MATHunesa
Jurnal Ilmiah Matematika
e-ISSN : 2716-506X | p-ISSN : 2301-9115

Volume 13 *No* 02
Tahun 2025

ANALISIS FAKTOR-FAKTOR YANG MEMENGARUHI DIABETES MELITUS DI JAWA BARAT MENGGUNAKAN *MULTIVARIATE ADAPTIVE REGRESSION SPLINES* (MARS)

Vera Maya Santi
Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta
e-mail: vmsanti@unj.ac.id*

Muhammad Rafi Azwar
Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta
e-mail: rafi.azwar16@gmail.com

Widyanti Rahayu
Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta
e-mail: wrahayu@unj.ac.id

Abstrak

Diabetes melitus adalah penyakit metabolik yang ditandai dengan kadar gula darah yang tinggi secara kronis akibat gangguan pada sekresi atau kerja insulin. Jumlah penderita diabetes melitus terus meningkat secara global, termasuk di Indonesia. Jawa Barat menempati posisi ke-dua sebagai provinsi dengan jumlah penderita diabetes melitus berdasarkan diagnosis dokter terbanyak di Indonesia. Jumlah penderita diabetes melitus pada tahun 2022 di provinsi ini menjadi perhatian khusus karena menjadi titik terendah, mengindikasikan adanya faktor-faktor yang secara signifikan memengaruhi penurunan jumlah penderita. Oleh karena itu, penting untuk mengetahui faktor-faktor apa saja yang memengaruhi jumlah penderita diabetes melitus di Jawa Barat pada tahun 2022. Akan tetapi, data riil jumlah penderitas diabetes melitus tersebut tidak menunjukkan kecenderungan pola data tertentu. Sehingga, salah satu pendekatan yang dapat digunakan untuk menganalisis faktor-faktor yang memengaruhi diabetes melitus di Jawa Barat adalah regresi nonparametrik yaitu *Multivariate Adaptive Regression Splines* (MARS). Metode ini menggunakan pendekatan yang fleksibel dengan membagi data menjadi segmen-segmen dan menerapkan regresi linier pada setiap segmen tersebut. Kriteria pemilihan model yang digunakan pada penelitian ini adalah *Generalized Cross Validation* (GCV) dan *Akaike Information Criterion* (AIC). Berdasarkan kriteria pemilihan model GCV diperoleh tujuh variabel yang signifikan, sedangkan dengan kriteria AIC diperoleh delapan variabel yang signifikan memengaruhi jumlah penderita diabetes melitus di Provinsi Jawa Barat. Perbedaan struktur model juga terlihat dari jumlah fungsi basis yang digunakan. Model GCV memiliki 13 fungsi basis, sementara model AIC memiliki 14 fungsi basis. Berdasarkan nilai R^2, model GCV memiliki nilai R^2 sebesar 0,994, sedangkan model AIC memiliki nilai R^2 sebesar 0,995.
Kata Kunci: AIC, diabetes melitus, GCV, *Multivariate Adaptive Regression Splines* (MARS), regresi nonparametrik.

Abstract

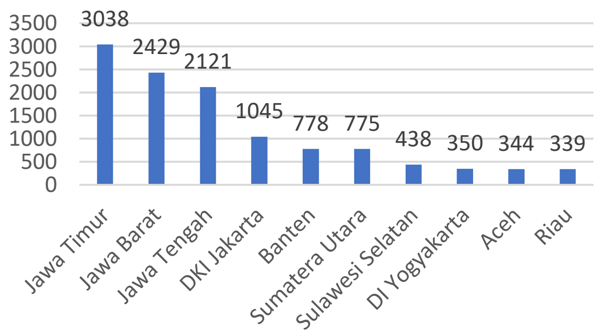
Diabetes mellitus is a metabolic disorder characterized by chronically elevated blood sugar levels due to impaired insulin secretion or action. West Java ranks second among Indonesian provinces with the highest number of diabetes mellitus cases based on medical diagnoses. In 2022, the number of diabetes mellitus cases in this province reached its lowest point in the past five years, suggesting the influence of significant factors contributing to this decline. Therefore, identifying the factors affecting diabetes mellitus prevalence in West Java in 2022 is essential. However, the data exhibit no clear pattern, necessitating a nonparametric regression approach for modeling these factors. This study employs Multivariate Adaptive Regression Splines (MARS), a flexible method that partitions data into segments and applies linear regression within each segment. Model selection criteria include Generalized Cross Validation (GCV) and Akaike Information Criterion (AIC). Based on GCV, seven significant variables were identified, whereas AIC indicated eight significant variables influencing diabetes mellitus prevalence in West Java. Structural differences between models are also observed in the number of basis functions: the GCV model utilizes 13 basis functions, while the AIC model employs 14. In terms of model performance, the GCV model achieves an R^2 *value of 0.994, whereas the AIC model attains an* R^2 *value of 0.995.*

Keywords: *AIC, diabetes mellitus, GCV, Multivariate Adaptive Regression Splines (MARS), nonparametric regression.*

PENDAHULUAN

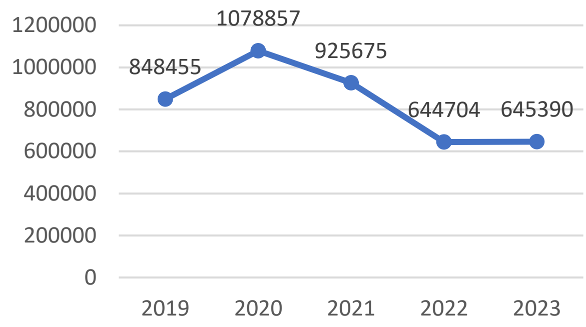
Diabetes merupakan salah satu masalah kesehatan global yang terus meningkat dari tahun ke tahun. Kondisi ini terjadi akibat tubuh tidak mampu memproduksi insulin atau tidak dapat menggunakannya secara efektif, yang menyebabkan kadar gula darah tinggi. Jika tidak dikelola dengan baik, diabetes dapat menyebabkan komplikasi serius pada berbagai organ tubuh, seperti jantung, ginjal, saraf, dan mata (International Diabetes Federation, 2017). Diabetes terbagi menjadi dua kategori utama, yaitu diabetes insipidus dan diabetes melitus. Diabetes insipidus terjadi akibat gangguan pada hormon antidiuretik (ADH) yang menyebabkan peningkatan produksi urin dan rasa haus yang berlebihan (Kemenkes, 2022). Sementara itu, diabetes melitus adalah penyakit metabolik kronis yang disebabkan oleh gangguan produksi atau kerja insulin, yang mengakibatkan kadar gula darah tinggi secara persisten (World Health Organization, 1999).

Jumlah penderita diabetes melitus terus meningkat secara signifikan di seluruh dunia, termasuk di Indonesia. Berdasarkan data dari Kementerian Kesehatan (2022), jumlah penderita diabetes melitus di Indonesia meningkat dari 10,7 juta jiwa pada tahun 2019 menjadi 19,47 juta orang pada tahun 2021. International Diabetes Federation (2021) memperkirakan jumlah penderita diabetes di Indonesia akan mencapai 28,57 juta jiwa pada tahun 2045, yang menunjukkan peningkatan sebesar 47% dibandingkan tahun 2021.



Gambar 1. Jumlah Penderita Diabetes Melitus Berdasarkan Diagnosis Dokter (2018)

Berdasarkan data Riset Kesehatan Dasar (2018), Jawa Barat menempati posisi ke-dua sebagai provinsi dengan jumlah penderita diabetes melitus berdasarkan diagnosis dokter terbanyak di Indonesia, dengan jumlah penderita mencapai 2429 jiwa. Diagnosis dokter pada data ini mencakup individu yang telah mendapatkan konfirmasi dari tenaga medis profesional bahwa mereka menderita Diabetes Melitus, baik melalui pemeriksaan langsung maupun berdasarkan riwayat pengobatan terkait diabetes melitus.



Gambar 2. Jumlah Penderita Diabetes Melitus di Jawa Barat (2019-2023)

Berdasarkan data dari Dinas Kesehatan Jawa Barat, jumlah penderita diabetes melitus di provinsi ini menunjukkan fluktuasi antara tahun 2019 hingga 2023. Gambar 2 juga memperlihatkan adanya tren penurunan yang terjadi pada tahun 2020 hingga 2022. Fenomena ini menarik untuk diteliti guna memahami faktor-faktor yang memengaruhi jumlah penderita diabetes melitus di Jawa Barat, khususnya pada tahun 2022, karena penderita diabetes melitus pada tahun tersebut merupakan titik terendah.

