

# Perbandingan Metode CV dan GCV pada Pemodelan MARS (Aplikasi Rata-Rata Lama Sekolah di Kabupaten Gianyar)

Luh Putu Safitri Pratiwi\*<sup>1</sup>, Ni Putu Meina Ayuningsih<sup>2</sup>, I Made Pasek Pradnyana Wijaya<sup>3</sup>  
<sup>1,2,3</sup>ITB STIKOM Bali;  
e-mail: \*[1putu\\_safitri@stikom-bali.ac.id](mailto:1putu_safitri@stikom-bali.ac.id), [2meina\\_ayuningsih@stikom-bali.ac.id](mailto:2meina_ayuningsih@stikom-bali.ac.id),  
[3pasek\\_pradnyana@stikom-bali.ac.id](mailto:3pasek_pradnyana@stikom-bali.ac.id)

## Abstrak

Penelitian ini dilakukan pemodelan rata-rata lama sekolah dengan menggunakan pendekatan MARS (Multivariate Adaptive Regression Spline). MARS adalah salah satu metode regresi nonparametric selain spline, kernel, dan lain-lain, dimana tidak ada informasi apapun tentang bentuk dari kurva regresi. MARS juga menggunakan parameter penghalus yaitu metode Cross Validation (CV) dan metode Generalized Cross Validation (GCV). Hasil yang diperoleh nilai RMSE dan  $R^2$  model MARS CV baik lebih baik dibandingkan dengan GCV karena memiliki nilai RMSE yang paling rendah dan  $R^2$  yang lebih besar. Hal tersebut bisa menunjukkan bahwa model MARS dengan CV adalah model yang lebih baik untuk memodelkan rata-rata lama sekolah di Kabupaten Gianyar. Model yang diperoleh dari metode CV dengan model yang diperoleh sebagai berikut:

$$y = 7.656 - 0.115 * BF1 - 0.366 * BF2 + 0.496 * BF3 + 0.205 * BF4$$

**Kata kunci** MARS, GCV, CV, Lama sekolah

## 1. PENDAHULUAN

Human Development Report (HDR) menggunakan pengukuran statistik yaitu indeks pembangunan manusia (IPM) yang telah dipublikasikan secara berkala dalam laporan tahunan HDR. UNDP memperkenalkan penghitungan IPM dengan metode yang baru. Metode ini menggunakan indikator baru dalam penghitungannya. Indikator dari metode baru tersebut ialah kesehatan, yang diukur dari angka harapan hidup saat kelahiran. Pendidikan, dihitung berdasarkan angka harapan sekolah dan angka rata-rata lama sekolah. Sedangkan standar hidup layak dari produk nasional bruto per kapita. Metode baru ini, baru diaplikasikan di Indonesia mulai tahun 2014 (BPS, 2014). Salah satu tujuan perubahan metode tersebut yaitu agar indikator yang digunakan lebih tepat sasaran sehingga target IPM terpenuhi.

IPM di Provinsi Bali sendiri dari tahun ke tahun konsisten dan mengalami peningkatan. Pada tahun 2018 Provinsi Bali masuk kedalam lima klasifikasi tertinggi IPM di seluruh provinsi yang ada di Indonesia. Salah satunya yaitu ditandainya dengan meningkatnya IPM pada tahun 2020 yang tercatat capaian sebesar 75,50 atau setara 0,12 poin dibanding tahun 2019 yang tercatat sebesar 75,38, akan tetapi pada tahun 2020 terjadi perlambatan pembangunan manusianya yang diindikasikan oleh pertumbuhan IPM yang sebesar 0,16 persen, lebih rendah dari pertumbuhan tahun 2019 sebesar 0,82 persen. Hal ini dikarenakan penurunan kemampuan masyarakat dalam memenuhi kebutuhan hidupnya akibat COVID-19, dapat dilihat dari komponen rata-rata pengeluaran per kapita disesuaikan (PPP) sebesar 13,93 juta rupiah menurun 217 ribu rupiah dari tahun 2019 (BPS, 2020). Sehingga hal ini juga berpengaruh terhadap segala aspek indikator IPM yang tidak hanya perlambatannya pada sektor ekonomi, kesehatan, namun di sector pendidikan pun terkena dampak besar akibat dari COVID-19 ini. Dilihat dari aspek indikator pendidikan biasanya seseorang yang memiliki pendidikan tinggi peluang diterima pada suatu perusahaan akan semakin besar. Semakin tinggi tingkat pendidikan seseorang maka capaian rata-rata lama sekolah akan semakin tinggi pula. Maka dari itu penulis ingin mengetahui pemodelan rata-rata lama sekolah dengan menggunakan pendekatan MARS (Multivariate

