

PENANGANAN KETIDAKSEIMBANGAN DATA PADA PEMODELAN RISIKO DIABETES TIPE 2 MENGGUNAKAN SMOTE-MARS

Canada Mewa Fizkadana¹, Nur Chamidah², Ardi Kurniawan², Naufal Ramadhan Al Akhwal Siregar¹

^{1,4}Magister Matematika, Departemen Matematika, Fakultas Sains dan Teknologi, Universitas Airlangga

^{2,3}Departemen Matematika, Fakultas Sains dan Teknologi, Universitas Airlangga

canada.mewa.fizkadana-2023@fst.unair.ac.id, nur-c@fst.unair.ac.id, ardi-k@fst.unair.ac.id,

naufal.ramadhan.al-2024@fst.unair.ac.id

ABSTRAK

Diabetes Melitus Tipe 2 (DMT2) merupakan tantangan kesehatan kritis di Indonesia, namun pemodelan faktor risikonya sering terhambat oleh ketidakseimbangan kelas di mana individu sehat melebihi jumlah kasus positif. Penelitian ini memodelkan risiko DMT2 menggunakan pendekatan hybrid yang menggabungkan Synthetic Minority Over-sampling Technique (SMOTE) dan Multivariate Adaptive Regression Splines (MARS). Penelitian ini menggunakan data medis dengan ketidakseimbangan awal 62% sehat dan 38% kasus DMT2. SMOTE diterapkan untuk menyeimbangkan distribusi training, diikuti dengan pemodelan MARS untuk menangkap interaksi non-linear. Hasil penelitian menunjukkan bahwa model SMOTE-MARS secara signifikan mengungguli MARS standar, mencapai Akurasi 0,70 dan AUC 0,73, dibandingkan dengan AUC model standar yang hanya 0,458. Model optimal mengidentifikasi Berat Badan, Tinggi Badan, dan Konsumsi Gula sebagai prediktor utama. Secara spesifik, berat badan di bawah 48 kg teridentifikasi sebagai faktor protektif yang kuat (OR 0,22), sedangkan tinggi badan kurang dari 165 cm (OR 1,22) dan interaksi berat badan berlebih (>62 kg) dengan konsumsi gula (OR 1,12) merupakan faktor risiko signifikan. Tingginya nilai sensitivitas mengindikasikan bahwa model SMOTE-MARS sangat potensial digunakan sebagai instrumen deteksi dini untuk mengidentifikasi pasien yang berisiko tinggi terkena diabetes.

Kata Kunci: Diabetes Melitus Tipe 2, Ketidakseimbangan Data, SMOTE, MARS, Pemodelan Risiko.

ABSTRACT

Type 2 Diabetes Mellitus (T2DM) is a critical health challenge in Indonesia, yet modeling its risk factors is often hindered by class imbalance where healthy individuals outnumber positive cases. This study models T2DM risk using a hybrid approach combining Synthetic Minority Over-sampling Technique (SMOTE) and Multivariate Adaptive Regression Splines (MARS). The study utilized medical data with an initial imbalance of 62% healthy and 38% T2DM cases. SMOTE was applied to balance the training distribution, followed by MARS modeling to capture non-linear interactions. The results demonstrated that the SMOTE-MARS model significantly outperformed standard MARS, achieving an Accuracy of 0.70 and an AUC of 0.73, compared to the standard model's poor AUC of 0.458. The optimal model identified Body Weight, Body Height, and Sugar Consumption as key predictors. Specifically, body weight under 48 kg was identified as a strong protective factor (Odds Ratio/OR 0.22), while body height less than 165 cm (OR 1.22) and the interaction of excess body weight (>62 kg) with sugar consumption (OR 1.12) were significant risk factors. The high sensitivity value indicates that the SMOTE-MARS model has significant potential as an early detection instrument for identifying patients at high risk of developing diabetes.

Key Words: Type 2 Diabetes Mellitus, Data Imbalance, SMOTE, MARS, Risk Modelling.

