

TESIS
PEMBELAJARAN MESIN UNTUK PREDIKSI STROKE
BERDASARKAN RANDOM FOREST DAN SHAP



disusun oleh
HARUMAWAN JATIPRASETYA
23.51.2530
Konsentrasi : Digital Transformation Intelligence

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026

TESIS
PEMBELAJARAN MESIN UNTUK PREDIKSI STROKE
BERDASARKAN RANDOM FOREST DAN SHAP

MACHINE LEARNING FOR STROKE PREDICTION
BASED ON RANDOM FOREST AND SHAP

Diajukan untuk memenuhi salah satu syarat mencapai derajat Pascasarjana
Program Studi Magister Teknik informatika



disusun oleh

HARUMAWAN JATIPRASETYA

23.51.2530

Konsentrasi : Digital Transformation Intelligence

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA

2026

HALAMAN PERSETUJUAN
PEMBELAJARAN MESIN UNTUK PREDIKSI STROKE
BERDASARKAN RANDOM FOREST DAN SHAP
MACHINE LEARNING FOR STROKE PREDICTION
BASED ON RANDOM FOREST AND SHAP

yang disusun dan diajukan oleh

Harumawan Jatiprasetya

23.51.2530

telah disetujui oleh Dosen Pembimbing Tesis
pada tanggal 4 Maret 2026

~~Dosen Pembimbing,~~


Prof. Dr. Ema Ummi, S.Si., M.Kom
NIK. 190302037

HALAMAN PENGESAHAN

**PEMBELAJARAN MESIN UNTUK PREDIKSI STROKE
BERDASARKAN RANDOM FOREST DAN SHAP**

MACHINE LEARNING FOR STROKE PREDICTION

BASED ON RANDOM FOREST AND SHAP

yang disusun dan diajukan oleh

Harumawan Jatiprasetya

23.51.2530

Telah dipertahankan di depan Dewan Penguji
pada tanggal 4 Maret 2026

Susunan Dewan Penguji

Nama Penguji

Dr. Kumara Ari Yuana, S.T., M.T.
NIK. 190302575

Dr. Andi Sunyoto, S.Kom, M.Kom.
NIK. 190302052

Prof. Dr. Ema Utami, S.Si., M.Kom
NIK. 190302037

Tanda Tangan



Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer
Tanggal 4 Maret 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : **Harumawan Jatiprasetya**
NIM : **23.51.2530**

Menyatakan bahwa Tesis dengan judul berikut:

PEMBELAJARAN MESIN UNTUK PREDIKSI STROKE BERDASARKAN RANDOM FOREST DAN SHAP

Dosen Pembimbing : Prof. Dr. Ema Utami, S.Si., M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 4 Maret 2026

Yang Menyatakan,



The image shows an official stamp of Universitas AMIKOM Yogyakarta. The stamp includes the university's logo, the text 'UNIVERSITAS AMIKOM YOGYAKARTA', and 'METERAI TEMPEL'. Below the stamp is a handwritten signature in black ink.

Harumawan Jatiprasetya

HALAMAN PERSEMBAHAN

Segala puji dan syukur saya panjatkan ke hadirat Allah SWT atas limpahan rahmat, taufik, dan hidayah-Nya. Dengan penuh kerendahan hati, karya ini saya dedikasikan kepada:

- Kedua orang tua dan istri tercinta, atas doa, kasih sayang, dan dukungan tiada henti.
- Seluruh keluarga besar yang menjadi sumber kekuatan dan inspirasi.
- Para dosen serta pembimbing yang dengan penuh ketulusan dan kesabaran telah membagikan ilmu.
- Sahabat dan rekan seperjuangan yang selalu memberikan dukungan dan semangat.
- Almater tercinta, tempat berharga dalam menimba ilmu dan pengalaman.

Semoga karya ini dapat memberi manfaat, khususnya dalam pengembangan ilmu pengetahuan serta kemaslahatan masyarakat.



KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas rahmat dan karunia-Nya sehingga tesis berjudul “Prediksi Stroke Menggunakan Random Forest dan SHAP” dapat diselesaikan. Tesis ini disusun sebagai salah satu persyaratan memperoleh gelar Magister Informatika pada Program Studi Magister Informatika Universitas AMIKOM Yogyakarta, dengan menerapkan algoritma Random Forest dan metode SHAP (Shapley Additive Explanations) untuk memprediksi serta menginterpretasikan risiko stroke berdasarkan data kesehatan.

Dalam proses penyusunan tesis ini, penulis menyadari bahwa banyak pihak yang telah memberikan bantuan, dukungan, serta bimbingan. Oleh karena itu, pada kesempatan ini penulis menyampaikan terima kasih yang sebesar-besarnya kepada:

- Prof. Dr. M. Suyanto, M.M., selaku rektor universitas amikom yogyakarta, yang telah memberikan kesempatan dan fasilitas kepada penulis untuk menempuh pendidikan di universitas ini.
- Prof. Dr. Ema Utami, S.Si., M.Kom., selaku dosen pembimbing tesis, yang telah memberikan bimbingan, arahan, serta masukan yang sangat berharga selama proses penelitian dan penyusunan tesis ini.
- Dr. Andi Sunyoto, S.Kom. M.Kom. dan Dr. Kumara Ari Yuana, S.T., M.T. selaku dosen penguji sidang tesis, yang telah memberikan berbagai saran, kritik, dan masukan yang konstruktif demi penyempurnaan penelitian ini.
- kedua orang tua tercinta, beserta istri yang senantiasa memberikan doa, dukungan, serta motivasi yang tidak pernah putus kepada penulis selama menempuh pendidikan.
- rekan-rekan mahasiswa magister informatika angkatan 31 kelas a, atas kebersamaan, dukungan, dan semangat selama menjalani proses perkuliahan hingga penyelesaian tesis ini.

Penulis menyadari bahwa tesis ini masih memiliki keterbatasan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan untuk pengembangan penelitian selanjutnya. Penulis berharap tesis ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya di bidang machine learning dan kesehatan, serta menjadi referensi bagi penelitian berikutnya.

Yogyakarta, 11 maret 2026

Penulis

DAFTAR ISI

HALAMAN JUDUL	1
HALAMAN JUDUL (INGGRIS)	2
HALAMAN PENGESAHAN	3
HALAMAN PERNYATAAN KEASLIAN	4
HALAMAN PERSEMBAHAN	5
KATA PENGANTAR	6
DAFTAR ISI	7
DAFTAR TABEL	10
DAFTAR GAMBAR	11
INTISARI	13
ABSTRACT	14
BAB 1 PENDAHULUAN	15
1.1 Latar Belakang	15
1.2 Rumusan Masalah	21
1.3 Batasan Masalah	22
1.4 Tujuan Penelitian	23
1.5 Manfaat Penelitian	25
BAB 2 TINJAUAN PUSTAKA	28
2.1 Tinjauan Pustaka	28
2.2 Keaslian Penelitian	40
2.3 Landasan Teori	50
2.3.1 Stroke	50

2.3.2 Machine Learning	50
2.3.3 Algoritma Random Forest	50
2.3.4 Random Forest untuk Prediksi Stroke	51
2.3.5 SHAP (Shapley Additive Explanations)	53
BAB 3 METODE PENELITIAN	59
3.1 Alur Penelitian	59
3.2 Jenis, Sifat, dan Pendekatan Penelitian	62
3.3 Metode Pengumpulan Data	63
3.4 Metode Analisis Data (Pipeline Penelitian)	63
3.4.1 Exploratory Data Analysis (EDA)	63
3.4.2 Pra-Pemrosesan Data (Data Preprocessing)	64
3.4.3 Penanganan Ketidakseimbangan Kelas (Class Imbalance)	65
3.4.4 Pembangunan, Optimasi, dan Validasi Model	67
3.4.5 Evaluasi Model	67
3.4.6 Interpretasi Model dan Audit Fairness	69
3.5 Lingkungan dan Perangkat Lunak	70
BAB 4 HASIL DAN PEMBAHASAN	71
4.1 Gambaran Umum dan Karakteristik Data Awal	71
4.2 Hasil Pra-Pemrosesan Data dan EDA	71
4.2.1 Penanganan Data dan Analisis Korelasi	71
4.2.2 Dampak SMOTE dan Pencegahan Data Leakage	74
4.3 Hasil Pembangunan dan Evaluasi Model	88
4.3.1 Hasil Optimasi Hyperparameter Random Forest	88
4.3.2 Validasi Silang dan Perbandingan Komprehensif Algoritma	89

4.3.3 Kinerja Generalisasi pada Hold-Out Test Set	93
4.4 Analisis Interpretabilitas Model dengan SHAP	103
4.4.1 Validasi Global: Summary Plot SHAP	103
4.4.2 SHAP Waterfall Plot	108
4.4.3 SHAP Clustering	110
4.4.4 SHAP Beeswarm Plot	112
4.4.5 SHAP Heatmap	114
4.4.6 SHAP Dependence Plot untuk BMI	115
4.4.7 SHAP Dependence Plot untuk Usia	117
4.4.8 SHAP Force Plot	120
4.5 Analisis Fairness dan Keterbatasan Model	121
4.5.1 Hasil Audit Fairness Berdasarkan Gender	121
4.5.2 Diskusi Keterbatasan dan Implikasi Validasi	122
4.6 Implikasi Klinis	122
4.6.1 Potensi Integrasi ke Sistem Klinik Nyata (EHR)	122
4.6.2 Studi Kasus Klinis	123
4.6.3 Tahap Akuisisi dan Input Data Klinis	126
BAB 5 PENUTUP	129
5.1 Kesimpulan	129
5.2 Saran	130
DAFTAR PUSTAKA	132

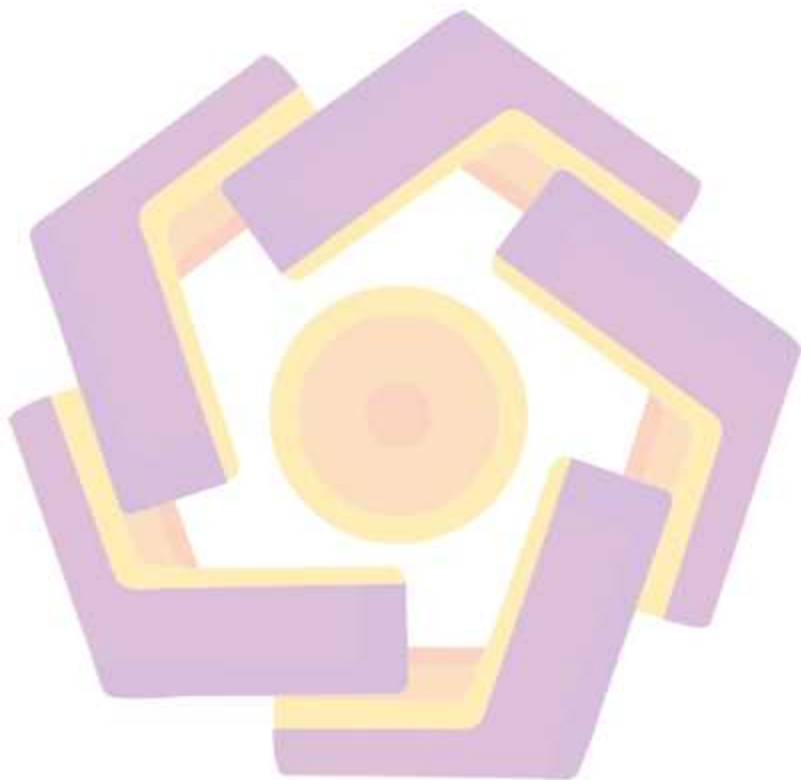
DAFTAR TABEL

Tabel 2.1	Matriks literatur review dan posisi penelitian	41
Tabel 2.2	KernelSHAP dan TreeSHAP	56
Tabel 3.1	Kinerja Validasi Random Forest	67
Tabel 4.1	Matriks BMI	73
Tabel 4.2	Matriks Jumlah Stroke	75
Tabel 4.3	Matriks Korelasi dalam Pemodelan Risiko Stroke	81
Tabel 4.4	Matriks SMOTE	87
Tabel 4.5	Konfigurasi Hyperparameter Terbaik Random Forest	88
Tabel 4.6	Hasil Validasi Silang 5-Fold untuk Berbagai Algoritma	89
Tabel 4.7	Matriks Algoritma	92
Tabel 4.8	Random Forest > XGBoost	93
Tabel 4.9	Karakteristik False Positive	95
Tabel 4.10	Karakteristik False Negative	95
Tabel 4.11	Implikasi Klinis	95
Tabel 4.12	Stabilitas Performa Random Forest pada 5-Fold CV	96
Tabel 4.13	Matriks Final Confusion Matrix	97
Tabel 4.14	Fitur Gini Importance	100
Tabel 4.15	Klasifikasi Random Forest pada Data Testing	102
Tabel 4.16	Hierarki Fitur Berdasarkan SHAP	107
Tabel 4.17	Variabel SHAP	113
Tabel 4.18	Audit Fairness Gender	121

DAFTAR GAMBAR

Gambar 2.1	Contoh Arsitektur Random Forest	51
Gambar 2.2	Arsitektur SHAP	58
Gambar 3.1	Alur Penelitian	59
Gambar 4.1	Dataset Awal	71
Gambar 4.2	Distribusi BMI Sebelum dan Setelah Imputasi	73
Gambar 4.3	Visualisasi Jumlah Kasus Stroke	74
Gambar 4.4	Distribusi Fitur Kategorikal	76
Gambar 4.5	Distribusi Kadar Glukosa	77
Gambar 4.6	Jumlah Kasus Stroke Menurut Jenis Kelamin	78
Gambar 4.7	Distribusi Kasus Stroke Berdasarkan Status Merokok	79
Gambar 4.8	Matriks Korelasi dalam Pemodelan Risiko Stroke	80
Gambar 4.9	Perbandingan Distribusi BMI Sebelum dan Sesudah Standardisasi	83
Gambar 4.10	Visualisasi SMOTE Sebelum dan Sesudah	85
Gambar 4.11	Perbandingan Kinerja Model pada Data Pelatihan dan Pengujian	91
Gambar 4.12	Confusion Matrix Random Forest (Test dan Train)	94
Gambar 4.13	ROC Curve Random Forest	98
Gambar 4.14	Feature Importance Random Forest	100
Gambar 4.15	SHAP Summary Plot	104
Gambar 4.16	SHAP Waterfall Plot	108
Gambar 4.17	SHAP Clustering	111
Gambar 4.18	SHAP Beeswarm Plot	112
Gambar 4.19	SHAP Heatmap	114

