PEMODELAN *MULTIVARIATE ADAPTIVE REGRESSION SPLINE* (MARS) PADA DATA MORBIDITAS PULAU SUMATERA

Skripsi

Oleh

MUHAMAD YUSUP SETIAWAN
NPM. 2117031115

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG

2026

ABSTRACT

MULTIVARIATE ADAPTIVE REGRESSION SPLINE (MARS) MODELING OF MORBIDITY DATA ON THE ISLAND OF SUMATERA

By

MUHAMAD YUSUP SETIAWAN

This study discusses the modeling of morbidity rates on the island of Sumatera using the Multivariate Adaptive Regression Spline (MARS) method. The research data consists of secondary data on morbidity rates in each regency/city on the island of Sumatra in 2023 as the response variable along with seven other predictor variables. The research steps include data preprocessing, building a MARS model based on the minimum GCV value, and determining the variables that affect the morbidity rate on the island of Sumatera.

The modeling resulting from this study is a MARS model with a combination of BF=14, MI=3, MO=15, which has a minimum GCV value of 8,778420 and an MSE of 6,64638. In addition, variable X_6 or the percentage of the population with health complaints was also obtained as the variable that contributed the most to building the MARS model.

Keywords: Modeling, Morbidity, Multivariate Adaptive Regression Spline, GCV.

ABSTRAK

PEMODELAN *MULTIVARIATE ADAPTIVE REGRESSION SPLINE* (MARS) PADA DATA MORBIDITAS PULAU SUMATERA

Oleh

MUHAMAD YUSUP SETIAWAN

Penelitian ini membahas mengenai pemodelan tingkat morbiditas di Pulau Sumatera dengan menggunakan metode *Multivariate Adaptive Regression Spline* (MARS). Data penelitian berupa data sekunder tingkat morbiditas setiap Kabupaten/Kota di Pulau Sumatera tahun 2023 sebagai variabel respon beserta tujuh variabel prediktor lainnya. Langkah penelitian meliputi pra-pemrosesan data, membangun model MARS berdasarkan nilai GCV minimum, dan menentukan variabel yang berpengaruh terhadap tingkat morbiditas Pulau Sumatera.

Pemodelan yang dihasilkan dari penelitian ini adalah model MARS dengan kombinasi BF=14, MI=3, MO=15 yang memiliki nilai GCV minimum sebesar 8,778420 dan MSE sebesar 6,64638. Selain itu, diperoleh juga variabel X_6 atau persentase penduduk yang memiliki keluhan kesehatan sebagai variabel yang memberikan kontribusi terbesar dalam membangun model MARS.

Kata kunci: Pemodelan, Morbiditas, *Multivariate Adaptive Regression Spline*, GCV.

PEMODELAN *MULTIVARIATE ADAPTIVE REGRESSION SPLINE* (MARS) PADA DATA MORBIDITAS PULAU SUMATERA

MUHAMAD YUSUP SETIAWAN

Skripsi

Sebagai Salah Satu Syarat untuk Memperoleh Gelar
SARJANA MATEMATIKA

Pada

Jurusan Matematika

Fakultas Matematika dan Ilmu Pengetahuan Alam

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2026

Judul Skripsi : **PEMODELAN *MULTIVARIATE ADAPTIVE REGRESSION SPLINE* (MARS) PADA DATA MORBIDITAS PULAU SUMATERA**

Nama Mahasiswa : *Muhamad Yusup Setiawan*

Nomor Pokok Mahasiswa : **2117031115**

Program Studi : **Matematika**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI

1. Komisi Pembimbing

Dr. Subian Saidi, S.Si., M.Si.
NIP 198008212008121001

Misgiyati, S.Pd., M.Si.
NIP 198509282023212032

2. Ketua Jurusan Matematika

Dr. Aang Nuryaman, S.Si.,M.Si.
NIP. 197403162005011001

MENGESAHKAN

1. Tim Penguji

Ketua : **Dr. Subian Saidi, S.Si., M.Si.** ....

Sekretaris : **Misgiyati, S.Pd., M.Si.** ....

Penguji
Bukan Pembimbing : **Drs. Nusyirwan, M.Si.** ....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Dr. Eng. Heri Satria, S.Si., M.Si.
NIP. 197110012005011002

Tanggal Lulus Ujian Skripsi: **28 Januari 2026**

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini:

Nama	:	**Muhamad Yusup Setiawan**
Nomor Pokok Mahasiswa	:	**2117031115**
Jurusan	:	**Matematika**
Judul Skripsi	:	**Pemodelan *Multivariate Adaptive Regression Spline* (MARS) Pada Data Morbiditas Pulau Sumatera**

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri. Apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 28 Januari 2026
Penulis,

METERAI TEMPEL
6F8A9ANX355648497

Muhamad Yusup Setiawan
NPM. 2117031115

RIWAYAT HIDUP

Penulis yang bernama lengkap Muhamad Yusup Setiawan lahir di Desa Tanjung Agung, Kecamatan Katibung, Lampung Selatan pada tanggal 17 Januari 2003. Penulis merupakan anak pertama dari tiga bersaudara, putra dari pasangan Bapak Suhaidi dan Ibu Suryani. Riwayat pendidikan penulis dimulai dari TK Al-Amanah yang diselesaikan pada tahun 2009. Setelah itu, penulis melanjutkan ke SD Negeri 2 Tanjung Ratu dan berhasil menamatkannya pada tahun 2015. Pendidikan berikutnya ditempuh di SMP Negeri 1 Katibung yang lulus pada tahun 2018, kemudian meneruskan pendidikan menengah atas di SMA Negeri 1 Katibung hingga lulus pada tahun 2021. Pada tahun yang sama, penulis diterima di Universitas Lampung, Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, untuk melanjutkan pendidikan tinggi. Selama masa perkuliahan, penulis aktif berpartisipasi dalam berbagai kegiatan organisasi dan kemahasiswaan, di antaranya sebagai anggota Himpunan Mahasiswa Matematika (HIMATIKA) pada periode 2022–2023 serta anggota BEM Universitas pada periode 2023–2024. Selain itu, Penulis juga pernah mengikuti kegiatan kepanitiaan di luar kampus.

KATA INSPIRASI

”Hasil tak selalu sejalan dengan usaha, tetapi berani berusaha selalu lebih bermakna daripada diam tanpa mencoba.”

PERSEMBAHAN

Dengan mengucap Alhamdulillah dan syukur kepada Allah SWT atas nikmat serta hidayah-Nya sehingga skripsi ini dapat terselesaikan dengan baik dan pada waktu yang tepat. Dengan rasa syukur dan Bahagia, saya persembahkan rasa terimakasih saya kepada:

Ayah dan Ibuku Tercinta

Ucapan terima kasih yang tak terhingga penulis sampaikan kepada kedua orang tua tercinta atas kasih sayang, motivasi, dukungan, serta doa yang tidak pernah putus, sehingga penulis dapat melalui berbagai tantangan selama masa perkuliahan. Setiap pelajaran yang diberikan menjadi penguat langkah bagi penulis hingga skripsi ini dapat diselesaikan dengan baik.

Dosen Pembimbing dan Pembahas

Terimakasih kepada dosen pembimbing dan pembahas yang sudah sangat membantu, memberikan motivasi, memberikan arahan serta ilmu yang berharga.

Sahabat-sahabatku

Terimakasih kepada semua orang-orang baik yang telah memberikan pengalaman, semangat, motivasinya, serta doa-doanya dan senantiasa memberikan dukungan dalam hal apapun.

Almamater Tercinta

Universitas Lampung

SANWACANA

Alhamdulillah, puji dan syukur penulis panjatkan kepada Allah SWT atas limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini yang berjudul ”Pemodelan *Multivariate Adaptive Regression Spline* (MARS) Pada Data Morbiditas Pulau Sumatera” dengan baik dan pada waktu yang tepat. Shalawat serta salam semoga senantiasa tercurahkan kepada Nabi Muhammad SAW.

Dalam proses penyusunan skripsi ini, banyak pihak yang telah membantu memberikan bimbingan, dukungan, arahan, motivasi serta saran sehingga skripsi ini dapat terselesaikan. Oleh karena itu, dalam kesempatan ini penulis mengucapkan terimakasih kepada:

1. Bapak Dr. Subian Saidi, S.Si., M.Si. selaku Pembimbing I yang telah banyak meluangkan waktunya untuk memberikan arahan, bimbingan, motivasi, saran serta dukungan kepada penulis sehingga dapat menyelesaikan skripsi ini.
2. Ibu Misgiyati, S.Pd., M.Si. selaku pembimbing II yang telah memberikan arahan, saran, motivasi, serta dukungan kepada penulis sehingga dapat menyelesaikan skripsi ini
3. Bapak Drs. Nusyirwan, M.Si. selaku Penguji yang telah bersedia memberikan kritik dan saran yang membangun serta evaluasi kepada penulis sehingga dapat menjadi lebih baik.
4. Bapak Tiryono, M.Sc., Ph.D. selaku dosen pembimbing akademik.
5. Bapak Dr. Ahmad Faisol, S.Si., M.Sc. selaku Sekretaris Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

6. Bapak Dr. Aang Nuryaman, S.Si., M.Si. selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

7. Ibu Anita selaku admin Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung yang telah membantu penulis dalam mempersiapkan berkas seminar maupun wisuda.

8. Seluruh dosen Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung yang tidak dapat penulis uraikan satu persatu karena telah memberikan ilmu serta masukan kepada penulis selama masa perkuliahan.