Beberapa penelitian terdahulu telah mengidentifikasi faktor-faktor yang memengaruhi diabetes melitus dan penerapan metode MARS dalam pemodelan data kesehatan. Dewi dkk. (2020) menggunakan regresi Cox untuk menganalisis faktor-faktor yang memengaruhi diabetes melitus di salah satu RSUD di Jawa Tengah dan menemukan bahwa usia serta kadar gula darah berpengaruh signifikan. Selain itu, penelitian yang dilakukan oleh Azizah dan Permatasari (2020) menggunakan MARS untuk memodelkan faktor-faktor yang memengaruhi *stunting* pada balita di provinsi NTT, didapatkan 2 dari 6 variabel prediktor yang berpengaruh secara signifikan, dengan variabel yang paling berpengaruh yaitu persentase ibu hamil yang berisiko kekurangan energi kronis. Fabregas (2020) memodelkan faktor-faktor risiko kematian pada kanker usus besar stadium 2 di Amerika Serikat menggunakan MARS, didapatkan 5 variabel berpengaruh dari 6 variabel yang digunakan dengan variabel yang paling berpengaruh adalah jumlah kelenjar getah bening.

Pendekatan yang biasa digunakan untuk memodelkan dan mengetahui faktor-faktor yang memengaruhi suatu variabel adalah pendekatan parametrik. Pendekatan ini mengasumsikan hubungan antara variabel respon dengan variabel prediktor harus mengikuti sebaran tertentu dan bentuk kurva regresi mengikuti beberapa pola tertentu, seperti linier, kuadratik, atau kubik (Hardle, 1994). Selain itu, pendekatan parametrik juga memerlukan asumsi tertentu, seperti normalitas, linieritas, multikolinieritas, dan homoskedastisitas (Montgomery dkk, 2012). Apabila terdapat asumsi yang tidak terpenuhi, maka pendekatan yang digunakan dalam menganalisis data adalah pendekatan nonparametrik untuk mengestimasi kurva regresi (Eubank, 1999).

Pendekatan nonparametrik mempunyai kelebihan dimana dalam mengestimasi bentuk kurva regresinya mempunyai fleksibilitas yang tinggi sehingga dapat menyesuaikan dengan data yang tidak diketahui bentuk kurva regresinya (Eubank, 1999; Santi dkk., 2022). Salah satu metode nonparametrik yang efektif dalam menganalisis hubungan kompleks antar variabel adalah *Multivariate Adaptive Regression Splines* (MARS). Metode ini merupakan pengembangan dari *Recursive Partitioning Regression* (RPR) yang dikombinasikan dengan pendekatan *spline* sehingga menghasilkan model yang lebih fleksibel dalam menangani data berdimensi tinggi dengan hubungan nonlinear (Friedman, 1991). MARS membagi data menjadi beberapa segmen berdasarkan titik knot yang dipilih secara otomatis dan menerapkan regresi linier pada setiap segmen, sehingga mampu mengidentifikasi interaksi antar variabel serta pola hubungan nonlinear yang kompleks (Hastie dkk, 2008).

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk memodelkan faktor-faktor yang memengaruhi jumlah penderita diabetes melitus di Jawa Barat menggunakan metode MARS dan mengetahui faktor-faktor apa saja yang memengaruhi diabetes melitus pada Kabupaten/Kota di Jawa Barat. Metode tersebut digunakan karena plot data masing-masing variabel prediktor terhadap variabel respon tidak menunjukkan kecenderungan pola data tertentu atau pola yang tidak jelas, maka akan digunakan pendekatan regresi nonparametrik. Selain itu, data yang digunakan memiliki sebelas variabel prediktor sehingga dapat dikatakan data berdimensi tinggi.

KAJIAN TEORI

MODEL REGRESI BERGANDA

Analisis regresi adalah teknik statistik untuk menyelidiki dan memodelkan hubungan antara variabel respon dan variabel prediktor. Bila dalam analisisnya hanya melibatkan sebuah variabel prediktor, maka analisis yang digunakan adalah analisis regresi linear sederhana. Sementara, apabila analisisnya melibatkan lebih dari satu variabel prediktor, maka analisis yang digunakan adalah analisis regresi linear berganda (Pasaribu dkk, 2015). Berikut merupakan bentuk umum persamaan regresi linier berganda.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i \qquad (1)$$

dengan y_i adalah nilai variabel respon untuk amatan ke-i, β_k parameter regresi yang bersesuaian dengan variabel prediktor x_k, x_{ik} nilai variabel prediktor ke-k untuk amatan ke-i, ε_i galat untuk amatan ke-i, dan k variabel prediktor.

REGRESI NONPARAMETRIK

Regresi nonparametrik merupakan suatu metode yang digunakan untuk mengetahui pola hubungan antara variabel prediktor dan respon yang tidak diketahui kurva regresinya atau tidak terdapat informasi masa lalu yang lengkap tentang bentuk pola data (Eubank, 1988). Secara umum model regresi nonparametrik dapat dituliskan sebagai berikut.

$$y_i = f(x_i) + \varepsilon_i \qquad (2)$$

dengan y_i = nilai variabel respon untuk amatan ke-i, $f(x_i)$ = fungsi dari kurva regresi yang bentuknya tidak diketahui, dan ε_i = galat untuk amatan ke-i.

REGRESI *SPLINE*

Regresi *spline* adalah salah satu metode regresi nonparametrik yang membagi data menjadi beberapa segmen dan menerapkan fungsi regresi yang berbeda pada setiap segmen. Fungsi *spline*, yang biasanya berupa polinomial digunakan untuk menghaluskan perubahan antar segmen sehingga model lebih fleksibel dalam menangkap pola nonlinier pada data. Pada titik-titik tertentu yang disebut knot, fungsi *spline* disambung sehingga tercipta kurva yang halus (James dkk, 2013). *Spline* merupakan polinomial orde q yang terdiri dari beberapa potongan (*piecewise*) dan memiliki turunan hingga orde $(q-1)$ pada titik knot (Friedman, 1991). Dalam *spline* dengan orde q dan k knot memiliki fungsi basis sebagai berikut.

$$1, x, \ldots, x^q, (x-t_1)_+^q, (x-t_2)_+^q, (x-t_k)_+^q \quad (3)$$

maka model *spline* dapat ditulis menjadi.

$$f(x_i) = \beta_0 + \sum_{i=1}^{q} \beta_i x_i + \sum_{k=1}^{k} \beta_{q+k} (x_i - t_k)_+^q \quad (4)$$

dengan β_0 = intersep, β_i = koefisien regresi yang bersesuiaian dengan variabel prediktor ke-i, β_{q+k} = koefisien regresi pada variabel prediktor x_i dengan *truncated* knot ke-$q+k$, x_i = variabel Prediktor ke-i, t_k = titik knot ke-k pada variabel prediktor, k = banyaknya knot, dan q = orde polinomial. Dengan fungsi truncated sebagai berikut.

$$(x-t_k)_+^q = \begin{cases} 0; x_i < t_k \\ (x_i - t_k)^q; x_i \geq t_k \end{cases} \quad (5)$$

RECURSIVE PARTITIONING REGRESSION (RPR)

Recursive Partitioning Regression (RPR) adalah salah satu metode komputasi yang bertujuan utama untuk mengestimasi data pada *subregion* dan parameter yang terkait dengan masing-masing *subregion* (Friedman, 1991). Fungsi basis dalam RPR memiliki bentuk sebagai berikut.

$$B_m(x) = \prod_{k=1}^{k_m} H\left[S_{km}.\left(x_{v(k,m)} - t_{km}\right)\right] \quad (6)$$

dengan B_m = fungsi Basis ke-m pada variabel prediktor, K_m = derajat interaksi pada fungsi basis ke-m, S_{km} = tanda pada titik knot (± 1), $x_{v(k,m)}$ = variabel prediktor, dan t_{km} = nilai knot dari variabel prediktor.