Gambar 4.20	SHAP Dependence Plot untuk BMI	116
Gambar 4.21	SHAP Dependence Plot untuk Usia	118
Gambar 4.22	SHAP Force Plot	120
Gambar 4.23	Visualisasi User Interface	124
Gambar 4.24	Alur Sistem Pendukung Keputusan Klinis Berbasis SHAP	125



INTISARI

Dengan lebih dari 12,2 juta kasus baru setiap tahun, stroke merupakan salah satu penyebab utama kematian dan disabilitas di dunia. Oleh karena itu, deteksi dini menjadi sangat penting untuk menurunkan angka mortalitas dan komplikasi jangka panjang. Untuk meningkatkan transparansi dan keandalan dalam proses pengambilan keputusan model, penelitian ini mengembangkan model prediksi stroke berbasis algoritma Random Forest yang dipadukan dengan metode interpretasi SHAP (Shapley Additive Explanations). Hasil evaluasi menunjukkan bahwa model memiliki kinerja klasifikasi yang sangat baik dengan nilai AUC sebesar $0,989 \pm 0,002$, akurasi sebesar 94,3%, presisi 92,5%, recall 96,5%, dan F1-score 94,5%. Selain menghasilkan prediksi yang akurat, pendekatan SHAP memungkinkan interpretasi kontribusi setiap fitur terhadap keputusan model. Visualisasi interpretasi seperti summary plot, dependence plot, dan force plot menunjukkan bahwa faktor usia, kadar glukosa darah, dan indeks massa tubuh (BMI) merupakan kontributor utama dalam meningkatkan risiko stroke. Analisis interaksi fitur juga menunjukkan bahwa kombinasi usia lanjut dan kadar glukosa darah tinggi secara signifikan meningkatkan probabilitas stroke hingga sekitar 85% dibandingkan dengan nilai dasar rata-rata model sebesar 0,4992. Dengan demikian, pendekatan ini tidak hanya menghasilkan model prediksi yang memiliki performa tinggi, tetapi juga menyediakan interpretasi yang dapat dipahami oleh tenaga medis, sehingga berpotensi mendukung pengambilan keputusan klinis secara lebih transparan dan berbasis data.

Kata kunci: Stroke, Pembelajaran Mesin, SHAP, Random Forest, Prediksi Dini, EHR.

ABSTRACT

With more than 12.2 million new cases each year, stroke is one of the leading causes of death and disability worldwide. Therefore, early detection is crucial to reduce mortality and long-term complications. To improve transparency and reliability in the model decision-making process, this study developed a stroke prediction model based on the Random Forest algorithm combined with the SHAP (Shapley Additive Explanations) interpretation method. The evaluation results show that the model has excellent classification performance with an AUC value of 0.989 ± 0.002 , accuracy of 94.3%, precision of 92.5%, recall of 96.5%, and an F1-score of 94.5%. In addition to producing accurate predictions, the SHAP approach allows for the interpretation of each feature's contribution to the model's decision. Visualizations such as summary-plots, dependence plots, and force plots show that age, blood glucose levels, and body mass index (BMI) are the main contributors to an increased risk of stroke. transparent and data-driven. Feature interaction analysis also shows that the combination of advanced age and high blood glucose levels significantly increases the probability of stroke by approximately 85% compared to the model's average baseline value of 0.4992. Thus, this approach not only produces a high-performance prediction model, but also provides interpretations that are understandable to medical personnel, thereby potentially supporting more transparent and data-driven clinical decision-making.

Keyword: Stroke, Machine Learning, SHAP, Random Forest, Early Prediction, EHR.

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Stroke merupakan salah satu penyakit kardiovaskular paling berbahaya dan berpotensi mengancam jiwa, dengan lebih dari 12,2 juta kasus baru tercatat setiap tahun di seluruh dunia (Umirzakova et al., 2023). Kondisi ini dapat menimbulkan kerusakan neurologis serius seperti kelumpuhan wajah, gangguan motorik, bahkan kematian (Ding et al., 2024). Deteksi dini stroke sangat krusial karena dapat mengurangi risiko disabilitas permanen maupun angka mortalitas (Hossain et al., 2025).

Kenyataannya lebih dari setengah kasus stroke gagal terdiagnosis dengan tepat oleh tenaga medis darurat, meskipun telah menggunakan instrumen skrining seperti Cincinnati Prehospital Stroke Scale (CPSS) (Kokkotis et al., 2022). Metode pencitraan Non-Contrast Computed Tomography (NCCT) selama ini digunakan sebagai standar pemeriksaan stroke (Choi et al., 2023). Akan tetapi, keterbatasan signifikan muncul terutama pada stroke iskemik, di mana kerusakan jaringan otak biasanya baru terlihat setelah enam jam onset (Borsos et al., 2024). Keterlambatan ini berdampak langsung pada efektivitas penanganan medis. Oleh sebab itu, diperlukan pendekatan alternatif berbasis data klinis dan riwayat pasien untuk mendukung deteksi yang lebih cepat dan akurat (Melnykova et al., 2025a).

Dalam konteks transformasi digital, pemanfaatan Electronic Health Records

(EHR) menjadi semakin penting. EHR berfungsi sebagai catatan medis yang komprehensif, real-time, dan berpusat pada pasien(Lee et al., 2023). Data klinis yang tercatat, seperti usia, jenis kelamin, tekanan darah, kadar glukosa, indeks massa tubuh (BMI), riwayat hipertensi, dan kebiasaan merokok, berperan penting sebagai faktor risiko stroke(Su et al., 2022). Kompleksitas dan besarnya data sering kali tidak dapat diolah secara optimal dengan metode analisis tradisional(Chakraborty et al., 2024a). Oleh karena itu, penerapan algoritma machine learning (ML) diperlukan untuk mengenali pola non-linear, mendeteksi interaksi variabel, dan menghasilkan prediksi dengan tingkat akurasi yang lebih tinggi(Heseltine-Carp et al., 2025a).

Apabila akurasi prediksi stroke tidak segera ditingkatkan, maka akan muncul implikasi negatif yang mungkin terjadi sangat serius(Shobayo et al., 2023). Pertama, meningkatnya angka disabilitas akibat diagnosis terlambat yang dapat memicu kelumpuhan permanen maupun gangguan kognitif. Kedua, bertambahnya angka kematian akibat intervensi medis yang tidak tepat waktu. Ketiga, munculnya beban ekonomi yang besar, karena pasien dengan komplikasi jangka panjang membutuhkan biaya perawatan yang tinggi(Sumwiza et al., 2023).

Sejumlah penelitian sebelumnya telah menggunakan algoritma ML seperti Support Vector Machine (SVM), Convolutional Neural Network (CNN), dan Logistic Regression untuk prediksi stroke(Guo et al., 2025a). Walaupun capaian akurasi cukup baik, persoalan besar muncul pada aspek interpretabilitas model(Gupta et al., 2025a). Banyak algoritma dianggap sebagai black box karena sulit menjelaskan dasar prediksi yang dihasilkan(Melnykova et al., 2025b). Dalam

dunia medis, keterbatasan ini menjadi hambatan serius, sebab dokter tidak hanya memerlukan prediksi yang akurat, tetapi juga penjelasan yang transparan agar keputusan klinis dapat dipertanggungjawabkan.

Untuk menjawab tantangan tersebut, penelitian ini mengusulkan penerapan Random Forest sebagai model utama, dikombinasikan dengan metode interpretabilitas SHAP (Shapley Additive Explanations). Random Forest dipilih karena kemampuannya menangani data multidimensi, baik numerik maupun kategorikal, serta kestabilannya dalam mengurangi risiko overfitting (Gupta et al., 2025b).

Dalam bidang explainable artificial intelligence (XAI), terdapat tantangan fundamental terkait interpretabilitas model ensemble berbasis pohon keputusan. Random Forest, meskipun unggul dalam akurasi prediktif, beroperasi sebagai "black box" yang sulit diinterpretasi oleh pengguna (Ali et al., 2023). Penggunaan Random Forest pada data klinis yang heterogen (numerik dan kategorikal), mampu bertahan terhadap outlier, kemampuan menangkap interaksi non-linear antar fitur, dan relatif mudah diinterpretasi fitur importance-nya dibanding model ensemble kompleks seperti deep learning (Resti et al., 2026).

SHAP (SHapley Additive exPlanations) muncul sebagai solusi dengan memberikan penjelasan berbasis teori permainan kooperatif. Namun demikian, penggunaan SHAP bukannya tanpa masalah. Tiga permasalahan utama mendasari penelitian ini. Pertama terkait dengan stabilitas eksplanasi, nilai SHAP pada Random Forest diketahui memiliki variabilitas tinggi akibat sifat stokastik dalam algoritma, mencakup bootstrap sampling dan random feature selection. Dua kali

pelatihan pada dataset yang sama dapat menghasilkan ranking fitur yang berbeda(Lange et al., 2025). Kedua terkait bias dalam feature Importance, seperti yang dijelaskan pada penelitian(Takefuji, 2025), membuktikan bahwa Gini importance yang digunakan Random Forest cenderung overestimate terhadap fitur kontinu dan fitur dengan banyak kategori. SHAP interaction diharapkan dapat mengoreksi bias ini, namun validasi empirisnya masih terbatas. Ketiga terkait dengan kompleksitas komputasi dimana Algoritma TreeSHAP memiliki kompleksitas $O(TLD^2 \cdot M)$ yang tidak linear terhadap kedalaman pohon. Pada dataset dengan dimensi menengah ($M=11$), komputasi eksak membutuhkan ~307 juta operasi, tidak praktis untuk implementasi real-time(Arenas et al., 2023).

Penelitian ini menggunakan prediksi stroke sebagai domain kasus untuk mengkaji tiga permasalahan fundamental dalam XAI tersebut. Dengan kata lain, penelitian ini tidak bertujuan membangun sistem prediksi stroke, melainkan mengkaji properti algoritmik SHAP pada Random Forest. SHAP berfungsi sebagai pendekatan interpretasi berbasis teori permainan kooperatif, yang memberikan kontribusi adil tiap fitur terhadap prediksi model(Guo et al., 2025b). Karakter aditif nilai SHAP memastikan konsistensi, transparansi, dan keadilan dalam menilai pengaruh faktor risiko seperti usia, BMI, kadar glukosa, maupun status merokok(Al Mamlook et al., 2024a).

Lebih lanjut, SHAP tidak hanya dipakai untuk menjelaskan hasil prediksi, tetapi juga berpotensi digunakan dalam kalibrasi ulang model guna mengurangi dominasi fitur tertentu, seperti usia(Kim, 2025a). Integrasi SHAP dengan sistem EHR bahkan memungkinkan dokter melakukan simulasi intervensi klinis secara

real-time(Ponce-Bobadilla et al., 2024a). Dengan demikian, SHAP tidak hanya menjadi alat analisis, melainkan juga bagian dari clinical decision support system (CDSS) yang relevan, adaptif, dan berorientasi pada pasien(McFall et al., 2023). Di sinilah SHAP berperan krusial. Berbasis teori permainan kooperatif Shapley, SHAP memberikan penjelasan yang adil, konsisten, dan lokal(Lamane et al., 2025a).

Sejumlah penelitian sebelumnya telah membuktikan efektivitas algoritma pembelajaran mesin dalam menganalisis data Electronic Health Records (EHR) untuk prediksi stroke. Misalnya, penelitian berjudul *A predictive analytics approach for stroke prediction using machine learning and neural networks* (Dev et al., 2022) menggunakan neural networks dan menemukan bahwa usia, penyakit jantung, kadar glukosa rata-rata, serta hipertensi merupakan faktor prediktor utama stroke, dengan capaian akurasi sekitar 78%. Selanjutnya, studi *A comparative analysis of machine learning classifiers for stroke prediction: A predictive analytics approach* (Biswas et al., 2022) melakukan perbandingan algoritma K-Nearest Neighbor (KNN) dan Decision Tree, di mana KNN menunjukkan hasil lebih unggul dengan akurasi mencapai 97,18%.

Penelitian lain, *Predicting stroke occurrences: a stacked machine learning approach with feature selection and data preprocessing*(Chakraborty et al., 2024b), melaporkan bahwa algoritma Random Forest mampu memberikan akurasi hingga 98,5%. Sementara itu, studi *Random forest-based prediction of stroke outcome*(Fernandez-Lozano et al., 2021a) menegaskan keunggulan Random Forest dalam memprediksi kemungkinan kematian pasien stroke berdasarkan variabel klinis maupun biokimia.

Di sisi lain, beberapa penelitian mencoba mengombinasikan algoritma prediksi dengan pendekatan interpretabilitas seperti SHAP (Shapley Additive Explanations) untuk meningkatkan transparansi model. Studi *Interpretable prediction of stroke prognosis: SHAP for SVM and nomogram for logistic regression* (Guo et al., 2025a) serta *Improving stroke risk prediction by integrating XGBoost, optimized principal component analysis, and explainable artificial intelligence* (Mochurad et al., 2025) menunjukkan bahwa penggunaan SHAP bersama algoritma XGBoost tidak hanya meningkatkan akurasi, tetapi juga memberikan tingkat transparansi yang tinggi dalam menjelaskan kontribusi masing-masing fitur terhadap hasil prediksi.

Kontribusi utama penelitian ini meliputi: (1) peningkatan akurasi prediksi stroke melalui optimasi parameter (grid search) dan reduksi dimensi dengan PCA, (2) identifikasi faktor risiko utama stroke melalui visualisasi SHAP seperti *summary plot*, *dependence plot*, dan *force plot*, serta (3) penerapan hasil interpretasi dalam mendukung pengambilan keputusan klinis secara transparan. Berdasarkan uraian tersebut, penelitian ini dirancang untuk menjawab dua kebutuhan mendesak: (1) meningkatkan akurasi prediksi stroke dengan algoritma tangguh seperti Random Forest, dan (2) memastikan transparansi serta interpretabilitas prediksi menggunakan SHAP. Kombinasi keduanya diharapkan menghasilkan model prediksi yang tidak hanya akurat, tetapi juga dapat dipercaya oleh praktisi medis, sehingga berguna dalam skrining dini, diagnosis pendukung, maupun upaya preventif berbasis data.