PENDAHULUAN

Status kesehatan Indonesia saat ini sedang mengalami transisi dari era penyakit menular ke penyakit tidak menular (PTM). Menurut World Health Organization (WHO) memperkirakan bahwa PTM menyumbang 70% dari seluruh kematian global, dengan lebih dari 75% kematian tersebut terjadi di

negara berpenghasilan rendah dan menengah (WHO, 2018). Secara global, Diabetes Melitus (DM) adalah PTM yang paling umum dan mematikan, berkontribusi pada jutaan kematian dini, stroke, gagal ginjal, dan kecacatan (Wahidin dkk., 2024). Secara khusus, Diabetes Melitus Tipe 2 (DMT2) adalah jenis yang paling dominan, mencakup sekitar 90% dari seluruh kasus diabetes.

DMT2 adalah gangguan metabolisme kronis yang ditandai dengan hiperglikemia akibat resistensi insulin dan defisiensi insulin relatif, di mana tubuh tidak dapat menggunakan insulin secara efektif (Rahman, 2021). Penyebaran DMT2 merupakan masalah luas karena prevalensinya yang tinggi, dampak ekonomi yang signifikan, dan morbiditas yang tinggi (Fatarona dkk., 2025; Rahmawati, 2024). Penelitian ini menganalisis hubungan antara status DMT2 dan faktor-faktor tersebut. Dalam penelitian ini, metode yang digunakan untuk memodelkan status DMT2 adalah regresi nonparametrik, karena bentuk kurva hubungan antara DMT2 dan faktor prediktor tidak mengikuti pola tertentu.

Regresi nonparametrik adalah pendekatan yang tidak mengasumsikan pola hubungan tertentu antara variabel respon dan prediktor (Chamidah & Lestari, 2022). Dalam pendekatan regresi nonparametrik, bentuk estimasi model ditentukan berdasarkan pola yang terdapat di dalam data. Kemampuan regresi nonparametrik untuk menemukan bentuk kurva regresi didukung oleh parameter dalam setiap jenis metode/pendekatan nonparametrik, yang membuat estimasi pola kurva regresi menjadi lebih fleksibel (Tjahjono dkk., 2018).

Namun, terdapat tantangan dalam pemodelan data medis, termasuk data DMT2, yaitu masalah ketidakseimbangan kelas. Seringkali, jumlah data pasien yang didiagnosis positif (kelas minoritas) jauh lebih sedikit dibandingkan dengan pasien sehat (kelas mayoritas) (Sidiq dkk., 2025). Ketidakseimbangan ini dapat menyebabkan model prediksi menjadi bias, cenderung memprediksi kelas mayoritas, sehingga mengurangi akurasi deteksi untuk kelas minoritas yang bisa dibilang lebih kritis. Untuk mengatasi masalah ini, Synthetic Minority Over-sampling Technique (SMOTE) diterapkan sebelum pemodelan dilakukan (Safitri dkk., 2025). SMOTE adalah metode untuk menangani ketidakseimbangan data yang bekerja dengan membangkitkan data sintesis baru untuk kelas minoritas berdasarkan prinsip k-nearest neighbor, sehingga menyeimbangkan distribusi data dan meningkatkan kinerja klasifikasi (Pradipta dkk., 2021).

Selanjutnya, metode regresi nonparametrik yang digunakan dalam penelitian ini adalah Multivariate Adaptive Regression Splines (MARS). Pemilihan metode estimasi ini mempertimbangkan kemampuan MARS untuk mendapatkan hasil prediksi yang baik dalam menangani data yang mengalami perubahan perilaku pada sub-interval tertentu, yang mengindikasikan adanya pergeseran pola perilaku data (Chamidah dkk., 2025; Hasyim dkk., 2025). MARS sangat cocok untuk struktur data penelitian ini dan untuk memodelkan status DMT2 guna mendapatkan hasil yang lebih akurat.

Banyak penelitian terdahulu mengenai pemodelan risiko DMT2 telah dilakukan. Diantaranya yaitu penelitian yang dilakukan oleh (Chang dkk., 2017; Jannoo dkk., 2017; Li dkk., 2021; Sayadi dkk., 2017). Kemudian, dari beberapa penelitian juga sudah meneliti terkait penerapan ketidakseimbangan kelas pada DMT2 diantaranya adalah Salam dkk. (2025) yang menggunakan metode hybrid k-means dan regresi logistik dengan menerapkan SMOTE, kemudian Wang dkk. (2021) yang membandingkan antara random forest, SVM, dan LASSO menggunakan SMOTE, dan Mushtaq dkk. (2022) yang menggunakan teknik *voting classification* untuk menangani *class imbalance*.