MULTIVARIATE ADAPTIVE REGRESSION SPLINES (MARS)

Multivariate Adaptive Regression Splines (MARS) merupakan metode regresi nonparametrik yang dirancang untuk mengatasi kelemahan regresi *spline*, dengan cara memilih titik knot secara otomatis berdasarkan data (adaptif) dan membagi data menjadi beberapa segmen. Model MARS dirancang untuk menangani data berdimensi tinggi, yaitu data dengan lebih dari tiga variabel prediktor (Hastie dkk, 2008). Menurut Friedman (1991) terdapat beberapa aspek yang harus diperhatikan dalam pemodelan menggunakan metode MARS, yaitu:

1. Knot. Knot merupakan titik yang menandai awal dari satu garis regresi dan akhir dari garis regresi lainnya. Titik ini merupakan tempat perpotongan fungsi, yang dapat diartikan sebagai titik yang mengindikasikan perubahan pola atau perilaku data dalam interval tertentu. Minimum observasi antar knot (MO) sebesar 0,1,2,3, ... sampai maksimum jumlah sampel.
2. Fungsi Basis atau *basis function* (BF). Fungsi basis adalah fungsi yang digunakan untuk menggambarkan hubungan antara variabel respon dan variabel prediktor, dengan tujuan menyajikan informasi yang melibatkan satu atau lebih variabel. Pada umumnya nilai fungsi basis yang digunakan berbentuk polinomial dengan turunan yang kontinu di setiap knot. Jumlah yang optimal untuk nilai maksimum fungsi basis adalah 2 hingga 4 kali jumlah variabel prediktor.
3. Interaksi. Interaksi merupakan hasil perkalian silang antara variabel-variabel prediktor yang saling berkorelasi. Jumlah maksimum interaksi (MI) yang diperbolehkan adalah 1, 2, atau 3. Jika jumlah interaksi melebihi 3, model akan menjadi semakin kompleks dan sulit untuk diinterpretasikan.

Model *Multivariate Adaptive Regression Splines* (MARS) adalah sebagai berikut.

$$\hat{f}(x) = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{k_m} \left[S_{km}.\left(x_{v(k,m)} - t_{km}\right)\right]_+ \quad (7)$$

Dengan a_0 = koefisien konstanta dari fungsi basis induk, a_m = koefisien dari fungsi basis ke-m, k_m = derajat interaksi pada fungsi basis ke-m, S_{km} = nilainya ± 1, $x_{v(k,m)}$ = variabel prediktor t_{km} = nilai knot dari variabel prediktor, v = banyaknya variabel prediktor, m = banyaknya fungsi basis, dan k = banyaknya interaksi.

GENERALIZED CROSS VALIDATION (GCV)

Pada metode MARS kriteria pemilihan model yang paling baik untuk seleksi model tebaik adalah *Generalized Cross Validation* (GCV). Model yang terbaik dapat ditentukan dengan melihat nilai GCV yang paling rendah (Friedman, 1991). Nilai GCV didefinisikan sebagai berikut.

$$GCV = \frac{\frac{1}{n}\sum_{i=1}^{n}\left[y_i - \hat{f}_m(x_i)\right]^2}{\left[1 - \frac{C(\widehat{M})}{n}\right]^2} \quad (8)$$

dengan y_i = variabel respon ke-i, $\hat{f}_m(x_i)$ = nilai taksiran variabel respon pada M fungsi basis, n = banyaknya data, $C(\widehat{M}) = C(M) + dM$, $C(M)$ = trace $[\boldsymbol{B}(\boldsymbol{B^T B})^{-1}] + 1$, dan d = nilai ketika setiap fungsi basis mencapai optimasi $(2 \leq d \leq 4)$.

Dalam menentukan model terbaik, jika terdapat lebih dari satu model yang memiliki nilai GCV terkecil, maka langkah selanjutnya adalah memeriksa nilai R^2 yang paling besar. Jika R^2 juga memiliki nilai yang sama, maka dapat menganut prinsip parsimoni model dengan memilih model yang memiliki nilai kombinasi fungsi basis (BF), maksimum interaksi (MI), dan minimum observasi antar knot (MO) terkecil (Azizah dan Permatasari, 2020).

AKAIKE INFORMATION CRITERION (AIC)

Akaike Information Criterion (AIC) adalah metode evaluasi model statistik yang mempertimbangkan keseimbangan antara kesesuaian model terhadap data dan kesederhanaan model. AIC menilai sebuah model berdasarkan seberapa dekat nilai yang cocok dengan nilai yang sebenarnya, sebagaimana dirangkum oleh jarak tertentu yang diharapkan di antara keduanya (Agresti, 2019). Persamaan AIC dituliskan sebagai berikut.

$$AIC = -2\ln\hat{L} + 2P \tag{9}$$

dengan $\hat{L}$ = nilai maksimum fungsi likelihood dan P = banyaknya variabel prediktor.

Model yang paling optimal adalah yang memiliki nilai AIC terendah. AIC dipilih karena memberikan penalti terhadap model yang memiliki jumlah parameter yang lebih besar, mencegah model menjadi terlalu kompleks tanpa mengorbankan keakuratan prediksi. Hal ini sesuai dengan prinsip seleksi model untuk menemukan model sederhana yang tetap informatif.

PENGUJIAN SIGNIFIKANSI MODEL MARS

Apabila model MARS terbaik telah ditemukan, langkah selanjutnya adalah melakukan pengujian untuk menilai signifikansi parameter guna mengevaluasi kecocokan model. Pengujian ini dilakukan dengan menguji koefisien regresi baik secara serentak maupun parsial.

1. Pengujian koefisien regresi serentak
   a. Rumusan hipotesis:
      H_0 : $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$ (model tidak signifikan)
      H_1 : minimal terdapat satu $\alpha_m \neq 0; m = 1,2,\ldots,M$ (model signifikan)
   b. Taraf signifikansi: α
   c. Statistik uji:
      $$F_{hitung} = \frac{\sum_{i=1}^{n}\frac{(\hat{y}_i-\bar{y})^2}{M}}{\sum_{i=1}^{n}\frac{(\hat{y}_i-\bar{y})^2}{N-M-1}} \tag{10}$$
      dengan N = banyaknya amatan, M = banyaknya variabel prediktor, $\hat{y}_i$ = nilai y ke-i yang diestimasi, y_i = nilai y ke-i yang diamati, dan $\bar{y}$ = nilai rata-rata y.
   d. Pengambilan keputusan:
      Tolak H_0 jika nilai $F_{hitung} > F_{\alpha(M;N-M-1)}$ atau $P-value < \alpha$.
2. Pengujian koefisien regresi parsial
   a. Rumusan hipotesis:
      H_0 : $\alpha_m = 0$ (koefisien α_m tidak berpengaruh terhadap model)
      H_1 : $\alpha_m \neq 0;$ untuk setiap m, $m = 1,2,\ldots,M$ (koefisien α_m berpengaruh terhadap model)
   b. Taraf signifikansi: α
   c. Statistik uji:
      $$t_{hitung} = \frac{\hat{a}_m}{Se(\hat{a}_m)} = \frac{\hat{a}_m}{\sqrt{var(\hat{a}_m)}} \tag{11}$$
      dengan $\hat{a}_m$ = nilai koefisien a_m dan $Se(\hat{a}_m)$ = standar error dari a_m.
   d. Pengambilan keputusan:
      Tolak H_0 jika nilai $|t\_hitung| > t_{\left(\frac{\alpha}{2},N-M\right)}$ atau $P-value < \alpha$.

METODE

Data pada penelitian ini merupakan data sekunder yang diperoleh dari situs Open Data Jabar dan Badan Pusat Statistik. Data yang digunakan berasal dari 27 Kabupaten/Kota di Jawa Barat yang merupakan data tahun 2022. Tahapan-tahapan yang dilakukan untuk menganalisis data adalah sebagai berikut:

1. Melakukan eksplorasi data untuk memperoleh gambaran secara umum mengenai karakteristik data.
2. Melakukan visualisasi data dengan membuat *scatterplot* variabel respon dengan variabel prediktor.
3. Menentukan jumlah maksimum fungsi basis (BF), maksimum fungsi basis yang diperbolehkan yaitu sebanyak dua sampai empat kali dari jumlah variabel prediktor yang digunakan. Dalam penelitian ini variabel prediktor yang digunakan sebanyak 11 variabel sehingga maksimum jumlah BF yang dapat digunakan adalah 22, 33, dan 44.
4. Menentukan batas jumlah maksimum interaksi, dalam penelitian ini jumlah maksimum interaksi (MI) adalah 1, 2, dan 3. MI bergantung pada

kemampuan model untuk menangkap hubungan antar variabel tanpa overfitting. Jika terdapat lebih dari 3 interaksi, maka interpretasi model menjadi terlalu kompleks.