Tantangan utama dalam komputasi prediksi stroke bukan sekadar mencapai

akurasi tinggi, melainkan memastikan stabilitas performa dan konsistensi interpretasi model saat menghadapi data klinis riil (EHR) yang heterogen, mengandung noise, dan tidak seimbang secara ekstrem. Karena penelitian sebelumnya sering kali hanya mengoptimasi metrik evaluasi tanpa menguji sensitivitas terhadap variasi pra-pemrosesan, penelitian ini difokuskan pada pengembangan pipeline machine learning yang tangguh guna menjamin keandalan dan interpretabilitas model dalam penerapan klinis yang nyata.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, permasalahan penelitian ini dirumuskan sebagai berikut:

a. Stabilitas SHAP

Bagaimana merancang dan mengevaluasi SHAP Stability Score (S^s) untuk mengukur konsistensi ranking fitur pada Random Forest, dan sejauh mana faktor bootstrap sampling serta random feature selection berkontribusi terhadap variabilitas Shapley value?

b. Bias Feature Importance

Sejauh mana SHAP interaction values mampu mengoreksi bias overestimasi pada Gini importance akibat multikolinieritas antar fitur?

c. Efisiensi Komputasi

Bagaimana trade-off antara bias estimasi SHAP dan efisiensi waktu pada berbagai strategi background sampling dalam TreeExplainer?

d. Representasi Ketidakpastian

Bagaimana merepresentasikan epistemic uncertainty dalam visualisasi SHAP melalui pendekatan bootstrapped confidence bands?

1.3 Batasan Masalah

- a. Sumber Data: Penelitian ini memanfaatkan dataset publik dari platform Kaggle yang terdiri atas 5.000 data pasien. Konsekuensinya, performa model yang dihasilkan sangat dipengaruhi oleh kualitas, distribusi, serta representasi data tersebut, sehingga belum tentu sepenuhnya mencerminkan kondisi populasi nyata di fasilitas kesehatan lokal.
- b. Faktor Risiko: Analisis hanya dilakukan pada variabel klinis dan demografis yang tersedia dalam dataset, meliputi usia, jenis kelamin, indeks massa tubuh (BMI), kadar glukosa darah, status merokok, serta riwayat hipertensi. Faktor eksternal lain seperti kebiasaan makan, riwayat keluarga, aktivitas fisik, maupun aspek genetik tidak termasuk dalam ruang lingkup penelitian ini.
- c. Algoritma yang Digunakan: Fokus penelitian diarahkan pada penggunaan Random Forest sebagai model utama dan Logistic Regression sebagai model pembanding. Walaupun terdapat algoritma lain seperti XGBoost, Support Vector Machine (SVM), atau Neural Network, penelitian ini tidak membahasnya guna menjaga konsistensi ruang lingkup kajian.

- d. **Evaluasi Model:** Kinerja model dievaluasi melalui metrik standar, yakni akurasi, presisi, recall, F1-score, dan AUC, dengan interpretasi hasil menggunakan metode SHAP. Pengujian tambahan seperti analisis ketahanan model terhadap noise, outlier, atau data tak seimbang tidak dibahas dalam penelitian ini.
- e. **Fairness Audit:** Audit keadilan model dibatasi pada variabel sensitif gender. Analisis fairness berdasarkan dimensi lain, seperti etnis, status sosial-ekonomi, atau tingkat pendidikan, tidak termasuk dalam ruang lingkup studi ini.
- f. **Lingkup Implementasi:** Model prediksi yang dikembangkan hanya diuji pada konteks penelitian menggunakan dataset Kaggle. Penerapan secara langsung dalam sistem layanan kesehatan nyata, termasuk integrasi dengan electronic health record (EHR) rumah sakit, belum dilakukan dan menjadi agenda penelitian lanjutan.

1.4 Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah yang telah diuraikan, penelitian ini memiliki tujuan utama untuk mengembangkan dan mengevaluasi sebuah pipeline prediksi risiko stroke yang akurat, stabil, dapat diinterpretasi, dan adil dengan mengintegrasikan algoritma Random Forest dan metode SHAP. Secara rinci, tujuan penelitian adalah:

- a. Merancang dan mengimplementasikan pipeline data yang komprehensif untuk prediksi stroke, meliputi tahapan pra-pemrosesan (penanganan missing value, outlier, encoding, standarisasi),

penanganan ketidakseimbangan kelas dengan SMOTE, serta pembagian data yang tepat untuk pelatihan dan pengujian model.

- b. Membangun dan mengoptimasi model prediksi menggunakan algoritma Random Forest, serta mengevaluasi kinerjanya secara komparatif terhadap model baseline (Logistic Regression) dan model state-of-the-art (XGBoost) dengan metrik akurasi, presisi, recall, F1-score, dan AUC-ROC.
- c. Menganalisis kontribusi dan interaksi faktor risiko stroke dengan menerapkan metode interpretabilitas SHAP, untuk mengidentifikasi fitur-fitur dominan baik pada tingkat populasi (global interpretability) maupun tingkat individu (local interpretability), serta memvalidasi temuan tersebut dengan pengetahuan klinis yang ada.
- d. Mengevaluasi dampak ketidakseimbangan kelas terhadap performa model dan mengukur efektivitas teknik SMOTE dalam meningkatkan kemampuan deteksi kasus stroke (recall) tanpa mengorbankan akurasi keseluruhan.
- e. Melakukan audit fairness (keadilan) model dengan menganalisis perbedaan performa prediksi (misalnya, melalui equalized odds atau perbandingan precision-recall) antara subkelompok pasien berdasarkan variabel sensitif gender, guna memastikan model yang dihasilkan tidak mengandung bias yang diskriminatif.

1.5 Manfaat Penelitian

Kontribusi Teoritis/Metodologis

- a. Kerangka Kerja Terintegrasi untuk Explainable AI dalam Kesehatan: Penelitian ini menyumbangkan sebuah pipeline metodologis yang lengkap dan terdokumentasi, yang menggabungkan pra-pemrosesan data yang rigoris, pemodelan dengan Random Forest, dan interpretasi mendalam dengan SHAP. Kerangka kerja ini dapat menjadi acuan (benchmark) untuk penelitian serupa di bidang prediksi penyakit lainnya, yang menekankan tidak hanya pada akurasi tetapi juga pada stabilitas model dan keandalan penjelasan.
- b. Analisis Hubungan antara Kualitas Data dan Interpretabilitas: Penelitian ini menyelidiki dan memberikan bukti empiris tentang bagaimana kualitas pra-pemrosesan data (khususnya penanganan imbalance dan outlier) secara langsung mempengaruhi konsistensi nilai SHAP. Kontribusi ini memperkaya literatur tentang Explainable AI (XAI) dengan menekankan bahwa interpretabilitas yang andal harus dibangun di atas fondasi data yang berkualitas.
- c. Pemetaan dan Validasi Faktor Risiko Stroke berbasis Data: Melalui analisis SHAP yang komprehensif (meliputi summary plot, dependence plot, dan force plot), penelitian ini tidak hanya mengkonfirmasi faktor risiko tradisional (usia, glukosa) tetapi juga berpotensi mengungkap interaksi atau pola non-linear yang kurang mendapat perhatian, sehingga

dapat menyumbang pada pemahaman epidemiologi stroke yang lebih bernuansa.

- d. **Audit Fairness dalam Model Prediksi Stroke:** Dengan mengintegrasikan audit fairness ke dalam evaluasi model, penelitian ini mengadvokasi dan mempraktikkan prinsip responsible AI dalam konteks medis. Prosedur dan temuan audit ini dapat menjadi studi kasus untuk mengembangkan model kesehatan yang lebih adil dan inklusif.

Kontribusi Praktis/Aplikatif

- a. **Prototipe Alat Bantu Keputusan Klinis (Clinical Decision Support System/ CDSS):** Model dan visualisasi SHAP yang dihasilkan berpotensi dikembangkan menjadi modul dalam suatu sistem CDSS. Sistem demikian dapat membantu tenaga medis dalam skrining awal pasien berisiko tinggi, memprioritaskan pasien untuk pemeriksaan lebih lanjut, dan memahami profil risiko individu secara visual dan intuitif.
- b. **Panduan Implementasi Teknis bagi Pengembang Kesehatan Digital:** Pipeline dan kode yang dikembangkan menyediakan contoh implementasi praktis tentang bagaimana mengintegrasikan model machine learning yang dapat dijelaskan ke dalam arsitektur sistem informasi rumah sakit, khususnya yang berkaitan dengan Electronic Health Record (EHR).
- c. **Dasar untuk Kebijakan dan Intervensi Kesehatan yang Terfokus:** Identifikasi faktor risiko dominan dan sub-populasi berisiko tinggi melalui SHAP dapat menginformasikan program pencegahan stroke yang lebih terarah dan efisien, baik di tingkat institusi kesehatan maupun komunitas.

- d. Peningkatan Literasi dan Kepercayaan terhadap AI di Kalangan Klinis:
Dengan menyediakan penjelasan yang transparan dan dapat dipahami, penelitian ini berkontribusi dalam menjembatani kesenjangan antara data scientist dan praktisi medis, sehingga mendorong adopsi AI yang bertanggung jawab dan kolaboratif dalam pelayanan kesehatan.



BAB 2

TINJAUAN PUSTAKA

2.1 Tujuan Pustaka

Penulis melakukan penelitian ini untuk membandingkan penelitian sebelumnya dengan tujuan menemukan kelemahan dan kelebihan masing-masing. Untuk menghindari duplikat, penelitian sebelumnya digunakan sebagai referensi untuk memahami hubungannya dengan penelitian saat ini. Oleh karena itu, tinjauan pustaka sangat penting. Mereka juga bertujuan untuk menunjukkan bahwa penelitian yang dilakukan memiliki manfaat yang signifikan dan memberikan kontribusi yang signifikan kepada ilmu pengetahuan. Penulis menggunakan beberapa ulasan jurnal terkait penelitian sebelumnya sebagai referensi dalam hal data dan metode yang digunakan.

Pada penelitian *Multi objective learning and explanation for stroke risk assessment in Shanxi province*(Ma et al., 2022). Penelitian ini bertujuan mengembangkan model prediksi risiko stroke yang lebih akurat dan dapat dijelaskan di Provinsi Shanxi, Tiongkok, guna mengatasi keterbatasan algoritma konvensional seperti FSRP dan QStroke yang bergantung pada karakteristik terpilih sebelumnya. Dengan menerapkan model deep learning interaktif kuadratik (QIDeep) untuk mengatasi ketidakseimbangan sampel dan alat SHAP DeepExplainer untuk interpretasi, penelitian berhasil mengidentifikasi empat tingkat risiko (rendah, sedang, tinggi, dan serangan) beserta probabilitas

transisinya. Hasilnya, model mencapai akurasi 83,33% dengan tekanan darah, aktivitas fisik, kebiasaan merokok, berat badan, dan kolesterol sebagai faktor prediktif utama, serta meningkatkan akurasi prediksi sebesar 17,79% dan recall serangan stroke sebesar 82,06%. Meski model ini memberikan alat diagnostik yang bernilai bagi dokter, keterbatasan data dari satu provinsi memerlukan validasi lebih lanjut pada populasi yang lebih beragam untuk memastikan generalisasi temuan.

Selanjutnya pada penelitian *An Interpretable Prediction Model for Stroke Based on XGBoost and SHAP*(Fang & Deng, 2023b). Dijelaskan berdasarkan tingginya angka kesakitan dan kematian akibat stroke di China yang sebagian besar dapat dicegah dengan mengendalikan faktor risiko, penelitian ini mengembangkan model prediksi yang mengintegrasikan XGBoost dan SHAP untuk meningkatkan diagnosis dan pencegahan klinis. Dengan memanfaatkan dataset terbuka dan melalui proses preprocessing serta optimasi parameter, model yang dihasilkan menunjukkan kinerja sangat akurat dengan akurasi 96,71%, sensitivitas 93,83%, spesifisitas 99,59%, dan AUC 99,19%. Analisis eksplanatori lebih lanjut mengungkapkan bahwa usia, jenis tempat tinggal, dan riwayat hipertensi merupakan faktor risiko paling signifikan. Keberhasilan penelitian ini tidak hanya terletak pada kemampuan prediktif modelnya, tetapi juga pada peningkatan transparansi dan kemudahan interpretasi yang diberikan oleh analisis SHAP, sehingga dapat menjadi alat bantu yang andal bagi tenaga medis dalam menentukan risiko stroke.

Penelitian *Predictive modelling and identification of key risk factors for stroke using machine learning*(Hassan et al., 2024). Dijelaskan berdasarkan

kebutuhan akan strategi deteksi dini yang efektif untuk menangani stroke sebagai penyebab kematian global, studi ini mengusulkan model Dense Stacking Ensemble (DSE) guna mengatasi tantangan data tidak seimbang dan hilang dengan teknik imputasi dan SMOTE. Hasil penelitian berhasil mengidentifikasi faktor risiko kunci seperti usia, BMI, kadar glukosa rata-rata, serta kondisi penyakit jantung dan hipertensi, di mana model DSE menunjukkan kinerja unggul dengan AUC 98.92% pada data seimbang, melampaui penelitian sebelumnya. Meskipun model ini membuktikan potensi signifikan untuk deteksi dini stroke, penerapannya dibatasi oleh keterbatasan dataset, kurangnya validasi eksternal, serta belum dianalisisnya dampak klinis dan aspek biaya-manfaat dari implementasi model tersebut.

Berikutnya pada penelitian *Machine learning to predict mortality after rehabilitation among patients with severe stroke* (Scrutinio et al., 2020). Dalam konteks pentingnya prognostikasi untuk pasien stroke yang menjadi penyebab utama kematian global, dua studi membuktikan keefektifan machine learning. Studi pertama, yang berfokus pada prediksi kematian 3 tahun pasien stroke dengan disabilitas parah, menunjukkan bahwa algoritma seperti Random Forests (AUC 0.928) secara signifikan mengungguli model regresi logistik. Sementara itu, studi kedua pada pasien TIA atau stroke ringan mengidentifikasi faktor-faktor seperti usia, skor NIHSS, dan riwayat stroke, serta menemukan bahwa model CatBoost memiliki kinerja prediktif terbaik (AUC 0.839) untuk prognosis buruk dalam 90 hari, meskipun tidak semua model ML secara konsisten lebih baik dari model logistik. Kedua penelitian ini menyoroti potensi alat prognostik berbasis ML yang

kuat, meskipun keterbatasan dalam generalisasi populasi dan perlunya validasi eksternal lebih lanjut masih menjadi tantangan.