Meskipun penelitian-penelitian tersebut menganalisis faktor risiko DM menggunakan berbagai metode, belum ada penelitian yang menggunakan metode MARS yang dikombinasikan dengan penanganan ketidakseimbangan data menggunakan SMOTE. Teknik MARS mampu menangani ukuran sampel yang besar dan banyak variabel, berfokus pada penyelesaian masalah berdimensi tinggi untuk menghasilkan prediksi respon yang presisi ketika diterapkan pada data berdimensi tinggi (Barron & Xiao, 1991; Friedman, 1991). Selain itu, MARS mengakomodasi interaksi antar variabel prediktor yang dinyatakan dalam fungsi basis dan mengklasifikasikan respon kategorik (Yasmirullah dkk., 2021). Dengan demikian, struktur data penelitian ini sangat cocok untuk dianalisis menggunakan metode MARS.

Namun, masalah ketidakseimbangan kelas yang melekat pada data medis DMT2, di mana jumlah pasien positif jauh lebih sedikit daripada pasien negatif, penggunaan MARS

saja tidak cukup. Oleh karena itu, penelitian ini mengintegrasikan *Synthetic Minority Over-sampling Technique* (SMOTE) sebagai langkah pra-pemrosesan untuk menyeimbangkan distribusi kelas sebelum pemodelan dilakukan. Penelitian ini bertujuan untuk memodelkan risiko Diabetes Melitus Tipe 2 menggunakan pendekatan *hybrid* SMOTE-MARS. Penerapan model SMOTE-MARS ini bertujuan untuk memprediksi status DM menggunakan beberapa faktor risiko utama secara lebih akurat, tanpa bias terhadap kelas mayoritas. Rekomendasi dari hasil penelitian ini adalah menyoroti potensi model ini untuk penggunaan praktis dalam sistem peringatan dini.

METODE PENELITIAN

Metode dalam penelitian ini menggunakan pendekatan *hybrid* yang menggabungkan *Synthetic Minority Over-sampling Technique* dan *Multivariate Adaptive Regression Splines* (SMOTE-MARS) untuk memodelkan risiko Diabetes Melitus Tipe 2 (DMT2). Metode *hybrid* ini dipilih karena mampu mengatasi dua tantangan mendasar dalam analisis data medis secara bersamaan, yaitu kompleksitas non-linear dari faktor-faktor risiko dan masalah ketidakseimbangan kelas. Selain itu, dibandingkan dengan penggunaan MARS saja, integrasi SMOTE-MARS lebih akurat untuk tugas klasifikasi di mana jumlah pasien sehat (mayoritas) secara signifikan lebih banyak daripada jumlah pasien diabetes (minoritas).

Fokus penelitian ini mencakup analisis faktor risiko DMT2 di Rumah Sakit Universitas Airlangga Surabaya (RSUA), pada tahun 2024. Data yang digunakan adalah data sekunder yang diperoleh dari rekam medis. Sampel penelitian terdiri dari total 100 pasien. Dari total tersebut, 38 pasien teridentifikasi positif DMT2 (kelas minoritas), sedangkan 62 pasien sisanya adalah non-DMT2 atau sehat (kelas mayoritas). Komposisi ini merupakan kelas tidak seimbang (*imbalanced*) yang dapat menyebabkan kinerja klasifikasi yang kurang optimal dan bias terhadap kelas mayoritas jika tidak ditangani.

Variabel penelitian ini terdiri dari variabel dependen atau respon dan independen atau prediktor. Variabel respon yang digunakan

dalam penelitian ini adalah status diabetes, yang dikategorikan 1 sebagai "Positif DMT2" dan 0 sebagai "Negatif DMT2". Kemudian, variabel prediktor yang digunakan meliputi berat badan, tinggi badan, konsumsi gula.