5. Menentukan minimum observasi antar knot (MO), yaitu 0, 1, 2, 3, 4, dan 5. MO = 0 menunjukkan jarak antar titik knot adalah 0, MO = 1 berarti jarak antar titik knot adalah 1, dan seterusnya.
6. Melakukan estimasi parameter model MARS menggunakan *Ordinary Least Square* (OLS).
7. Menentukan model MARS berdasarkan kombinasi antara BF, MI, dan MO yang menghasilkan nilai *Generalized Cross Validation* (GCV) terkecil.
8. Menentukan model MARS berdasarkan kombinasi antara BF, MI, dan MO yang menghasilkan nilai *Akaike Information Criterion* (AIC) terkecil.
9. Menguji signifikansi model MARS melalui pengujian koefisien regresi, baik secara serentak menggunakan Uji F maupun secara parsial menggunakan Uji t.
10. Melakukan interpretasi model terbaik dan menginterpretasikan variabel-variabel yang berpengaruh dalam model tersebut.

HASIL DAN PEMBAHASAN

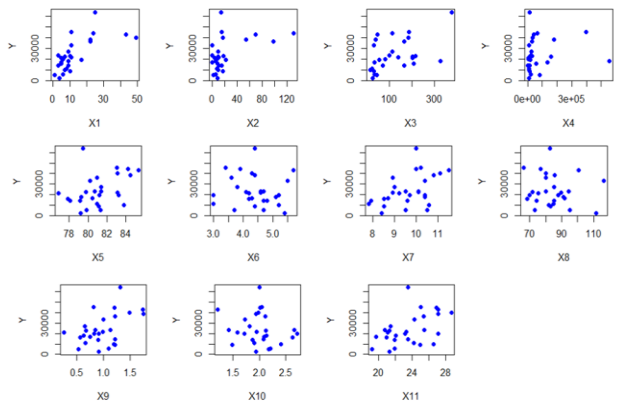
DESKRIPSI DATA

Deskripsi data bertujuan untuk memaparkan atau menggambarkan data secara jelas agar informasi yang disajikan lebih mudah dipahami oleh pembaca. Data yang akan dijelaskan terdiri dari satu variabel respon, yaitu jumlah penderita diabetes melitus (Y), dan sebelas variabel prediktor, yaitu: jumlah rumah sakit umum (X_1), jumlah klinik (X_2), jumlah puskesmas (X_3), jumlah penderita hipertensi (X_4), indeks kesehatan (X_5), tingkat konsumsi gula (X_6), tingkat konsumsi minyak dan lemak (X_7), tingkat konsumsi sayur dan buah-buahan (X_8), tingkat konsumsi teh (X_9), tingkat konsumsi kopi (X_{10}), dan persentase perokok (X_{11}).

Tabel 1. Statistik Deskriptif Variabel Respon dan Prediktor

Variabel	Min	Max	Range	Median	Mean
Y	2.377	63.978	61.601	21.129	23.877,93
X_1	1	49	48	9	12,67
X_2	0	131	131	9	20,85
X_3	15	377	362	101	120,11
X_4	2.914	556.178	553.264	22.026	72.459,89
X_5	76,85	85,35	8,5	80,97	81,11
X_6	3	5,7	2,7	4,4	4,4
X_7	7,8	11,5	3,7	9,7	9,64
X_8	66,4	116,3	49,9	83,2	84,56
X_9	0,26	1,74	1,48	0,91	0,98
X_{10}	1,22	2,72	1,5	2,01	2
X_{11}	19,21	28,68	9,47	23,49	23,62

Berdasarkan Tabel 1, jumlah penderita diabetes melitus (Y) tertinggi sebesar 63.978 kasus yang terdapat di Kabupaten Bogor dan terendah sebesar 2.377 kasus yang terdapat di Kota Banjar. Jumlah rumah sakit umum (X_1) tertinggi terdapat di Kabupaten Bekasi dan terendah di Kabupaten Pangandaran. Jumlah klinik (X_2) tertinggi terdapat di Kota Bandung, sementara terdapat kabupaten tanpa klinik sama sekali yaitu Kabupaten Ciamis, Sumedang, dan Indramayu. Jumlah puskesmas (X_3) tertinggi terdapat di Kabupaten Bogor dan terendah di Kota Cimahi. Jumlah penderita hipertensi (X_4) tertinggi terdapat di Kabupaten Sukabumi dan terendah di Kabupaten Purwakarta. Indeks kesehatan (X_5) tertinggi terdapat di Kota Bekasi dan terendah di Kabupaten Tasikmalaya. Tingkat konsumsi gula (X_6) tertinggi terdapat di Kota Bekasi dan terendah di Kota Bogor dan Kota Tasikmalaya. Tingkat konsumsi minyak dan lemak (X_7) tertinggi terdapat di Kota Bekasi dan terendah di Kota Tasikmalaya. Tingkat konsumsi sayur dan buah-buahan (X_8) tertinggi terdapat di Kabupaten Indramayu dan terendah di Kabupaten Bandung. Tingkat konsumsi teh (X_9) tertinggi terdapat di Kota Depok dan terendah di Kabupaten Tasikmalaya. Tingkat konsumsi kopi (X_{10}) tertinggi terdapat di Kabupaten Bandung Barat dan terendah di Kota Bekasi. Sementara itu, persentase perokok (X_{11}) tertinggi terdapat di Kabupaten Bekasi dan terendah di Kabupaten Pangandaran.



Gambar 3. *scatterplot* antara masing-masing variabel prediktor dengan variabel respon

Langkah berikutnya adalah melihat hubungan antara variabel respon dengan variabel prediktor yang dilakukan dengan mengamati pola sebaran data menggunakan *scatterplot*. Berdasarkan Gambar 3 dapat dilihat bahwa hubungan antara semua variabel prediktor dengan variabel respon tidak memiliki pola tertentu, seperti linier, kuadratik, atau kubik sehingga diasumsikan bahwa semua variabel prediktor sudah sesuai untuk dilakukan pemodelan dengan pendekatan nonparametrik.

SPESIFIKASI MODEL MARS

Model MARS terbaik ditentukan berdasarkan nilai GCV dan AIC terkecil. Nilai minimum ini diperoleh melalui proses *trial and error* dengan mengombinasikan jumlah fungsi basis (BF), tingkat interaksi maksimum (MI), dan jumlah observasi minimum (MO). Variasi nilai BF yang digunakan adalah dua hingga empat kali variabel prediktor yaitu 22, 33, dan 44, nilai MI yang digunakan sebesar 1, 2, dan 3, dan nilai MO yang digunakan sebesar 0, 1, 2, 3. Hasil perhitungan nilai GCV dan AIC untuk semua kombinasi model disajikan berikut ini.

Tabel 2. Hasil seleksi Model MARS Berdasarkan GCV

Model	BF	MI	MO	GCV	R^2
22	33	3	1	20.054	0,994

Model terbaik yang terpilih merupakan model yang terbentuk dari kombinasi BF=33, MI=3, dan MO=1 dengan nilai GCV sebesar 20.054.

Tabel 3. Hasil seleksi Model MARS Berdasarkan AIC

Model	BF	MI	MO	AIC	R^2
21	33	3	0	476,253	0,995

Model terbaik yang terpilih merupakan model yang terbentuk dari kombinasi BF=33, MI=3, dan MO=0 dengan nilai AIC sebesar 476,253.

MODEL MARS TERBAIK BERDASARKAN GCV

Model MARS terbaik berdasarkan nilai GCV adalah model yang tebentuk dari kombinasi kombinasi BF=33, MI=3, dan MO=1 dengan nilai GCV sebesar 20.054 . Langkah selanjutnya yaitu melakukan estimasi parameter model menggunakan metode *Ordinary Least Square* (OLS).