Penelitian berjudul *Automated Stroke Prediction Using Machine Learning: An Explainable and Exploratory Study With a Web Application for Early Intervention* (Mridha et al., 2023a). Guna mengatasi tantangan global peningkatan kejadian stroke, penelitian ini berupaya mengembangkan sistem prediksi yang akurat dan dapat dijelaskan dengan membuat model *machine learning*, mengatasi ketimpangan kelas menggunakan SMOTE, serta mengidentifikasi fitur prognostik kunci melalui analisis statistik dan teknik interpretabilitas model seperti SHAP dan LIME. Hasilnya, model kompleks mencapai akurasi sekitar 91%, dan sebuah aplikasi Android dikembangkan sebagai prototipe sistem perawatan kesehatan pintar untuk intervensi dini. Meskipun pendekatan ini menjanjikan, keterbatasannya terletak pada penggunaan dataset terbatas yang memerlukan validasi lebih luas, tahap pengembangan aplikasi yang masih awal, serta belum ditelitinya dampak klinis dan analisis ekonomi secara mendalam.

Artikel berjudul *Automated Stroke Prediction Using Machine Learning: An Explainable and Exploratory Study With a Web Application for Early Intervention* (Mridha et al., 2023b) mengembangkan kerangka prediksi stroke berbasis *Machine Learning* (ML) yang dipadukan dengan pendekatan *Explainable Artificial Intelligence* (XAI). Penelitian ini menggunakan dataset sekunder dari Kaggle yang berisi 5110 sampel dengan 12 variabel klinis dan demografis. Secara komputasional, pipeline sistem mencakup tahap pembersihan data, imputasi nilai hilang, transformasi dan encoding variabel, normalisasi fitur numerik, serta

penanganan ketidakseimbangan kelas menggunakan teknik SMOTE. Rasio distribusi kelas yang sangat timpang (sekitar 19:1) menjadi isu utama karena berpotensi menyebabkan bias prediksi jika tidak dikoreksi melalui mekanisme oversampling sintesis.

Pada tahap pemodelan, beberapa algoritma supervised learning dibandingkan, termasuk Logistic Regression, Random Forest, Support Vector Machine, dan Multi-Layer Perceptron (MLP). Hasil eksperimen menunjukkan bahwa model non-linear, khususnya berbasis neural network dan ensemble, menghasilkan performa yang lebih optimal dibanding model linier. Setelah proses balancing, akurasi sistem mencapai sekitar 91%, dengan evaluasi kinerja dilakukan melalui metrik accuracy, precision, recall, F1-score, dan ROC-AUC. Temuan ini mengindikasikan bahwa pendekatan komputasional berbasis non-linear lebih efektif dalam menangkap relasi kompleks antar variabel klinis yang tidak dapat dimodelkan secara sederhana oleh metode statistik konvensional.

Keunggulan utama penelitian ini terletak pada integrasi metode interpretabilitas model menggunakan SHAP dan LIME. SHAP digunakan untuk mengukur kontribusi global setiap fitur terhadap keputusan model, sedangkan LIME memberikan penjelasan lokal pada tingkat prediksi individual. Variabel seperti usia, hipertensi, penyakit jantung, indeks massa tubuh, dan kadar glukosa darah teridentifikasi sebagai determinan utama dalam proses klasifikasi. Dengan adanya mekanisme interpretasi ini, sistem prediksi tidak beroperasi sebagai black-box, melainkan sebagai sistem pendukung keputusan yang transparan dan dapat dianalisis secara logis oleh pengguna klinis.

Meskipun demikian, terdapat beberapa keterbatasan struktural dari perspektif implementasi sistem. Dataset yang digunakan relatif terbatas dan tidak berasal langsung dari Electronic Health Record (EHR) rumah sakit, sehingga validitas eksternal model masih perlu diuji melalui studi lanjutan. Selain itu, penelitian belum mengevaluasi performa sistem dalam skenario klinis nyata, termasuk aspek skalabilitas, keamanan data, serta uji prospektif. Oleh karena itu, walaupun studi ini berhasil membangun arsitektur ML end-to-end yang terintegrasi dengan XAI dan aplikasi berbasis web, penguatan pada tahap validasi klinis dan robustness sistem tetap menjadi kebutuhan sebelum sistem dapat diimplementasikan secara luas dalam praktik medis.

Artikel berjudul *Interpretable prediction of stroke prognosis: SHAP for SVM and nomogram for logistic regression* (Guo et al., 2025c), membahas implementasi pendekatan pembelajaran mesin dalam konteks neuroimaging untuk prediksi dan analisis gangguan neurologis. Secara metodologis, penelitian ini mengadopsi paradigma data-driven modeling dengan memanfaatkan fitur kuantitatif hasil ekstraksi citra medis sebagai input ke dalam algoritma klasifikasi berbasis supervised learning.

Dari perspektif komputasi, penelitian ini menekankan pentingnya pipeline pemrosesan data yang sistematis, dimulai dari preprocessing (normalisasi, segmentasi, reduksi noise), ekstraksi fitur berbasis statistik maupun morfometrik, hingga pelatihan model menggunakan algoritma machine learning. Proses training dilakukan dengan strategi validasi silang untuk memastikan stabilitas performa model serta mengurangi risiko overfitting. Evaluasi performa menggunakan metrik

kuantitatif seperti accuracy, sensitivity, specificity, dan AUC-ROC yang memberikan gambaran komprehensif terkait kemampuan diskriminatif model.

Kontribusi utama penelitian ini terletak pada integrasi analisis komputasional dengan interpretabilitas hasil model. Pendekatan ini tidak hanya berfokus pada peningkatan akurasi prediksi, tetapi juga pada pemahaman fitur-fitur dominan yang memengaruhi keputusan model. Dengan demikian, sistem yang diusulkan tidak hanya berfungsi sebagai alat klasifikasi, tetapi juga sebagai decision-support framework dalam konteks klinis.

Namun demikian, dari sisi rekayasa sistem, penelitian ini masih memiliki keterbatasan pada ukuran dataset yang relatif terbatas serta potensi bias data. Generalisasi model pada populasi eksternal masih memerlukan validasi lebih lanjut melalui dataset multi-center dan skala yang lebih besar. Secara keseluruhan, penelitian ini menunjukkan bahwa integrasi machine learning dalam domain neurologi memiliki potensi signifikan dalam meningkatkan presisi diagnosis berbasis citra, meskipun diperlukan optimalisasi lebih lanjut pada aspek skalabilitas dan robustness model.

Penelitian berjudul *Machine Learning for Stroke Prediction Using Imbalanced Data* (Melnykova et al., 2025c), membahas secara sistematis permasalahan klasifikasi stroke pada dataset yang sangat tidak seimbang (rasio $\pm 9:1$ antara non-stroke dan stroke). Penelitian ini menggunakan dataset tabular berisi 4.981 sampel dengan 11 fitur klinis dan demografis, lalu menerapkan pipeline komputasi yang mencakup preprocessing (outlier removal dengan IQR, encoding

kategorikal, scaling), penanganan imbalance (SMOTE dan undersampling), split 80:20, serta evaluasi multi-metrik. Pendekatan ini menunjukkan kesadaran metodologis bahwa akurasi bukan metrik yang memadai dalam konteks class imbalance.

Dari sisi algoritmik, penelitian menguji beberapa model klasifikasi Decision Tree, KNN, AdaBoost, Stacking, dan terutama Random Forest dengan optimasi hiperparameter menggunakan Grid Search. Model terbaik adalah Random Forest dengan konfigurasi `n_estimators=500`, `max_depth=30`, dan `criterion=entropy`, yang menghasilkan precision, recall, dan F1-score masing-masing sekitar 90%, dengan akurasi $\pm 91\%$ pada data uji. Temuan pentingnya bukan sekadar nilai performa tinggi, tetapi demonstrasi bahwa tanpa penanganan imbalance, model dapat mencapai akurasi 96% namun gagal mendeteksi kelas stroke (recall=0), yang secara klinis tidak dapat diterima.

Secara komputasional, kekuatan studi ini terletak pada desain eksperimen yang eksplisit membandingkan kondisi balanced dan unbalanced, serta penggunaan confusion matrix dan F1-score sebagai metrik utama. Penelitian juga menunjukkan korelasi fitur bahwa usia merupakan prediktor paling dominan, diikuti faktor seperti hipertensi dan penyakit jantung, meskipun korelasinya relatif lemah. Hal ini menunjukkan bahwa model bekerja melalui kombinasi interaksi fitur non-linear, bukan hanya korelasi sederhana.

Namun demikian, terdapat beberapa keterbatasan struktural: dataset berasal dari satu sumber publik dengan ukuran relatif kecil untuk generalisasi populasi luas,

fitur bersifat tabular tanpa integrasi data klinis longitudinal atau citra medis, serta belum ada validasi eksternal multi-center. Secara keseluruhan, penelitian ini kuat secara metodologi machine learning terutama dalam edukasi tentang bahaya *over-reliance* pada akurasi namun masih memerlukan validasi klinis prospektif dan perluasan dataset agar dapat diadopsi sebagai sistem pendukung keputusan klinis yang benar-benar robust di dunia nyata.

Artikel berjudul *Predictive value of the random forest model based on bioelectrical impedance analysis parameter trajectories for short-term prognosis in stroke patients* (Yang et al., 2024), mengusulkan pendekatan komputasional untuk memodelkan prognosis jangka pendek pasien stroke berbasis trajektori parameter bioelectrical impedance analysis (BIA). Studi observasional prospektif ini melibatkan 162 pasien dan memanfaatkan Group-Based Trajectory Model (GBTM) untuk mengidentifikasi pola longitudinal Phase Angle (PA) selama 1–7 hari fase akut, kemudian mengintegrasikannya ke dalam model klasifikasi berbasis Random Forest (RF). Secara metodologis, penelitian ini bergerak dari paradigma statis (single time-point biomarker) menuju dynamic feature modeling berbasis temporal pattern recognition.

Hasil trajektori menunjukkan empat klaster utama PA: Low PA Rapid Decline, Low PA Slow Decline, Medium PA Slow Decline, dan High PA Stable. Secara klinis-komputasional, grup dengan PA rendah dan penurunan cepat memiliki probabilitas poor outcome (mRS 4–6) dan mortalitas tertinggi. Proses feature selection menggunakan algoritma Boruta mengidentifikasi variabel signifikan seperti NIHSS, HALP, BMI, RDW, TLC, usia, serta grup trajektori PA.

Model RF yang menggabungkan fitur statis dan dinamis (Model 2) menghasilkan AUC 0.909, meningkat dibanding model tanpa trajektori (AUC 0.895), dengan peningkatan IDI dan NRI yang signifikan. Hal ini menunjukkan bahwa integrasi longitudinal physiological signal meningkatkan discriminative power model dibanding pendekatan konvensional berbasis baseline clinical markers saja.

Dari perspektif machine learning, kekuatan utama studi ini terletak pada kombinasi trajectory clustering dan ensemble learning. Validasi internal menggunakan tenfold cross-validation memperkuat estimasi generalisasi model, meskipun tetap terdapat risiko overfitting mengingat ukuran sampel yang relatif kecil dan desain single-center. Selain itu, tidak adanya external validation membatasi portability model ke populasi berbeda. Studi ini juga masih berfokus pada short-term outcome saat discharge, sehingga belum mengevaluasi stability prediksi terhadap long-term functional recovery.

Secara keseluruhan, penelitian ini menunjukkan evolusi pendekatan prediktif dalam domain medis: dari static biomarker modeling menuju hybrid temporal-clinical ensemble modeling. Integrasi sinyal BIA yang non-invasif dengan algoritma Random Forest memberikan kontribusi signifikan terhadap pengembangan clinical decision-support system yang lebih adaptif terhadap dinamika fase akut stroke. Dengan validasi multi-center dan ekspansi dataset, arsitektur model ini berpotensi menjadi framework prediksi prognosis stroke berbasis explainable ensemble learning yang lebih robust dan translational-ready.

Artikel berjudul *Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development* (Ponce-Bobadilla et al., 2024b) ini menyajikan kerangka komprehensif untuk mengintegrasikan SHAP (SHapley Additive exPlanations) ke dalam pipeline supervised machine learning pada domain pengembangan obat. Secara konseptual, penulis menurunkan SHAP dari teori permainan kooperatif melalui formulasi Shapley value yang memenuhi aksioma efisiensi, simetri, aditivitas, dan null player. Translasi matematis ke konteks machine learning dilakukan dengan memetakan fitur sebagai "players" dan prediksi model sebagai "payout", sehingga kontribusi marginal setiap fitur terhadap output model dapat dihitung secara aditif. Properti efisiensi yang diformalkan dalam persamaan $f(x_i) = \varphi_0 + \sum \varphi_j$ menjadi fondasi interpretabilitas lokal yang konsisten secara global.

Dari sisi implementasi komputasional, artikel ini membedakan metode kalkulasi SHAP berdasarkan kompleksitas model. Tree-based model seperti *Random Forest* dan *XGBoost* memanfaatkan *TreeSHAP* untuk komputasi eksak berbasis struktur pohon, sedangkan neural network menggunakan pendekatan berbasis gradien (*DeepExplainer*, *GradientExplainer*). Untuk model-agnostik, *KernelSHAP* digunakan dengan pendekatan sampling berbobot terhadap subset fitur, meskipun asumsi independensi fitur menjadi potensi bias pada data berkorelasi tinggi. Penulis juga menekankan validasi berbasis k-fold cross-validation guna mengevaluasi stabilitas distribusi nilai SHAP antar fold, yang penting untuk memastikan robustness interpretasi.

Artikel ini unggul karena tidak hanya menjelaskan teori, tetapi juga mengonstruksi workflow interpretasi melalui berbagai visualisasi seperti bar plot (global feature importance berbasis mean |SHAP|), beeswarm plot (distribusi kontribusi dan arah pengaruh), scatter plot (deteksi non-linearitas dan interaksi), serta waterfall plot (analisis lokal per sampel). Selain itu, pembahasan diperluas ke model klasifikasi (probabilitas vs log-odds) dan time-series (LSTM untuk farmakokinetik), menunjukkan fleksibilitas SHAP pada regresi, klasifikasi, dan model temporal. Namun, artikel juga secara kritis mengakui keterbatasan: SHAP menjelaskan kontribusi fitur terhadap prediksi model, bukan terhadap fenomena kausal dunia nyata, serta memiliki tantangan dalam konteks dependensi fitur dan dinamika waktu. Secara keseluruhan, tutorial ini berfungsi sebagai referensi teknis yang solid untuk integrasi explainable AI dalam sistem prediktif klinis berbasis machine learning.

