

PENERAPAN METODE *MULTIVARIATE ADAPTIVE REGRESSION SPLINE* UNTUK MEMAHAMI DINAMIKA KEMISKINAN DI INDONESIA

APPLICATION OF MULTIVARIATE ADAPTIVE REGRESSION SOKINE METHOD TO UNDERSTAND POVERTY DYNAMICS IN INDONESIA

Nurviqotun Khasanah¹, Zilrahmi², Syafriandi³

^{1,2,3}Universitas Negeri Padang

¹nurvina739@gmail.com, ²zilrahmi@fmipa.unp.ac.id, ³syafriandi@fmipa.unp.ac.id

Abstrak

Kemiskinan masih menjadi tantangan besar bagi pembangunan di negara berkembang, khususnya Indonesia. Berbagai faktor seperti pendidikan, kesehatan dan pendapatan masyarakat diketahui mempengaruhi tingkat kemiskinan, namun hubungan antar faktor tidak sederhana. Studi ini dilakukan untuk memprediksi Presentase Penduduk Miskin Di Indonesia berdasarkan faktor sosial ekonomi menggunakan metode *Multivariate Adaptive Regression Spline* yang mampu menangkap hubungan nonlinear dan interaksi antar variabel. Penelitian menggunakan Data dan Informasi Kemiskinan Kab/Kota di Indonesia Tahun 2023 dari publikasi Badan Pusat Statistik (BPS) yang telah melalui proses *Pre-processing data*. Model terbaik dibangun dari 0.8 data *training* dan 0.2 data *testing* dengan kombinasi BF=26, MI=3, MO=1 dengan *Generalized Cross Validation* (GCV) terkecil sebesar 0.160211 dan dari 13 variabel prediktor yang diteliti menunjukkan bahwa variabel Persentase Pengeluaran Rata-Rata per Orang untuk Makanan Kategori Miskin dan Tidak Miskin (X_5) dan variabel Persentase Pengeluaran Rata-Rata per Orang untuk Makanan Kategori Miskin dan Tidak Miskin (X_6) yang mempunyai skor tertinggi sebesar 100% untuk menurunkan nilai GCV model dan menurunkan *Residual Sum of Squares* (RSS) pada model. Selain itu, model MARS mampu menjelaskan variasi tingkat kemiskinan dengan nilai *R-squared* sebesar 83,7% yang mengidentifikasi prediksi cukup akurat.

KATA KUNCI : Kemiskinan, MARS, GCV

Abstract

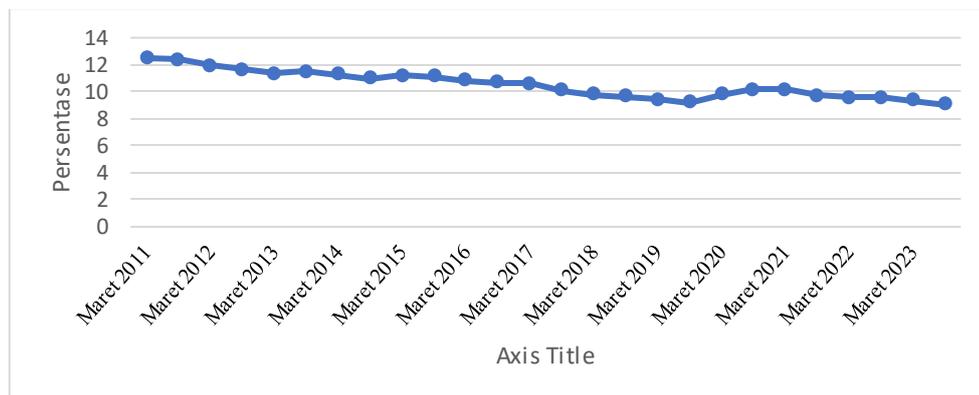
Poverty remains a major challenge for development in developing countries, especially Indonesia. Various factors such as education, health and income are known to affect the poverty rate, but the relationship between factors is not simple. This study aims to predict the percentage of poor people in Indonesia based on socioeconomic factors using the *Multivariate Adaptive Regression Spline* method which is able to capture nonlinear relationships and interactions between variables. The research uses data and information on poverty in districts / cities in Indonesia in 2023 obtained from the Central Statistics Agency (BPS) which has gone through a process of cleaning, standardisation and handling outliers. The best model was built from

0.8 training data and 0.2 testing data with a combination of $BF=26$, $MI=3$, $MO=1$ with the smallest Generalised Cross Validation (GCV) of 0.160211 and of the 13 predictor variables studied showed that the variable Percentage of Average Expenditure per Person on Food for Poor and Non-Poor Categories (X5) and the variable Percentage of Average Expenditure per Person on Food for Poor and Non-Poor Categories (X6) which had the highest score of 100% to reduce the GCV value of the model and reduce the Residual Sum of Squares (RSS) in the model. In addition, the MARS model is able to explain the variation in poverty rates with an R-squared value of 83.7%, which identifies a fairly accurate prediction.

Keywords: Poverty, MARS, GCV

PENDAHULUAN

Kemiskinan adalah salah satu permasalahan sosial yang masih menjadi fokus utama dalam pembangunan di Indonesia. Badan Pusat Statistik mencatat rata-rata proporsi penduduk miskin dari 1999-2014 sebesar 14,97%. Namun rata-rata proporsi penduduk miskin tahun 2015-2024 terjadi penurunan yang signifikan di sekitar 10% seperti di Gambar 1 (Reno, 2024).