Seluruh data yang digunakan telah diproses untuk menjamin anonimitas dan kerahasiaan data pasien. Analisis data dilakukan menggunakan perangkat lunak R, dengan memanfaatkan *package* "themis" dan "recipes" untuk penyeimbangan data serta paket "earth" untuk pemodelan MARS. Tahapan analisis dalam penelitian ini meliputi melakukan analisis statistika deskriptif, penyeimbangan data dengan SMOTE, estimasi parameter menggunakan MARS, mencari faktor paling signifikan yang memengaruhi DMT2, dan evaluasi model.

HASIL DAN PEMBAHASAN

Bab ini memaparkan hasil analisis data yang telah dilakukan sesuai dengan tahapan metodologi penelitian. Pembahasan diawali dengan gambaran umum karakteristik data pasien melalui statistika deskriptif, dilanjutkan dengan proses penyeimbangan data menggunakan SMOTE, serta pembentukan dan interpretasi model MARS.

Statistika Deskriptif

Statistika deskriptif digunakan untuk memberikan gambaran umum mengenai karakteristik data melalui ukuran pemusatan dan penyebaran. Pemilihan metode deskriptif ini disesuaikan dengan skala pengukuran variabel, baik itu kategorik maupun kontinu. Karakteristik data digambarkan melalui nilai rata-rata (*mean*), ragam (*variance*), serta nilai minimum dan maksimum. Rincian statistika deskriptif untuk variabel prediktor kontinu dalam penelitian ini disajikan pada **Tabel 1** sebagai berikut:

Tabel 1. Statistika Deskriptif

Variabel	DMT2	Mean	Var	Min	Max
Berat	1	67.5	250.3	31	117
Badan	0	62.2	126.5	43	82
Tinggi	1	161.00	53.57	146	175
Badan	0	157.84	59.49	141	175
Konsumsi	1	11.13	64.39	0.12	35.56
Gula	0	8.22	51.54	0.2	34.97

Berdasarkan **Tabel 1**, data dipisahkan berdasarkan status DMT2, yaitu Tipe 0 untuk pasien Negatif/Sehat dan Tipe 1 untuk pasien

Positif/DMT2. Pada kelompok pasien dengan DMT2 menunjukkan rata-rata berat badan yang lebih tinggi (67,46 kg) dibandingkan dengan kelompok non-DMT2, yang memiliki rata-rata berat badan 62,25 kg. Selanjutnya, pada variabel konsumsi gula, rata-rata konsumsi gula pada kelompok DMT2 (11,13) lebih tinggi daripada kelompok non-DMT2 (8,22). Pasien yang didiagnosis dengan DMT2 biasanya berada di bawah saran medis untuk membatasi asupan gula guna mengontrol kadar glukosa darah, sedangkan populasi non-diabetes mungkin tidak berada di bawah batasan diet ketat tersebut. Varians pada kelompok DMT2 untuk konsumsi gula juga lebih tinggi (64,39).

Pra-pemrosesan menggunakan SMOTE

Dalam penelitian ini, langkah pertama melibatkan pembagian dataset menjadi dua bagian, 80 data (80%) untuk *data training* dan 20 data (20%) untuk data *testing*. Pembagian *data training* dan uji dilakukan menggunakan stratified K-fold, dengan 5 fold dipilih untuk proses validasi. Setelah data dibagi menjadi himpunan latih dan uji, langkah selanjutnya adalah pra-pemrosesan data sebelum membangun model menggunakan *data training*. Pada tahap ini, ditemukan ketidakseimbangan antara kelas 0 dan kelas 1 pada *data training*, yang ditangani menggunakan Synthetic Minority Over-Sampling Technique (SMOTE). Ketidakseimbangan kelas dapat memengaruhi kinerja model klasifikasi secara signifikan, terutama pada kelas minoritas. Berdasarkan kategori diabetes dalam data *training*, **Tabel 2** di bawah ini menyajikan distribusi variabel respon sebelum dan sesudah SMOTE diterapkan.