*Adaptive Regression Spline*). MARS adalah salah satu metode regresi nonparametrik selain spline, kernel, dan lain-lain, dimana tidak ada informasi apapun tentang bentuk dari kurva regresi, untuk kasus yang terdiri dari beberapa variabel prediktor, pendekatan regresi nonparametrik dapat menggunakan plot data hubungan variabel prediktor dan respon menunjukkan pola data yang tidak jelas. Karena didukung oleh penelitian-penelitian sebelumnya yang menjelaskan kelebihan MARS dibanding metode-metode yang lainnya, maka hal tersebut mendorong penulis untuk menggunakan metode MARS. Penelitian Utami (2021) mengenai pemodelan MARS pada kasus diare di Provinsi Jawa Tengah-Jawa Barat tahun 2019 menggunakan nilai Generalized Cross Validation (GCV) terkecil yang dihasilkan tiap model kombinasi dari minimum observasi (MO), basis function (BF) dan Interaksi (MI). Penelitian Darmawanti, Suparti, dan Safitri (2014) menggunakan pemodelan MARS yang digunakan untuk menganalisis pergerakan IHSG dengan metode CV yang diperoleh model terbaik yang terjadi pada kombinasi antara BF=9, MI=2, dan MO=1

MARS juga menggunakan parameter penghalus beberapa metode untuk memilih parameter penghalus yaitu, Cross Validation (CV) (Craven dan Wahba, 1979) dan Wahba (1990) memberikan suatu metode Generalized Cross Validation (GCV). Penelitian mengenai perbandingan CV dan GCV pernah diteliti oleh Razak (2019) penelitian tersebut menyimpulkan metode CV lebih baik dalam pemilihan bandwidth Optimal yang dapat menghasilkan estimasi mendekati data aslinya dalam pemilihan bandwidth optimal pada pemodelan regresi nonparametrik kernel *gaussian*. Dalam beberapa kasus, CV memiliki kaitan yang erat dengan GCV. GCV merupakan modifikasi dari CV yang didapat dengan meminimumkan fungsi CV, sehingga menarik untuk dilakukan suatu perbandingan antara metode CV maupun GCV (Eubank, 1999).

## 2. METODE PENELITIAN

Sumber data yang digunakan dalam penelitian ini adalah data sekunder diperoleh dari Nasional (SUSENAS) di Direktorat Diseminasi Statistik BPS di Kabupaten Gianyar. Data memuat informasi tentang lama sekolah dan faktor-faktor yang mempengaruhi lama sekolah, dimana akan digunakan sebagai variabel penelitian. Adapun gambaran umum variabel penelitian yang digunakan sebagai berikut:

a. Variabel Respon

Y Lama Sekolah Penduduk Usia Sekolah

b. Variabel bebasnya

X<sub>1</sub> IPM (Indeks Pembangunan Manusia)

X<sub>2</sub> Persentase pengeluaran perkapita (ribu rupiah)

X<sub>3</sub> AMH (Angka Melek Huruf)

### Alur Analisis

Adapun tahapan analisis data yang akan dilakukan dalam penelitian ini adalah sebagai berikut:

1. Mengambil data SUSENAS di Direktorat Diseminasi Statistik BPS.
2. Mengumpulkan data-data semua variabel yang akan digunakan dalam pembentukan model.
3. Sebelum dilakukan pengolahan data, terlebih dahulu dilakukan pem-plot-an data variabel respon dengan setiap variabel-variabel prediktornya.