Gambar 1. Persentase Penduduk Miskin di Indonesia antara 2015-2024

Berdasarkan Gambar 1. pada Maret 2024 Badan Pusat Statistik (BPS) melaporkan bahwa Persentase Penduduk Miskin di Indonesia tercatat sebesar 9,03% atau sekitar 25,22 juta jiwa, angka ini menunjukkan tren penurunan dalam beberapa tahun. namun angka ini masih jauh dari target *ideal* angka kemiskinan nasional Indonesia menurut Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2020 hingga 2024 sebesar 6,5 – 7,5%. Hal ini memperjelas adanya ketimpangan yang cukup signifikan antar wilayah di Indonesia, sehingga perlu menjadi fokus kebijakan yang lebih tepat sasaran.

Pemerintah telah mengusahakan berbagai program sebagai upaya pengentasan kemiskinan, seperti Bantuan Pangan Non Tunai (BPNT), Program Keluarga Harapan (PKH), serta pembangunan berbagai infrastruktur dasar di daerah tertinggal. Hasil dari program-program tersebut lumayan efektif secara nasional dalam menekan jumlah penduduk miskin. Akan tetapi, hasilnya belum merata di seluruh wilayah, terutama di daerah-daerah yang memiliki aksesibilitas

dan kondisi sosial-ekonomi lebih kompleks. Salah satu tantangan utama dalam pemerataan pembangunan adalah adanya ketimpangan regional yang signifikan. Ketimpangan ini terlihat jelas antara wilayah perkotaan dan pedesaan. Di perkotaan, infrastruktur lebih memadai, akses ke layanan kesehatan, pendidikan, dan lapangan kerja relatif lebih baik, sehingga tingkat kemiskinan cenderung lebih rendah. Sebaliknya, di wilayah pedesaan, banyak masyarakat yang masih menghadapi keterbatasan dalam hal akses pendidikan, fasilitas kesehatan, serta kesempatan ekonomi, sehingga membuat mereka lebih rentan terhadap kemiskinan.

Selain itu, ketimpangan juga terjadi antara pulau Jawa dan luar Jawa, serta antara wilayah bagian barat dan timur Indonesia. Pulau Jawa, sebagai pusat pertumbuhan ekonomi dan populasi terbesar, umumnya memiliki tingkat pembangunan yang lebih maju dibandingkan wilayah timur seperti Papua. Menurut data BPS (2024), tingkat kemiskinan di beberapa wilayah luar Jawa dan Indonesia timur masih berada di atas rata-rata nasional. Misalnya, persentase penduduk miskin di Kab Nduga, Papua mencapai 37.09%, sedangkan di DKI Jakarta hanya sekitar 4.68%. Perbedaan ini mencerminkan adanya disparitas dalam distribusi sumber daya, pembangunan, serta kualitas infrastruktur dan layanan publik. Ketimpangan antar wilayah ini menjadi salah satu tantangan besar dalam pencapaian tujuan pembangunan inklusif dan berkelanjutan. Oleh karena itu, perlu adanya strategi yang lebih spesifik dan disesuaikan dengan karakteristik lokal untuk memastikan bahwa penanggulangan kemiskinan tidak hanya berhasil secara makro, tetapi juga menyentuh masyarakat di wilayah yang paling tertinggal.

Untuk merancang strategi penanggulangan kemiskinan yang lebih efektif, penting untuk memahami variabel-variabel yang memengaruhi penduduk miskin. Sayangnya, hubungan antar prediktor penyebab kemiskinan sering kali bersifat tidak *linear* dan melibatkan interaksi yang rumit. Pendekatan regresi tradisional yang mengasumsi linieritas kurang mampu menggambarkan dinamika sebenarnya. Oleh karena itu, metode statistik yang lebih fleksibel seperti *Multivariate Adaptive Regression Splines* (MARS) menjadi solusi alternatif yang relevan. MARS merupakan teknik non-parametrik yang dapat secara otomatis mendeteksi pola non-linear dan interaksi antar variabel melalui basis fungsi adaptif. Penerapan model MARS, bertujuan untuk mengali lebih dalam bagaimana interaksi dan kontribusi setiap variabel prediktor terhadap dinamika kemiskinan di Indonesia tahun 2023. Hasil yang diharapkan dapat memberikan gambaran tentang mekanisme multidimensional penyebab kemiskinan. Sehingga menjadi dasar kuat dalam pengambilan keputusan dan perumusan kebijakan yang lebih efektif dalam penurunan persentase angka kemiskinan di Indonesia.

KAJIAN TEORI

Salah satu metode statistik utama yang digunakan untuk menganalisis pengaruh variabel prediktor terhadap variabel respon adalah regresi. Model regresi yang paling sederhana dan umum digunakan adalah regresi linear. Tapi regresi linear memiliki keterbatasan dalam menangkap pola hubungan yang bersifat non-linear dan interaksi kompleks antar variabel. Untuk mengatasi keterbatasan itu, regresi non-linear parametrik menjadi solusi dalam melihat hubungan antar variabel dengan fungsi matematis tertentu seperti fungsi kuadrat, eksponensial atau logaritmik.

Regresi *spline* merupakan salah satu pendekatan yang fleksibel dalam

menangani non-linearitas. Metode ini menggunakan fungsi polinomial yang dibagi menjadi beberapa segmen pada titik-titik tertentu yang disebut knot. Dengan membagi domain variabel prediktor menjadi *interval-interval*, regresi *spline* mampu menangkap perubahan pola hubungan secara lokal. Meskipun demikian, regresi *spline* konvensional memiliki keterbatasan dalam hal penentuan posisi knot yang biasanya dilakukan secara manual atau berdasarkan heuristik, serta kurang mampu secara otomatis menangani interaksi antar variabel yang kompleks.