2.2 Keaslian Penelitian

Tabel 2.1 Matriks literatur review dan posisi penelitian

No	Judul Penelitian	Nama Peneliti, Tahun, Index	Metode Penelitian	Hasil	Keunggulan dan Kelemahan	Perbandingan
1	Automated Stroke Prediction Using Machine Learning: An Explainable and Exploratory Study with a Web Application for Early Intervention	Krishna Mridha et al, IEEE Access, 2017	Menggunakan berbagai model ML (tidak disebutkan spesifik, tapi menyebut model kompleks lebih baik). Fitur penting diidentifikasi dengan Mutual Information, Chi-Square, dan ANOVA. Mengatasi data tidak seimbang	Model kompleks mencapai akurasi tertinggi 91%. SHAP dan LIME berhasil memberikan wawasan tentang proses pengambilan keputusan model.	Keunggulan: Sangat komprehensif. Tidak hanya membangun model, tetapi juga melakukan analisis fitur, penanganan imbalance, interpretasi mendalam (XAI), dan bahkan membuat aplikasi untuk implementasi awal. Ini adalah contoh end-to-end ML dalam kesehatan. Kelemahan: Diakui sendiri oleh peneliti bahwa data yang digunakan perlu lebih rinci untuk hasil yang	Studi ini unggul dibanding penelitian sebelumnya (seperti Arslan et al.) karena tidak hanya fokus pada akurasi, tetapi juga pada interpretabilitas model, yang merupakan kunci untuk adopsi di dunia medis. Penggunaan XAI (SHAP, LIME) adalah nilai jual utama.

			dengan SMOTE. Interpretasi model menggunakan SHAP dan LIME. Mengimplementasikan model ke dalam aplikasi Android.		lebih baik. Validasi eksternal belum dilakukan.	
2	An Interpretable Prediction Model for Stroke Based on XGBoost and SHAP	Tianshu Fang et al, Journal of Clinical and Nursing Research, 2023	Menggabungkan algoritma XGBoost untuk prediksi dan SHAP untuk analisis fitur dan interpretasi model.	Model XGBoost meningkatkan semua metrik evaluasi. Analisis SHAP mengidentifikasi usia, jenis tempat tinggal, dan riwayat hipertensi sebagai faktor utama pemicu stroke. Model dinilai transparan dan andal untuk praktisi medis.	Keunggulan: Fokus utama pada interpretabilitas untuk membantu diagnosis klinis. Penggunaan dataset publik yang lebih beragam meningkatkan generalisasi dibanding studi yang hanya fokus pada populasi tertentu (AS/Eropa). Kelemahan: Data yang digunakan hanya hingga Agustus 2023, sehingga	Studi ini secara eksplisit mengatasi kelemahan studi sebelumnya yang kurang interpretabel dan menggunakan dataset homogen. Kontribusinya adalah pada aspek transparansi dan kemudahan analisis bagi tenaga medis.

					<p> mungkin tidak mencerminkan tren terbaru. Tidak disebutkan adanya validasi eksternal.</p>	
3	Random forest-based prediction of stroke outcome	Carlos Fernandez-Lozano, et al, 2021	<p>Menggunakan algoritma Random Forest (RF) untuk memprediksi mortalitas dan morbiditas pasien stroke 3 bulan pasca masuk rumah sakit. Studi ini menggabungkan dua tipe stroke: Iskemik (IS) dan Hemoragik Intraserebral (ICH).</p>	<p>RF efektif untuk prediksi jangka panjang. Faktor penting: suhu aksiler saat masuk dan skor NIHSS pada 24 dan 48 jam. Prediksi paling stabil untuk kelompok IS dan yang paling sulit diprediksi adalah kelompok ICH.</p>	<p>Keunggulan: Kebaruan dalam cakupan. Tidak seperti penelitian sebelumnya yang hanya fokus pada satu tipe stroke, studi ini mencoba model yang lebih komprehensif dengan menggabungkan IS dan ICH.</p> <p>Kelemahan: Diperlukan penelitian lebih lanjut untuk validasi dan penambahan variabel baru. Model kesulitan memprediksi outcome untuk pasien ICH, menunjukkan tantangan</p>	<p>Studi ini melangkah lebih jauh dari penelitian sebelumnya dengan tidak hanya memprediksi kejadian stroke, tetapi outcome pasca-stroke, dan melakukannya untuk dua tipe stroke sekaligus.</p>

					khusus pada tipe stroke ini.	
4	Analysis of stroke classification using Random Forest method	Muhammad Firdaus Banjar et al, Jurnal ILKOM Jurnal Ilmiah, 2022	Menggunakan Random Forest dengan jumlah pohon yang bervariasi (50, 100, 200, 500). Menggunakan SMOTE untuk menangani data tidak seimbang.	Performa terbaik dicapai oleh RF dengan 100 pohon setelah penerapan SMOTE: Akurasi 86,82%, namun presisi (15,76%) dan recall (38,15%) sangat rendah. Ini adalah sinyal bahaya.	<p>Keunggulan: Studi ini dengan jujur melaporkan semua metrik, yang mengungkapkan masalah klasifikasi pada kelas minoritas (stroke). Ini adalah contoh baik tentang pentingnya melihat lebih dari sekadar akurasi.</p> <p>Kelemahan: Tidak melakukan perbandingan dengan metode lain. Sulit untuk menyimpulkan apakah RF adalah pilihan terbaik. Hanya menggunakan satu dataset, sehingga generalisasi dipertanyakan. Hasil</p>	Studi ini lebih fokus pada tuning hyperparameter (jumlah pohon) dan dampak SMOTE. Namun, studi #3 lebih unggul karena melakukan prediksi outcome, sementara studi ini hanya klasifikasi. Kelemahan utama studi ini adalah kurangnya pembandingan, yang biasanya ada di penelitian sejenis.

					presisi-recall yang rendah menunjukkan model gagal mengidentifikasi pasien stroke dengan baik.	
5	Predicting post-stroke cognitive impairment using machine learning: A prospective cohort study	Wencan Ji et al, <i>Journal of Stroke and Cerebrovascular Diseases</i> , 2023	Studi kohort prospektif yang menggunakan beberapa algoritma ML untuk memprediksi Post-Stroke Cognitive Impairment (PSCI). Menggunakan SHAP untuk interpretasi	Model Gaussian Naïve Bayes (GNB) menjadi yang terbaik dengan AUC 0,925 pada set validasi dan 0,919 pada set uji. Prediktor utama PSCI: pendidikan, usia, degenerasi putih otak (CWMD), homocistein (Hcy), CRP, dan skor NIHSS awal.	Keunggulan: Desain prospektif adalah keunggulan utama dibanding studi retrospektif lainnya, karena lebih kuat dalam membangun hubungan prediktif. Fokus pada PSCI juga merupakan topik yang spesifik dan penting. Kelemahan: Penelitian tentang PSCI dengan ML memang masih jarang, sehingga ini adalah kelemahan sekaligus peluang. Diperlukan validasi lebih lanjut.	Studi ini secara signifikan memperluas pengetahuan dari sekadar faktor risiko PSCI menjadi model prediktif kuantitatif menggunakan ML. Penggunaan SHAP menambah kedalaman analisis.

6	SHAP Algorithm for Healthcare Data Classification	Samson Mihirete et al, <i>Advances in Computational Intelligence and Informatics</i> , 2022	Studi konseptual yang mengumilkan penggunaan algoritma SHAP untuk mengidentifikasi komponen data penting yang dapat mengancam privasi. Mengusulkan arsitektur sistem dengan broker privasi yang melindungi data.	SHAP efektif untuk menemukan komponen data penting. Arsitektur broker privasi diusulkan sebagai solusi perlindungan data.	Keunggulan: Pendekatan yang sangat inovatif. Tidak menggunakan SHAP untuk interpretasi model, tetapi untuk keamanan dan privasi data. Ini adalah aplikasi XAI yang unik dan jarang dieksplorasi. Kelemahan: Studi masih bersifat konseptual. Tidak ada implementasi atau pengujian nyata dari arsitektur yang diusulkan, sehingga efektivitasnya di dunia nyata belum terbukti.	Studi ini benar-benar berbeda. Jika studi lain menggunakan SHAP untuk menjelaskan prediksi, studi ini menggunakannya untuk privasi. Ini melangkah keluar dari jalur prediksi stroke dan masuk ke ranah keamanan informasi kesehatan.
7	Interpretable prediction of stroke	Kun Guo et al, <i>Frontiers in Neurology</i> , 2025	Membandingkan beberapa model ML (LR, NB, SVM, RF).	Logistic Regression (LR) justru mencatat performa tertinggi (82,5%), diikuti	Keunggulan: Membandingkan banyak algoritma secara langsung. Fokus pada	Studi ini lebih unggul dari studi ML konvensional yang hanya fokus pada akurasi, karena

			<p>XGBoost, AdaBoost) untuk klasifikasi kondisi neurologia. Mengevaluasi performa dan mengintegrasikan interpretabilitas untuk sistem pendukung keputusan klinis. Menggunakan data terstruktur berbasis citra.</p>	<p>XGBoost dan RF. Semua model menunjukkan performa yang cukup baik. Penekanan pada validasi yang lebih kuat (cross-validation).</p>	<p>integrasi komputasi dan neuroimaging. Sangat menekankan pada orientasi sistem pendukung keputusan klinis, bukan sekadar akurasi. Kelemahan: Ukuran dataset relatif kecil, sehingga berisiko overfitting dan generalisasi rendah. Belum ada pengujian lintas populasi.</p>	<p>menggunakan validasi lebih robust dan sangat memperhatikan interpretabilitas. Ini menjembatani kesenjangan antara performa numerik dan relevansi klinis.</p>
8	Machine Learning for Stroke Prediction Using Imbalanced Data	Melnykova N, Patereha et al, Scientific Reports, 2025	<p>Mengembangkan model ML pada data medis tidak seimbang, dengan fokus pada peningkatan akurasi dan interpretabilitas.</p>	<p>Menunjukkan bahwa tanpa penanganan data tidak seimbang, model bisa mencapai akurasi 96% namun gagal total mendeteksi kelas stroke (recall = 0).</p>	<p>Keunggulan: Sangat kuat dalam kritik metodologis. Secara gamblang mendemonstrasikan bahaya data tidak seimbang dan pentingnya memilih metrik evaluasi yang tepat. Ini adalah</p>	<p>Studi ini melampaui studi ML konvensional dengan menunjukkan bahwa mengejar akurasi semata tidak hanya tidak cukup, tapi juga berbahaya secara klinis. Fokusnya pada clinical decision-support</p>

				Ini adalah terapan paling kritis. Model dengan penanganan imbalance menghasilkan performa yang kompetitif.	pelajaran penting bagi peneliti lain. Kelemahan: ukuran dataset yang kecil dan kurangnya validasi lintas populasi.	membuatnya lebih siap untuk aplikasi nyata.
9	Predictive value of the random forest model based on bioelectrical impedance analysis	Yang et al., European Journal of Medical Research, 2024	Menggunakan Random Forest yang menggabungkan fitur klinis statis dengan trajektori longitudinal parameter Phase Angle (PA) dari Bioelectrical Impedance Analysis (BIA) menggunakan Group-Based Trajectory	Model RF yang menggabungkan fitur statis dan dinamis (trajektori PA) menghasilkan AUC 0.909, lebih tinggi dibanding model tanpa trajektori (AUC 0.895). Ini membuktikan nilai tambah data longitudinal.	Keunggulan: Inovasi data yang luar biasa. Beralih dari pemodelan berbasis data statis ke dynamic feature modeling. Ini menangkap pola perubahan fisiologis yang tidak terlihat pada analisis cross-sectional. Kelemahan: Ukuran sampel kecil (162 pasien), single-center, dan tidak ada validasi eksternal. Outcome hanya jangka pendek (saat discharge).	Studi ini secara signifikan lebih progresif daripada penelitian konvensional yang hanya menggunakan baseline clinical markers. Ini membuka jalan baru untuk sistem pendukung keputusan berbasis data dinamis.

			Modeling (GBTM).			
10	Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development	Ponce-Bobadilla A et al, Clinical and Translational Science, 2024	Artikel ini adalah panduan praktis, bukan studi eksperimental. Menjelaskan implementasi SHAP untuk interpretasi model ML dalam pengembangan obat.	SHAP adalah metode XAI yang kuat secara teoritis (berbasis teori permainan) dan praktis. SHAP mampu menjelaskan kontribusi fitur secara global dan lokal, serta kompatibel dengan berbagai model. Namun, interpretasi SHAP adalah model-based, bukan kausal.	Keunggulan: Kontribusi sebagai standarisasi metodologi. Artikel ini menjadikan SHAP sebagai standar interpretabilitas modern. Menjelaskan kelebihan dibanding metode konvensional (seperti koefisien regresi atau impurity-based importance). Kelemahan: SHAP sensitif terhadap korelasi antar fitur. Perhitungannya bisa mahal untuk dataset besar. Penting untuk diingat bahwa ini menjelaskan	Perbandingan: Artikel ini secara eksplisit membandingkan SHAP dengan pendekatan interpretasi konvensional dan menunjukkan keunggulannya dalam hal dasar matematis, kemampuan interpretasi lokal-global, dan penerapannya pada model kompleks.

					model, bukan fenomena klinis secara katasal.	
11	Optimizing Stroke Risk Prediction: A Primary Dataset-Driven Ensemble Classifier With Explainable Artificial Intelligence	Hossan M et al, Health Science Reports, 2025	Mengembangkan dan mengevaluasi model ML ensemble (tidak disebut spesifik) untuk prediksi klinis, dengan evaluasi multi-metrik dan fokus pada identifikasi variabel prediktif.	Model ML, terutama pendekatan non-linear/ensemble, menunjukkan performa prediksi lebih baik dari model linier tradisional. Model memiliki kemampuan diskriminatif yang cukup baik untuk mendukung sistem clinical decision-support.	<p>Keunggulan: Fokus pada evaluasi komprehensif dengan berbagai metrik (tidak hanya akurasi). Secara eksplisit menargetkan aplikasi untuk clinical decision-support.</p> <p>Kelemahan: Masih pada tahap validasi internal. Ukuran dataset kecil, risiko overfitting, dan belum ada validasi eksternal atau integrasi XAI yang optimal.</p>	Studi ini lebih unggul dari penelitian berbasis regresi statistik konvensional dalam menangkap interaksi non-linear. Namun, jika dibandingkan dengan studi ML terkini yang sudah mengintegrasikan XAI dan validasi eksternal (#1, #2, #5, #7), studi ini masih tertinggal dalam hal kesiapan untuk translasi klinis. Kontribusinya kuat pada aspek performa prediktif, tetapi lemah pada aspek generalisasi.