Tabel 2. Perbandingan Distribusi Data training dengan Proses Penyeimbangan SMOTE

Perlakuan Data	Kejadian DMT2				Total
	0: Non DMT2		1: DMT2		
	<i>n</i>	Persentase	<i>n</i>	Persentase	
Sebelum SMOTE	50	62,5%	30	37,5%	80
Sesudah SMOTE	50	50%	50	50%	100

Seperti disajikan pada **Tabel 2** bagian "Setelah SMOTE", jumlah kejadian DMT2 meningkat dari 30 menjadi 50, menyamai

jumlah kejadian Non-DMT2. Akibatnya, total *data training* meningkat menjadi 100 observasi, menghasilkan distribusi yang seimbang sempurna sebesar 50% untuk setiap kelas. Dataset yang seimbang ini memberikan landasan yang lebih kuat untuk pemodelan MARS selanjutnya, memastikan bahwa algoritma mempelajari karakteristik kasus DMT2 positif seefektif mempelajari kasus negatif.

Pemodelan Menggunakan SMOTE-MARS

Sebelum mengembangkan model MARS, proses *tuning* parameter MARS dilakukan menggunakan perangkat lunak R dengan paket "earth". Dua parameter utama yang perlu di-*tuning* dalam model MARS adalah *nprune* dan *degree*. Parameter *nprune* mengontrol jumlah maksimum fungsi basis yang dipertahankan dalam model akhir setelah proses pemangkasan (*pruning*), biasanya ditetapkan sebesar 2 hingga 4 kali jumlah variabel prediktor. Sementara itu, parameter *degree* menentukan tingkat interaksi yang diperbolehkan antar variabel prediktor, umumnya berkisar antara 2 hingga 3. Hasil tuning parameter MARS dapat dilihat pada **Tabel 3**.

Tabel 3. Parameter Tuning MARS

Degree	nprune	Akurasi	Kappa
2	6	0.67	0,34
	9	0.67	0,34
	12	0.67	0,34
3	6	0.68	0,36
	9	0.68	0,36
	12	0.68	0,36

Hasil tuning terbaik dicapai dengan akurasi dan nilai *kappa* tertinggi dengan *degree* ditetapkan sebesar 3 dan *nprune* sebesar 6. Setelah mengidentifikasi parameter optimal dalam model MARS, proses pemangkasan (*pruning*) akan dilakukan untuk mencegah *overfitting* dan mengurangi kompleksitas model. Parameter *nprune* secara langsung memengaruhi kompleksitas model.

Kemudian, berdasarkan dari hasil tuning parameter terbaik, prosedur estimasi dilakukan berdasarkan *5-fold cross-validation*, yang dapat dilihat pada **Tabel 4**.

Tabel 4. Kinerja MARS Menggunakan Pemilihan Parameter Terbaik untuk 5-fold cross-validation

Metrik Evaluasi	Fold [1]	Fold [2]	Fold [3]	Fold [4]	Fold [5]
Akurasi	0.8	0.6	0.7	0.65	0.6
Sensitivitas	0.8	0.6	0.6	0.5	0.7
Spesifisitas	0.8	0.6	0.8	0.8	0.5
AUC	0.85	0.475	0.75	0.79	0.65

Fold terbaik diperoleh pada fold-1, yang menghasilkan metrik evaluasi tertinggi. Setelah itu, pada tahap pemangkasan mundur (*backward pruning*), pemilihan fungsi basis (BF) optimal didasarkan pada kriteria *Generalized Cross-Validation (GCV)* minimum yaitu 0,227. Hasilnya, model optimal terdiri dari 6 *terms*, yang mencakup 5 fungsi basis dan 1 intersep.

Estimasi parameter dengan model optimal dapat dilihat pada **Tabel 5**.

Tabel 5. Hasil Estimasi Parameter MARS

Terms	Notation	Estimation	OR
Intercept	$\hat{\alpha}_0$	-2.63	-
Basis Fungsi 1 $BF_1 = h(48 - x_1)$	$\hat{\alpha}_1$	-1.51	0.22
Basis Fungsi 2 $BF_2 = h(165 - x_2)$	$\hat{\alpha}_2$	0.20	1.22
Basis Fungsi 3 $BF_3 = (59,97 - x_1) * h(14,78 - x_3)$	$\hat{\alpha}_3$	0.05	1.05
Basis Fungsi 4 $BF_4 = h(x_1 - 62) * h(14,78 - x_3)$	$\hat{\alpha}_4$	0.11	1.12
Basis Fungsi 5 $BF_5 = h(x_1 - 65) * h(14,78 - x_3)$	$\hat{\alpha}_5$	-0.12	0.87