Menurut Fatmawati (2017), *Multivariate Adaptive Regression Spline* (MARS) hadir sebagai metode regresi non-parametrik yang mengatasi keterbatasan regresi *spline* konvensional. MARS membangun model melalui kombinasi linear dari fungsi basis. Keunggulan utama MARS adalah otomatis memilih variabel, menentukan posisi knot, dan menangkap interaksi antar variabel tanpa perlu asumsi bentuk fungsi yang eksplisit. Berikut tahapan analisis MARS :

Pre-Processing Data

Preprocessing data merupakan proses transformasi atau pembersihan data menjadi lebih berstruktur untuk di analisis. Proses ini sangat penting dalam mengatasi masalah *missing values* (nilai hilang), *outlier* dan lain-lain. *Pre-processing* data diawali dengan data *cleaning* (menghapus atau mengisi nilai yang hilang). Kemudian dilakukan transformasi data (normalisasi atau standarisasi data dengan menyamakan skala data numerik).

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

Dengan:

- x : Nilai individu
- μ : Rata-rata populasi
- σ : Deviasi Standar

Setelah dilakukan *Pre-processing data*, dilakukan penanganan *outlier* dalam data numerik dengan teknik *Winsorizing*. Dimana nilai ekstrem tidak dihapus, tetapi diganti dengan nilai tertentu, biasanya digunakan persentil ke-5 dan ke-95.

$$x_{winsor} = \begin{cases} Q_{0.05} & \text{jika } x < Q_{0.05} \\ Q_{0.95} & \text{jika } x > Q_{0.95} \\ x & \text{lainnya} \end{cases} \quad (2)$$

Pembagian Data

Menurut Adinugroho (2022), proses ini bertujuan untuk membagi dataset ke dalam dua bagian, yaitu *training set* (data pelatihan) dan *testing set* (data pengujian). Proses ini dilakukan untuk mengevaluasi kinerja model dalam memproses *training set*. Pendekatan ini didasarkan pada prinsip Pareto (8:2), 0.8 data untuk membangun model, sementara 0.2 sisanya untuk menguji akurasi prediksi dari model.

Penentuan Parameter MARS

Dalam penerapan metode MARS, terdapat tiga parameter utama yang berperan mengontrol kompleksitas dan *fleksibilitas* model. Parameter-parameter ini harus ditentukan secara tepat agar model yang dibangun tidak hanya akurat, tetapi juga bisa menangkap pola hubungan yang *kompleks* antar variabel.

Basis Functions (BF), yaitu jumlah maksimum fungsi basis dalam model. jumlah maksimum Basis Fungsi yang disarankan biasanya berkisar antara 2 hingga

4 kali jumlah seluruh variabel prediktor yang ada dalam model, agar tetap menjaga keseimbangan antara *fleksibilitas* dan risiko *overfitting*.

Minspan (MO), yaitu jarak Minimal Observasi antar Knot untuk menghindari pemilihan titik yang terlalu dekat. Jarak minimum antar knot atau jumlah observasi minimal di antara dua knot dapat diatur sesuai kebutuhan, umumnya bernilai 0, 1, 2, atau 3. Nilai MO ini memengaruhi *fleksibilitas* model dalam menangkap pola data serta berkontribusi dalam mengontrol kompleksitas model agar tidak terjadi *overfitting*.

Max Interactions (MI), yaitu tingkat Interaksi Maksimum antar variabel. Menurut Friedman (1991), jumlah MI yang disarankan dalam model adalah sebanyak 1, 2, atau maksimal 3. Batasan ini ditetapkan untuk menjaga agar struktur model tetap *interpretable* dan tidak terlalu rumit.

Forward Selection

Proses *forward selection* merupakan tahapan awal yang penting saat pembentukan model MARS. Pada proses ini dilakukan penambahan satu per satu fungsi-fungsi secara bertahap, tujuannya untuk meningkatkan kinerja model dengan menurunkan galat prediksi (*prediction error*) semaksimal mungkin. Fungsi basis yang digunakan yaitu fungsi *Hinge*. Fungsi *Hinge* yang ditambahkan berbentuk (Friedman, 2007):

$$\begin{aligned} h(x) &= \max(0, x - c) \\ &\text{atau} \\ h(x) &= \max(0, c - x) \end{aligned} \quad (3)$$

Dengan :

- x : Variabel prediktor
- c : Titik potong yang ditentukan

Sehingga model MARS secara sistematis dinyatakan seperti berikut

$$f(x) = \beta_0 + \sum_{m=1}^M \beta_m \cdot h_m(x) \quad (4)$$

Dengan :

- β_0 : *Intersep*
- β_m : Koefisien dari fungsi basis ke-m
- $h_m(x)$: Fungsi *Hinge*

Backward Pruning

Menurut Friedman (2007), pada tahapan ini terjadi proses pengurangan kompleksitas model dengan menghapus fungsi paling kecil untuk menghindari *overfitting*, menggunakan kriteria *Generalized Cross Validation* (GCV). Nilai GCV semakin mendekati 0, semakin baik kinerja prediktif model yang dihasilkan namun tetap sederhana. Rumus GCV dirumuskan sebagai berikut :

$$\begin{aligned} RSS &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ GCV &= \frac{RSS}{n \cdot \left(1 - \frac{d}{n}\right)^2} \end{aligned} \quad (5)$$

Dengan :

- RSS : Jumlah kuadrat galat (*residual sum of squares*)
- n : Jumlah total observasi dalam data

- y_i : Nilai aktual
 \hat{y}_i : Nilai prediksi
 d : Derajat kebebasan model (basis fungsi) yang digunakan setelah proses seleksi

Pemilihan Model Terbaik

Menurut Friedman (2007), model dibangun dengan berbagai kombinasi ketiga parameter, kemudian pemilihan model dengan menggunakan kriteria GCV seperti di atas. Model dengan nilai GCV terkecil akan diambil sebagai model terbaik, karena dianggap mampu menghindari *overfitting* dan mempertahankan akurasi prediksi yang tertinggi. Setelah model terbaik terpilih, dilakukan interpretasi dengan mengidentifikasi variabel-variabel yang memiliki signifikan tertinggi melalui visualisasi *Partial Dependence Plots* (PDP). Tahapan ini sangat perlu untuk pengambilan keputusan (Septian et al., 2023).