Untuk mendapatkan model lama sekolah di Kabupaten Gianyar dengan pendekatan MARS, langkah-langkah yang harus dilakukan adalah sebagai berikut:

1. Untuk mendapatkan model terbaik dilakukan dengan cara *trial and error* sampai didapat model dengan nilai CV dan GCV minimum. Tahapannya sebagai berikut:
  - a. Menentukan maksimum fungsi basis
  - b. Menentukan jumlah maksimum interaksi

- c. Menentukan minimal jumlah pengamatan setiap knots
2. Mendapatkan hasil estimasi parameter.
3. Menetapkan model terbaik dengan kriteria seperti pada langkah 1.

Model MARS digunakan untuk mengatasi kelemahan RPR yaitu menghasilkan model yang kontinu pada knot. Perbaikan yang dilakukan untuk mengatasi keterbatasan RPR, antara lain menghasilkan fungsi basis menjadi:

$$B_m^{(q)}(\mathbf{x}) = \prod_{k=1}^{K_m} [s_{km} \cdot (x_{v(k,m)} - t_{km})]_+^q$$

Estimasi dari kurva regresi  $f(\mathbf{x})$  secara umum didapatkan melalui *penalized least square* (PLS), yaitu meminimumkan persamaan berikut:

$$\frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 + \eta \int_a^b (f^{(n)}(x))^2 dx \quad (1)$$

Setelah dilakukan modifikasi model RPR dan dikombinasikan dengan *spline*, estimator model MARS dapat ditulis sebagai berikut:

$$\hat{f}(x) = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} [s_{km} (x_{v(k,m)} - t_{km})] \quad (2)$$

dengan

- $a_0$  = konstanta
- $a_m$  = koefisien dari fungsi basis ke- $m$
- $M$  = banyaknya fungsi basis (*nonconstant basis function*)
- $K_m$  = derajat interaksi
- $x_{v(k,m)}$  = variabel prediktor
- $t_{km}$  = nilai *knot* dari variabel prediktor  $x_{v(k,m)}$

Dengan menggunakan persamaan (2), maka model MARS adalah :

$$y_i = \alpha_0 + \sum_{m=1}^M \alpha_m \prod_{k=1}^{K_m} [s_{km} \cdot (x_{v(k,m)} - t_{km})] + \varepsilon_i = \alpha_0 + \sum_{m=1}^M \alpha_m B_m(\mathbf{x}) + \varepsilon_i \quad (3)$$

dengan,  $B_m(\mathbf{x}) = \prod_{k=1}^{K_m} [s_{km} \cdot (x_{v(k,m)} - t_{km})]$

Dalam bentuk matrik dapat ditulis menjadi :

$$\mathbf{Y} = \mathbf{B}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \quad (4)$$

dengan,

$$\mathbf{Y} = (Y_1, \dots, Y_n)^T, \quad \boldsymbol{\alpha} = (\alpha_0, \dots, \alpha_m)^T, \quad \boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$$

$$\mathbf{B} = \begin{pmatrix} 1 & \prod_{k=1}^{K_1} s_{1m}(x_{1(1,m)} - t_{1m}) & \cdots & \prod_{k=1}^{K_M} s_{Mm}(x_{1(M,m)} - t_{Mm}) \\ 1 & \prod_{k=1}^{K_1} s_{1m}(x_{2(1,m)} - t_{1m}) & \cdots & \prod_{k=1}^{K_M} s_{Mm}(x_{2(M,m)} - t_{Mm}) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \prod_{k=1}^{K_1} s_{1m}(x_{n(1,m)} - t_{1m}) & \cdots & \prod_{k=1}^{K_M} s_{Mm}(x_{n(M,m)} - t_{Mm}) \end{pmatrix}$$

Prosedur forward dan backward menghasilkan sebuah model (3) dan penjabaran dapat disajikan sebagai berikut :

$$\begin{aligned} \hat{f}(x) = & \alpha_0 + \sum_{m=1}^M \alpha_m [s_{1m} \cdot (x_{v(1,m)} - t_{1m})] \\ & + \sum_{m=1}^M \alpha_m [s_{1m} \cdot (x_{v(1,m)} - t_{1m})][s_{2m} \cdot (x_{v(2,m)} - t_{2m})] \\ & + \sum_{m=1}^M \alpha_m [s_{1m} \cdot (x_{v(1,m)} - t_{1m})][s_{2m} \cdot (x_{v(2,m)} - t_{2m})][s_{3m} \cdot (x_{v(3,m)} - t_{3m})] \\ & + \dots \end{aligned} \quad (5)$$