9. Seluruh karyawan dan staff jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

10. Teristimewa untuk kedua orangtua tercinta, Papa Suhaidi yang senantiasa berusaha sekuat tenaga memenuhi kebutuhan penulis selama menghadapi lika-liku drama kehidupan pada masa perkuliahan serta Mama Suryani yang tiada hentinya memberikan dukungan berupa doa, motivasi, dan selalu hadir disaat Penulis kehilangan semangat dalam menjalani perkuliahan. Meskipun tidak pernah merasakan duduk di bangku perkuliahan, namun mereka sangat mengusahakan supaya anaknya mengenyam pendidikan lebih baik dibanding yang mereka alami sebelumnya. Penulis bangga dan sangat berterima kasih kepada mereka. Tanpa pengorbanan yang mereka berikan, mungkin saja penulis tidak dapat merasakan semua yang didapat pada saat ini. Oleh karena itu, karya ini penulis persembahkan khusus untuk mereka sebagai bukti kalau pengorbanan yang telah mereka berikan tidak sia-sia.

11. Kedua adik Penulis Dila Putri Apriani dan Marvel Agus Rhomadanu yang telah memberikan dukungan kepada penulis dalam menjalani perkuliahan dari awal sampai akhir.

12. Seluruh keluarga besar yang selalu memberikan doa, motivasi, dan semangat selama masa perkuliahan.

13. Rekan seperjuangan, Ilham, Yudhi Darmawan, Rhizanu Akbar, dan Zulfakhri

Dwiantara yang telah hadir dan membersamai penulis sejak awal perkuliahan hingga terselesaikannya skripsi ini, penulis menyampaikan terima kasih atas segala dukungan, motivasi, serta bantuan yang telah diberikan. Semoga kalian senantiasa memperoleh kemudahan dalam meraih cita-cita dan kesuksesan di masa mendatang.

14. Mahasiswa Jurusan Matematika, terutama teman-teman angkatan 2021 yang telah membersamai Penulis selama perkuliahan di kampus.
15. Semua pihak yang telah memberikan dukungan dalam berbagai bentuk selama proses penyusunan skripsi ini, namun tidak dapat disebutkan satu per satu.
16. Almamater Universitas Lampung yang telah menjadi tempat untuk belajar, berkembang, mencari pengalaman dan menimba ilmu selama masa perkuliahan.

Semoga skripsi ini dapat bermanfaat bagi kita semua. Penulis menyadari bahwa skripsi ini masih jauh dari sempurna, sehingga penulis mengharapkan kritik dan saran yang membangun untuk menjadikan skripsi ini lebih baik lagi.

Bandar Lampung, 28 Januari 2026

Muhamad Yusup Setiawan

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB I

PENDAHULUAN

1.1 Latar Belakang & Masalah

Analisis regresi adalah metode populer dalam statistika yang digunakan untuk mengetahui pola hubungan antara variabel respon dan variabel prediktor. Tujuan utamanya adalah membuat sebuah model atau fungsi yang sesuai untuk melakukan prediksi berdasarkan pengaruh variabel prediktor terhadap variabel respon (Mendenhall & Sincich, 2012). Banyak bidang ilmu pengetahuan seperti bidang ekonomi, kesehatan bahkan sosial menerapkan regresi sebagai alat untuk melakukan analisis data. Dalam membangun model, analisis regresi pada umumnya menggunakan dua pendekatan yaitu pendekatan secara parametrik dan nonparametrik. Pendekatan parametrik memerlukan banyak asumsi yang harus dipenuhi, terkhusus bentuk kurva perlu diketahui dari awal. Sedangkan dalam dataset nyata, tidak semua bentuk kurva diketahui polanya dari awal. Oleh karena itu, pendekatan nonparametrik lebih baik digunakan pada kasus tersebut (Montgomery *et al.*, 2021).

Regresi nonparametrik merupakan pendekatan regresi yang dipakai ketika bentuk pola kurvanya tidak diketahui dari awal. Diperlukan sebuah fungsi tertentu untuk mengestimasi bentuk kurva dengan pola yang tidak beraturan. Pendekatan fungsi yang biasanya digunakan untuk memodelkan regresi nonparametrik yaitu fungsi *spline*, fungsi *kernel*, dan deret *fourier* (Rahmawati *et al.*, 2021). Salah satu pendekatan fungsi untuk memodelkan bentuk kurva regresi nonparametrik yang banyak digunakan adalah regresi nonparametrik dengan pendekatan fungsi *spline*.

Regresi *spline* adalah metode analisis regresi nonparametrik yang dapat digunakan untuk mengestimasi model nonparametrik dengan data yang memiliki pola berubah-ubah pada subinterval tertentu. *Spline* merupakan suatu fungsi yang berasal dari potongan-potongan (*piecewise*) fungsi polinomial di mana untuk masing-masing potongan fungsi polinomial yang tersegmen dibentuk oleh suatu titik knot. Titik knot sendiri adalah titik pertemuan di mana terjadinya perubahan perilaku pola kurva dalam interval yang berbeda. Model *spline* terbaik dengan titik knot optimal ditunjukkan oleh nilai GCV (*Generalized Cross Validation*) yang minimum (Widyastuti *et al.*, 2021). Berdasarkan kelebihannya, regresi nonparametrik menggunakan pendekatan fungsi *spline* lebih fleksibel dibandingkan dengan fungsi polinomial pada umumnya.

Seiring berjalannya waktu, metode regresi nonparametrik dengan pendekatan fungsi *spline* semakin berkembang. Salah satu metode yang juga menggunakan fungsi *spline* adalah metode *Multivariate Adaptive Regression Spline* (MARS). Diperkenalkan pertama kali oleh Friedman pada tahun 1991, MARS adalah metode regresi nonparametrik yang memiliki bentuk fleksibel karena dapat menangani perubahan perilaku data pada subinterval tertentu dengan adanya titik knot menggunakan kombinasi antara fungsi *spline* dan Regresi Partisi Rekursif (RPR) (Sriningsih *et al.*, 2021). RPR memiliki keterbatasan yaitu dalam menghasilkan model yang tidak kontinu pada titik knot. Dalam hal ini, metode MARS dapat menutupi keterbatasan tersebut dan mengidentifikasi adanya fungsi linear aditif. Pemilihan model MARS terbaik dapat menggunakan algoritma *stepwise* (*forward* dan *backward*). Tujuan dari algoritma ini adalah membangkitkan model dan kemudian memangkas model kembali ke model yang lebih efektif (Damayanti & Sunendiari, 2021). Selain itu, metode MARS juga berguna untuk mengatasi data yang berdimensi tinggi dengan variabel prediktor sebanyak $3 \leq x \leq 20$ dan data sampel sebanyak $50 \leq n \leq 1000$ (Mattalunru *et al.*, 2022).

Menurut BPS RI (2019), morbiditas dimaknai sebagai persentase penduduk suatu wilayah yang memiliki gangguan terhadap kondisi fisik termasuk karena kecelakaan atau hal lain yang menyebabkan terganggunya kegiatan sehari-hari. Pada umumnya keluhan utama yang banyak dialami oleh penduduk adalah panas, sakit kepala, batuk, flu, diare, sesak nafas, dan sakit gigi. Berdasarkan data yang tercatat pada BPS RI, Pulau Sumatera menyumbang 20,77% penyebaran morbiditas dari total jumlah penduduk Indonesia yang mengalami keluhan kesehatan. Dengan kata lain, ada sebanyak 6.382.118 orang penduduk di Pulau Sumatera yang mengalami keluhan kesehatan pada tahun 2023 dari total 30.727.693 orang Indonesia yang mengalami keluhan kesehatan. Angka tersebut tepat berada di bawah Pulau Jawa dengan 17.451.033 orang penduduk yang mengalami keluhan kesehatan atau 56,79% menyumbang angka morbiditas dari total morbiditas nasional. Jumlah yang mencapai enam juta penduduk yang mengalami keluhan kesehatan bukanlah angka yang kecil, sehingga perlu ada langkah serius untuk menurunkan angka tersebut. Dibandingkan dengan Pulau Jawa, Pulau Sumatera memiliki keragaman kondisi sosial, ekonomi, lingkungan, dan akses terhadap pelayanan kesehatan yang lebih tinggi. Oleh karena itu, penelitian mengenai morbiditas di Pulau Sumatera penting untuk dilakukan.

Penelitian sebelumnya dengan menggunakan metode MARS pernah dilakukan oleh Mattalunru *et al.* (2022) untuk mengetahui faktor yang memengaruhi curah hujan di Kota Makassar. Penelitian tersebut menghasilkan model terbaik dari kombinasi BF = 12, MI = 1, MO = 1 dengan nilai GCV minimum sebesar 31,14 dan nilai R^2 sebesar 81,20%. Diperoleh pula faktor yang berpengaruh secara signifikan terhadap curah hujan di Kota Makassar yaitu suhu udara, kelembaban udara, kecepatan angin, dan tekanan udara. Sementara itu, penelitian sebelumnya yang mengambil topik pembahasan tentang data morbiditas di Jawa Tengah pernah dilakukan oleh Rosanti & Budiantara menggunakan metode regresi nonparametrik *spline*. Penelitian tersebut menghasilkan model dengan kombinasi titik knot 2,3,2,3,3,3 serta variabel yang memengaruhinya. Penelitian berikutnya yang membahas data morbiditas Provinsi Sumatera Utara adalah penelitian yang dilakukan oleh Haquel *et al.* (2024) dengan menggunakan metode regresi nonparametrik *spline*. Hasil yang diperoleh dari penelitian tersebut adalah model *spline* dengan kombinasi titik knot 3,3,1,3,3,3,3 serta variabel yang memiliki peran penting dalam meningkatnya morbiditas di daerah Provinsi Sumatera Utara.