Evaluasi Model

R-squared adalah ukuran seberapa baik variabel prediktor memprediksi variabel respon. Menurut Awaludin (2010), *R-squared* dianggap baik jika bernilai di atas 0.5 karena nilai *R-squared* berkisar antara 0 hingga 1. Hal ini menunjukkan bahwa model mampu menjelaskan lebih dari 50% variansi data. Dengan itu evaluasi model dilakukan untuk mengukur performa model dengan menghitung nilai koefisien determinasi (*R-squared*), yang menjelaskan seberapa besar proporsi *variabilitas* variabel prediktor dapat dijelaskan oleh model. Nilai R^2 dihitung dengan rumus :

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

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (6)$$

Dengan :

- \bar{y} : Rata-rata nilai aktual
 y_i : Nilai aktual
 \hat{y}_i : Nilai prediksi

METODE

Penelitian ini merupakan studi kuantitatif yang menggunakan pendekatan analisis statistik untuk mengidentifikasi berbagai prediktor yang mempengaruhi dinamika kemiskinan di Indonesia. Pendekatan utama yang diterapkan pada studi ini adalah metode *Multivariate Adaptive Regression Spline* (MARS), yaitu teknik regresi non-parametrik yang memiliki kemampuan untuk menangkap hubungan nonlinear serta interaksi antar variabel dengan cara yang adaptif dan fleksibel. Selain itu, penelitian ini memiliki sifat *eksploratif* dengan tujuan untuk menemukan pola-pola tertentu, mengukur pengaruh berbagai faktor, sekaligus memberikan prediksi terkait tingkat kemiskinan berdasarkan sejumlah variabel prediktor. (Sinurat, 2023)

Sumber data yang digunakan dalam penelitian ini berasal dari data sekunder yang diperoleh dari publikasi Badan Pusat Statistik (BPS) berjudul Data dan Informasi Tingkat Kemiskinan Kabupaten/Kota di Indonesia Tahun 2023. Berikut variabel dalam penelitian ini:

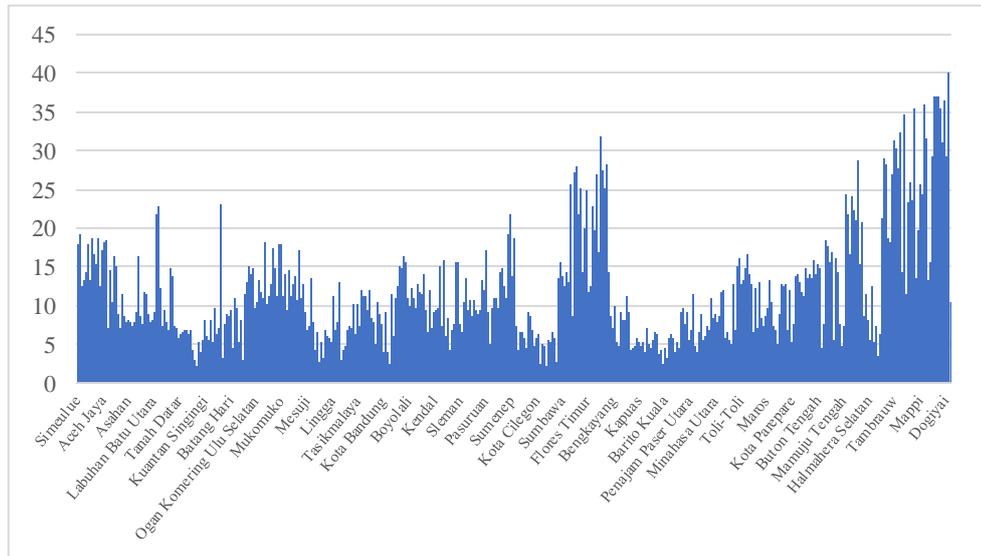
Tabel 1. Variabel Penelitian

Variabel	Keterangan
Y	Persentase Tingkat Kemiskinan
X ₁	Persentase Rumah Tangga Miskin yang Menggunakan Air Layak
X ₂	Persentase Rumah Tangga Miskin yang Menggunakan Jamban Sendiri
X ₃	Persentase Pengeluaran Rata-Rata per Orang untuk Makanan Kategori Miskin
X ₄	Persentase Pengeluaran Rata-Rata per Orang untuk Makanan Kategori Tidak Miskin
X ₅	Persentase Pengeluaran Rata-Rata per Orang untuk Makanan Kategori Miskin dan Tidak Miskin
X ₆	Persentase Penduduk Miskin Usia 15 Tahun ke Atas yang Tidak Bekerja
X ₇	Persentase Penduduk Miskin Usia 15 Tahun ke Atas yang Bekerja Di Sektor Pertanian
X ₈	Persentase Penduduk Miskin Usia 15 Tahun ke Atas yang Bekerja Bukan Di Sektor Pertanian
X ₉	Persentase Penduduk Miskin Usia 15 Tahun ke Atas yang Bekerja Di Sektor Informal
X ₁₀	Persentase Penduduk Miskin Usia 15 Tahun ke Atas yang Bekerja Di Sektor Formal
X ₁₁	Persentase Angka Partisipasi Sekolah (APS) Penduduk Miskin Umur 7-12 Tahun
X ₁₂	Persentase Angka Partisipasi Sekolah (APS) Penduduk Miskin Umur 13-15 Tahun
X ₁₃	Rata-Rata Lama Sekolah