Sehingga, model MARS dapat ditulis sebagai berikut.

$$Y_i = \alpha_0 + \sum_{m=1}^M \alpha_m \prod_{k=1}^{K_m} [s_{km} (X_{v(k,m)} - t_{km})] + \varepsilon_i = \alpha_0 + \sum_{m=1}^M \alpha_m B_m(X) + \varepsilon_i \quad (6)$$

dengan:

- $\alpha_0$  = fungsi basis induk
- $\alpha_m$  = koefisien dari fungsi basis ke-m
- $M$  = maksimum fungsi basis (nonkonstan fungsi basis)
- $K_m$  = derajat interaksi
- $S_{km}$  = nilainya  $\pm 1$
- $X_{v(k,m)}$  = variabel independen
- $t_{km}$  = nilai knot dari variabel independen  $X_{v(k,m)}$

GCV merupakan modifikasi dari CV yang didapat dengan meminimumkan fungsi CV, Dasar dari metode CV adalah untuk memilih nilai  $\mathbf{k}$  (knot) yang meminimumkan  $CV(\mathbf{k})$ . Langkah awal dalam metode CV adalah memperhatikan bahwa nilai bergantung secara linear pada data dengan persamaan 7.

$$\hat{y} = \mathbf{A}(\mathbf{k})y \quad (7)$$

dengan

$$\mathbf{A}(\mathbf{k}) = \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$$

Hasil yang diperoleh dari pengembangan perhitungan nilai CV diberikan sebagai berikut (Green dan Silverman, 1994).

$$CV(\mathbf{k}) = \frac{1}{n} \sum_{i=1}^n \left[ \frac{y_i - \hat{y}_i}{1 - \mathbf{A}_{ii}(\mathbf{k})} \right]^2 \quad (8)$$

Metode CV mengasumsikan bahwa fungsi diperoleh tanpa pengamatan ke-i dari data. Metode CV biasa disebut sebagai metode hapus-satu, yaitu suatu metode yang bertujuan untuk meminimumkan jumlah kuadrat dari error prediksi untuk variabel respon, dimana prediktor untuk respon tersebut didasarkan pada estimator yang menggunakan seluruh data kecuali data (Andrews, 1991). Sedangkan Nilai GCV didefinisikan sebagai berikut. (Friedman and Silverman, 1989).

$$LOF(\hat{f}_M) = GCV(M) = \frac{\frac{1}{N} \sum_{i=1}^N [y_i - \hat{f}_M(X_i)]^2}{\left[ 1 - \frac{\tilde{C}(M)}{N} \right]^2} \quad i=1,2,3,\dots,N \quad (9)$$

dengan:

LOF = *Loss Of Function*

ASR =  $\frac{1}{N} \sum_{i=1}^N [y_i - \hat{f}_M(X_i)]^2$  dan

N = banyaknya pengamatan

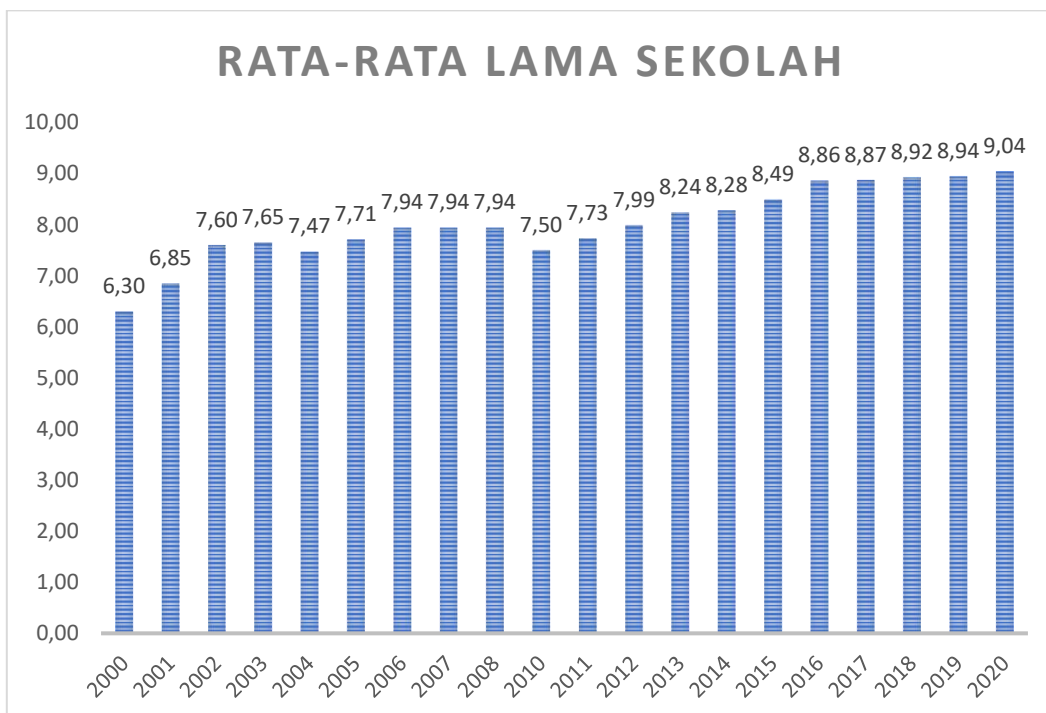
C(M) = Trace[B(B'B)<sup>-1</sup>B'] + 1 adalah banyaknya parameter yang diestimasi

$\tilde{C}(M)$  = C(M) + d.M, nilai d yang terbaik berada dalam interval  $2 \leq d \leq 4$ .

### 3. HASIL DAN PEMBAHASAN

Rata-rata lama sekolah digunakan untuk mengidentifikasi jenjang kelulusan pendidikan penduduk suatu daerah. Rata-rata lama sekolah dapat diartikan lamanya waktu tempuh pendidikan yang dilakukan oleh seseorang. Sebagai gambaran, seseorang yang telah menamatkan pendidikan sampai tingkat SD maka ia telah memiliki lama sekolah sebanyak 6 tahun. Kegunaan rata-rata lama sekolah dapat digunakan untuk monitoring pelaksanaan Program Wajib Belajar 9 Tahun yang dicanangkan oleh pemerintah (BPS Aceh, 2015), yang berarti untuk melewati target program tersebut maka rata-rata lama sekolah harus sudah mencapai 9 tahun.

Perkembangan angka lama sekolah di Kabupaten Gianyar dapat dilihat pada gambar 1 dibawah ini.



**Gambar 1** Angka lama sekolah di Kabupaten Gianyar Tahun 2000-2020

Pada Gambar 1, terlihat bahwa rata-rata lama sekolah di Kabupaten Gianyar paling tinggi terjadi pada tahun 2020 yaitu sebesar 9,04 atau minimal sampai jenjang SMP. Data lebih lengkap mengenai rata-rata lama sekolah dan faktor mempengaruhinya dapat dilihat pada Tabel 1.

Tabel 1 Analisis Deskriptif

Analisis Deskriptif					
Variabel	N	Minimum	Maximum	Rata-rata	Std. Deviasi
Y	20	6.300	9.040	8.013	0.721
X <sub>1</sub>	20	67.200	77.360	72.463	3.309
X <sub>2</sub>	20	11.122	15.888	12.949	1.324
X <sub>3</sub>	20	79.20	95.36	87.77	4.97

Berdasarkan Tabel 1, diketahui bahwa pada karakteristik variabel X<sub>1</sub> yaitu mengenai IPM memiliki rata-rata sebesar 72,463 dan standar deviasi sebesar 3,309. Jumlah maksimum IPM sebesar 77,360 pada tahun 2020 dan minimum pada tahun 2000 dan 2001.

Pada karakteristik variabel X<sub>2</sub> yaitu mengenai persentase pengeluaran perkapita (ribu rupiah) diperoleh rata-rata sebesar 12,949. Presentase pengeluaran perkapita maksimum sebesar 15,888 pada tahun 2020 dan 20 pengeluaran perkapita sebesar minimum sebesar 11,112 pada tahun 2000. Memiliki standar deviasi sebesar 1,324 yang menunjukkan rentang variasi data pengeluaran perkapita antar tahun menunjukkan angka yang cukup rendah.