Persamaan MARS yang dihasilkan dari estimasi parameter disajikan dalam persamaan (1).

$$\text{logit}(\hat{\pi}(x)) = -2,63 - 1,51 BF_1 + 0,20 BF_2 + 0,05 BF_3 + 0,11 BF_4 - 0,12 BF_5 \quad (1)$$

Berdasarkan hasil estimasi parameter pada Tabel 5, model MARS terbaik terbentuk dari kombinasi satu intersep dan lima fungsi basis (BF) yang melibatkan variabel Berat Badan (x_1), Tinggi Badan (x_2), dan Konsumsi Gula (x_3). Analisis terhadap efek utama menunjukkan bahwa berat badan di bawah 48 kg (BF_1) dengan koefisien -1,51 dan Odds Ratio (OR) sebesar 0,22, hal ini

mengindikasikan bahwa individu dengan berat badan sangat rendah memiliki risiko diabetes yang jauh lebih kecil dengan menganggap variabel lain bernilai konstan. Sebaliknya, fungsi basis kedua (BF_2) menyoroti faktor risiko terkait tinggi badan, di mana individu dengan tinggi badan kurang dari 165 cm memiliki kecenderungan risiko 1,22 kali lebih tinggi untuk terkena DMT2. Selain efek tunggal, model juga menangkap pola interaksi non-linier antara berat badan dan konsumsi gula. Secara spesifik, BF_4 menunjukkan bahwa individu dengan berat badan lebih dari 62 kg memiliki peningkatan risiko sebesar 1,12 kali lipat terkena DMT2.

Evaluasi Hasil Pemodelan SMOTE-MARS

Setelah mendapatkan model terbaik, metrik evaluasi MARS tanpa SMOTE dan MARS dengan SMOTE akan dibandingkan menggunakan data *testing*, seperti disajikan pada **Tabel 6**.

Tabel 6. Evaluasi Kinerja Model MARS dan MARS dengan Peningkatan SMOTE

Metrik Evaluasi	MARS	SMOTE-MARS
Akurasi	0,5	0,7
Sensitivitas	0,667	0,875
Spesifisitas	0,25	0,58
AUC	0,46	0,73

Berdasarkan Tabel 6, pada model MARS tanpa SMOTE, ketika diterapkan pada dataset yang tidak seimbang tanpa pra-pemrosesan, memberikan hasil yang kurang optimal. Model ini hanya mencapai akurasi 0,500 dan AUC 0,458. Nilai AUC di bawah 0,5 menunjukkan bahwa kemampuan prediksi model MARS lebih buruk. Sebaliknya, pada model SMOTE-MARS menunjukkan peningkatan substansial di seluruh metrik utama. Akurasi keseluruhan meningkat secara signifikan menjadi 0,70, menunjukkan bahwa model *hybrid* mengklasifikasikan 70% kasus uji dengan benar. Pada kriteria AUC meningkat drastis menjadi 0,73, memindahkan kinerja model ke dalam kategori klasifikasi "baik". Hasil ini menyiratkan bahwa model SMOTE-MARS mengidentifikasi individu sehat dan tidak dengan baik. Hasil ini juga menunjukkan bahwa menyeimbangkan data dengan

SMOTE memungkinkan algoritma MARS untuk menetapkan batas keputusan yang lebih tepat, sehingga menyelesaikan masalah ketidakseimbangan kelas dan menghasilkan prediksi risiko DMT2 dengan baik.