HASIL DAN PEMBAHASAN

Deskriptif Data

Kemiskinan adalah fenomena sosial ekonomi yang mengacu pada kondisi ketika individu atau kelompok masyarakat mengalami kesulitan untuk mengakses sumber daya dasar yang diperlukan untuk memenuhi kebutuhan dasar hidup. Kebutuhan tersebut meliputi pangan, layanan kesehatan, pendidikan dan layanan dasar lainnya. Persentase tingkat kemiskinan dihitung dari persentase penduduk dibawah garis kemiskinan. Garis kemiskinan sendiri merepresentasikan batas minimum pengeluaran atau pendapatan yang dibutuhkan agar seseorang dapat menjalani kehidupan yang memenuhi standar kelayakan hidup minimum. Dalam perhitungan garis kemiskinan, terdapat dua faktor utama, yaitu Garis Kemiskinan Makanan (GKM), mencakup kebutuhan konsumsi pokok, dan Garis Kemiskinan Non-Makanan (GKNM), mencakup pengeluaran untuk pendidikan, perumahan, kesehatan serta berbagai kebutuhan non-pangan lainnya. Untuk memberikan pemahaman yang lebih mendalam mengenai dinamika masalah kemiskinan di Indonesia, berikut grafik yang menggambarkan persentase penduduk miskin di seluruh Kab/kota di Indonesia. (Matdoan, 2020)



Gambar 2. Grafik Persentase Penduduk Miskin Kab/Kota di Indonesia Tahun 2023

Gambar 1. menunjukkan bahwa persentase penduduk miskin bervariasi secara signifikan antar daerah. Persentase Tingkat Kemiskinan terendah ada di Kota Sawahlunto sebesar 2.27%, sedangkan Persentase Tingkat Kemiskinan tertinggi berada di Kab. Intan Jaya sebesar 40.01%. Beberapa wilayah memiliki tingkat kemiskinan yang relatif rendah (di bawah 10%), sementara beberapa lainnya mencapai angka yang lebih tinggi, bahkan melebihi 35%. Wilayah seperti Irian Jaya, Paniai, Deiyai dan Nduga tercatat memiliki tingkat kemiskinan tertinggi, dengan persentase penduduk miskin yang sangat besar. Sebaliknya, wilayah seperti Deli Serdang, Kota Solok, Kota Tangerang dan Kota Pontianak memiliki tingkat kemiskinan yang rendah dibandingkan daerah lainnya. Pola ini mengindikasikan ketidaksetaraan sosial-ekonomi yang terlihat jelas di antara Kabupaten/Kota di Indonesia, dengan beberapa daerah yang masih menghadapi tantangan ekstrem dalam upaya memerangi kemiskinan (Afriliana & Wahyudi, 2022).

Pemodelan MARS

Dalam studi ini, model MARS diterapkan untuk mengidentifikasi variabel-variabel yang berpengaruh terhadap presentase penduduk miskin di Indonesia, dengan memilih kombinasi parameter *Basis Functions* (BF), *Maximum Interaksi* (MI), serta *Minimum Number of Observations* (MO). Jumlah BF yang ditentukan antara 2 sampai 4 kali dari jumlah prediktor, guna memastikan kemampuan model dalam menangkap pola data yang kompleks. Parameter MI diuji pada tingkatan 1, 2, dan 3 untuk mengidentifikasi pengaruh interaksi antarvariabel, sementara MO diatur pada rentang 0 sampai 3 sebagai batas minimum jumlah observasi (RahmadhanI, 2019).

Tabel 2. Kombinasi Parameter MARS dan Nilai GCV

BF	MI	MO	GCV	BF	MI	MO	GCV
26	3	1	0.160211	39	2	1	0.229646
26	3	2	0.16037	39	2	2	0.231602
26	1	0	0.210034	39	2	0	0.256581

26	1	1	0.210065	39	3	0	0.256581
26	1	2	0.210223	39	2	3	0.260154
26	1	3	0.214226	39	3	3	0.331989
26	2	1	0.229646	52	3	1	0.160211
26	2	2	0.231602	52	3	2	0.16037
26	2	0	0.256581	52	1	0	0.210034
26	3	0	0.256581	52	1	1	0.210065
26	2	3	0.260154	52	1	2	0.210223
26	3	3	0.331989	52	1	3	0.214226
39	3	1	0.160211	52	2	1	0.229646
39	3	2	0.16037	52	2	2	0.231602
39	1	0	0.210034	52	2	0	0.256581
39	1	1	0.210065	52	3	0	0.256581
39	1	2	0.210223	52	2	3	0.260154
39	1	3	0.214226	52	3	3	0.331989