Berdasarkan penelitian sebelumnya, pembahasan data morbiditas menggunakan metode MARS di Pulau Sumatera secara umum belum pernah dilakukan. Selain itu, kondisi sosial, ekonomi, lingkungan, dan akses ke pelayanan kesehatan yang cenderung heterogen dibanding Pulau Jawa juga turut menjadi alasan yang melatarbelakangi penulis melakukan penelitian dengan judul ”Pemodelan *Multivariate Adaptive Regression Spline* (MARS) pada Data Morbiditas Pulau Sumatera”. Hasil penelitian ini diharapkan dapat digunakan oleh pemangku kebijakan sebagai petunjuk dalam meningkatkan derajat kesehatan masyarakat.

1.2 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Membuat model *Multivariate Adaptive Regression Spline* (MARS) pada data morbiditas Pulau Sumatera.
2. Mengetahui faktor-faktor yang memengaruhi tingkat morbiditas di Pulau Sumatera.

1.3 Manfaat Penelitian

Hasil dari penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Sebagai pengetahuan dan wawasan mengenai penerapan metode *Multivariate Adaptive Regression Spline* (MARS) dalam membuat model pada data morbiditas Pulau Sumatera, serta mengetahui faktor- faktor yang memengaruhinya.
2. Dapat digunakan sebagai referensi bagi penelitian selanjutnya.

BAB II

TINJAUAN PUSTAKA

2.1 Analisis Regresi

Analisis regresi merupakan salah satu teknik dalam statistika yang digunakan untuk mengetahui pola antara variabel respon dan variabel prediktor. Tujuan utama dari melakukan analisis regresi adalah untuk membangun suatu fungsi atau model yang mengestimasi bentuk kurva regresi dengan tepat. Terdapat dua pendekatan yang digunakan dalam analisis regresi, yaitu pendekatan parametrik dan pendekatan nonparametrik (Mendenhall & Sincich, 2021). Berikut adalah bentuk umum persamaan regresi parametrik.

$$Y = \beta_0 + \beta_i X_i + \varepsilon_i, \quad i = 1, 2, \cdots, n \tag{2.1.1}$$

keterangan:

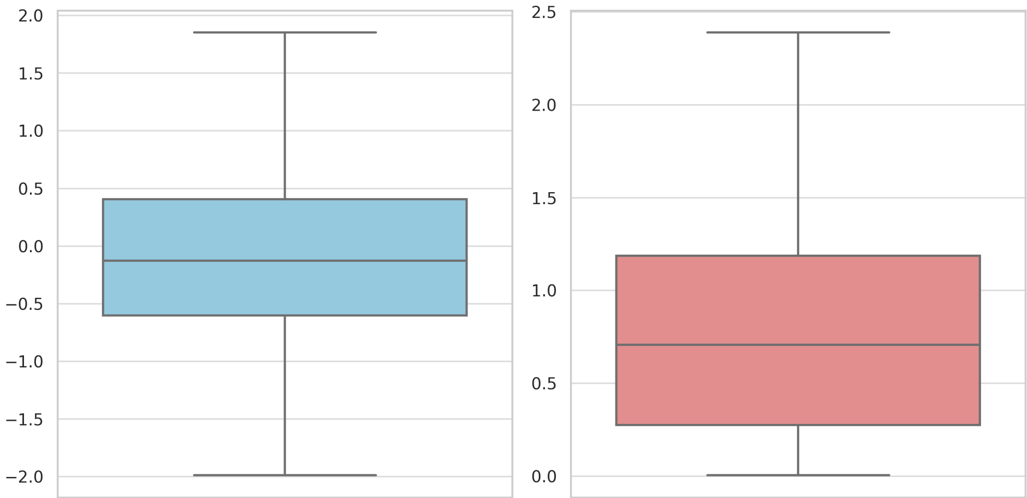
Y = variabel respon,
X_i = variabel prediktor,
β_0 = konstanta,
β_i = parameter (koefisian regresi),
ε_i = galat yang diasumsikan identik, independen, dan berdistribusi normal dengan $\mu = 0$ dan σ^2.

Pendekatan parametrik mengasumsikan bahwa pola kurva diketahui bentuk awalnya seperti linear, kuadratik, kubik, eksponensial dan sebagainya. Selain itu, pada pendekatan parametrik ada beberapa asumsi yang harus dipenuhi yaitu sebagai berikut (Mardiatmoko, 2020):

1. Normalitas

   Uji Normalitas adalah prosedur statistik yang digunakan untuk melihat apakah data sampel yang diambil dari suatu populasi mengikuti distribusi normal. Kegagalan dalam mengidentifikasi normalitas dapat menyebabkan hasil dan kesimpulan tidak akurat. Terdapat dua metode dalam melakukan uji normalitas yaitu dengan metode grafis atau visual dan uji normalitas secara statistik (Paramasivam *et al*, 2024).

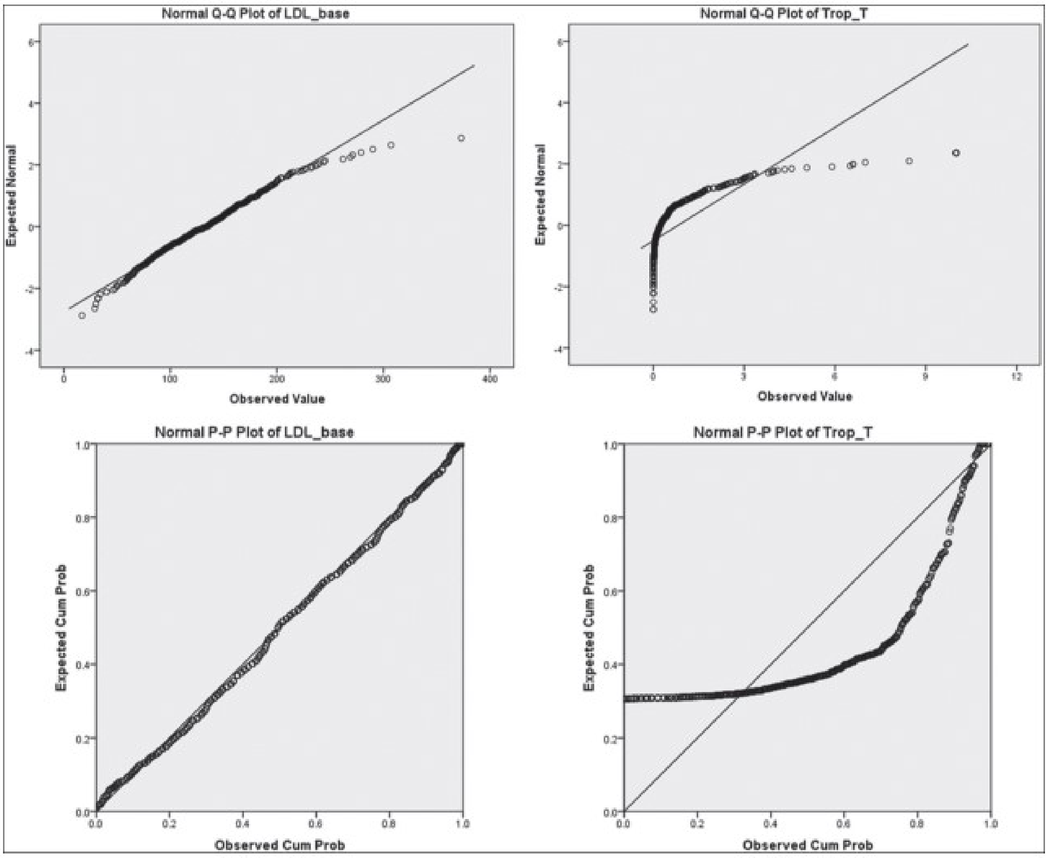
   Dalam metode grafis dapat menggunakan Box-Plot untuk melihat kenormalitasan data. Plot ini memberikan lima informasi dari dataset yaitu kuartil pertama (Q_1), median (Q_2), kuartil ketiga (Q_3), nilai minimum, dan nilai maksimum. Plot ini juga dapat menampilkan pencilan pada data.



Gambar 2.1 Ilustrasi Box-Plot

   Box-Plot yang disajikan pada Gambar 2.1 sebelah kiri menjelaskan bahwa data berdistribusi normal dengan median yang berada di tengah kotak dan garis penyebaran data cenderung simetris. Sedangkan Gambar 2.1 sebelah kanan menjelaskan bahwa data memiliki penyebaran yang tidak simetris walaupun median berada di tengah kotak.

   Metode grafis yang kedua adalah dengan plot kuantil-kuantil atau Q-Q Plot dan plot persentil-persentil atau P-P Plot. Cara menginterpretasi kedua plot ini adalah dengan melihat penyebaran titik-titik apakah sejajar dengan garis lurus pada plot. Ilustrasi Q-Q Plot dan P-P Plot disajikan pada Gambar 2.2 di bawah ini.