SIMPULAN DAN SARAN

Penelitian ini menyimpulkan bahwa pendekatan hybrid SMOTE-MARS terbukti efektif dalam mengatasi ketidakseimbangan kelas pada pemodelan risiko Diabetes Mellitus Tipe 2, yang ditunjukkan oleh peningkatan kinerja model dengan akurasi mencapai 0,7 dan nilai AUC sebesar 0,73, serta peningkatan sensitivitas yang signifikan menjadi 0,875 dibandingkan metode MARS standar. Secara spesifik, bahwa individu dengan berat badan lebih dari 62 kg memiliki peningkatan risiko sebesar 1,12 kali lipat terkena DMT2. Dengan kemampuan sensitivitas yang tinggi dalam mendeteksi kasus positif, model ini menawarkan potensi besar sebagai instrumen deteksi dini yang andal. Namun, penelitian ini menggunakan sampel terbatas dari satu rumah sakit. Untuk penelitian selanjutnya, disarankan memperbesar ukuran sampel dan memperluas cakupan pengambilan data dari berbagai pusat kesehatan (*multicenter study*). Hal ini bertujuan untuk meningkatkan kemampuan generalisasi model dan menangkap variasi karakteristik populasi yang lebih luas.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada pihak Rumah Sakit Universitas Airlangga (RSUA) Surabaya yang telah memberikan izin serta memfasilitasi akses data rekam medis yang digunakan dalam penelitian ini. Penulis juga menyampaikan apresiasi kepada Departemen Matematika, Fakultas Sains dan Teknologi, Universitas Airlangga atas dukungan fasilitas yang diberikan selama penyusunan artikel ilmiah ini.

DAFTAR PUSTAKA

Barron, A. R., & Xiao, X. (1991). Discussion: Multivariate adaptive regression splines. *The Annals of Statistics*, 19(1), 67–82.

Chamidah, N., Hasyim, M., Saifudin, T., & Lestari, B. (2025). Nonparametric Bi-Response Ordinal Logistic Regression

Model for Diabetes Mellitus and Hypertension Risks Based on Multivariate Adaptive Regression Spline. *Data and Metadata*, 4, 912.

- Chamidah, N., & Lestari, B. (2022). *Analisis Regresi Nonparametrik dengan Perangkat Lunak R*. Airlangga University Press.
- Chang, H.-Y., Li, X., & Karniadakis, G. E. (2017). Modeling of biomechanics and biorheology of red blood cells in type 2 diabetes mellitus. *Biophysical journal*, 113(2), 481–490.
- Fatarona, A., Cahyono, H. D., & Martiana, I. (2025). Upaya Pengendalian Diabetes Mellitus (DM) dengan Education Implementation pada Kelompok DM dan Kelompok Beresiko DM pada Masyarakat Urban. *Bhakti Patrika*, 1(1), 8–13.
- Friedman, J. H. (1991). Multivariate adaptive regression splines. *The annals of statistics*, 19(1), 1–67.
- Hasyim, M., Chamidah, N., & Saifudin, T. (2025). Hypertension modelling using nonparametric ordinal logistic regression based on multivariate adaptive regression spline. *Commun. Math. Biol. Neurosci.*, 2025, Article-ID.
- Jannoo, Z., Wah, Y. B., Lazim, A. M., & Hassali, M. A. (2017). Examining diabetes distress, medication adherence, diabetes self-care activities, diabetes-specific quality of life and health-related quality of life among type 2 diabetes mellitus patients. *Journal of clinical & translational endocrinology*, 9, 48–54.
- Li, J., Bao, Y., Chen, X., & Tian, L. (2021). Decision models in type 2 diabetes mellitus: a systematic review. *Acta Diabetologica*, 58(11), 1451–1469.
- Mushtaq, Z., Ramzan, M. F., Ali, S., Baseer, S., Samad, A., & Husnain, M. (2022). Voting Classification-Based Diabetes Mellitus Prediction Using Hypertuned