Keterangan : *) Model terbaik

Berdasarkan Tabel 2. diperoleh kriteria pemilihan model terbaik pada MARS yaitu model dari BF = 26, MI = 3, MO = 1; BF = 39, MI = 3, MO = 1 ; BF = 52, MI = 3, MO = 1. Model MARS terbaik memiliki bentuk sistematis dari BF = 26, MI = 3, MO = 1 sebagai berikut :

$$\begin{aligned} \hat{f}(x) = & 1.445 - 0.94 BF_1 + 12.95 BF_2 - 7.61 BF_3 - 13.32 BF_4 + 8.52 BF_5 \\ & - 21.55 BF_6 - 3.44 BF_7 - 0.08 BF_8 + 12.68 BF_9 - 0.23 BF_{10} \\ & + 0.251 BF_{11} + 1.17 BF_{12} - 2.95 BF_{13} - 0.97 BF_{14} + 1.06 BF_{15} \end{aligned}$$

Dengan :

- $BF_1 : \max(0, X_5 - 1.203)$
 $BF_2 : \max(0, 1.222 - X_6)$
 $BF_3 : \max(0, X_6 - 1.222)$
 $BF_4 : \max(0, 1.199 - X_7)$
 $BF_5 : \max(0, X_7 - 1.199)$
 $BF_6 : \max(0, X_{13} - 0.317)$
 $BF_7 : \max(0, X_7 - 1.199) \cdot X_{10}$
 $BF_8 : \max(0, 0.944867 - X_5) \cdot \max(0, 1.398 - X_9)$
 $BF_9 : \max(0, X_5 - 1.25) \cdot \max(0, X_{13} - 0.32)$
 $BF_{10} : \max(0, 0.49 - X_5) \cdot \max(0, 0.32 - X_{13})$
 $BF_{11} : \max(0, X_5 - 0.487) \cdot \max(0, 0.32 - X_{13})$
 $BF_{12} : \max(0, X_6 - 0.12) \cdot \max(0, 0.32 - X_{13})$
 $BF_{13} : \max(0, -1.43 - X_4) \cdot \max(0, X_6 - 0.12) \cdot \max(0, 0.32 - X_{13})$
 $BF_{14} : \max(0, X_4 - 1.43) \cdot \max(0, X_6 - 0.116) \cdot \max(0, 0.32 - X_{13})$
 $BF_{15} : \max(0, X_6 - 0.12) \cdot \max(0, -0.75 - X_{12}) \cdot \max(0, 0.32 - X_{13})$

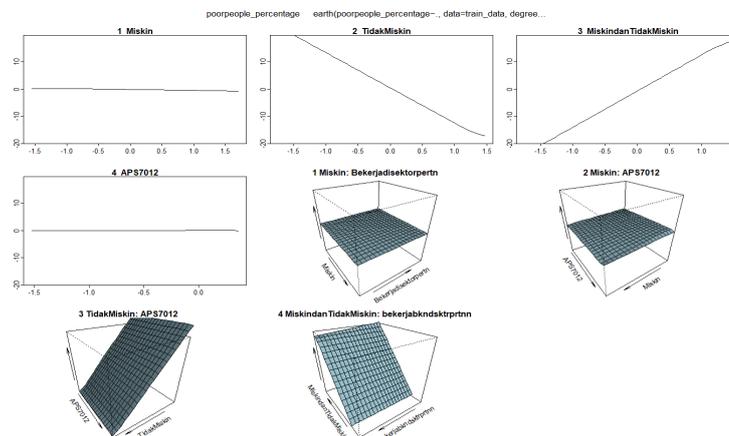
Berdasarkan model di atas, diketahui bahwa ada 8 variabel prediktor yang signifikan dengan Persentase Penduduk Miskin Kab/Kota di Indonesia Tahun 2023. Interpretasi model MARS sama untuk setiap *Basis Functions*, misalnya $BF_1 : \max(0, X_5 - 1.203)$ yang artinya, koefisien BF_1 bernilai negatif (-0.94), yang berarti jika

nilai $X_5 > 1.203$, maka setiap kenaikan satu satuan X_5 dapat menurunkan persentase penduduk berusia 15 tahun keatas yang miskin sebesar 0.94%. (Wicaksono et al., 2014)

Tabel 3. Pentingnya Variabel Prediktor Berdasarkan GCV dan RSS

No	Variabel	<i>nsubsets</i>	GCV	RSS
1	X_4	14	100	100
2	X_5	14	100	100
3	X_{10}	12	73.7	74.7
4	X_{11}	12	73.7	74.7
5	X_3	12	56.9	59.2
6	X_8	12	56.9	59.2
7	X_2	8	21.5	27.6
8	X_7	7	72.1	73.1

Berdasarkan tabel 3. variabel X_4 dan X_5 dengan skor tertinggi sebesar 100% dapat menurunkan nilai GCV model dan menurunkan *Residual Sum of Squares* (RSS) pada model. Selain memanfaatkan model *evimp*, kita bisa menggunakan *Plot Partial Dependence* (PPD) untuk melihat hubungan antar variabel. Plot ini mempermudah dalam memahami hubungan antara prediktor dan fitur bersifat linear, monoton atau lebih kompleks (Christoph Molnar, 2025). Berikut plot *partialnya* : (Vania Grace Sianturi et al., 2021)