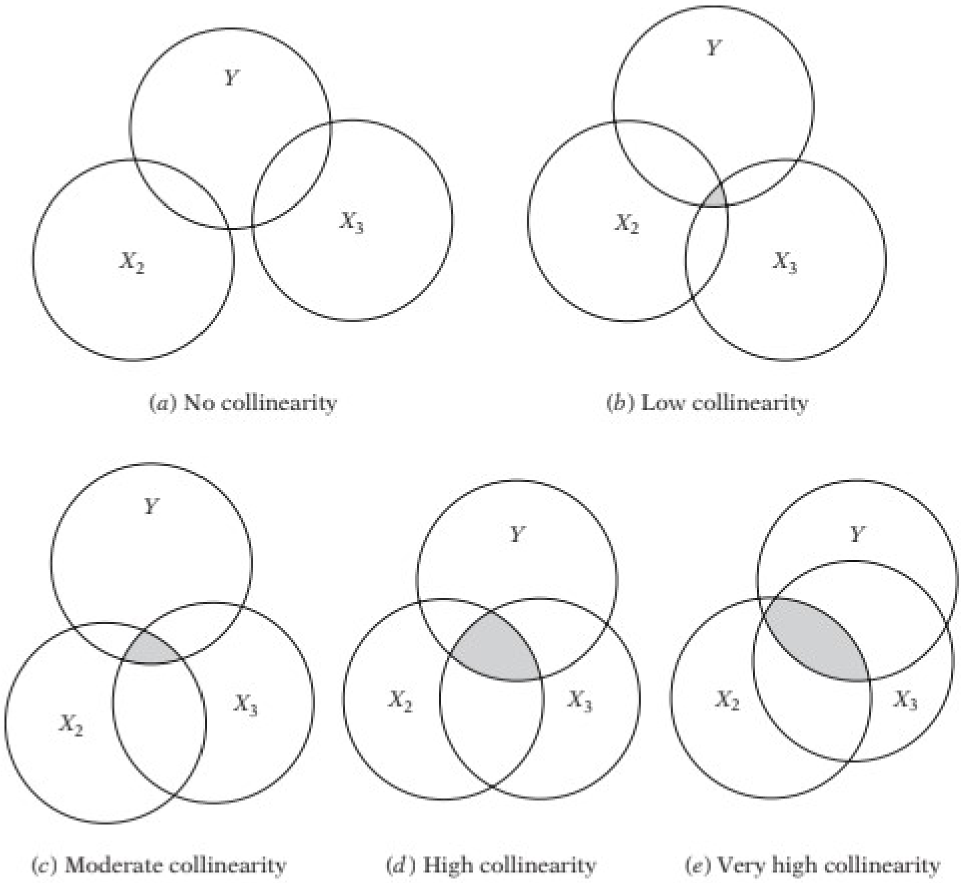
Gambar 2.2 Ilustrasi Q-Q Plot & P-P Plot

Pada Gambar 2.2 sebelah kiri menampilkan penyebaran titik-titik pada plot sejajar dan mendekati garis. Hal ini menyimpulkan bahwa data mengikuti distribusi normal. Sedangkan pada Gambar 2.2 sebelah kanan menunjukkan data yang tidak berdistribusi normal karena penyebaran titik-titik pada plot tidak sejajar dan jauh dari garis.

Metode kedua dalam menguji kenormalitasan pada data adalah dengan melakukan uji secara statistik. Uji statistik yang umum digunakan untuk normalitas adalah uji Shapiro-Wilk, Anderson-Darling, dan Kolmogorov-Smirnov. Hipotesis nol yang digunakan dalam uji hipotesis normalitas adalah data berdistribusi normal dengan hipotesis alternatifnya adalah data tidak berdistribusi normal. Jika *p-value* yang diperoleh lebih besar dari taraf signifikansi, maka hipotesis nol gagal ditolak, sehingga bisa ditarik kesimpulan bahwa data mengikuti distribusi normal.

2. Multikolinearitas

   Multikolinearitas adalah keadaan di mana terdapat hubungan linear yang kuat antar variabel prediktor dalam model. Ilustrasi untuk melihat apakah suatu model mengalami multikolinearitas atau tidak digambarkan dengan diagram Venn yang disajikan pada Gambar 2.3 berikut.



Gambar 2.3 Ilustrasi Multikolinearitas

Pada Gambar 2.3 bagian (a) tidak terjadi irisan antara X_2 dan X_3 yang artinya model tidak terindikasi multikolinearitas. Sedangkan pada bagian (b) hingga bagian (e) terjadi irisan antara X_2 dan X_3 yang menandakan terjadinya multikolinearitas (semakin besar irisan, maka semakin besar pula multikolinearitas yang terjadi) (Gujarati & Porter, 2009).

Adanya multikolinearitas dapat menyebabkan selang kepercayaan menjadi lebih lebar dan menghasilkan nilai *likelihood* yang kurang kuat untuk variabel prediktor. Hal ini akan menyebabkan model yang dihasilkan tidak dapat dipercaya (Shresta, 2020). Beberapa cara yang dapat dilakukan untuk mendeteksi multikolinearitas adalah dengan koefisien korelasi, *Variance Inflation Vector* (VIF) dan metode nilai eigen. Salah satu cara yang umum digunakan untuk mendeteksi multikolinearitas adalah dengan nilai VIF. Persamaan VIF ditampilkan dalam persamaan (2.1.2) sebagai berikut.

$$VIF = \frac{1}{1 - R^2} \tag{2.1.2}$$

dengan R^2 adalah koefisien determinasi dengan persamaan

$$R^2 = \frac{\Sigma_{i=1}^{n}(\hat{X}_i - \bar{X})^2}{\Sigma_{i=1}^{n}(X_i - \bar{X})^2} \tag{2.1.3}$$

Apabila nilai VIF menunjukkan angka lebih besar dari 10, maka mengindikasikan bahwa terdapat hubungan antar variabel prediktor atau terjadi multikolinearitas (Mendenhall & Sincich, 2021).

3. Heteroskedastisitas
   Salah satu asumsi dalam analisis regresi yang penting adalah memiliki varians galat yang konsisten atau dikenal dengan istilah homoskedastisitas. Kondisi dimana galat dalam model regresi memiliki varians yang berbeda disebut heteroskedastisitas. Heteroskedastisitas dapat dideteksi dengan berbagai metode pengujian seperti uji Glejser, uji korelasi rank Spearman, uji Goldfeld-Quandt, uji Koenker-Basselt, Uji White, dan uji Breusch-Pagan-Godfrey (Gujarati & Porter, 2009). Metode yang paling umum digunakan dalam uji heteroskedastisitas adalah uji White dan uji Breusch-Pagan-Godfrey. Salah satu cara untuk mengatasi heteoskedastisitas adalah dengan melakukan transformasi log pada variabel respon (Idowu *et al*, 2024).

4. Autokorelasi
   Autokorelasi adalah keadaan terjadinya korelasi antara galat pada suatu periode dan galat pada periode sebelumnya. Adanya autokorelasi pada galat akan menghasilkan estimator yang tidak efisien, yaitu selang kepercayaan menjadi lebih lebar. Uji yang umum digunakan untuk mendeteksi autokorelasi yaitu uji Durbin-Watson yang ditampilkan pada persamaan (2.1.4) di bawah ini (Gujarati & Porter, 2009).

$$d = \frac{\Sigma_{t=2}^{n}(\varepsilon_t - \varepsilon_{t-1})^2}{\Sigma_{t=1}^{n}\varepsilon_t^2} \tag{2.1.4}$$

keterangan:

ε_t = galat pada periode t,

ε_{t-1} = galat pada periode $t - 1$,

n = jumlah pengamatan.

Pada Uji Durbin-Watson terdapat batas atas d_U dan batas bawah d_L. Kriteria pengujian Durbin-Watson disajikan pada Tabel 2.1 berikut.

Tabel 2.1 Kriteria Keputusan Durbin-Watson

Hipotesis Nol	Keputusan	Jika
Tidak ada autokorelasi positif	Tolak	$0 < d < d_L$
Tidak ada autokorelasi positif	Tidak ada keputusan	$d_L < d < d_U$
Tidak ada autokorelasi negatif	Tolak	$4 - d_L < d < 4$
Tidak ada autokorelasi negatif	Tidak ada keputusan	$4 - d_U \leq d \leq 4 - d_L$
Tidak ada autokorelasi positif atau negatif	Tidak tolak	$d_U < d < 4 - d_U$

Gujarati (2009) dalam bukunya menjelaskan metode untuk mengatasi autokorelasi yaitu dengan metode *Generalized Least Square* (GLS) dan metode Newey-West.

Berbeda dengan pendekatan parametrik, pada pendekatan nonparametrik tidak memerlukan asumsi yang spesifik seperti pada pendekatan regresi secara parametrik.

2.2 Regresi Nonparametrik

Regresi nonparametrik adalah salah satu pendekatan yang digunakan untuk mengetahui pola hubungan antara variabel respon dan prediktor yang tidak diketahui bentuk pola dari kurva regresinya atau tidak terdapat informasi masa lalu tentang bentuk pola data. Model regresi nonparametrik dapat ditunjukkan pada persamaan sebagai berikut (Al-Azies & Hapsery, 2019).