Tabel 4. Hasil Estimasi Parameter Model Berdasarkan GCV

Parameter	Estimasi	$p-value$	S.E.
a_0	26.854,240	0,0000*	1,468
a_1	18.831,127	0,0000*	5,983
a_2	8.495,916	0,0000*	1,956
a_3	−186,746	0,0000*	2,185
a_4	49.918,146	0,0000*	1,372
a_5	7.844,747	0,0000*	4,675
a_6	−53.398,833	0,0012*	1,911
a_7	−218,202	0,0000*	3,837
a_8	10.718,658	0,0000*	1,498
a_9	0,184	0,0025*	4,944
a_{10}	−2.329,621	0,0000*	2,069
a_{11}	−299,931	0,0002*	5,349
a_{12}	−2.682,950	0,0007*	2,328
a_{13}	−0,010	0,0221*	8,396

Berdasarkan hasil pada Tabel 4, maka dugaan model Model MARS terbaik dapat dituliskan sebagai berikut.

$$\begin{aligned}\hat{Y} = {} & 26.854{,}240 + 18.831{,}127 * \mathrm{BF}_1 + 8.495{,}916 \\ & * \mathrm{BF}_2 - 186{,}746 * \mathrm{BF}_3 + 49.918{,}146 * \mathrm{BF}_4 \\ & + 7.844{,}747 * \mathrm{BF}_5 - 53.398{,}833 * \mathrm{BF}_6 - 218{,}202 \\ & * \mathrm{BF}_7 + 10.718{,}658 * \mathrm{BF}_8 + 0{,}184 * \mathrm{BF}_9 \\ & - 2.329{,}621 * \mathrm{BF}_{10} - 299{,}931 * \mathrm{BF}_{11} - 2.682{,}950 \\ & * \mathrm{BF}_{12} - 0{,}010 * \mathrm{BF}_{13}\end{aligned}$$

dengan $\mathrm{BF}_1 = \max(0, 10 - X_1)$, $\mathrm{BF}_2 = \max(0, X_1 - 10)$, $\mathrm{BF}_3 = \max(0, 101 - X_3)$, $\mathrm{BF}_4 = \max(0, X_7 - 9{,}7)$, $\mathrm{BF}_5 = \max(0, 10{,}1 - X_7)$, $\mathrm{BF}_6 = \max(0, X_7 - 10{,}1)$, $\mathrm{BF}_7 = \max(0, 80{,}3 - X_8)$, $\mathrm{BF}_8 = \max(0, 22{,}03 - \mathrm{X}_{11})$, $\mathrm{BF}_9 = \max(0, 10 - X_1) * X_4$, $\mathrm{BF}_{10} = \max(0, 10 - X_1) * X_7$, $\mathrm{BF}_{11} = \max(0, X_1 - 10) * X_{11}$, $\mathrm{BF}_{12} = X_6 * \max(0, 10{,}1 - X_7)$, dan $\mathrm{BF}_{13} = \max(0, 10 - X_1) * X_4 * X_{11}$.

Interpretasi model MARS terbaik berdasarkan GCV adalah sebagi berikut.

1. Koefisien $\mathrm{BF}_1 = 18.831{,}127$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_1 akan meningkatkan jumlah penderita diabetes melitus sebesar 18.831,127 pada daerah dengan nilai baku jumlah rumah sakit kurang (X_1) dari 10. Hasil ini sejalan dengan penelitian yang dilakukkan oleh Mulyani dkk (2023) bahwa rumah sakit berhubungan dengan peningkatan deteksi dan pengelolaan diabetes melitus, serta menunjukkan bahwa akses ke layanan kesehatan yang memadai dapat membantu dalam mengurangi keparahan penyakit melalui pengelolaan yang lebih baik.
2. Koefisien $\mathrm{BF}_2 = 8.495{,}916$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_2 akan meningkatkan jumlah penderita diabetes melitus sebesar 8.495,916 pada daerah dengan nilai baku jumlah rumah sakit (X_1) lebih dari 10. Hasil ini sejalan dengan penelitian yang dilakukkan oleh Mulyani dkk (2023) bahwa rumah sakit berhubungan dengan peningkatan deteksi dan pengelolaan diabetes melitus, serta menunjukkan bahwa akses ke layanan kesehatan yang memadai dapat membantu dalam mengurangi keparahan penyakit melalui pengelolaan yang lebih baik.
3. Koefisien $\mathrm{BF}_3 = -186{,}746$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_3 akan menurunkan jumlah penderita diabetes melitus sebesar 186,746 pada daerah dengan nilai baku jumlah puskesmas (X_3) lebih dari 101. Hasil ini sejalan dengan penelitian yang dilakukkan oleh Sujana (2019) bahwa Puskesmas berperan penting dalam deteksi dini, melakukan perawatan, membuat rujukan dan melakukan penyuluhan mengenai diabetes melitus.
4. Koefisien $\mathrm{BF}_4 = 49.918{,}146$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_4 akan meningkatkan jumlah penderita diabetes melitus sebesar 49.918,146 pada daerah dengan nilai baku tingkat konsumsi minyak dan lemak (X_7) lebih dari 9,7. Hasil ini sejalan dengan penelitian yang dilakukkan oleh Dewi dkk (2018) bahwa konsumsi lemak memiliki kaitan erat dengan kadar glukosa darah karena tinggi rendahnya glukosa darah berpengaruh terhadap diabetes melitus.
5. Koefisien $\mathrm{BF}_5 = 7.844{,}747$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_5 akan meningkatkan jumlah penderita diabetes melitus sebesar 7.844,747 pada daerah dengan nilai baku tingkat konsumsi minyak dan lemak (X_7) kurang dari 10,1. Hasil ini sejalan dengan penelitian yang dilakukkan oleh Dewi dkk (2018) bahwa konsumsi lemak memiliki kaitan erat dengan kadar glukosa darah karena tinggi rendahnya glukosa darah berpengaruh terhadap diabetes melitus.
6. Koefisien $\mathrm{BF}_6 = -53.398{,}833$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_6 akan menurunkan jumlah penderita diabetes melitus sebesar 53.398,833 pada daerah dengan nilai baku tingkat konsumsi minyak dan lemak (X_7) lebih dari 10,1. Hasil ini sejalan dengan penelitian yang dilakukkan oleh Dewi dkk (2018) bahwa konsumsi lemak memiliki kaitan erat dengan kadar glukosa darah karena tinggi rendahnya glukosa darah berpengaruh terhadap diabetes melitus.
7. Koefisien $\mathrm{BF}_7 = -218{,}202$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_7 akan menurunkan jumlah penderita diabetes melitus sebesar 218,202 pada daerah dengan nilai baku tingkat konsumsi sayur dan buah-buahan (X_8) kurang dari 80,3. Hasil ini sejalan dengan penelitian yang dilakukkan oleh Cooper dkk (2012) bahwa dengan semakin beragam konsumsi buah, sayur, dan kombinasi keduanya, semakin rendah risiko terkena diabetes melitus.
8. Koefisien $\mathrm{BF}_8 = 10.718{,}658$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_8 akan meningkatkan jumlah penderita diabetes melitus sebesar 10.718,658 pada daerah dengan nilai baku persentase perokok (X_{11}) kurang dari 22,03. Penelitian ini sejalan dengan penelitian yang dilakukkan oleh Choirunnisa (2022) yang menunjukkan zat nikotin dalam rokok dapat menyebabkan resistensi insulin dan menurunkan sekresi insulin pada pankreas. Seseorang yang memiliki kebiasaan merokok dan mampu menghabiskan 20 batang rokok dalam 1 hari memiliki resiko penyakit diabetes melitus 62% lebih tinggi dibandingkan dengan orang yang tidak memiliki kebiasaan merokok.
9. Koefisien $\mathrm{BF}_9 = 0{,}184$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_9 akan

meningkatkan jumlah penderita diabetes melitus sebesar 0,184 pada daerah dengan nilai baku jumlah rumah sakit (X_1) kurang dari 10, yang berinteraksi dengan jumlah penderita hipertensi (X_4).

10. Koefisien $\mathrm{BF}_{10} = -2.329{,}621$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_{10} akan menurunkan jumlah penderita diabetes melitus sebesar 2.329,621 pada daerah dengan nilai baku jumlah rumah sakit (X_1) kurang dari 10, yang berinteraksi dengan tingkat konsumsi minyak dan lemak (X_7).
11. Koefisien $\mathrm{BF}_{11} = -299{,}931$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_{11} akan menurunkan jumlah penderita diabetes melitus sebesar 299,931 pada daerah dengan nilai baku jumlah rumah sakit (X_1) lebih dari 10, yang berinteraksi dengan persentase perokok (X_{11}).
12. Koefisien $\mathrm{BF}_{12} = -2.682{,}950$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_{12} akan menurunkan jumlah penderita diabetes melitus sebesar 2.682,950 pada daerah di mana tingkat konsumsi gula (X_6) yang berinteraksi dengan tingkat konsumsi minyak dan lemak (X_7) yang kurang dari 10,1.
13. Koefisien $\mathrm{BF}_{13} = -0{,}010$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_{13} akan menurunkan jumlah penderita diabetes melitus sebesar 0,010 pada daerah dengan nilai baku jumlah rumah sakit (X_1) kurang dari 10, yang berinteraksi dengan jumlah penderita hipertensi (X_4) dan persentase perokok (X_{11}).

