variabel prediktor yang digunakan pada penelitian ini ada sebanyak tiga variabel maka banyaknya fungsi basis yang digunakan adalah sebanyak 6, 9, dan 12. Maximum Interaction yang digunakan adalah 1, 2, dan 3, menurut Friedman (1991) jika lebih dari 3 maka nilai GCV semakin meningkat dan model yang dihasilkan semakin kompleks. Nilai dari Minimum Observation sebesar 0, 1, 2, dan 3, oleh karena itu akan ada 36 model yang terbentuk. Tabel 2 menunjukkan hasil trial and error pembentukan Model MARS sebagai berikut.

Tabel 2. Trial and Error pembentukan model MARS

Model	BF	MI	MO	GCV	Model	BF	MI	MO	GCV
1	6	1	0	0.06861	19	9	2	2	0.07285
2	6	1	1	0.06611	20	9	2	3	0.07504
3	6	1	2	0.06642	21	9	3	0	0.0706
4	6	1	3	0.06672	22	9	3	1	0.07459
5	6	2	0	0.07324	23	9	3	2	0.07285
6	6	2	1	0.07281	24	9	3	3	0.07504
7	6	2	2	0.07311	25	12	1	0	0.06944
8	6	2	3	0.07311	26	12	1	1	0.06699
9	6	3	0	0.07324	27	12	1	2	0.06744
10	6	3	1	0.07281	28	12	1	3	0.06773
11	6	3	2	0.07311	29	12	2	0	0.07523
12	6	3	3	0.07311	30	12	2	1	0.07565
13	9	1	0	0.07087	31	12	2	2	0.07612
14	9	1	1	0.06614	32	12	2	3	0.07611
15	9	1	2	0.06659	33	12	3	0	0.07523
16	9	1	3	0.06688	34	12	3	1	0.07565
17	9	2	0	0.0706	35	12	3	2	0.07612

18	9	2	1	0.07459	36	12	3	3	0.07611
----	---	---	---	---------	----	----	---	---	---------

Sumber: Data Penelitian yang Sudah Diolah, 2021.

Model MARS terbaik adalah model yang menghasilkan nilai GCV terkecil. Nilai GCV dihasilkan dari mengkombinasikan jumlah *basis function* (BF), *maximum interaction* (MI), dan *minimum observation* (MO) secara *trial and error* sehingga dari 36 model yang terbentuk dihasilkan model terbaiknya yaitu dari kombinasi BF=6, MI=1, dan MO=1 karena memiliki nilai GCV minimum sebesar 0,06611, dengan persamaan model sebagai berikut.

$$Y = 2.51884 + 0.0497355 * BF2 - 0.0873117 * BF3 + 0.459984 * BF4 - 0.211774 * BF5;$$

Dengan:

$$BF2 = \max(0; 56,9583 - X_3);$$

$$BF3 = \max(0; X_2 - 6,56);$$

$$BF4 = \max(0; 6,56 - X_2);$$

$$BF5 = \max(0; X_3 - 72,4418);$$

Interpretasi MARS pada persamaan terbaik tersebut sebagai berikut.

1. $BF2 = \max(0; 56,9583 - X_3);$

Dengan koefisien 0,0497355

$$+ 0,0497355 h(56,9583 - X_3) = \begin{cases} 56,9583 - X_3; & \text{Jika } X_3 < 56,9583 \\ 0 & ; \text{Jika } X_3 \geq 56,9583 \end{cases}$$

Artinya bahwa setiap kenaikan BF2 sebesar satu satuan akan meningkatkan *Total Fertility Rate* (TFR) sebesar 0,0497355 atau dengan kata lain jika variabel penggunaan alat kontrasepsi (X_3) lebih kecil dari 56,9583 akan meningkatkan *Total Fertility Rate* (TFR) sebesar 0,0497355.

2. $BF3 = \max(0; X_2 - 6,56);$

Dengan koefisien -0,0873117

$$- 0,0873117 h(X_2 - 6,56) = \begin{cases} X_2 - 6,56; & \text{Jika } X_2 > 6,56 \\ 0 & ; \text{Jika } X_2 \leq 6,56 \end{cases}$$

Artinya bahwa setiap kenaikan BF3 sebesar satu satuan akan menurunkan *Total Fertility Rate* (TFR) sebesar 0,0873117 atau jika variabel Pendidikan (X_2) lebih besar dari 6,56 maka akan menurunkan *Total Fertility Rate* (TFR) sebesar 0,0873117.

3. $BF4 = \max(0; 6,56 - X_2);$

Dengan koefisien 0,459984

$$+ 0,459984 h(6,56 - X_2) = \begin{cases} 6,56 - X_2; & \text{Jika } X_2 < 6,56 \\ 0 & ; \text{Jika } X_2 \geq 6,56 \end{cases}$$

Artinya bahwa setiap kenaikan BF4 sebesar satu satuan akan meningkatkan *Total Fertility Rate* (TFR) sebesar 0,459984 atau jika variabel Pendidikan (X_2) lebih kecil dari 6,56 maka akan meningkatkan *Total Fertility Rate* (TFR) sebesar 0,459984.