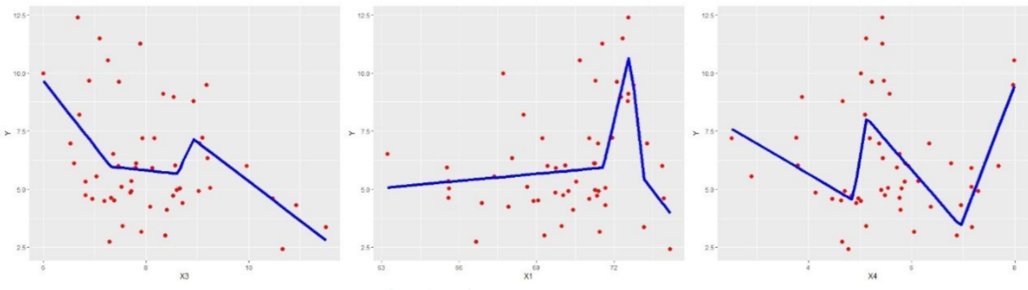
$$y_i = f(x_i) + \varepsilon_i,\ i = 1, 2, ..., n \tag{2.2.5}$$

keterangan:
y_i = variabel respon,
$f(x_i)$ = fungsi pemulusan yang tidak diketahui ke-i,
x_i = variabel prediktor,
ε_i = galat yang diasumsikan identik, independen, dan berdistribusi normal dengan mean nol dan varians σ^2.

Ada beberapa model pendekatan dalam regresi nonparametrik untuk mengestimasi bentuk kurva diantaranya yaitu, *Kernel*, *Spline*, *Wavelet*, deret *Fourier*, polinomial lokal, dan lain-lain. Salah satu pendekatan regresi nonparametrik yang memiliki interpretasi statistik dan visual yang cukup baik adalah dengan menggunakan model *spline* (Rahmawati *et al.*, 2021).

2.3 Regresi Nonparametrik *Spline*

Spline merupakan salah satu jenis *piecewise* polinomial yang memiliki sifat tersegmen atau terpotong-potong. Sifat tersegmen dari model *spline* ini memberikan fleksibilitas tinggi sehingga mampu menangani pola data yang menunjukkan pola naik atau turun secara tajam dengan bantuan titik-titik knot dan menghasilkan kurva yang relatif mulus serta dapat menggambarkan perubahan pola-pola dari fungsi subinterval tertentu (Wahyuni *et al.*, 2020). Membangun model *spline* perlu memperhatikan beberapa hal, yakni menentukan jumlah derajat dari model regresi, jumlah titik knot, dan lokasi titik knot berada.



Gambar 2.4 Ilustrasi Model *Spline* dengan 3 titik knot

Berdasarkan Gambar 2.4, pendekatan *spline* dapat memberikan gambaran mengenai bentuk pola dari data. Gambar menunjukkan model dengan derajat satu dan memiliki tiga titik knot. Semakin tinggi derajat polinomial model, maka semakin mulus pula perubahan pola data yang terbentuk. Secara umum, model regresi *spline* ditunjukkan pada persamaan sebagai berikut (Al-Azies & Hapsery, 2019).

$$f(x_i) = \sum_{j=0}^{q} \beta_j x_i^j + \sum_{k=1}^{K} \beta_{k+q}(x_i - t_k)_+^q \tag{2.3.6}$$

dengan,

$$(x_i - K_k)_+^q = \begin{cases} (x_i - t_k)^q, & x_i \geq t_k \\ 0, & x_i < t_k \end{cases} \tag{2.3.7}$$

keterangan:
$f(x_i)$ = fungsi regresi *spline*,
x_i = variabel prediktor,
q = orde polinomial,
$t_1, t_2, ..., t_K$ = titik knot,
β = konstanta.

2.4 *Multivariate Adaptive Regression Spline* (MARS)

Multivariate Adaptive Regression Spline (MARS) diperkenalkan pertama kali oleh Jerome H. Friedman (1991). MARS adalah salah satu metode regresi nonparametrik yang digunakan untuk memprediksi suatu nilai dari variabel respon numerik dan juga dapat digunakan untuk klasifikasi variabel respon kategorik dari sekumpulan variabel prediktor numerik. MARS merupakan model regresi yang diperoleh meskipun hubungan antara variabel prediktor dan variabel respon sulit untuk didekati dengan model parametrik. Model MARS berfokus untuk mengatasi permasalahan data berdimensi tinggi (*curse of dimensionality*), yaitu data yang memiliki variabel prediktor sebanyak $3 \leq x \leq 20$ dengan ukuran sampel sebesar $50 \leq n \leq 1000$. MARS menghasilkan model yang kontinu dalam knot. Dalam pemilihan model MARS dengan knot optimum dapat dilihat berdasarkan nilai GCV (*Generalized Cross Validation*)(Yasmirullah *et al.*, 2020).

MARS merupakan salah satu pendekatan untuk regresi nonparametrik multivariat antara variabel respon dan beberapa variabel prediktor pada regresi yang memiliki sifat tersegmen (*piecewise regression*). Jika suatu garis regresi tidak dapat menjelaskan data secara keseluruhan, maka beberapa garis regresi atau region digunakan untuk menjelaskan data secara keseluruhan dari variabel respon. Perubahan pola dari satu region ke region lainnya dihubungkan oleh titik knot. Dengan kata lain, knot adalah akhir dari suatu region dan awal dari region lainnya. Pada setiap knot yang terbentuk diharapkan adanya kontinuitas antar satu region dengan region lainnya (Friedman, 1991).

Terdapat beberapa istilah penting dalam membangun model MARS antara lain sebagai berikut (Damayanti & Sunendiari, 2021):

1. Knot
   Knot adalah nilai variabel prediktor ketika *slope* (kemiringan) suatu garis regresi mengalami perubahan pola. Dengan kata lain knot dapat didefinisikan juga sebagai akhir dari suatu garis regresi dan awal bagi garis regresi yang lain.

Knot berperan penting dalam menentukan bagaimana suatu model akan menyesuaikan data. Jarak minimum antar knot atau Minimum Observasi (MO) berperan signifikan dalam menentukan kompleksitas model. Penentuan nilai MO dapat menggunakan persamaan yang dikemukakan oleh Friedmann (1991) sebagai berikut.

$$MO = 3 - log_2(\frac{\alpha}{n}) \tag{2.4.8}$$

dimana α merupakan taraf signifikansi $0,05$ dan n adalah jumlah pengamatan.

2. Basis Fungsi (BF)
   Basis fungsi merupakan suatu fungsi untuk menjelaskan hubungan antara variabel respon dan prediktor yang terdiri dari satu atau lebih variabel prediktor. Pada umumnya, basis fungsi yang dipilih berbentuk fungsi linear yang kontinu pada setiap titik knot. Nilai maksimum basis fungsi yang diperbolehkan adalah dua sampai empat kali jumlah variabel prediktor yang digunakan.

3. Interaksi
   Interaksi adalah perkalian silang (*cross product*) antar variabel yang saling berkorelasi. Jumlah maksimum interaksi (MI) yang disarankan yaitu 1, 2, dan 3. Apabila MI lebih dari 3 maka akan menyebabkan model semakin kompleks dan sulit untuk diinterpretasikan dan meningkatkan nilai GCV.

MARS merupakan kombinasi yang kompleks antara metode *spline* dan *Recursive Partitioning Regression* (RPR). Model MARS digunakan untuk mengatasi kelemahan dari RPR yaitu menghasilkan model yang kontinu pada knot dan dapat mengidentifikasi adanya fungsi linier dan aditif (Hafifi *et al.*, 2021). Hasil modifikasi model RPR dengan kombinasi fungsi *spline* adalah model MARS yang disajikan pada persamaan (2.4.9) berikut ini.

$$\hat{f}(x) = a_0 + \sum_{m=1}^{M} a_m B_m(x) \tag{2.4.9}$$

keterangan:

a_0 = konstanta,
M = basis maksimum,
a_m = koefisien fungsi basis ke-m,
$B_m(x)$ = basis fungsi ke-m pada variabel x.

Dengan basis fungsi $B_m(x)$ yang didefinisikan dalam persamaan (2.4.10) berikut.

$$B_m(x) = \prod_{k=1}^{K_m} [S_{km}(x_{v(k,m)} - t_{km})]_+ \tag{2.4.10}$$

keterangan:

$B_m(x)$ = basis fungsi ke-m pada variabel x,
K_m = derajat interaksi pada basis fungsi ke-m,
$x_{v(k,m)}$ = variabel prediktor,
S_{km} = nilainya ± 1, jika knotnya terletak di kanan subregion maka nilainya $+1$ atau jika knotnya terletak di kiri subregion maka nilainya -1,
t_{km} = nilai knot dari variabel prediktor $x_{v(k,m)}$,
v = banyaknya variabel prediktor,
k = banyaknya interaksi,
m = banyaknya basis fungsi.

dengan fungsi,

$$(x_{v(k,m)} - t_{km})_+ = \begin{cases} (x_{v(k,m)} - t_{km})_+, & x_{v(k,m)} > t_{km} \\ 0, & x_{v(k,m)} \leq t_{km} \end{cases} \qquad (2.4.11).$$

Dengan menggunakan estimator MARS, maka model regresi yang terbentuk adalah sebagai berikut.

$$y_i = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{K_m} [S_{km}(x_{v(k,m)} - t_{km})] + \varepsilon_i \qquad (2.4.12).$$

Menurut Hasanah (2021), model MARS pada persamaan (2.4.12) dapat diuraikan menjadi:

$$\begin{aligned} \hat{f}(x) = a_0 &+ \sum_{m=1}^{M} a_m [S_{1m}(x_{v(1,m)} - t_{1m})] \\ &+ \sum_{m=1}^{M} a_m [S_{1m}(x_{v(1,m)} - t_{1m})][S_{2m}(x_{v(2,m)} - t_{2m})] \\ &+ \sum_{m=1}^{M} a_m [S_{1m}(x_{v(1,m)} - t_{1m})][S_{2m}(x_{v(2,m)} - t_{2m})] \\ &[S_{3m}(x_{v(3,m)} - t_{3m})] + \cdots \end{aligned} \qquad (2.4.13).$$

Penguraian pada persamaan (2.4.13) dapat juga ditulis dalam bentuk berikut.

$$\begin{aligned}\hat{f}(x) = a_0 + \sum_{Km=1} f_i(x_i) + \sum_{Km=2} f_{i,j}(x_i, x_j) \\ + \sum_{Km=3} f_{i,j,k}(x_i, x_j, x_k) + \cdots\end{aligned} \quad (2.4.14).$$

Pada persamaan (2.4.14) penjumlahan pertama mencakup semua fungsi basis yang hanya melibatkan satu variabel saja. Penjumlahan kedua mencakup semua fungsi basis yang melibatkan tepat dua variabel, yang merepresentasikan interaksi dua variabel. Demikian pula, penjumlahan ketiga merepresentasikan kontribusi dari interaksi tiga variabel, dan seterusnya.