UJI SIGNIFIKANSI MODEL BERDASARKAN GCV

1. Uji koefisien regresi serentak

a. Rumusan hipotesis:

H_0 : $a_1 = a_2 = \cdots = a_{13} = 0$ (model tidak signifikan)

H_1 : minimal terdapat satu $a_m \neq 0; m = 1,2,\ldots,13$ (model signifikan)

b. Taraf signifikansi:

$\alpha = 0{,}05$

c. Statistik uji:

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

d. Daerah kritis:

Tolak H_0 jika nilai $F_{hitung} > F_{tabel}$ atau $p-value < \alpha$.

e. Keputusan:

Berdasarkan hasil pengolahan data, diperoleh $p-value$ sebesar 0,0000 dan $F_{Hitung} = 180{,}8$ serta nilai $F_{0,05(11:27)} = 2{,}26$. Sehingga, $p-value < \alpha$ atau $F_{hitung} > F_{0,05(11:27)}$ maka tolak H_0, yang berarti model signifikan.

2. Uji koefisien regresi parsial

a. Rumusan hipotesis:

H_0 : $a_m = 0$ (koefisien a_m tidak berpengaruh terhadap model)

H_1 : minimal terdapat satu $a_m \neq 0; m = 1,2,\ldots,13$ (koefisien a_m berpengaruh terhadap model)

b. Taraf signifikansi:

$\alpha = 0{,}05$

c. Statistik uji:

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

d. Daerah kritis:

Tolak H_0 jika nilai $|t_{hitung}| > t_{tabel}$ atau $p-value < \alpha$.

Tabel 5. Uji Parsial Model MARS Berdasarkan GCV

Parameter	Estimasi	S.E.	t_{hitung}	$p-value$
Constant	26.854,240	1,468	18,293	0,0000*
BF_1	18.831,127	5,983	14,201	0,0000*
BF_2	8.495,916	1,956	9,627	0,0000*
BF_3	−186,746	2,185	−13,729	0,0000*
BF_4	49.918,146	1,372	−13,609	0,0000*
BF_5	7.844,747	4,675	−11,422	0,0000*
BF_6	−53.398,833	1,911	4,106	0,0012*
BF_7	−218,202	3,837	13,010	0,0000*
BF_8	10.718,658	1,498	7,158	0,0000*
BF_9	0,184	4,944	3,727	0,0025*
BF_{10}	−2.329,621	2,069	−11,259	0,0000*
BF_{11}	−299,931	5,349	−5,015	0,0002*
BF_{12}	−2.682,950	2,328	−4,385	0,0007*
BF_{13}	−0,010	8,396	−2,599	0,0221*

e. Keputusan:

Berdasarkan Tabel 5, dapat dilihat $p-value$ pada setiap $m < \alpha$, dengan $t_{\left(\frac{\alpha}{2},27\right)} = 2{,}052$ maka setiap $|t_{hiung}| > t_{\left(\frac{\alpha}{2},27\right)}$. Sehingga tolak H_0, yang berarti koefisien $a_1, a_2, \ldots, a_{13}$ berpengaruh signifikan terhadap model.

TINGKAT KEPENTINGAN VARIABEL PREDIKTOR PADA MODEL TERBAIK GCV

Tabel 6. Tingkat Kepentingan Variabel Prediktor (GCV)

Variabel	Tingkat Kepentingan
X_1	100%
X_{11}	80,5%
X_3	66,6%
X_7	54,3%
X_4	49,6%
X_6	8,6%
X_8	1,7%

Berdasarkan Tabel 6, dapat dilihat tingkat kepentingan masing-masing variabel prediktor yang memengaruhi jumlah penderita diabetes melitus. Variabel yang memiliki tingkat kepentingan tertinggi adalah jumlah rumah sakit (X_1), sedangkan variabel yang memiliki tingkat kepentingan terendah adalah tingkat konsumsi sayur dan buah-buahan (X_8).

MODEL MARS TERBAIK BERDASARKAN AIC

Model MARS terbaik berdasarkan nilai AIC adalah model yang tebentuk dari kombinasi kombinasi BF=33, MI=3, dan MO=0 dengan nilai AIC sebesar 476,253 . Langkah selanjutnya yaitu melakukan estimasi parameter model menggunakan metode *Ordinary Least Square* (OLS).

Tabel 7. Hasil Estimasi Parameter Model Berdasarkan AIC

Parameter	Estimasi	$p-value$	S.E.
a_0	35.595,913	0,0000*	1,025
a_1	−1.471,300	0,0000*	9,723
a_2	14.543,513	0,0000*	1,678
a_3	−1.063,393	0,0000*	2,430
a_4	395,018	0,0000*	5,019
a_5	12.937,535	0,0000*	1,046
a_6	16.414,237	0,0000*	2,316
a_7	−730,972	0,0000*	1,447
a_8	−10.979,179	0,0000*	1,062
a_9	−348,009	0,0000*	3,005
a_{10}	−555,698	0,0000*	8,415
a_{11}	−305,368	0,0000*	1,265
a_{12}	−89,339	0,0000*	124,599
a_{13}	10,336	0,0000*	3,377
a_{14}	−456,612	0,0005*	2,366

Berdasarkan hasil pada Tabel 7, maka dugaan model Model MARS terbaik dapat dituliskan sebagai berikut.

$$\hat{Y} = 35.595{,}913 - 1.471{,}300 * \mathrm{BF}_1 + 14.543{,}513 * \mathrm{BF}_2 - 1.063{,}393 * \mathrm{BF}_3 + 395{,}018 * \mathrm{BF}_4 + 12.937{,}535 * \mathrm{BF}_5 + 16.414{,}237 * \mathrm{BF}_6 - 730{,}972 * \mathrm{BF}_7 - 10.979{,}179 * \mathrm{BF}_8 - 348{,}009 * \mathrm{BF}_9 - 555{,}698 * \mathrm{BF}_{10} - 305{,}368 * \mathrm{BF}_{11} - 89{,}339 * \mathrm{BF}_{12} + 10{,}336 * \mathrm{BF}_{13} - 456{,}612 * \mathrm{BF}_{14}$$

dengan $\mathrm{BF}_1 = \max(0, 10 - X_1)$, $\mathrm{BF}_2 = \max(0, X_1 - 10)$, $\mathrm{BF}_3 = \max(0, 114 - X_3)$, $\mathrm{BF}_4 = \max(0, X_3 - 114)$, $\mathrm{BF}_5 = \max(0, 10 - X_7)$, $\mathrm{BF}_6 = \max(0, X_7 - 10)$, $\mathrm{BF}_7 = \max(0, X_8 - 85{,}1)$, $\mathrm{BF}_8 = \max(0, 0{,}906 - \mathrm{X}_9)$, $\mathrm{BF}_9 = \max(0, 10 - X_1) * X_2$, $\mathrm{BF}_{10} = \max(0, X_1 - 10) * X_7$, $\mathrm{BF}_{11} = \max(0, X_1 - 10) * X_{11}$, $\mathrm{BF}_{12} = \max(0, X_3 - 114) * X_6$, $\mathrm{BF}_{13} = \max(0, 114 - X_3) * X_8$, dan $\mathrm{BF}_{14} = X_1 * X_6 * \max(0, 10 - X_7)$.

Interpretasi model MARS terbaik berdasarkan AIC adalah sebagi berikut.

1. Koefisien $\mathrm{BF}_1 = -1.471{,}300$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_1 akan menurunkan jumlah penderita diabetes melitus sebesar 1.471,300 pada daerah dengan nilai baku jumlah rumah sakit (X_1) kurang dari 10 . Hasil ini sejalan dengan penelitian yang dilakukkan oleh Mulyani dkk (2023) bahwa rumah sakit berhubungan dengan peningkatan deteksi dan pengelolaan diabetes melitus, serta menunjukkan bahwa akses ke layanan kesehatan yang memadai dapat membantu dalam mengurangi keparahan penyakit melalui pengelolaan yang lebih baik.
2. Koefisien $\mathrm{BF}_2 = 14.543{,}513$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_2 akan meningkatkan jumlah penderita diabetes melitus sebesar 14.543,513 pada daerah dengan nilai baku jumlah rumah sakit (X_1) lebih dari 10 . Hasil ini sejalan dengan penelitian yang dilakukkan oleh Mulyani dkk (2023) bahwa rumah sakit berhubungan dengan peningkatan deteksi dan pengelolaan diabetes melitus, serta menunjukkan bahwa akses ke layanan kesehatan yang memadai dapat membantu dalam mengurangi keparahan penyakit melalui pengelolaan yang lebih baik.
3. Koefisien $\mathrm{BF}_3 = -1.063{,}393$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_3 akan menurunkan jumlah penderita diabetes melitus sebesar 1.063,393 pada daerah dengan nilai baku jumlah puskesmas (X_3)

kurang dari 114. Hasil ini sejalan dengan penelitian yang dilakukkan oleh Sujana (2019) bahwa Puskesmas berperan penting dalam deteksi dini, melakukan perawatan, membuat rujukan dan melakukan penyuluhan mengenai diabetes melitus.