Pada karakteristik variabel X<sub>3</sub> yaitu mengenai AMH diperoleh rata-rata penduduk sebesar 95,36. Rata-rata AMH 95,36 menunjukkan bahwa AMH sudah cukup baik dan harus terus dilakukan perbaikan untuk meningkatkan kualitas AMH. AMH maksimum sebesar 95,36 berada pada tahun 2020 dan AMH minimum sebesar 79,20 berada pada tahun 2000. Memiliki standar deviasi sebesar 4,97 yang menunjukkan bahwa rentang variasi AMH menunjukkan angka yang cukup rendah.

Menurut Friedman (1991) dalam analisis regresi spline, jika diberikan data berdimensi tinggi yaitu data dengan variabel prediktor dimana  $3 \leq p \leq 20$  dengan ukuran sampel  $50 \leq n \leq 1000$  menggunakan metode Multivariate Adaptive Regression Splines (MARS). Hal yang diperhatikan dalam penggunaan model MARS adalah (Nisa' dan Budiantara, 2012):

1. Knot yaitu akhir dari sebuah garis regresi dan awal dari sebuah regresi yang lainnya.
2. Basis Function/Fungsi Basis (BF) yaitu suatu fungsi yang dipergunakan untuk menjelaskan hubungan antar variabel respon dan variable prediktor. Jumlah basis fungsi (BF) sebesar 2-4 kali variabel prediktor. Interaksi yaitu hubungan korelasi antar variabel dengan jumlah maksimum interaksi (MI) sebesar 1, 2, dan
3. Pemodelan MARS ditentukan berdasarkan *trial and error* untuk kombinasi BF, MI, dan MO untuk mendapatkan nilai dari parameter pemulus yang minimum. MO yaitu minimum jarak antar knot atau minimum observasi antar knot (MO) sebesar 0, 1, 2, dan 3 (Nisa' dan Budiantara, 2012).

Berikut disajikan seleksi model MARS dengan metode GCV pada Tabel 2.

Tabel 2 Hasil Seleksi Model MARS Menggunakan GCV dengan BF = 9, dan Kombinasi Nilai MI dan MO

No	BF	MI	MO	GCV
1	9	1	0	0.565
2	9	2	1	0,073
3	9	3	2	0.151
4	9	1	3	0.109
5	9	2	0	0.086
6	9	3	1	0.235
7	9	1	2	0.111
8	9	2	3	0,095
9	9	3	0	0.090
10	9	1	1	0.100
11	9	2	2	0.201
12	9	3	3	0.085

Pada Tabel 2, model terbaik yaitu pada BF = 9, MI = 2, MO = 1 memiliki nilai GCV terkecil yaitu sebesar 0.0733 dengan model sebagai berikut:

$$y = 7.770 - 0.205 \cdot BF1 + 0.206 \cdot BF2$$

dengan

$$BF1 = \max(0, 85 - x_3)$$

$$BF2 = \max(0, x_1 - 71.45)$$

Intepretasi model:

$$BF1 = \max(0, 85 - x_3)$$

Artinya setiap kenaikan BF1 akan mengurangi nilai rata-rata lama sekolah sebesar 0,205 untuk data rata-rata lama sekolah Jika menggunakan data asli sebesar 0,205 dikali standar deviasi dari data asli rata-rata lama sekolah. Perubahan tersebut terjadi dikarenakan AMH yang kurang dari 0,85 untuk data yang sudah distandarkan. Jika nilai AMH kurang dari data tersebut maka nilai BF1 tidak bermakna atau bernilai 0.

$$BF2 = \max(0, x_1 - 71.45)$$

Artinya setiap kenaikan BF2 akan menambah nilai rata-rata lama sekolah sebesar 0,206 untuk data rata-rata lama sekolah yang sudah distandarkan. Jika menggunakan data asli sebesar 0,206 dikali standar deviasi dari data asli rata-rata lama sekolah. Perubahan tersebut terjadi dikarenakan IPM yang kurang dari -71,45 untuk data yang sudah distandarkan. Jika nilai IPM kurang dari data tersebut maka nilai BF2 tidak bermakna atau bernilai 0.