Diberikan $V(m) = \{v(k,m)\}_1^{Km}$ adalah himpunan variabel yang diasosiasikan dengan fungsi basis ke - m. Sehingga setiap fungsi pada persamaan (2.4.14) dapat dijabarkan sebagai berikut.

$$f_i(x_i) = \sum_{Km=1} a_m B_m(x_i), \quad i \in V(m) \quad (2.4.15).$$

Persamaan (2.4.15) adalah penjumlahan semua fungsi basis satu variabel yang hanya melibatkan x_i dan merupakan representasi *spline* dengan $q = 1$ yang merepresentasikan fungsi satu variabel (univariat). Kemudian untuk setiap fungsi bivariat dapat dinyatakan dalam persamaan sebagai berikut.

$$f_{ij}(x_i, x_j) = \sum_{Km=2} a_m B_m(x_i, x_j), \quad (i,j) \in V(m) \quad (2.4.16).$$

Selanjutnya fungsi trivariat dari persamaan (2.4.14) dapat dinyatakan sebagai berikut.

$$f_{ijk}(x_i, x_j, x_k) = \sum_{Km=3} a_m B_m(x_i, x_j, x_k), \quad (i,j,k) \in V(m) \quad (2.4.17).$$

Penggunaan terminologi ANOVA untuk menyatakan fungsi dari satu variabel sebagai efek utama, fungsi dari dua variabel sebagai efek dari interaksi dua faktor, dan seterusnya. Persamaan (2.4.15) sampai dengan persamaan (2.4.16) merupakan proses dalam dekomposisi analisis varians (ANOVA) dari model MARS. Dekomposisi ANOVA adalah penjumlahan dari fungsi aditif. Interpretasi melalui dekomposisi ANOVA adalah merepresentasikan variabel yang masuk dalam model, baik untuk satu variabel maupun interaksi antar variabel, yang kemudian dapat merepresentasikannya dalam grafik.

Selanjutnya untuk mempermudah dalam menginterpretasikan model MARS, maka persamaan (2.4.12) dapat disederhanakan sebagai berikut.

$$\hat{f}(x) = a_0 + a_1 BF_1 + a_2 BF_2 + \cdots + a_M BF_M \tag{2.4.18}$$

keterangan :

$\hat{f}(x)$ = variabel respon,
a_0 = konstanta,
a_M = koefisien untuk basis fungsi ke - M,
BF_M = basis fungsi ke - M.

Pendekatan metode MARS memberikan suatu kepentingan variabel relatif. Kepentingan variabel relatif adalah ukuran yang digunakan untuk menilai seberapa besar kontribusi masing-masing variabel prediktor dalam menjelaskan keragaman dari variabel respon dalam suatu model statistik. Menentukan seberapa besar kepentingan variabel prediktor terhadap variabel respon menjadi sangat penting dalam sebagian besar kasus untuk memaksimalkan kegunaan dari sistem penyeleksian.

2.5 Estimasi Parameter

Metode yang digunakan untuk mengestimasi parameter model MARS adalah metode kuadrat terkecil (Darwin & Zurimi, 2019). Pada persamaan (2.4.9), a_m adalah koefisien regresi atau koefisien basis fungsi B_m dengan $m = 1, 2, ..., M$ yang akan diestimasi menggunakan metode kuadrat terkecil (Hasanah, 2021). Bentuk model MARS dapat ditulis dalam bentuk matriks berikut.

$$\mathbf{y} = \mathbf{Ba} + \boldsymbol{\varepsilon} \tag{2.5.19}$$

dengan

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{a} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_M \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 1 & \prod_{k=1}^{K_1}[S_{1m}(x_{1(1,m)} - t_{1m})] & \cdots & \prod_{k=1}^{K_m}[S_{k_m}(x_{1(k,m)} - t_{k_m})] \\ 1 & \prod_{k=1}^{K_1}[S_{1m}(x_{2(1,m)} - t_{1m})] & \cdots & \prod_{k=1}^{K_m}[S_{k_m}(x_{2(k,m)} - t_{k_m})] \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \prod_{k=1}^{K_1}[S_{1m}(x_{v(1,m)} - t_{1m})] & \cdots & \prod_{k=1}^{K_m}[S_{k_m}(x_{v(k,m)} - t_{k_m})] \end{bmatrix},$$

Agar mendapatkan estimator **a** menggunakan kuadrat terkecil, dilakukan dengan meminimumkan nilai galat $\varepsilon = \mathbf{y} - \mathbf{Ba}$ dan mengkuadratkan persamaan:

$$\begin{aligned} \sum_{i=1}^{n} \varepsilon_i^2 &= \varepsilon' \varepsilon \\ &= (\mathbf{y} - \mathbf{Ba})'(\mathbf{y} - \mathbf{Ba}) \\ &= (\mathbf{y}' - \mathbf{B}'\mathbf{a}')(\mathbf{y} - \mathbf{Ba}) \\ &= \mathbf{y'y} - \mathbf{y'Ba} - \mathbf{a}'\mathbf{B'y} + \mathbf{a}'\mathbf{B'Ba} \\ &= \mathbf{y'y} - 2\mathbf{a}'\mathbf{B'y} + \mathbf{a}'\mathbf{B'Ba} \end{aligned} \qquad (2.5.20).$$

Cara berikutnya adalah mencari turunan parsial dari persamaan (2.5.20) terhadap **a** dan menyamadenganakan nol.

$$\frac{\partial(\varepsilon' \varepsilon)}{\partial \mathbf{a}} = 0 \qquad (2.5.21)$$

sehingga diperoleh

$$\begin{aligned} -2\mathbf{B'y} + 2\mathbf{B'Ba} &= 0 \\ \mathbf{B'Ba} &= \mathbf{B'y} \\ (\mathbf{B'B})^{-1}(\mathbf{B'B})\mathbf{a} &= (\mathbf{B'B})^{-1}\mathbf{B'y} \\ \hat{\mathbf{a}} &= (\mathbf{B'B})^{-1}\mathbf{B'y} \end{aligned} \qquad (2.5.22).$$

2.6 Algoritma *Forward Backward Stepwise*

Menurut Mattalunru (2022), terdapat dua pendekatan untuk memilih model MARS terbaik, yaitu *forward stepwise* dan *backward stepwise*. Pendekatan *forward stepwise* digunakan untuk memperoleh jumlah fungsi basis dengan cara meminimumkan *Average sum of Square Residual* (ASR). Langkah-langkah yang dilakukan untuk algoritma *forward stepwise* adalah sebagai berikut.

1. Menentukan nilai basis awal yang merupakan fungsi basis konstan.
2. Membentuk basis fungsi hingga basis fungsi ke M berdasarkan kombinasi

variabel prediktor dan titik knot.

3. Membentuk interaksi antar basis fungsi sehingga menghasilkan GCV minimum.

4. Mengulangi kembali langkah ketiga hingga diperoleh model MARS sebanyak maksimum basis fungsi.

Sedangkan pendekatan dengan *backward stepwise* digunakan untuk memenuhi konsep parsimoni (model yang sederhana) dengan cara menghilangkan fungsi basis yang memiliki kontribusi kecil terhadap respon dari *forward stepwise* dengan meminimumkan *Generalized Cross Validation* (GCV) (Friedman, 1991). Langkah-langkah yang dilakukan untuk algoritma *forward stepwise* adalah sebagai berikut.

1. Hapus variabel yang memiliki pengaruh paling kecil terhadap nilai GCV.

2. Ulangi langkah pertama sampai tidak ada lagi variabel yang dapat dihapus.

2.7 Pemilihan Model Terbaik

Pemilihan model MARS yang terbaik ditentukan berdasarkan nilai GCV yang paling kecil. Berikut adalah fungsi GCV untuk memilih model terbaik (Friedman, 1991).

$$GCV = \frac{\frac{1}{n}\sum_{i=1}^{n}[y_i - \hat{f}_M(x_i)]^2}{[1 - \frac{C(\hat{M})}{n}]^2} \tag{2.7.23}$$

dimana:
$C(\hat{M}) = C(M) + M$,
$C(M) = tr(\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T) + 1$.

Keterangan:
n = ukuran sampel,
M = jumlah kombinasi fungsi basis,
y_i = variabel respon,
$\hat{f}_M$ = penduga fungsi f dengan M fungsi basis,
$C(M)$ = banyaknya parameter yang diestimasi.

2.8 Pengujian Kelayakan Model

Pada model MARS dilakukan uji signifikansi model yang terdiri dari uji simultan dan uji parsial. Uji signifikansi model dilakukan untuk mengetahui apakah suatu variabel memberikan pengaruh secara signifikan terhadap model atau tidak. Pengujian parameter dapat dilakukan secara simultan ataupun secara parsial (Rosanti & Budiantara, 2020).