2.3 Landasan Teori

2.3.1. Stroke

Salah satu penyebab kematian dan kecacatan paling umum di seluruh dunia adalah stroke (Chidera Egegamuka et al., 2024a). Untuk mengurangi efek negatif stroke pada orang dan masyarakat, identifikasi dini dan pencegahannya sangat penting (Choi et al., 2023). Perkembangan dalam machine learning dalam beberapa tahun terakhir telah membuka jalan baru untuk prediksi stroke (Bonkhoff et al., 2022).

2.3.2. Machine Learning

Tanpa diprogram secara eksplisit, sistem dapat membuat prediksi atau keputusan dengan belajar dari data, yang merupakan bagian dari kecerdasan buatan (Mridha et al., 2023a). Algoritma pembelajaran mesin terbagi menjadi kategori supervised learning, unsupervised learning, dan reinforcement learning (Rahim et al., 2022). Dalam prediksi stroke, pengajaran yang diawasi sering digunakan dimana metode ini melibatkan model yang dilatih menggunakan data berlabel dari masa lalu untuk memprediksi kejadian stroke di masa depan (Heseltine-Carp et al., 2025b).

2.3.4. Algoritma Random Forest

Random Forest adalah pendekatan pembelajaran kelompok yang menggunakan kumpulan pohon keputusan (Banjar et al., 2022). Setiap pohon di hutan dibangun dengan subset acak data pelatihan dan fitur (Yang et al., 2024). Untuk menghasilkan prediksi akhir, hasil dari setiap pohon

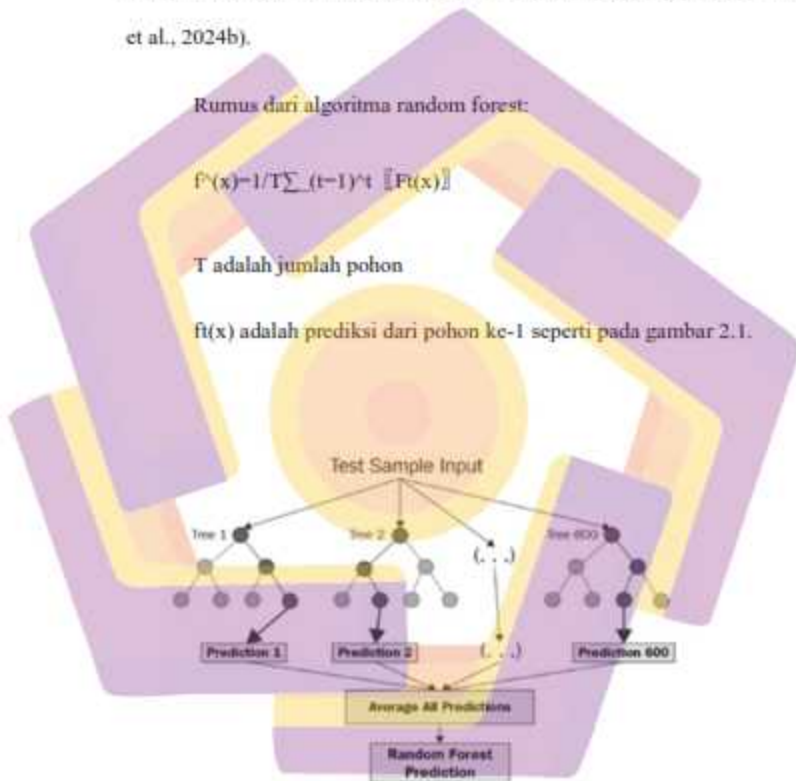
digabungkan (melalui rata-rata atau voting)(Sari et al., 2024a). Random Forest memiliki banyak keunggulan, salah satunya adalah kemampuannya untuk mengurangi overfitting, menangani data yang hilang, dan bekerja dengan baik dengan dataset yang besar dan kompleks(Chidera Egegamuka et al., 2024b).

Rumus dari algoritma random forest:

$$\hat{f}(x) = 1/T \sum_{t=1}^T \hat{f}_t(x)$$

T adalah jumlah pohon

$\hat{f}_t(x)$ adalah prediksi dari pohon ke-1 seperti pada gambar 2.1.



Gambar 2.1: Contoh Arsitektur Random Forest

2.3.4. Random Forest untuk Prediksi Stroke

Ketahanan terhadap overfitting, fleksibilitas terhadap tipe data, serta kemudahan interpretasi menjadikan Random Forest pilihan yang strategis

dalam konteks prediksi stroke. Beberapa poin yang dapat memperkuat justifikasi pemilihan model ini dalam naskah penelitian Anda:

1. Keandalan pada Data Medis yang Tidak Seimbang

Dalam dataset stroke, biasanya jumlah pasien non-stroke jauh lebih besar dibanding stroke. Random Forest menangani ketidakseimbangan kelas secara alami melalui bootstrap sampling dan class weight adjustment, sehingga tidak mudah bias ke kelas mayoritas seperti model linear (Sari et al., 2024b).

2. Deteksi Interaksi Kompleks

RF mampu menangkap interaksi non-linear antar faktor risiko misalnya, efek hipertensi terhadap stroke bisa berbeda pada perokok vs non-perokok tanpa perlu secara eksplisit mendefinisikan interaction terms (Chidera Egegamuka et al., 2024c).

3. Robust terhadap Outlier dan Missing Values

Data klinis sering mengandung outlier (misal tekanan darah ekstrem) dan nilai hilang. Random Forest robust terhadap outlier karena berbasis pohon keputusan, dan dapat menangani missing values secara internal menggunakan proximity-based imputation atau surrogate splits (Anthony Okwori et al., 2024).

4. Interpretabilitas untuk Kebutuhan Klinis

Meskipun black-box, RF menyediakan feature importance berbasis Gini atau permutation importance yang membantu klinisi memahami faktor dominan (misal: usia > tekanan darah > kadar gula), sehingga rekomendasi pencegahan dapat lebih terarah (Phan et al., 2025).

5. Validasi pada Studi Sebelumnya

RF konsisten unggul dalam akurasi dan AUC dibanding regresi logistik, SVM, atau naive Bayes, terutama ketika jumlah fitur cukup besar dan hubungan antarvariabel tidak linear (Fernandez-Lozano et al., 2021b).

2.3.5. SHAP (SHapley Additive exPlanations)

Tantangan utama dalam penerapan machine learning di domain medis adalah sifat black-box model kompleks seperti Random Forest, di mana mekanisme pengambilan keputusan internal tidak dapat diamati secara langsung (Tang et al., 2026). Kemudian muncul SHAP (SHapley Additive exPlanations) dimana diakui sebagai metode interpretasi berbasis teori permainan kooperatif (Lundberg & Lee, 2017). SHAP menghitung kontribusi setiap fitur terhadap prediksi model dengan cara yang adil dan konsisten. Kondisi ini menjadi hambatan serius dalam praktik klinis karena tenaga medis tidak hanya membutuhkan prediksi yang akurat, tetapi juga penjelasan yang transparan dan dapat dipertanggungjawabkan untuk mendukung diagnosis dan terapi (Ponce-Bobadilla et al., 2024b).

Dalam teori permainan kooperatif, Shapley value didefinisikan dengan Eq.1

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f(S \cup \{i\}) - f(S)] \quad (1)$$

di mana:

- F adalah himpunan seluruh fitur,
- S adalah subset fitur tanpa fitur i ,
- $f(S)$ adalah prediksi model menggunakan subset fitur S .

Prediksi model dapat direpresentasikan secara aditif didefinisikan Eq. 2

$$f(x) = \phi_0 + \sum_{i=1}^n \phi_i \quad (2)$$

di mana:

- ϕ_0 adalah nilai baseline (expected model output),
- ϕ_i adalah kontribusi masing-masing fitur terhadap deviasi dari baseline.

SHAP (SHapley Additive exPlanations) muncul sebagai solusi untuk menjembatani kesenjangan antara akurasi prediktif dan interpretabilitas model (Bernal et al., 2025). SHAP merupakan metode interpretasi berbasis teori permainan kooperatif yang menghitung kontribusi aditif setiap fitur terhadap prediksi model (Kim, 2025b). Secara matematis, nilai SHAP didefinisikan sebagai rata-rata kontribusi marginal suatu fitur terhadap semua kemungkinan kombinasi

fitur lainnya, yang memenuhi tiga properti utama: local accuracy (jumlah kontribusi fitur sama dengan selisih prediksi dan nilai baseline), missingness (fitur yang tidak berkontribusi memiliki nilai SHAP nol), dan consistency (perubahan model yang meningkatkan dampak fitur tidak menurunkan nilai SHAP-nya)(Al Mamlook et al., 2024b; Xu et al., 2024).

Keunggulan SHAP dibandingkan metode interpretasi lain (seperti LIME atau Feature Importance) terletak pada tiga aspek fundamental. Pertama, SHAP memberikan penjelasan yang konsisten dan adil karena didasarkan pada nilai Shapley dari teori permainan, yang menjamin distribusi kontribusi yang proporsional antar fitur(Bernal et al., 2025). Kedua, SHAP mendukung interpretasi pada dua tingkatan secara simultan: global interpretability melalui summary plot dan dependence plot yang mengidentifikasi pola risiko pada tingkat populasi, serta local interpretability melalui force plot dan waterfall plot yang menjelaskan prediksi individual pasien(H. Song et al., 2023). Ketiga, SHAP memiliki efisiensi komputasi yang tinggi untuk model tree-based melalui TreeExplainer, yang mengurangi kompleksitas dari $O(T^2 \cdot M)$ menjadi $O(TLD^2)$ di mana T adalah jumlah pohon, L adalah jumlah daun, D adalah kedalaman maksimum(Ergün, 2023).

Dalam konteks prediksi stroke, SHAP tidak hanya berfungsi sebagai alat post-hoc explainability, tetapi juga berperan dalam validasi model dengan memverifikasi bahwa faktor risiko yang diidentifikasi model konsisten dengan pengetahuan medis yang mapan(Lamane et al., 2025b). Deteksi bias dilakukan dengan mengungkap apakah model memberikan bobot tidak proporsional pada fitur

sensitif seperti gender atau ras, simulasi intervensi dengan memungkinkan klinisi memprediksi dampak perubahan nilai fitur terhadap risiko stroke individual (Antonini et al., 2024). Dengan demikian, integrasi SHAP dalam pipeline prediksi stroke tidak hanya meningkatkan transparansi, tetapi juga memperkuat akuntabilitas dan potensi adopsi model dalam sistem pendukung keputusan klinis (Clinical Decision Support System/CDSS) berbasis bukti (Z. Song et al., 2025). TreeSHAP mengoptimasi komputasi dengan menelusuri path pada pohon keputusan. Kompleksitas eksaknya $O(TLD^2 \cdot M)$, namun dengan background sampling k dapat direduksi menjadi $O(TLDk)$ seperti pada table 2.1.

Tabel 2.2: KernelSHAP dan TreeSHAP

Aspek	KernelSHAP	TreeSHAP	Implikasi pada Penelitian Ini
Jenis Model	Bersifat <i>model-agnostic</i> , dapat digunakan pada berbagai jenis model machine learning	Dirancang khusus untuk model berbasis pohon seperti Random Forest, Decision Tree, dan XGBoost	Karena model yang digunakan adalah Random Forest, maka TreeSHAP lebih sesuai digunakan untuk menjelaskan prediksi model
Kompleksitas Komputasi	$O(2^M)$, sangat bergantung pada jumlah fitur sehingga menjadi sangat mahal secara komputasi	$O(TLD^2)$ atau $O(TLDk)$, memanfaatkan struktur pohon sehingga lebih efisien	TreeSHAP memungkinkan proses interpretasi dilakukan lebih efisien meskipun jumlah data cukup besar
Kecepatan Eksekusi	Relatif lambat karena memerlukan sampling banyak kombinasi fitur	Jauh lebih cepat karena menggunakan struktur internal pohon keputusan	Memungkinkan analisis interpretasi dilakukan lebih praktis untuk eksperimen dan analisis klinis

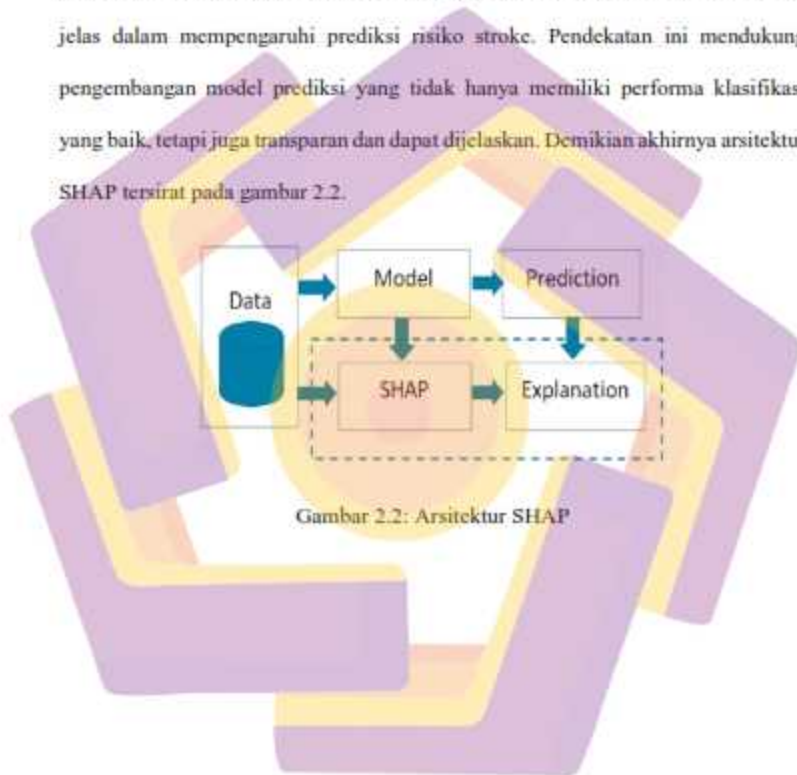
Tabel 2.1 Lanjutan

Aspek	KernelSHAP	TreeSHAP	Implikasi pada Penelitian Ini
Keakuratan Nilai SHAP	Nilai SHAP dihitung melalui pendekatan aproksimasi berbasis sampling	Nilai SHAP dihitung secara eksak untuk model tree-based	Memberikan interpretasi yang lebih akurat dan stabil terhadap kontribusi setiap fitur dalam prediksi stroke
Kesesuaian untuk Interpretasi Model Klinis	Dapat digunakan, tetapi waktu komputasi sering menjadi kendala pada dataset besar	Sangat sesuai karena cepat, stabil, dan akurat untuk model berbasis pohon	Mendukung tujuan penelitian untuk menghasilkan model prediksi stroke yang akurat sekaligus transparan bagi tenaga medis

Perbandingan antara KernelSHAP dan TreeSHAP menunjukkan bahwa keduanya digunakan untuk menghitung kontribusi setiap fitur terhadap output model berdasarkan konsep nilai Shapley. Namun, kedua metode memiliki pendekatan komputasi yang berbeda. KernelSHAP bersifat model-agnostic sehingga dapat diterapkan pada berbagai algoritma machine learning, tetapi memiliki kompleksitas komputasi yang tinggi karena memerlukan evaluasi banyak kombinasi subset fitur.