Tabel 3 Hasil Seleksi Model MARS Menggunakan CV dengan BF = 12, dan Kombinasi Nilai MI dan MO

No	BF	MI	MO	CV
1	12	1	0	0.096
2	12	2	1	0.087
3	12	3	2	0.072
4	12	1	3	0.082
5	12	2	0	0.178
6	12	3	1	0.173
7	12	1	2	0.081
8	12	2	3	0,084
9	12	3	0	0.194
10	12	1	1	0.180
11	12	2	2	0.080
12	12	3	3	0.200

Berdasarkan Tabel 3, model terbaik yaitu pada BF = 12, MI = 3, MO = 2 memiliki nilai CV terkecil sebesar 0.072 dengan model sebagai berikut:

$$y = 7.656 - 0.115 \cdot BF1 - 0.366 \cdot BF2 + 0.496 \cdot BF3 + 0.205 \cdot BF4$$

dengan

$$BF1 = \max(0, x_3 - 85)$$

$$BF2 = \max(0, 85 - x_3)$$

$$BF3 = \max(0, x_1 - 71.45)$$

$$BF4 = \max(0, 71.45 - x_1)$$

Interpretasi model:

$$BF1 = \max(0, x_3 - 85)$$

Artinya setiap kenaikan BF1 akan mengurangi rata-rata lama sekolah sebesar 0,115 untuk data rata-rata lama sekolah yang sudah distandarkan. Jika menggunakan data asli sebesar 0,115 dikali standar deviasi dari data asli rata-rata lama sekolah. Perubahan tersebut menyebabkan AMH yang lebih dari 85 untuk data yang sudah distandarkan.. Jika nilai AMH kurang dari data tersebut maka nilai BF1 tidak bermakna atau bernilai 0.

$$BF2 = \max(0, 85-x3)$$

Artinya setiap kenaikan BF2 akan mengurangi nilai rata-rata lama sekolah sebesar 0,366 untuk data rata-rata lama sekolah yang sudah distandarkan. Jika menggunakan data asli sebesar 0,366 dikali standar deviasi dari data asli rata-rata lama sekolah. Perubahan tersebut menyebabkan AMH yang kurang dari 0,85 untuk data yang sudah distandarkan,. Jika nilai AMH kurang dari data tersebut maka nilai BF2 tidak bermakna atau bernilai 0.

$$BF3 = \max(0, x1-71.45)$$

Artinya setiap kenaikan BF3 akan menambah nilai rata-rata lama sekolah sebesar 0,496 untuk data rata-rata lama sekolah yang sudah distandarkan. Jika menggunakan data asli sebesar 0,496 dikali standar deviasi dari data asli rata-rata lama sekolah. Perubahan tersebut terjadi IPM yang lebih dari -71,45 untuk data yang sudah distandarkan,. Jika nilai IPM kurang dari data tersebut maka nilai BF3 tidak bermakna atau bernilai 0

$$BF4 = \max(0, 71.45-x1)$$

Artinya setiap kenaikan BF4 akan mengurangi nilai rata-rata lama sekolah sebesar 0,204 untuk data rata-rata lama sekolah yang sudah distandarkan. Jika menggunakan data asli sebesar 0,204 dikali standar deviasi dari data asli rata-rata lama sekolah. Perubahan tersebut terjadi IPM yang kurang dari 71,45 untuk data yang sudah distandarkan, Jika nilai IPM kurang dari data tersebut maka nilai BF4 tidak bermakna atau bernilai 0.

#### Perbandingan Nilai RMSE MARS dengan GCV dan CV

Nilai RMSE digunakan untuk melihat tingkat ketepatan peramalan. Model dikatakan baik jika nilai RMSE kecil dan memiliki R<sup>2</sup> yang besar berikut dijabarkan pada Tabel 4.