2.8.1 Uji Simultan

Menurut Darwin dan Zurimi (2019), uji simultan bertujuan untuk menilai apakah seluruh fungsi basis yang terbentuk dalam model MARS secara bersama-sama memiliki pengaruh yang signifikan terhadap variabel respon. Dengan kata lain, uji ini digunakan untuk menguji keberartian model secara keseluruhan, sehingga dapat diketahui apakah model yang dihasilkan mampu menjelaskan variasi pada data dengan baik. Pengujian dilakukan dengan menggunakan uji F, yang membandingkan variasi yang dijelaskan oleh model dengan variasi yang tidak dijelaskan (galat).

Pengujian signifikansi simultan dilakukan melalui uji hipotesis sebagai berikut:

a. Rumusan hipotesis:
   $H_0 : a_1 = a_2 = \cdots = a_m = 0$ (model tidak signifikan)
   H_1 : minimal terdapat satu $a_m \neq 0; m = 1, 2, \cdots, M$ (model signifikan)

b. Menentukan taraf signifikansi α

c. Statistik uji:

$$F_{hitung} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2/k}{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2/n-k-1} \tag{2.8.24}$$

   dengan n adalah sampel acak dan k adalah banyaknya fungsi basis yang berkontribusi terhadap model.

d. Daerah kritis:
   Tolak H_0 jika $F_{hitung} > F_{\alpha(k,n-k-1)}$ atau *p-value* $< \alpha$.

2.8.2 Uji Parsial

Uji parsial bertujuan untuk menilai pengaruh masing-masing fungsi basis yang terbentuk dalam model MARS secara individual terhadap variabel respon. Pengujian ini dilakukan untuk mengetahui apakah setiap fungsi basis memberikan kontribusi yang signifikan dalam menjelaskan variasi variabel respon, setelah mempertimbangkan keberadaan fungsi basis lainnya dalam model. Dengan demikian, uji parsial membantu mengidentifikasi fungsi basis mana yang benar-benar memiliki peran penting dalam pembentukan model dan mana yang kontribusinya tidak signifikan secara statistik. Pengujian ini dilakukan dengan uji t (Darwin & Zurimi, 2019).

Pengujian signifikansi parsial dilakukan melalui uji hipotesis sebagai berikut:

a. Rumusan hipotesis:
   $H_0 : a_m = 0, m = 1, 2, \cdots, M$ (koefisien a_m tidak berpengaruh terhadap model)
   $H_1 : a_m \neq 0, m = 1, 2, \cdots, M$ (koefisien a_m berpengaruh terhadap model)

b. Menetapkan taraf signifikansi α

c. Statistik uji:

$$t_{hitung} = \frac{a_m}{Se_{a_m}} \tag{2.8.25}$$

   dimana S_{a_j} merupakan standar galat a_j yang diperoleh dari:

$$S_{a_m} = \sqrt{\left(\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1}\right)C_{jj}}$$

   C_{jj} adalah elemen-elemen pada diagonal utama matriks $(\mathbf{B}^T\mathbf{B})^{-1}$.

d. Daerah kritis:
   Tolak H_0 jika $t_{hitung} > t_{\frac{\alpha}{2}(n-k)}$ atau *p-value* $< \alpha$.

2.9 Variabel Berpengaruh dalam Model MARS

Penentuan variabel yang berpengaruh dalam pembentukan model MARS merupakan salah satu hal yang penting. Melalui proses ini akan diketahui variabel prediktor mana yang memberikan pengaruh paling besar dalam pembentukan model. Nilai variabel yang berpengaruh dalam pembentukan model MARS diperoleh setelah melakukan langkah sebagai berikut (Otok *et al*, 2023).

1. Menentukan model MARS yang akan dicari nilai pengaruh variabel beserta nilai GCV nya.

2. Hapus salah satu kombinasi basis fungsi dari model tersebut kemudian tentukan nilai GCV model yang telah dihapus satu kombinasi basis fungsi. Lakukan pada semua kombinasi basis fungsi yang terbentuk dalam model, kemudian tentukan juga nilai GCV nya.

3. Bandingkan nilai GCV dari setiap model yang dihapus satu kombinasi basis fungsi pada langkah kedua, kemudian pilih opsi model yang memiliki nilai GCV terendah untuk di eliminasi dari model awal. Sehingga didapatkan model dengan satu kombinasi basis fungsi yang dihapus.

4. Ulangi langkah kedua dan ketiga pada model dengan satu kombinasi basis fungsi yang dihapus sampai dengan model yang hanya berisi konstanta.

5. Hitung selisih nilai GCV antara model dengan satu kombinasi basis fungsi yang dihapus dan model penuh sampai pada selisih GCV antara model yang hanya berisi konstanta dan model yang hanya berisi satu kombinasi basis fungsi.

6. Hitung nilai kumulatif selisih GCV pada langkah kelima untuk setiap variabel prediktor yang terlibat.

7. Tentukan nilai maksimum dari nilai kumulatif selisih GCV yang diperoleh dari langkah keenam.

8. Hitung nilai variabel yang berpengaruh menggunakan persaman sebagai berikut.

$$VI(X_i) = \sqrt{\frac{Nilai\,Kumulatif\;Selisih\,GCV\;(X_i)}{Nilai\,Kumulatif\;Selisih\,GCV\;maksimum}}.100 \qquad (2.9.26)$$

dengan,
$VI(X_i)$ = nilai pengaruh variabel prediktor X_i,
X_i = variabel prediktor ke - i.

2.10 Morbiditas

Definisi morbiditas menurut BPS adalah gangguan terhadap kondisi fisik termasuk karena kecelakaan atau hal lain yang menyebabkan terganggunya kegiatan sehari-hari. Pada umumnya keluhan utama yang banyak dialami oleh penduduk adalah panas, sakit kepala, batuk, flu, diare, sesak nafas, dan sakit gigi. Kondisi yang demikian dapat menyebabkan terganggunya aktivitas sehari-hari seperti bekerja, mengurus rumah tangga, dan lainnya. Jika angka morbiditas semakin tinggi, maka hal tersebut menunjukkan bahwa semakin buruk derajat kesehatan penduduknya. Sebaliknya, jika angka morbiditas semakin rendah, maka derajat kesehatan penduduknya semakin baik. Oleh karena itu, morbiditas menjadi salah satu faktor penting dalam mengukur derajat kesehatan masyarakat di suatu daerah.

Menurut Haquel *et al* (2024), ada beberapa faktor yang memengaruhi angka morbiditas di suatu wilayah, diantaranya yaitu kepadatan penduduk, persentase penduduk miskin, rata-rata lama sekolah, persentase rumah tangga yang memiliki akses terhadap sanitasi layak, persentase rumah tangga yang memiliki akses terhadap air minum layak, persentase penduduk yang memiliki keluhan kesehatan, dan tingkat pengangguran terbuka. Angka morbiditas dapat dihitung dengan persamaan berikut (Rosanti & Budiantara, 2020).

$$AM = \frac{JPKK}{JP} \times 100\% \tag{2.10.27}$$

keterangan:
AM = angka morbiditas,
$JPKK$ = jumlah penduduk yang mengalami keluhan kesehatan,
JP = jumlah penduduk.

BAB III

METODE PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada Semester Genap tahun ajaran 2024/2025 di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampuung.

3.2 Data Penelitian

Data yang digunakan pada penelitian ini adalah data sekunder atau data yang didapatkan dari sumber yang sudah tersedia. Data yang digunakan pada penelitian ini yaitu data dari publikasi Statistik Kesejahteraan Rakyat setiap provinsi di Pulau Sumatera yang diterbitkan oleh Badan Pusat Statistik (BPS) pada tahun 2023. Statistik Kesejahteraan Rakyat berisi data sosial yang mencakup bidang pendidikan, kesehatan/gizi, perumahan, sosial ekonomi, kegiatan sosial budaya, konsumsi/pengeluaran dan pendapatan rumah tangga, serta pendapatan masyarakat mengenai kesejahteraan rumah tangga. Data diakses pada tanggal 10 November 2024.

Dalam penelitian ini digunakan satu variabel respon dan tujuh variabel prediktor yang semuanya bersifat kontinu. Berikut ini adalah variabel respon (Y) dan variabel prediktor (X_i) yang digunakan dalam penelitian ini dan disajikan dalam Tabel 3.1.

Tabel 3.1 Variabel Penelitian

Variabel	Nama Variabel	Satuan
Y	Morbiditas	Persen
X_1	Kepadatan Penduduk	Jiwa/Km^2
X_2	Persentase Penduduk Miskin	Persen
X_3	Rata-rata Lama Sekolah	Tahun
X_4	Persentase Rumah Tangga yang Memiliki Akses Sanitasi Layak	Persen
X_5	Persentase Rumah Tangga yang Memiliki Akses Air Minum Layak	Persen
X_6	Persentase Penduduk yang Memiliki Keluhan Kesehatan	Persen
X_7	Tingkat Pengangguran Terbuka	Persen

3.3 Metode Penelitian

Adapun langkah-langkah yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Pengumpulan data yang terdiri dari satu variabel respon dan tujuh variabel prediktor.
2. Melakukan analisis deskriptif untuk mengetahui gambaran umum mengenai data morboditas Pulau Sumatera.
3. Melakukan prapemrosesan data yang meliputi mendeteksi nilai hilang dan duplikasi data.
4. Melakukan uji multikolinearitas dengan menggunakan ukuran *Variance Inflation Factor* (VIF).
5. Melakukan pemodelan MARS
   a. Menentukan kombinasi terbaik dari Basis Fungsi (BF), Maksimum Interaksi (MI), dan Minimum Observasi (MO) dengan mempertimbangkan nilai *Generalized Cross Validation* (GCV) terkecil.
   b. Memodelkan kombinasi yang telah ditentukan ke dalam data penelitian.
   c. Melakukan interpretasi model MARS yang diperoleh.
6. Melakukan uji signifikansi basis fungsi yang telah diperoleh.
   a. Melakukan uji simultan menggunakan uji-F.
   b. Melakukan uji parsial menggunakan uji-t.