4. Koefisien $\text{BF}_4 = 395{,}018$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_4 akan meningkatkan jumlah penderita diabetes melitus sebesar 395,018 pada daerah dengan nilai baku jumlah puskesmas (X_3) lebih dari 114. Hasil ini sejalan dengan penelitian yang dilakukkan oleh Sujana (2019) bahwa Puskesmas berperan penting dalam deteksi dini, melakukan perawatan, membuat rujukan dan melakukan penyuluhan mengenai diabetes melitus.
5. Koefisien $\text{BF}_5 = 12.937{,}535$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_5 akan meningkatkan jumlah penderita diabetes melitus sebesar 12.937,535 pada daerah dengan nilai baku tingkat konsumsi minyak dan lemak (X_7) kurang dari 10. Hasil ini sejalan dengan penelitian yang dilakukkan oleh Dewi dkk (2018) bahwa konsumsi lemak memiliki kaitan erat dengan kadar glukosa darah karena tinggi rendahnya glukosa darah berpengaruh terhadap diabetes melitus.
6. Koefisien $\text{BF}_6 = 16.414{,}237$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_6 akan meningkatkan jumlah penderita diabetes melitus sebesar 16.414,237 pada daerah dengan nilai baku tingkat konsumsi minyak dan lemak (X_7) lebih dari 10. Hasil ini sejalan dengan penelitian yang dilakukkan oleh Dewi dkk (2018) bahwa konsumsi lemak memiliki kaitan erat dengan kadar glukosa darah karena tinggi rendahnya glukosa darah berpengaruh terhadap diabetes melitus.
7. Koefisien $\text{BF}_7 = -730{,}972$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_7 akan menurunkan jumlah penderita diabetes melitus sebesar 730,972 pada daerah dengan nilai baku tingkat konsumsi sayur dan buah-buahan (X_8) lebih dari 85,1. Hasil ini sejalan dengan penelitian yang dilakukkan oleh Cooper dkk (2012) bahwa dengan semakin beragam konsumsi buah, sayur, dan kombinasi keduanya, semakin rendah risiko terkena diabetes melitus.
8. Koefisien $\text{BF}_8 = -10.979{,}179$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_8 akan menurunkan jumlah penderita diabetes melitus sebesar 10.979,179 pada daerah dengan nilai baku tingkat konsumsi teh (X_9) kurang dari 0,906. Hasil ini sejalan dengan penelitian yang dilakukkan oleh Meiwati (2021) bahwa kebiasaan minum teh dapat mendukung pengendalian gula darah, sehingga bermanfaat bagi individu yang berisiko atau telah didiagnosis dengan diabetes.
9. Koefisien $\text{BF}_9 = -348{,}009$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_9 akan menurunkan jumlah penderita diabetes melitus sebesar 348,009 pada daerah dengan nilai baku jumlah rumah sakit (X_1) kurang dari 10, yang berinteraksi dengan jumlah klinik (X_2).
10. Koefisien $\text{BF}_{10} = -555{,}698$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_{10} akan menurunkan jumlah penderita diabetes melitus sebesar 555,698 pada daerah dengan nilai baku jumlah rumah sakit (X_1) lebih dari 10, yang berinteraksi dengan tingkat konsumsi minyak dan lemak (X_7).
11. Koefisien $\text{BF}_{11} = -305{,}368$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_{11} akan menurunkan jumlah penderita diabetes melitus sebesar 305,368 pada daerah dengan nilai baku jumlah rumah sakit (X_1) lebih dari 10, yang berinteraksi dengan persentase perokok (X_{11}).
12. Koefisien $\text{BF}_{12} = -89{,}339$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_{12} akan menurunkan jumlah penderita diabetes melitus sebesar 89,339 pada daerah dengan nilai baku jumlah puskesmas (X_3) lebih dari 114, yang berinteraksi dengan tingkat konsumsi gula (X_6).
13. Koefisien $\text{BF}_{13} = 10{,}336$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_{13} akan meningkatkan jumlah penderita diabetes melitus sebesar 10,336 pada daerah dengan nilai baku jumlah puskesmas (X_3) kurang dari 114, yang berinteraksi dengan tingkat konsumsi sayur dan buah-buahan (X_8).
14. Koefisien $\text{BF}_{14} = -456{,}612$ pada model memiliki arti bahwa setiap kenaikan satu satuan BF_{14} akan menurunkan jumlah penderita

diabetes melitus sebesar 456,612 pada daerah dengan nilai baku jumlah rumah sakit (X_1) yang berinteraksi dengan tingkat konsumsi gula (X_1), dan tingkat konsumsi minyak dan lemak (X_7) kurang dari 10.

UJI SIGNIFIKANSI MODEL BERDASARKAN AIC

1. Uji koefisien regresi serentak
a. Rumusan hipotesis:
H_0 : $a_1 = a_2 = \cdots = a_{14} = 0$ (model tidak signifikan)
H_1 : minimal terdapat satu $a_m \neq 0; m = 1,2,\dots,14$ (model signifikan)
b. Taraf signifikansi:
$\alpha = 0{,}05$
c. Statistik uji:

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

d. Daerah kritis:
Tolak H_0 jika nilai $F_{hitung} > F_{tabel}$ atau $p-value < \alpha$.
e. Keputusan:
Berdasarkan hasil pengolahan data, diperoleh $p-value$ sebesar 0,0000 dan $F_{Hitung} = 204{,}4$ serta nilai $F_{0,05(11:27)} = 2{,}26$. Sehingga, $p-value < \alpha$ atau $F_{hitung} > F_{0,05(11:27)}$ maka tolak H_0, yang berarti model signifikan.

2. Uji koefisien regresi parsial
a. Rumusan hipotesis:
H_0 : $a_m = 0$ (koefisien a_m tidak berpengaruh terhadap model)
H_1 : minimal terdapat satu $a_m \neq 0; m = 1,2,\dots,14$ (koefisien a_m berpengaruh terhadap model)
b. Taraf signifikansi:
$\alpha = 0{,}05$
c. Statistik uji:

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

d. Daerah kritis:
Tolak H_0 jika nilai $|t_{hitung}| > t_{tabel}$ atau $p-value < \alpha$.

Tabel 8. Uji Parsial Model MARS Berdasarkan AIC

Parameter	Estimasi	S.E.	t_{hitung}	$p-value$
Constant	35.595,913	1,025	34,734	0,0000*
BF_1	−1.471,300	9,723	14,958	0,0000*
BF_2	14.543,513	1,678	−8,768	0,0000*
BF_3	−1.063,393	2,430	−12,569	0,0000*
BF_4	395,018	5,019	7,870	0,0000*
BF_5	12.937,535	1,046	−10,162	0,0000*
BF_6	16.414,237	2,316	7,086	0,0000*
BF_7	−730,972	1,447	8,941	0,0000*
BF_8	−10.979,179	1,062	−8,416	0,0000*
BF_9	−348,009	3,005	−11,581	0,0000*
BF_{10}	−555,698	8,415	−6,603	0,0000*
BF_{11}	−305,368	1,265	8,172	0,0000*
BF_{12}	−89,339	124,599	−5,867	0,0000*
BF_{13}	10,336	3,377	−13,521	0,0000*
BF_{14}	−456,612	2,366	−4,640	0,0005*

e. Keputusan:
Berdasarkan Tabel 8, dapat dilihat $p-value$ pada setiap $m < \alpha$, dengan $t_{\left(\frac{\alpha}{2},27\right)} = 2{,}052$ maka setiap $|t_{hiung}| > t_{\left(\frac{\alpha}{2},27\right)}$. Sehingga tolak H_0, yang berarti koefisien $a_1, a_2, \dots, a_{14}$ berpengaruh signifikan terhadap model.

TINGKAT KEPENTINGAN VARIABEL PREDIKTOR PADA MODEL TERBAIK AIC

Tabel 9. Tingkat Kepentingan Variiabel Prediktor (AIC)

Variabel	Tingkat Kepentingan
X_1	100%
X_{11}	79,9%
X_3	65,1%
X_6	52,8%
X_7	52,8%
X_8	49%
X_2	47,3%
X_9	28,8%

Berdasarkan Tabel 9, dapat dilihat tingkat kepentingan masing-masing variabel prediktor yang memengaruhi jumlah penderita diabetes melitus. Variabel yang memiliki tingkat kepentingan tertinggi adalah jumlah rumah sakit (X_1), sedangkan variabel yang memiliki tingkat kepentingan terendah adalah tingkat konsumsi teh (X_9).