Gambar 3. Plot Partial Dependence

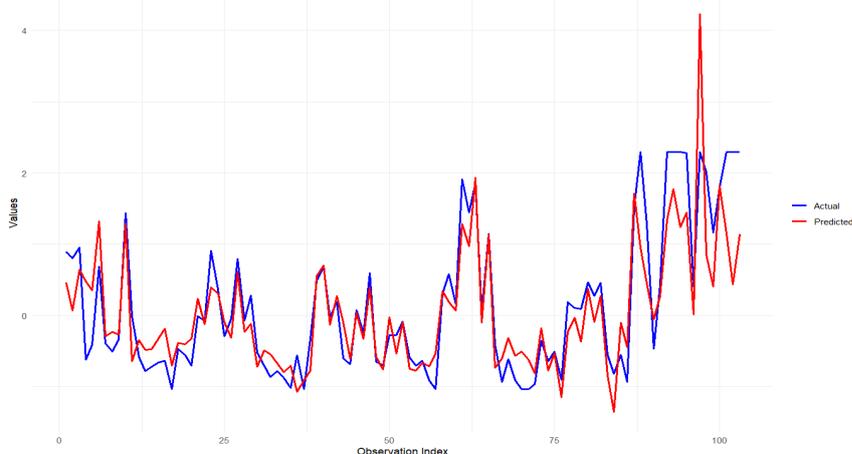
Berdasarkan Gambar 2. bagian atas terdapat plot satu variabel (1D) yang menunjukkan bagaimana perubahan satu fitur seperti (Miskin, Tidak Miskin, dll) mempengaruhi hasil prediksi model Ketika variabel lain dianggap tetap. Sedangkan bagian bawah Gambar 2. Menampilkan plot dua variabel (2D) dalam bentuk grafik 3 dimensi, Dimana memperlihatkan interaksi antar 2 fitur sekaligus misalnya antara variabel X_3 dan X_7 atau variabel X_3 dan X_{11} . Semakin miring plot 2D nya seperti plot variabel Tidak Miskin dan X_{11} menunjukkan bahwa perubahan nilai fitur memberikan pengaruh yang lebih besar terhadap prediksi model, serta menandakan adanya hubungan yang kuat antar fitur dengan respon model. PPD dapat mengidentifikasi fitur mana yang paling berpengaruh dan bagaimana hubungan

antar fitur mempengaruhi hasil prediksi, sehingga dapat membantu dalam interpretasi model MARS.

Evaluasi Akurasi Model

Hasil evaluasi akurasi model MARS memperlihatkan bahwa koefisien determinasi (*R-squared*) mencapai 83.7% dari variansi total variabel prediktor. Angka ini menggambarkan kesesuaian model yang baik dengan data, sehingga hasil prediksi yang didapat dianggap *representatif* dengan pola sesungguhnya. Hanya sekitar 16.3% variasi respon yang tidak ter jelaskan dalam model, yang kemungkinan dipengaruhi oleh faktor lain di luar prediktor dalam model.

Selain *R-squared*, evaluasi model dapat dilihat melalui visualisasi perbandingan antara nilai *actual* dan nilai *predicted* yang disajikan dalam bentuk :



Gambar 4. Perbandingan Nilai *Actual* dan Nilai *Predicted* Data *Testing*

Gambar 3. menunjukkan perbandingan antara nilai aktual (warna biru) dan nilai prediksi (warna merah) terhadap 20% data *testing* Persentase Tingkat Kemiskinan Kab/Kota di Indonesia Tahun 2023. Secara umum, kedua garis mengikuti pola yang sama, menandakan bahwa model MARS mampu menangkap pola tren utama dari data aktual. Meskipun terdapat beberapa titik yang menyimpang dari data aktual Persentase Tingkat Kemiskinan Kab/Kota Indonesia tahun 2023 (Afriliana & Wahyudi, 2022).

Berdasarkan hasil evaluasi di atas, model MARS yang dikembangkan memiliki tingkat akurasi yang memadai untuk memprediksi Persentase Tingkat Kemiskinan Kab/Kota di Indonesia Tahun 2023. Nilai *R-squared* yang baik serta pola prediksi yang mendekati nilai aktual menunjukkan bahwa model dapat digunakan sebagai alat bantu dalam analisis kebijakan dan intervensi pengentasan kemiskinan (Annur et al., 2015).

SIMPULAN DAN SARAN

Dari hasil studi yang dilakukan menggunakan metode *Multivariate Adaptive Regression Spline* (MARS), diperoleh model terbaik untuk memahami dinamika Persentase Tingkat Kemiskinan Kab/Kota di Indonesia dengan parameter $BF = 26$, $MI = 3$, dan $MO = 1$ dengan GCV minimum sebesar 0.160211. Dari model didapat variabel yang berpengaruh secara signifikan terhadap Presentasi Tingkat Kemiskinan Kab/Kota di Indonesia Tahun 2023 yaitu variabel X_5 , X_6 , X_{11} dan X_{12} .

Diharapkan hasil penelitian ini bisa menjadi pertimbangan dalam menentukan program untuk penuntasan masalah kemiskinan di Indonesia (Rahma et al., 2024).

Saran untuk penelitian selanjutnya yaitu mengembangkan model dengan menambah variabel prediktor. Agar model dapat menjelaskan hampir semua faktor penyebab tingginya Persentase Tingkat Kemiskinan Kab/Kota di Indonesia. Selain itu, peneli selanjutnya diharapkan untuk lebih memperhatikan variabel-variabel prediktor yang tidak atau kurang berpengaruh terhadap model (Ardesfira et al., 2022).