Tabel 4 Nilai RMSE dan R<sup>2</sup> MARS GCV dan CV

Model	RMSE	R <sup>2</sup>
MARS GCV	0.1910	0,8892
MARS CV	0.1492	0,9324

Berdasarkan tabel 4, nilai RMSE dan R<sup>2</sup> model MARS CV baik lebih baik dibandingkan dengan GCV karena memiliki nilai RMSE yang paling rendah dan R<sup>2</sup> yang lebih besar. Hal tersebut bisa menunjukkan bahwa model MARS dengan CV adalah model yang lebih baik untuk memodelkan rata-rata lama sekolah di Kabupaten Gianyar.

#### 4. KESIMPULAN

Nilai RMSE dan R<sup>2</sup> model MARS CV baik lebih baik dibandingkan dengan GCV karena memiliki nilai RMSE yang paling rendah dan R<sup>2</sup> yang lebih besar. Hal tersebut bisa menunjukkan bahwa model MARS dengan CV adalah model yang lebih baik untuk memodelkan rata-rata lama sekolah di Kabupaten Gianyar. Model yang diperoleh dari metode CV sebagai berikut:

$$y = 7.656 - 0.115*BF1 - 0.366*BF2 + 0.496*BF3 + 0.205*BF4$$

dengan

$$BF1 = \max(0, x3-85)$$

$$BF2 = \max(0, 85-x3)$$

$$BF3 = \max(0, x1-71.45)$$

$$BF4 = \max(0, 71.45-x1)$$

#### DAFTAR PUSTAKA



- Andrews, Y. Ng., 1991, *Preventing “Overfitting” of Cross-Validation Data*, School of Computer Science, Carnegie Mellon University, Pittsburgh PA.
- BPS, 2014, *Indeks Pembangunan Manusia 2014*, Badan Pusat Statistik, Jakarta.
- BPS, 2018, *Indikator Pendidikan 2018*, Badan Pusat Statistik, Jakarta.
- BPS, 2020, *Statistik Pendidikan 2020*, Badan Pusat Statistik, Jakarta
- BPS Provinsi Aceh, 2015, *Indikator Pendidikan Provinsi Aceh 2014*, BPS Aceh, Aceh
- Craven, P. and Wahba, G., 1979, Smoothing Noisy Data with Spline Functions. Estimating The Correct Degree of Smoothing by The Method of Generalized Cross Validation, *Numer. Math.* (31) 317-403.
- Darmawanti, N. D., Suparti., dan Safitri, D., 2014, Pemodelan Indeks Harga Saham Gabungan (IHSG) Menggunakan *Multivariate Adaptive Regression Splines* (Mars), *Jurnal Gaussian*, Vol 3, No 4, Hal 771 – 780
- Eubank, R.L., 1999, *Nonparametric Regression and Spline Smoothing. Second Edition*, Marcel Dekker, Inc, New York.
- Friedman, J. H., 1991, *Multivariate Adaptive Regression Splines*, The Annals of Statistics, Vol. 19, No. 1 (Mar., 1991), pp. 1-67, Institute of Mathematical Statistics.
- Friedman, J.H. dan Silverman, B.W., 1989, Flexible parsimony smoothing and additive modeling, *Technometrics*, **31**, 3 – 39.
- Nisa', S. F., dan Budiantara, I. N., 2006, Analisis Survival dengan Pendekatan Multivariate Adaptive Regression Splines pada Kasus Demam Berdarah Dengue (DBD). *Jurnal Sains dan Seni ITS*, Vol. 1, No. 1, (Sept.2012). Surabaya: Institut Teknologi Sepuluh November.
- Razak, R. A., 2019, Perbandingan Metode Generalized Cross Validation (GCV) Dan Cross Validation (CV) Dalam Pemilihan Bandwidth Optimal Pada Pemodelan Regresi Nonparametrik Kernel Gaussian (Studi Kasus : Gizi Buruk Pada Balita Di Indonesia), *Skripsi*, Program Studi Statistika. Universitas Muhammadiyah Semarang.
- Utami, A'. N., 2021, Pemodelan Multivariate Adaptive Regression Spline (MARS) (Studi Kasus: Kasus Diare Pada Balita di Provinsi Jawa Barat-Jawa Tengah Tahun 2019), *Skripsi*, Program Studi Matematika Fakultas Sains Dan Teknologi Uin Sunan Kalijaga Yogyakarta
- Wahba, G., 1990, *Spline Models for Observational Data*, SIAM, Pennsylvania.