- Machine-Learning Techniques. *Mobile Information Systems*, 2022(1), 6521532.
- Organization, W. H. (2018). Noncommunicable diseases country profiles 2018. Dalam *Noncommunicable diseases country profiles 2018*.
- Pradipta, G. A., Wardoyo, R., Musdholifah, A., Sanjaya, I. N. H., & Ismail, M. (2021). SMOTE for handling imbalanced data problem: A review. *2021 sixth international conference on informatics and computing (ICIC)*, 1–8.
- Rahman, S. (2021). [BUKU] *Diabetes Melites Tipe 2 dan Akses Pelayanan Kesehatan*.
- Rahmawati, D. (2024). Kualitas Hidup Pasien Diabetes Melitus dan Hipertensi dalam Program Penyakit Kronis (Prolanis) di Indonesia: Narative Review. *Jurnal Mandala Pharmacoin Indonesia*, 10(1), 116–122.
- Safitri, L. R., Chamidah, N., & Saifudin, T. (2025). Modeling of stroke risk using synthetic minority oversampling technique in multivariate adaptive regression spline model. *Commun. Math. Biol. Neurosci.*, 2025, Article-ID.
- Salam, A., Azhari, L., Septarini, R. S., & Heriyani, N. (2025). Pendekatan Hybrid K-Means SMOTE dan Logistic Regression Untuk Deteksi Dini Diabetes Mellitus Pada Imbalanced Data. *Bulletin of Computer Science Research*, 5(3), 219–227.
- Sayadi, M., Zibaenezhad, M. J., & AYATOLLAHI, S. M. T. (2017). *Simple prediction of type 2 diabetes mellitus via decision tree modeling*.
- Sidiq, S., Alfian, A., & Mabrur, N. S. (2025). Pengembangan Model Prediksi Risiko Diabetes Menggunakan Pendekatan AdaBoost dan Teknik Oversampling SMOTE. *Jurnal Ilmiah Informatika dan Ilmu Komputer (JIMA-ILKOM)*, 4(1), 13–23.
- Tjahjono, E., Mardianto, M. F. F., & Chamidah, N. (2018). Prediction of electricity consumption using Fourier series estimator in bi-response nonparametric regression model. *Far East J. Math. Sci*, 103, 1251–1263.
- Wahidin, M., Achadi, A., Besral, B., Kosen, S., Nadjib, M., Nurwahyuni, A., Ronoatmodjo, S., Rahajeng, E., Pane, M., & Kusuma, D. (2024). Projection of diabetes morbidity and mortality till 2045 in Indonesia based on risk factors and NCD prevention and control programs. *Scientific reports*, 14(1), 5424.
- Wang, X., Zhai, M., Ren, Z., Ren, H., Li, M., Quan, D., Chen, L., & Qiu, L. (2021). Exploratory study on classification of diabetes mellitus through a combined Random Forest Classifier. *BMC medical informatics and decision making*, 21(1), 105.
- Yasmirullah, S. D. P., Otok, B. W., Purnomo, J. D. T., & Prastyo, D. D. (2021). Modification of multivariate adaptive regression spline (MARS). *Journal of Physics: Conference Series*, 1863(1), 012078.

Biografi Penulis

Canada Mewa Fizkadana merupakan mahasiswa dan peneliti dari Program Studi Magister Matematika, Fakultas Sains dan Teknologi, Universitas Airlangga. Fokus penelitiannya meliputi statistika terapan, khususnya dalam penggunaan metode regresi nonparametrik untuk analisis data kesehatan dan sosial.





Nur Chamidah merupakan Guru Besar di Departemen Matematika, Fakultas Sains dan Teknologi, Universitas Airlangga. Beliau menyelesaikan pendidikan Doktorat (S3) di Institut Teknologi Sepuluh Nopember (ITS) Surabaya. Bidang keahlian dan minat penelitian utamanya adalah statistika nonparametrik dan semiparametrik, khususnya pengembangan estimator Spline dan Local Polynomial, serta aplikasinya pada pemodelan pertumbuhan anak (stunting) dan data lingkungan.



Ardi Kurniawan merupakan staf dosen pengajar di Departemen Matematika, Fakultas Sains dan Teknologi, Universitas Airlangga. Beliau menempuh pendidikan Doktorat di Universitas Airlangga dengan spesialisasi statistika. Fokus penelitiannya mencakup biostatistika, analisis survival, serta metode statistika lainnya yang diterapkan pada ilmu kehayatan dan kesehatan.



Naufal Ramadhan Al Akhwal Siregar adalah peneliti muda dan mahasiswa Program Studi Magister Matematika, Fakultas Sains

dan Teknologi, Universitas Airlangga. Ia aktif mempublikasikan karya ilmiah di jurnal nasional maupun prosiding internasional. Fokus penelitiannya (research interest) meliputi pemodelan statistika, khususnya Regresi Nonparametrik (Spline) dan aplikasinya dalam analisis data kesehatan (seperti risiko penyakit jantung) serta pemodelan dampak pandemi COVID-19.