DAFTAR RUJUKAN

- Adinugroho, R. (2022). Perbandingan Rasio Split Data Training Dan Data Testing Menggunakan Metode Lstm Dalam Memprediksi Harga Indeks Saham Asia. In *Repository.Uinjkt.Ac.Id*.
[https://repository.uinjkt.ac.id/dspace/handle/123456789/67314%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/67314/1/RAHMADHAN ADINUGROHO-FST.pdf](https://repository.uinjkt.ac.id/dspace/handle/123456789/67314%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/67314/1/RAHMADHAN%20ADINUGROHO-FST.pdf)
- Afriliana, S. N., & Wahyudi, S. T. (2022). Analisis Pertumbuhan Ekonomi Inklusif: Studi Komparasi Antar Provinsi Di Indonesia. *Journal of Development Economic and Social Studies*, 1(1), 44–57.
<https://doi.org/10.21776/jdess.2022.01.1.5>
- Annur, M., Dahlan, J. A., & Agustina, F. (2015). Penerapan Metode *Multivariate Adaptive Regression Spline* (MARS) untuk Menentukan Faktor yang Mempengaruhi Masa Studi Mahasiswa FPMIPA UPI. *Jurnal EurekaMatika*, 3(1), 135–155.
- Ardesfira, G., Zedha, H. F., Fazana, I., Rahmadhiyanti, J., Rahima, S., Anwar, S., Statistika, J., Kuala, U. S., Aceh, B., & Tukar, N. (2022). Jambura Journal Of Probability And Statistics Volume 3 Nomor 2, November 2022. *Jambura Journal of Probability and Statistics*, 3(November 2022).
- Awaludin, H. N. (2010). analisis pengaruh kualitas produk dan kepercayaan merek terhadap keputusan pembelian (Studi kasus pada penggunaan SIM CARD CDMA ESIA). *Jurnal Manajemen Ekonomi Dan Bisnis*.
- Badan Pusat Statistik. (2023). Data dan Informasi Kemiskinan Kabupaten/Kota 2023.
- Fatmawati, B. (2017). *Multivariate Adaptive Regression Spline Untuk Prakiraan Cuaca Jangka Pendek Dengan Pra-Pemrosesan Independent Component Analysis Untuk Prakiraan Cuaca Jangka Pendek Dengan Pra-Pemrosesan Independent Component Analysis*.
- Friedman, J.H. 1991. *Multivariate Adaptive Regression Spline (With Discussion)*, *The Annals of Statistics*, Vol. 19, hal. 1-141.
- Friedman, J. H. (2007). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1). <https://doi.org/10.1214/aos/1176347963>
- Kasse, I. (2017). Memodelkan Faktor-Faktor Yang Mempengaruhi Gizi Buruk Balita Dengan Metode *Multivariate Adaptive Regression Spline* (MARS). *Matematika dan Statistika serta Aplikasinya* Vol 5, No 1.
- M, R. (CNBC I. (n.d.). *Orang Miskin RI Tak Banyak Berkurang Selama 10 Tahun Jokowi*. CNBC Indonesia.

- <https://www.cnbcindonesia.com/research/20240703143812-128-551462/orang-miskin-ri-tak-banyak-berkurang-selama-10-tahun-jokowi>
- Matdoan, Y. (2020). Pemodelan *Multivariate Adaptive Regression Spline* (MARS) Pada Faktor-Faktor yang Mempengaruhi Kemiskinan di Provinsi Maluku dan Maluku Utara. *J Statistika: Jurnal Ilmiah Teori Dan Aplikasi Statistika*, 13(1), 8–14. <https://doi.org/10.36456/jstat.vol13.no1.a3267>
- Nelli, F. (2015). “*Python Data Analytics : Data Analysis and Science Using Pandas, matplotlib, and the Python Programming Language*”. California: Apress Media.
- Nur Fitriyanti Rahmadhani, N. I. M. 15610009. (2019). *Multivariate Adaptive Regression Spline (Mars) Dalam Menentukan Faktor-Faktor Kecelakaan Lalu Lintas Di Kota Yogyakarta*. 163. <https://digilib.uin-suka.ac.id/id/eprint/37188/>
- Rahma, D., Nonong Amalita, Yenni Kurniawati, & Zamahsary Martha. (2024). Application of Multivariate Adaptive Regression Splines for Modeling Stunting Toddler on The Island of Java. *UNP Journal of Statistics and Data Science*, 2(3), 338–343. <https://doi.org/10.24036/ujsds/vol2-iss3/205>
- Rahmadhani, N. F. (2019). *Multivariate Adaptive Regression Spline* (MARS) dalam Menentukan Faktor-Faktor Kecelakaan Lalu Lintas di Kota Yogyakarta. Yogyakarta: UIN Sunan Kalijaga Yogyakarta.
- Septian, M., Martha, S., & Miftahul Huda, ainul. (2023). Analisis Bootstrap Aggregating Multivariate Adaptive Regression Splines Untuk Akreditasi Sekolah SMA/MA di Kalimantan Barat. *Buletin Ilmiah Math. Stat. Dan Terapannya (Bimaster)*, 12(2), 143–150.
- Sinurat, R. P. P. (2023). Analisis Faktor-Faktor Penyebab Kemiskinan Sebagai Upaya Penanggulangan Kemiskinan Di Indonesia. *Jurnal Registratie*, 5(2), 87–103. <https://doi.org/10.33701/jurnalregistratie.v5i2.3554>
- Vania Grace Sianturi, M. Syafii, & Ahmad Albar Tanjung. (2021). Analisis Determinasi Kemiskinan di Indonesia Studi Kasus (2016-2019). *Jurnal Samudra Ekonomika*, 5(2), 125–133. <https://doi.org/10.33059/jse.v5i2.4270>
- Wicaksono, W., Wilandari, Y., Suparti, & Pengajar Jurusan Statistika, S. (2014). Pemodelan *Multivariate Adaptive Regression Splines* (Mars) Pada Faktor-Faktor Resiko Angka Kesakitan Diare (Studi Kasus : Angka Kesakitan Diare Di Jawa Tengah, Jawa Timur Dan Daerah Istimewa Yogyakarta Tahun 2011). *Jurnal Gaussian*, 3(2), 253–262. <http://ejournal-s1.undip.ac.id/index.php/gaussian>