4. $BF5 = \max(0; X_3 - 72,4418);$

Dengan Koefisien -0,211774

$$- 0,211774 h(X_3 - 72,4418) = \begin{cases} X_3 - 72,4418; & \text{Jika } X_3 > 72,4418 \\ 0 & ; \text{Jika } X_3 \leq 72,4418 \end{cases}$$

Artinya bahwa setiap kenaikan BF5 sebesar satu satuan akan menurunkan *Total Fertility Rate* (TFR) sebesar 0,211774 atau jika variabel penggunaan alat kontrasepsi (X_3) lebih besar dari 72,4418 maka akan menurunkan *Total Fertility Rate* (TFR) sebesar 0,211774.

4. Kesimpulan

Berdasarkan hasil analisis dengan penerapan *Multivariate Adaptive Regression Spline* (MARS) dapat disimpulkan sebagai berikut.

1. Model *Multivariate Adaptive Regression Spline* (MARS) pada variabel prediktor untuk *Total Fertility Rate* (TFR) dengan persamaan model sebagai berikut.

$$Y = 2.51884 + 0.0497355 * BF2 - 0.0873117 * BF3 + 0.459984 * BF4 - 0.211774 * BF5$$

Basis Fungsi yang digunakan dalam model ialah: BF2,BF3,BF4, dan BF5

Model MARS terbaik di atas diperoleh dari kombinasi fungsi basis (BF)=6, maksimum interaksi (MI)=1,dan minimum observasi (MO)=1 secara *trial and error* dengan nilai GCV terkecil sebesar 0,06611. Dimana hanya dua variabel yang masuk didalam model yaitu Penggunaan Alat Kontrasepsi (X_3) dan Pendidikan (X_2). Sedangkan, Pendapatan (X_1) hilang dan tidak terdapat di dalam persamaan model terbaik yang diperoleh.

2. Variabel prediktor yang mempengaruhi *Total Fertility Rate* (TFR) di Pulau Kalimantan yaitu variabel Pendidikan (X_2) dan variabel Penggunaan Alat Kontrasepsi (X_3) sedangkan variabel Pendapatan (X_1) tidak berpengaruh, karena tidak terdapat di dalam model terbaik. Besar tingkat pentingnya variabel prediktor terhadap model terbaik yang mempengaruhi *Total Fertility Rate* (TFR) adalah Penggunaan Alat Kontrasepsi (X_3) dengan tingkat kepentingan 100% dan Pendidikan (X_2) dengan tingkat kepentingannya 71,68%. Penggunaan Alat Kontrasepsi (X_3) dengan pentingnya 100% artinya bahwa penggunaan alat kontrasepsi yang memiliki tingkat pentingnya sangat mempengaruhi *Total Fertility Rate* (TFR).

Acknowledge

Terima kasih kepada seluruh pihak yang membantu memberi saran maupun masukan sampai terlaksananya penelitian ini.

Daftar Pustaka

- [1] Friedman, J. H. (1991). *Multivariate Adaptive Regression Spline* (With Discussion). In *The Annals of Statistics* (pp. 1–141).
- [2] Hatmadji, S. H. (1981). *Fertilitas. Dalam Dasar-Dasar Demografi*. Lembaga Demografi Fakultas Ekonomi Universitas Indonesia.
- [3] Karyana, Y., Remi, S. ., Yusuf, A. ., & Purnagunawan, M. (2020). Fertilitas dan Keputusan Menambah Anak di Indonesia. In S. S. Widyaputra & C. B. Kartasasmita (Eds.), *PANDEMI COVID-19: SAATNYA KAJI ULANG ARAH PENELITIAN DAN PENDIDIKAN KESEHATAN* (p. 397). Unpad Press.
- [4] Mahendra, A. (2016). Faktor-Faktor yang Mempengaruhi Fertilitas di Indonesia Timur. *Jrak*, 3(2), 223–242. <http://103.76.21.184/index.php/JRAK/article/download/448/478>
- [5] Suniarti, S., Gunarto, T., & Saimul. (2015). *Analisis Faktor-Faktor Yang Mempengaruhi Tingkat Kelahiran Total Di Provinsi Lampung Suatu Kajian Sosial Ekonomi*. <http://digilib.unila.ac.id/12653/>
- [6] Utama Muhammad Bangkit Riksa, Hajarisman Nusar. (2021). *Metode Pemilihan Variabel pada Model Regresi Poisson Menggunakan Metode Nordberg*. *Jurnal Riset Statistika*, 1(1), 35-42.