PEMBAHASAN UMUM

Dalam penelitian ini digunakan dua kriteria pemilihan model untuk menentukan model terbaik dalam metode *Multivariate Adaptive Regression Splines* (MARS), yaitu *Generalized Cross Validation* (GCV) dan

Akaike Information Criterion (AIC). Kedua kriteria tersebut memberikan hasil yang berbeda karena adanya perbedaan pendekatan dalam mengevaluasi model. Perbedaan hasil dari kedua kriteria pemilihan model tersebut disajikan sebagai berikut.

Tabel 10. Perbedaan Model Terbaik Berdasarkan GCV dan AIC

Kriteria	Jumlah fungsi basis	R^2
GCV	13	0,994
AIC	14	0,995

Perbedaan struktur model terlihat dari jumlah fungsi basis, di mana model GCV memiliki 13 fungsi basis, sementara model AIC memiliki 14 fungsi basis. Selain itu, GCV menghasilkan tujuh variabel yang berpengaruh, sedangkan AIC menghasilkan delapan variabel yang berpengaruh. Perbedaan ini menunjukkan bahwa GCV cenderung menghasilkan model yang lebih sederhana. Dari sisi nilai R^2, model GCV memiliki nilai R^2 sebesar 0,994, sedangkan model AIC memiliki nilai R^2 sebesar 0,995. Meskipun terdapat perbedaan kecil pada nilai R^2, keduanya mencerminkan kemampuan yang sangat baik dalam memodelkan faktor-faktor yang memengaruhi jumlah penderita diabetes melitus di Jawa Barat.

PENUTUP

SIMPULAN

Berdasarkan hasil dan pembahasan yang telah dijelaskan sebelumnya diperoleh kesimpulan sebagai berikut:

1. Metode *Multivariate Adaptive Regression Splines* (MARS) dapat digunakan untuk memodelkan jumlah penderita diabetes melitus di provinsi Jawa Barat, dimana data yang digunakan tidak memiliki pola tertentu dan menunjukkan bahwa kriteria pemilihan model AIC menghasilkan nilai R^2 yang tinggi sehingga dapat dikatakan kriteria ini dapat digunakan sebagai alternatif dalam pemilihan model MARS.
2. Berdasarkan kriteria pemilihan model GCV diperoleh tujuh variabel prediktor yang signifikan mempengaruhi variabel respon, yaitu jumlah rumah sakit umum (X_1), jumlah puskesmas (X_3), jumlah penderita hipertensi (X_4), tingkat konsumsi gula (X_6), tingkat konsumsi minyak dan lemak (X_7), tingkat konsumsi sayur dan buah-buahan (X_8), dan persentase perokok (X_{11}).
3. Berdasarkan kriteria pemilihan model AIC diperoleh tujuh variabel prediktor yang signifikan mempengaruhi variabel respon, yaitu jumlah rumah sakit umum (X_1), jumlah klinik (X_2), jumlah puskesmas (X_3), tingkat konsumsi gula (X_6), tingkat konsumsi minyak dan lemak (X_7), tingkat konsumsi sayur dan buah-buahan (X_8), tingkat konsumsi teh (X_9), dan persentase perokok (X_{11}).

SARAN

1. Penelitian selanjutnya disarankan untuk menambahkan lebih banyak faktor yang dapat memengaruhi jumlah penderita diabetes melitus.
2. Penelitian berikutnya diharapkan menggunakan kriteria pemilihan model lainnya seperti *Schwarz Bayes information Criterion* (SBC).

DAFTAR PUSTAKA

Agresti, A. (2019). *An introduction to categorical data analysis* (3rd ed.). Wiley.

Azizah, D. M., & Permatasari, E. O. (2020). Modeling of toddler stunting in the Province of East Nusa Tenggara using Multivariate Adaptive Regression Splines (MARS) method. *Journal of Physics: Conference Series*, 1490(1), 012013. https://doi.org/10.1088/1742-6596/1490/1/012013

Badan Penelitian dan Pengembangan Kesehatan. (2018). *Laporan Nasional Riskesdas 2018*. Kementerian Kesehatan Republik Indonesia.

Choirunnisa, N. H. (2022). Konsumsi gula dan kebiasaan merokok dengan kejadian diabetes melitus di Puskesmas Gading Surabaya. *JIK (Jurnal Ilmu Kesehatan)*, 6(2), 469.

Cooper, A. J., Sharp, S. J., Lentjes, M. A., Luben, R. N., Khaw, K. T., Wareham, N. J., & Forouhi, N. G. (2012). A prospective study of the association between quantity and variety of fruit and vegetable intake and incident type 2 diabetes. *Diabetes Care*, 35(6), 1293–1300. https://doi.org/10.2337/dc11-2388

Dewi, A. Y., Dwidayati, N. K., & Agoestanto, A. (2020). Analisis survival model regresi Cox dengan metode MLE untuk penderita diabetes mellitus. *Unnes Journal of Mathematics*, 9(1), 31–40.

Dewi, P. A. T., Moviana, Y., & Cintari, L. (2018).

Konsumsi lemak dan kadar kolesterol HDL pasien DM Tipe 2 rawat jalan di RSUP Sanglah Denpasar. *Jurnal Ilmu Gizi: Journal of Nutrition Science*, 7(1), 1–5.

Dinas Kesehatan Jawa Barat. (2023). Diakses 21 September, 2024, dari website: https://opendata.jabarprov.go.id/id/dataset/jumlah-penderita-diabetes-melitus-berdasarkan-kabupatenkota-di-jawa-barat

Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*. Marcel Dekker Inc.

Fabregas, J. (2021). SO-22 Importance order of risk factors for mortality in colon cancer stage II: A multivariate adaptive regression spline (MARS) analysis. *Annals of Oncology*, 32(S211–S212).

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, 19(1), 1–141.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2008). *The elements of statistical learning*. Springer.

Härdle, W. (1994). Applied nonparametric regression. Humboldt-Universität zu Berlin.

International Diabetes Federation. (2017). *IDF diabetes atlas* (8th ed.). International Diabetes Federation.

International Diabetes Federation. (2021). IDF diabetes atlas: Global estimates of diabetes prevalence for 2021 and projections for 2045. Diakses 21 September, 2024, dari website: https://diabetesatlas.org/data/en/country/94/id.html

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). Springer.

Kementerian Kesehatan Republik Indonesia. (2022). *Diabetes insipidus*. Direktorat Jenderal Pelayanan Kesehatan. Diakses 20 September, 2024, dari website: https://yankes.kemkes.go.id/view_artikel/193/diabetes-insipidus

Kemenkes RI, K. K. R. I. (2022). Profil Kesehatan Indonesia Tahun 2021. P2PTM Kementerian Kesehatan Republik Indonesia.

Meiwati, I., Novitasari, D., & Maryoto, M. (2021). Pengaruh konsumsi seduhan teh hijau (Camellia sinensis) terhadap kadar glukosa darah dan kolesterol penderita diabetes melitus. *Seminar Nasional Penelitian dan Pengabdian kepada Masyarakat*, 961–967.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (5th ed.). Wiley.

Mulyani, A. Y., Arman, A., & Patimah, S. (2023). Analisis faktor yang mempengaruhi kualitas hidup pasien diabetes melitus tipe II di Rumah Sakit Umum Daerah Lasinrang Kabupaten Pinrang tahun 2022. *Journal of Muslim Community Health*, 4(4), 345–357.

Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric regression*. Cambridge University Press.

Santi, V. M., Ningsih, N. R., & Ladayya, F. (2022). Analyzing Open Unemployment Rate in Java Using Penalized Spline Nonparametric Regression. *International Journal of Applied Science and Sustainable Development* (IJASSD), 4(2), 29–36.

Sujana, T. (2019). Peran puskesmas dalam identifikasi dini penyakit diabetes melitus pada lansia. *Jurnal Kesehatan Bakti Tunas Husada: Jurnal Ilmu-ilmu Keperawatan, Analis Kesehatan dan Farmasi*, 19(1).

Tripena, A. (2011). Analisis regresi spline kuadratik. *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika*.

World Health Organization. (1999). *Definition, diagnosis and classification of diabetes mellitus and its complications: Report of a WHO consultation*. World Health Organization.