7. Menentukan variabel yang memiliki pengaruh paling signifikan terhadap model MARS.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah diuraikan pada bab sebelumnya, dapat ditarik kesimpulan sebagai berikut.

1. Persamaan model yang diperoleh menggunakan metode *Multivariate Adaptive Regression Spline* (MARS) pada data morbiditas setiap kabupaten/kota di Pulau Sumatera adalah model dengan kombinasi BF = 14, MI = 3, dan MO = 15 sebagai berikut.

$$\begin{aligned} Y =& 7,873944 + 0,5409366.BF1 + 0,4030349.BF2 + 0,2872119.BF3 \\ &+ 0,1120214.BF4 - 0,1102888.BF2.BF5 + 0,1055271.BF6.BF4 \\ &- 0,01838955.BF7.BF4 - 0,0007936193.BF8.BF6.BF4 \\ &- 0,001259145.BF9.BF10.BF4 \end{aligned}$$

   dengan:

$$\begin{aligned} BF1 &= h(8 - X_2) \\ BF2 &= h(X_2 - 8) \\ BF3 &= h(X_5 - 89,85) \\ BF4 &= h(X_6 - 20,03) \\ BF5 &= h(X_7 - 3,89) \\ BF6 &= h(10,12 - X_3) \\ BF7 &= h(X_4 - 85,30) \\ BF8 &= h(143,26 - X_1) \\ BF9 &= h(11,38 - X_2) \\ BF10 &= h(85,30 - X_4) \end{aligned}$$

   Dengan nilai GCV dan MSE minimum yang diperoleh dari model tersebut secara berturut-turut yaitu 8,778420 dan 6,646380.

2. Variabel X_1 hingga X_7 berkontribusi dalam pembentukan model MARS untuk memprediksi angka morbiditas di Pulau Sumatera, meskipun dengan tingkat pengaruh yang berbeda-beda. Variabel dengan kontribusi terbesar adalah persentase penduduk yang memiliki keluhan kesehatan (X_6) dengan nilai kontribusi 100%. Sedangkan variabel dengan kontribusi terkecil adalah Tingkat Pengangguran Terbuka (X_7) dengan kontribusi 13,86%.

5.2 Saran

Dalam rangka pengembangan dan penyempurnaan penelitian selanjutnya, penulis menyarankan untuk mempertimbangkan penggunaan metode analisis yang berbeda guna memungkinkan dilakukannya perbandingan tingkat akurasi dengan metode lain yang relevan. Selain itu, disarankan juga untuk menambahkan variabel-variabel lain yang berpotensi memengaruhi tingkat morbiditas, seperti faktor lingkungan, kualitas pelayanan kesehatan, dan tingkat pendidikan masyarakat, guna memperoleh model yang lebih akurat.

DAFTAR PUSTAKA

Al-Azies, H., & Hapsery, A. 2019. Spline Nonparametric Regression Approach for Modelling Factors Affecting Vocational National Exam Result in Surabaya, hlm. 244 – 252. *Proceedings of the* 1^{th} *Steeem*, Yogyakarta.

Ananda, R.F., Harsyiah, L., & Alfian, M.R. 2023. Classification of Perceptions of the Covid-19 Vaccine Using Multivariate Adaptive Regression Spline. *Jurnal Varian*. **6**(2): 137 – 148.

Damayanti, Y., & Sunendiari, S. 2021. Pemodelan Faktor-Faktor yang Mempengaruhi Penyakit Tuberkulosis Menggunakan Metode Multivariate Adaptive Regression Spline (MARS): Studi Kasus Provinsi Jawa Barat, Jawa Timur, dan Jawa Tengah Tahun 2019. *Statistika*. **7**(2): 676 – 683.

Darwin, & Zurimi, S. 2019. Analisis Model Aplikatif Multivariate Adaptive Regression Spline (MARS) Terhadap Klasifikasi Faktor yang Mempengaruhi Masa Studi Mahasiswa FKIP Universitas Darussalam Ambon. *Jurnal Simetrik*. **9**(2): 250 – 255.

Friedman, J.H. 1991. Multivariate Adaptive Regression Splines. *The Annals Statistics*. **19**(1):1 – 67.

Gujarati, D.N., & Porter, D.C. 2009. *Basic Econometrics*. 5^{th} Edition. The McGraw-Hill Companies Inc., New York.

Hafifi, O.V., Sudarno, Maruddani, D.A.I., Ernayanti, T., Setiyawan, A., Siniwi, L.M., & Rozan, M.H. 2021. Analisis Survival Penderita Gagal Ginjal dengan Pendekatan Multivariate Adaptive Regression Spline. *Jurnal Endurance: Kajian Ilmiah Problema Kesehatan*. **6**(3): 504 – 511.

Haquel, H., Nurhasanah, & Fitriana. 2024. Pemodelan Regresi Spline pada Faktor-faktor yang Mempengaruhi Angka Morbiditas di Provinsi Sumatera Utara. *Jurnal Peluang*. **12**(1): 21 – 34.

Hasanah, S.H. 2021. Multivariate Adaptive Regression Spline (MARS) for Modelling the Student Status at Universitas Terbuka. *Jurnal Matematika MANTIK*. **7**(1): 51 – 58.

Idowu, E.O., Ikegwu, E.M., Fadiji, A.A., & Evro, M.U. 2024. Detection and Correction of Heteroscedasticity and Its Effect on Modelling of Nigerian Economic Data. *International Journal of Mathematical Analysis and Modelling*. **7**(2): 153 – 158.

Mardiatmoko, G. 2020. The Importance of the Classical Assumption Test in Multiple Linear Regression Analysis (A Case Study of the Preparation of the Allometric Equation of Young Walnuts). *Barekeng: Jurnal Ilmu Matematika dan Terapan*. **14**(3): 333 – 342.

Mattalunru, M.R., Annas, S., & Aidid, M.K. 2022. Application of Multivariate Adaptive Regression Spline (MARS) to Determine the Factors that Affecting Rainfall in Makassar City. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*. **4**(1): 9 – 19.

Mendenhall, W., & Sincich, T. 2012. *A Second Course in Statistics Regression Analysis*. 7^{th} Edition. Pearson Education Inc., Boston.

Montgomery, D.C., Peck, E.A., & Vining, G.G. 2021. *Introduction to Linear Regression Analysis*. 6^{th} Edition. John Wiley & Sons, New York.

Otok, B.W., Rumiati, A.T., Apulembang, A.P., & Azies, H. 2023. Decomposition and Importance Variable Process in Multivariate Adaptive Regression Spline Model. *International Journal on Advance Science Engineering Information Technology*. **13**(3): 928 – 934.

Paramasivam, G., Rao, I.R., & Prabhu, M.A. 2024. Normality Testing in Statistics: What Clinician - Researchers Should Know. *Heart Failure Journal of India*. **2**(1): 55 – 60.

Rahma, D., Amalita, N., Kurniawati, Y., & Martha, Z. 2024. Application of Multivariate Adaptive Regression Spline for Modelling Stunting Toddler on the Island of Java. *UNP Journal of Statistics and Data Science*. **2**(3): 338 – 343.

Rahmawati, D.P., Budiantara, I.N., Prastyo, D.D., & Octavanny, M.A.D. 2021. Mixed Spline Smoothing and Kernel Estimator in Biresponse Nonparametric Regression. *International Journal of Mathematics and Mathematical Sciences*. **2021**(1): 1 – 14.

Rosanti, I.W., & Budiantara, I.N. 2020. Pemodelan Faktor-Faktor yang Mempengaruhi Morbiditas di Jawa Tengah Menggunakan Regresi Nonparametrik Spline Truncated. *INFERENSI*. **3**(2): 107 – 114.

Shresta, N. 2020. Detecting Multicollinearity in Regression Analysis. *American Journal of Applied Mathematics and Statistics*. **8**(2): 39 - 42.

Sriningsih, R., Otok, B.W., & Sutikno. 2021. Factors Affecting the Number of Dengue Fever Cases in West Sumatra Province using the Multivariate Adaptive Regression Spline (MARS), hlm. 1 – 6. *Journal of Physics: Conference Series, Surabaya*.

Wahyuni, S.A., Ratnawati, Indriyani, & Fajri, M. 2020. Spline Regression Analysis to Modelling the Open Unemployment Rate in Sulawesi. *Natural Science: Journal of Science and Technology*. **9**(2): 34 – 39.

Widyastuti, D.A., Fernandes, A.A.R., & Pramoedyo, H. 2021. Spline Estimation Method in Nonparametric Regression Using Truncted Spline Approach, hlm. 1 – 9. *Journal of Physics: Conference Series, Malang*.

Yasmirullah, S.D.P., Otok, B.W., Purnomo, J.D.T., & Prastyo, D.D. 2021. Modification of Multivariate Adaptive Regression Spline (MARS), hlm. 1 – 10. *Journal of Physics: Conference Series, Surabaya*.