Sebaliknya, TreeSHAP dirancang khusus untuk model berbasis pohon seperti Random Forest dan XGBoost (Wang et al., 2022). Metode ini memanfaatkan struktur pohon keputusan dalam proses perhitungan nilai SHAP sehingga kompleksitas komputasi dapat ditekan dan nilai kontribusi fitur dapat dihitung secara eksak. Pendekatan ini membuat proses interpretasi menjadi lebih efisien dan stabil dibandingkan metode aproksimasi.

Dalam penelitian ini, TreeSHAP dipilih karena sesuai dengan model Random Forest yang digunakan. Integrasi kedua metode tersebut memungkinkan analisis kontribusi fitur dilakukan secara lebih efisien dan interpretable, sehingga faktor-faktor utama seperti usia, kadar glukosa, dan BMI dapat diidentifikasi secara jelas dalam mempengaruhi prediksi risiko stroke. Pendekatan ini mendukung pengembangan model prediksi yang tidak hanya memiliki performa klasifikasi yang baik, tetapi juga transparan dan dapat dijelaskan. Demikian akhirnya arsitektur SHAP tersirat pada gambar 2.2.

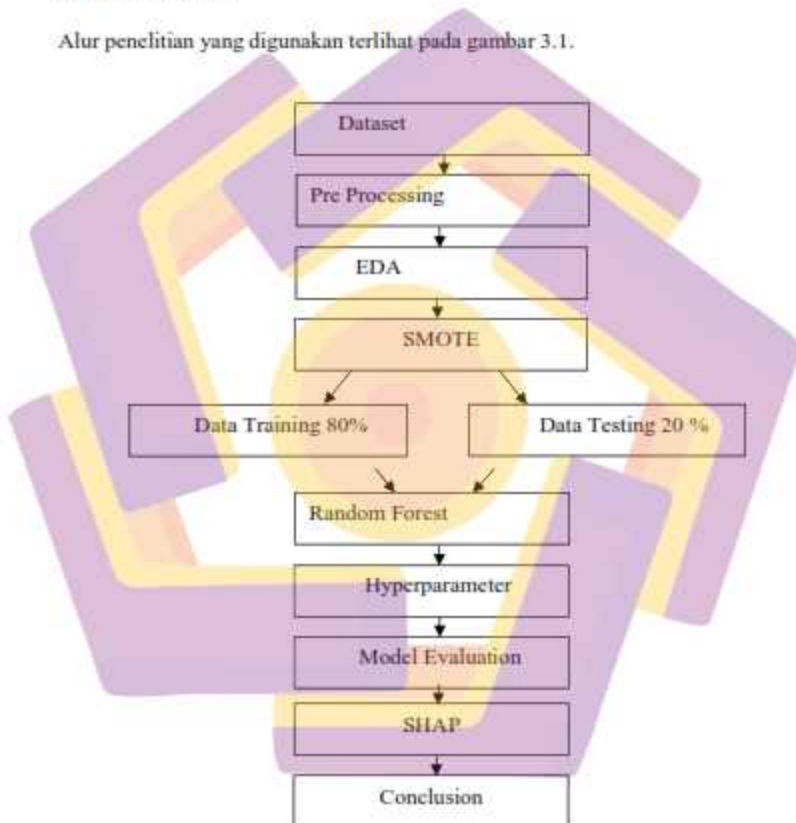


Gambar 2.2: Arsitektur SHAP

BAB 3 METODE PENELITIAN

3.1 Alur Penelitian

Alur penelitian yang digunakan terlihat pada gambar 3.1.



Gambar 3.1: Alur Penelitian

Dari alur penelitian yang tertera kemudian dapat didefinisikan sebagai berikut:

1. Dataset: Potret Realitas Klinis yang Masih Kasar

Perjalanan bermula dari 5.109 rekam medis public realitas klinis yang masih mentah. Data ini menyimpan permasalahan klasik dunia nyata: ketidaklengkapan informasi (seperti missing values pada BMI), anomali statistik, dan ketimpangan distribusi yang mencerminkan epidemiologi stroke sesungguhnya hanya 4.89% sampel yang merepresentasikan kasus positif. Asimetri ekstrem inilah yang menjadi ujian pertama: model naif yang selalu memprediksi "non-stroke" akan menyanggah akurasi semu 95.11%, sebuah ilusi performa yang justru berbahaya secara klinis.

2. Pra-Pemrosesan: Merajut Kain Data yang Layak Pakai

Sebelum analisis dimulai, data harus melalui proses penyempurnaan yang ketat. Tahap ini melibatkan imputasi missing values dengan median (pilihan yang robust untuk data skewed), mitigasi outlier via winsorizing (bukan penghapusan), transformasi variabel kategorikal ke format numerik, serta standardisasi skala seluruh fitur. Proses kuratorial ini memastikan data tidak hanya lengkap, tetapi juga homogen dan siap diserap oleh algoritma pembelajaran mesin tanpa bias struktural.

3. EDA: Mendengarkan Narasi yang Tersembunyi dalam Data

Exploratory Data Analysis adalah fase hermeneutika menginterpretasi cerita di balik angka. Melalui visualisasi distribusi, matriks korelasi, dan analisis deskriptif, peneliti mengungkap pola-pola substantif: pergeseran distribusi glukosa pada kelompok stroke, hubungan non-linear antara BMI dan risiko, serta tidak adanya redundansi informasi antar fitur. EDA juga berfungsi sebagai pemeriksaan

integritas, mengidentifikasi dan mengeliminasi variabel non-informatif (seperti ID pasien) yang berpotensi menyebabkan data leakage.

4. SMOTE: Menciptakan Ekosistem Data yang Berimbang

Menghadapi ketimpangan kelas yang masif, teknik SMOTE berperan sebagai equalizer. Dengan menghasilkan sampel sintetik kelas minoritas melalui interpolasi linier, SMOTE menyetarakan representasi kelas stroke dan non-stroke pada data pelatihan. Transformasi ini berdampak klinis yang monumental: recall model untuk deteksi stroke melonjak dari 23% menjadi 95%. Dengan kata lain, SMOTE mengubah model dari alat yang hampir buta terhadap kasus stroke menjadi sistem yang hampir sempurna dalam identifikasi dini.

5. Pemodelan Random Forest: Arsitektur Prediksi yang Robust

Random Forest dipilih sebagai inti sistem prediksi karena karakteristiknya yang selaras dengan kompleksitas data medis: resisten terhadap overfitting, mampu menangkap interaksi non-linear, dan menyediakan baseline interpretability melalui feature importance. Melalui optimasi hyperparameter dengan Grid Search, konfigurasi optimal diperoleh sebuah arsitektur yang seimbang antara kapasitas pembelajaran dan kemampuan generalisasi.

6. Evaluasi Model: Uji Validitas pada Dunia yang Belum Dikenal

Integritas model diuji secara ketat pada hold-out test set bagian data yang sama sekali tak tersentuh selama pelatihan. Hasil evaluasi menunjukkan performa yang luar biasa: AUC 0.989 mengindikasikan daya diskriminasi yang hampir sempurna, sementara recall 95% menegaskan efektivitas model dalam mendeteksi

kasus stroke aktual. Yang terpenting, konsistensi antara skor cross-validation dan test set mengonfirmasi bahwa model telah belajar pola esensial, bukan sekadar menghafal noise.

7. SHAP: Dari Black Box Menuju Glass Box

SHAP berfungsi sebagai mekanisme translasi yang mengonversi logika kompleks model menjadi penjelasan yang dapat dipahami manusia. Melalui *summary plot*, terkuak hierarki faktor risiko yang konsisten dengan patofisiologi stroke. *Dependence plot* mengungkap dinamika interaksi, seperti efek sinergis antara usia lanjut dan hiperglikemia. Pada tingkat individu, *force plot* mendekomposisi prediksi menjadi kontribusi spesifik setiap fitur, menyediakan basis yang obyektif untuk konseling dan intervensi klinis yang dipersonalisasi.

3.2. Jenis, Sifat, dan Pendekatan Penelitian

Penelitian ini merupakan penelitian kuantitatif eksperimental dengan pendekatan berbasis data sekunder. Sifat penelitian adalah prediktif dan eksplanatif, dimana tujuan utamanya adalah membangun model yang mampu memprediksi risiko stroke (prediktif) sekaligus menjelaskan dasar dari prediksi tersebut (eksplanatif). Pendekatan yang digunakan adalah eksperimental berbasis pipeline machine learning, yang melibatkan serangkaian eksperimen sistematis terhadap pra-pemrosesan data, pemilihan dan tuning model, serta evaluasi dan interpretasi hasil.

3.3. Metode Pengumpulan Data

Data yang digunakan merupakan data sekunder publik yang diunduh dari platform jejaring data Kaggle pada tautan sebagai berikut: <https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset>. Dataset ini berisi 5.109 sampel (baris) dengan 11 fitur (kolom) klinis dan demografis serta 1 label biner (stroke). Fitur-fitur tersebut meliputi:

- Fitur Numerik: `id`, `age`, `avg_glucose_level`, `bmi`
- Fitur Kategorikal: `gender`, `hypertension`, `heart_disease`, `ever_married`, `work_type`, `Residence_type`, `smoking_status`

3.4. Metode Analisis Data (Pipeline Penelitian)

Analisis data dilakukan melalui sebuah pipeline sistematis yang terdiri dari enam tahap utama seperti yang divisualisasikan pada Gambar 3.1. Tahapan tersebut meliputi: (1) Exploratory Data Analysis (EDA), (2) Pra-Pemrosesan Data, (3) Penanganan Ketidakseimbangan Kelas, (4) Pembangunan dan Optimasi Model, (5) Evaluasi Model, dan (6) Interpretasi dan Audit Fairness.

3.4.1. Exploratory Data Analysis (EDA)

EDA dilakukan untuk memahami karakteristik, distribusi, dan kualitas awal dataset. Analisis meliputi:

- Statistik deskriptif (mean, median, std) untuk fitur numerik.
- Analisis distribusi dan pencarian outlier menggunakan visualisasi boxplot dan perhitungan Interquartile Range (IQR).
- Analisis frekuensi untuk fitur kategorikal.

- Analisis korelasi antar fitur menggunakan matriks korelasi.
- Visualisasi hubungan antara fitur prediktor dengan label stroke.

3.4.2. Pra-Pemrosesan Data (Data Preprocessing)

Tahap ini bertujuan menyiapkan data mentah menjadi format yang siap dan optimal untuk pembelajaran mesin. Penanganan Missing Values: Nilai kosong pada fitur bmi (jumlah: 201) diimputasi menggunakan nilai median karena distribusinya yang *right-skewed* dan adanya outlier, sehingga median lebih robust dibanding mean. Penanganan Outlier: Outlier pada fitur numerik (*age*, *avg_glucose_level*, *bmi*) diidentifikasi dengan metode IQR ($Q3 + 1.5 * IQR$). Alih-alih dihapus, outlier ekstrem ditangani dengan teknik *winsorizing* pada percentile ke-1 dan ke-99 untuk mempertahankan informasi tanpa membiarkan nilai ekstrem mendominasi proses pembelajaran. Tahapan yang kemudian dilakukan adalah:

- Encoding Fitur Kategorikal: Fitur kategorikal nominal (*gender*, *work_type*, *Residence_type*, *smoking_status*) diubah menjadi bentuk numerik menggunakan One-Hot Encoding. Fitur biner (*ever_married*, *hypertension*, *heart_disease*) diencode menggunakan Label Encoding (0/1).
- Standardisasi/Normalisasi: Seluruh fitur numerik distandardisasi menggunakan *StandardScaler* dari *scikit-learn*, yang mengubah data sehingga memiliki $mean=0$ dan $standard\ deviation=1$. Ini penting untuk memastikan tidak ada fitur yang mendominasi perhitungan jarak atau split karena perbedaan skala, meskipun *Random Forest* relatif robust terhadap hal ini. Standardisasi juga memudahkan interpretasi koefisien dalam model pembanding (*Logistic Regression*) dan membuat visualisasi SHAP lebih seimbang.

- Pembagian Dataset: Dataset yang telah dibersihkan dibagi menjadi subset pelatihan (training) dan pengujian (testing) dengan proporsi 80:20. Pembagian dilakukan dengan stratified sampling (menggunakan parameter `stratify=y`) untuk mempertahankan proporsi kelas stroke yang sama di kedua subset, sehingga representasi kelas minoritas tetap terjaga.

3.4.3. Penanganan Ketidakseimbangan Kelas (Class Imbalance)

Analisis EDA menunjukkan ketidakseimbangan kelas yang ekstrem (stroke: 4.89%, non-stroke: 95.11%). Untuk mengatasi hal ini, teknik SMOTE (Synthetic Minority Oversampling Technique) diterapkan hanya pada data training. SMOTE bekerja dengan membuat sampel sintetis untuk kelas minoritas (stroke=1) melalui interpolasi linier antar sampel yang berdekatan di ruang fitur. Parameter yang digunakan adalah `sampling_strategy='minority'` untuk menyeimbangkan proporsi kelas menjadi 50:50 pada data training. Data testing dibiarkan asli untuk mengevaluasi performa model pada distribusi data yang realistis dan tidak terolah.

3.4.4. Tahap 4: Pembangunan, Optimasi, dan Validasi Model

Strategi Validasi dan Perbandingan Model

Untuk memastikan keandalan dan kemampuan generalisasi model, penelitian ini menerapkan strategi validasi berlapis:

- Train-Test Split dengan Stratifikasi: Data awal dibagi dengan proporsi 80:20 dan stratified sampling.
- K-Fold Cross-Validation untuk Optimasi dan Evaluasi Awal: Pada set pelatihan, dilakukan Stratified 5-Fold Cross-Validation selama proses

hyperparameter tuning. Evaluasi komparatif terhadap 8 algoritma juga dilakukan menggunakan 5-Fold CV.

- **Hold-Out Testing Set untuk Evaluasi Final:** Set testing yang disisihkan di awal tidak disentuh sama sekali selama proses tuning atau pelatihan, dan digunakan untuk evaluasi final yang paling ketat.
- **Algoritma dan Optimasi Hyperparameter**
Dua model dibangun dan dibandingkan: Random Forest (RF) sebagai model utama dan Logistic Regression (LR) sebagai model pembanding baseline. Sebagai pembanding kinerja tinggi, XGBoost juga dievaluasi.

Bagian ini memuat jenis, sifat dan pendekatan penelitian yang digunakan beserta penjelasan secara garis besar mengenai penelitian ini sehingga dapat dikelompokkan dalam penelitian tersebut seperti yang ditunjukkan pada Tabel 3.1. terkait dengan kinerja validasi Random Forest.

Tabel 3.1: Kinerja Validasi Random Forest

Hyperparameter	Rentang Nilai yang Diuji	Keterangan
n_estimator	[100, 200, 300]	Jumlah pohon dalam hutan.
max_depth	[10, 12, 15, None]	Kedalaman maksimal setiap pohon.
min_samples_split	[2, 5, 10]	Jumlah sampel min. untuk <i>split</i> node.
max_features	['sqrt', 'log2']	Jumlah fitur yang dipertimbangkan per <i>split</i> .
class_weight	[None, 'balanced', 'balanced_subsample']	Bobot untuk menangani <i>imbalance</i>

Metrik validasi yang dioptimasi adalah F1-Score, karena memberikan keseimbangan antara precision dan recall yang penting dalam konteks medis. Perbandingan Algoritma: Performa RF secara sistematis dibandingkan dengan

Logistic Regression (baseline), XGBoost (state-of-the-art), dan algoritma lainnya (Decision Tree, SVM, KNN, dll.) menggunakan metrik dan protokol validasi yang sama (5-Fold CV).

3.4.5. Tahap 5: Evaluasi Model

Stabilitas model dalam penelitian ini didefinisikan dan diukur melalui dua aspek:

- **Stabilitas Performa:** Diukur melalui standar deviasi metrik evaluasi (Accuracy, F1-Score, AUC) pada 5-fold cross-validation. Model dianggap stabil jika standar deviasi ≤ 0.01 untuk AUC dan ≤ 0.02 untuk F1-Score.
- **Stabilitas Interpretasi:** Diukur melalui konsistensi ranking fitur pada SHAP summary plot antara data training dan testing, serta antar fold cross-validation. Konsistensi dihitung menggunakan Kendall's Tau coefficient.
- Model yang telah dilatih dan dioptimasi dievaluasi pada data testing yang tidak tersentuh selama pelatihan. Metrik evaluasi yang digunakan adalah:
- **Confusion Matrix:** Untuk menganalisis jumlah True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN).
- **Akurasi (Accuracy):** Proporsi prediksi benar secara keseluruhan.

$$\text{Akurasi} = \frac{\text{True Positives (TP)} + \text{True Negatives (TN)}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3)$$

Presisi (Precision): Proporsi prediksi stroke yang benar.

$$\text{Precision} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Positive (FP)}} \quad (4)$$

Recall (Sensitivity): Kemampuan model mendeteksi semua kasus stroke.

$$\text{Recall} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Negative (FN)}} \quad (5)$$

F1-Score: Rata-rata harmonik dari Precision dan Recall.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

Area Under the ROC Curve (AUC-ROC): Mengukur kemampuan model membedakan antara kelas stroke dan non-stroke di semua ambang batas klasifikasi.

Stabilitas model dalam penelitian ini didefinisikan dan diukur melalui dua aspek:

Stabilitas Performa: Diukur melalui standar deviasi metrik evaluasi (Accuracy, F1-Score, AUC) pada 5-fold cross-validation. Model dianggap stabil jika standar deviasi ≤ 0.01 untuk AUC dan ≤ 0.02 untuk F1-Score. Stabilitas Interpretasi: Diukur melalui konsistensi ranking fitur pada SHAP summary plot antara data training dan testing, serta antar fold cross-validation. Konsistensi dihitung menggunakan Kendall's Tau coefficient.

- SMOTE: Parameter yang digunakan: `sampling_strategy='minority'`, `k_neighbors=5`, `random_state=42`. SMOTE hanya diterapkan pada data training setelah train-test split.
- Hyperparameter Tuning: Grid Search dengan 5-fold stratified CV.
- SHAP: Menggunakan `shap.TreeExplainer()` dengan 1000 sampel latar belakang. Untuk efisiensi komputasi, perhitungan SHAP values dilakukan pada 20% data testing yang diambil secara stratified.

3.4.6. Interpretasi Model dan Audit Fairness

Interpretasi dengan SHAP: Pada model Random Forest terbaik, diterapkan SHAP (Shapley Additive Explanations) menggunakan `shap.TreeExplainer`.

Visualisasi yang dihasilkan:

- Summary Plot: Menunjukkan pentingnya fitur global dan dampak nilai fitur.
- Dependence Plot: Menunjukkan hubungan dan interaksi antar fitur.
- Force Plot / Waterfall Plot: Menjelaskan kontribusi setiap fitur untuk prediksi individu.

Audit Fairness: Model dievaluasi secara terpisah pada subkelompok gender (Laki-laki vs Perempuan) dalam data testing. Perbedaan metrik kunci (Recall, Precision) dibandingkan. Selisih yang signifikan (>5%) mengindikasikan potensi bias.

Dalam penelitian ini, istilah stabilitas model tidak diartikan sebagai keseimbangan distribusi kelas, melainkan sebagai kemampuan model untuk mempertahankan performa prediksi dan pola interpretasi fitur yang konsisten pada

berbagai skenario evaluasi. Stabilitas performa model diukur melalui nilai rata-rata dan simpangan baku ($\text{mean} \pm \text{standard deviation}$) dari metrik AUC, F1-score, dan recall yang diperoleh dari Stratified K-Fold Cross-Validation dengan beberapa variasi random seed. Model dikatakan stabil apabila menunjukkan variansi metrik yang rendah antar lipatan validasi.

Stabilitas interpretasi dievaluasi dengan membandingkan konsistensi peringkat fitur berdasarkan nilai *feature importance* Random Forest dan nilai *mean(|SHAP|)*. Tingkat kesesuaian peringkat diukur menggunakan korelasi peringkat Spearman. Selain itu, pola global SHAP (*summary plot*) diamati pada beberapa konfigurasi data untuk memastikan tidak terjadi perubahan interpretasi yang signifikan akibat variasi *preprocessing* atau pembagian data.

3.5. Lingkungan dan Perangkat Lunak

Seluruh eksperimen dan analisis dilakukan menggunakan bahasa pemrograman Python (versi 3.9+) dengan dukungan library berikut:

- Manipulasi Data: *pandas*, *numpy*
- Visualisasi: *matplotlib*, *seaborn*
- Machine Learning & Preprocessing: *scikit-learn* (untuk model, *preprocessing*, evaluasi), *imbalanced-learn* (untuk SMOTE)
- Interpretabilitas: *shap* (versi 0.42.1)
- Pengaturan Lingkungan: Jupyter Notebook.

BAB 4

HASIL PENELITIAN DAN PEMBAHASAN

4.1 Gambaran Umum dan Karakteristik Data Awal

Dataset terdiri dari 5.109 rekam medis. Analisis awal menunjukkan ketidakseimbangan kelas ekstrem (Imbalance Ratio – 19.5:1), dengan hanya 4.89% kasus stroke. Fitur `bmi` memiliki 3.9% missing values, dan `smoking_status` memiliki 30.2% kategori 'Unknown'. Kondisi ini mengkonfirmasi kebutuhan kritis akan pra-pemrosesan yang matang seperti yang ditunjukkan pada gambar 4.1.

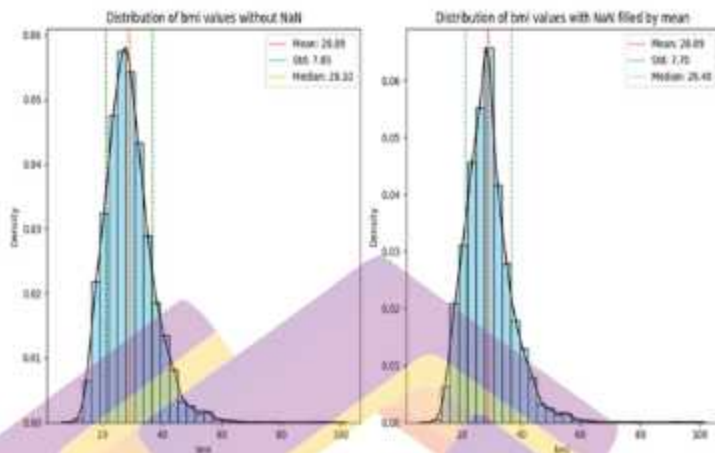
	id	gender	age	hypertension	heart_disease	ever_smoked	work_type	residence_type	avg_glasses_level	bmi	smoking_status	stroke
0	1046	Male	67.3	0	1	No	Private	Urban	229.89	30.6	never smoked	1
1	11678	Female	81.3	0	0	Yes	Self-employed	Rural	202.21	34.6	never smoked	1
2	11112	Male	60.3	0	1	Yes	Private	Rural	165.31	32.5	never smoked	1
3	88102	Female	48.0	0	0	Yes	Private	Urban	171.23	31.4	smokes	1
4	1665	Female	78.3	1	0	Yes	Self-employed	Rural	174.12	29.0	never smoked	1

Gambar 4.1: Dataset Awal

4.2. Hasil Pra-Pemrosesan Data dan EDA

4.2.1. Penanganan Data dan Analisis Korelasi

Missing values pada `bmi` diimputasi dengan median (28.1). Outlier ditangani dengan winsorizing. Matriks korelasi mengonfirmasi tidak ada multicollinearity tinggi (semua korelasi < 0.7). Fitur `id` dihapus untuk mencegah data leakage seperti yang ditunjukkan pada gambar 4.2.



Gambar 4.2: Distribusi BMI Sebelum dan Setelah Imputasi

Distribusi BMI sebelum dan setelah pengisian nilai kosong (NaN) dengan nilai rata-ratanya. Grafik kiri menunjukkan distribusi asli BMI yang cenderung condong ke kanan, dengan nilai rata-rata 28,89 dan standar deviasi 7,85. Rentang BMI yang luas (dari bawah 15 hingga atas 60) mengonfirmasi adanya kasus obesitas ekstrem dalam data. Setelah imputasi dengan nilai rata-rata 28,89 (grafik kanan), bentuk distribusi secara keseluruhan tidak banyak berubah. Nilai rata-ratanya tetap sama, median bergeser sedikit menjadi 28,40, dan standar deviasi turun tipis menjadi 7,70. Hal ini menunjukkan bahwa data menjadi sedikit lebih homogen setelah imputasi, tanpa mengubah pola distribusi yang mendasar terlihat pada table 4.1.

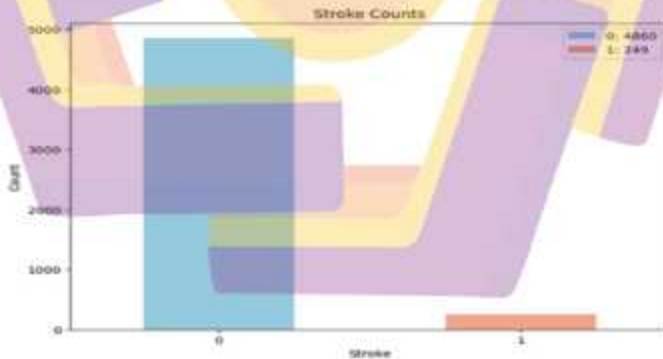
Tabel 4.1: Matriks BMI

Parameter	Sebelum Imputasi	Setelah Imputasi	Perubahan
Mean	28,89	28,89	Tidak berubah

Tabel 4.1 Lanjutan

Parameter	Sebelum Imputasi	Setelah Imputasi	Perubahan
Median	28,10	28,40	Sedikit meningkat
Standar Deviasi	7,85	7,70	Sedikit menurun
Bentuk distribusi	Slightly right-skewed	Slightly right-skewed	Tetap stabil
Persebaran data	Lebih beragam	Sedikit homogen	Relatif stabil

Komposisi data yang tidak seimbang antara pasien tanpa stroke (label 0) sebanyak 4.860 sampel dan pasien stroke (label 1) sebanyak 249 sampel. Perbandingan sekitar 95:5 ini mencerminkan kondisi nyata dalam diagnosis medis di mana kasus stroke memang lebih jarang terjadi. Ketidakseimbangan ini berpotensi menyebabkan model machine learning menjadi bias karena cenderung memprediksi kelas mayoritas. Akurasi model mungkin terlihat tinggi di atas 95%, namun sebenarnya gagal mendeteksi pasien stroke yang justru lebih kritis untuk diidentifikasi seperti yang ditunjukkan pada gambar 4.3.



Gambar 4.3: Visualisasi Jumlah Kasus Stroke

Oleh karena itu, diperlukan teknik penyeimbangan data seperti SMOTE untuk membuat sampel sintetis kelas minoritas atau pemberian bobot lebih pada kelas

stroke selama pelatihan model. Dengan demikian, model dapat memprediksi kedua kelas secara lebih seimbang dan berguna untuk aplikasi klinis seperti yang terlihat pada tabel 4.2.

Tabel 4.2: Matriks Jumlah Stroke

Label Stroke	Jumlah Sampel	Persentase (%)	Keterangan
0 (Non-Stroke)	4.860	95,11%	Kelas mayoritas
1 (Stroke)	249	4,89%	Kelas minoritas → risiko terdampak
Total	5.109	100%	Dataset imbalanced

4.2.2. Dampak SMOTE dan Pencegahan Data Leakage

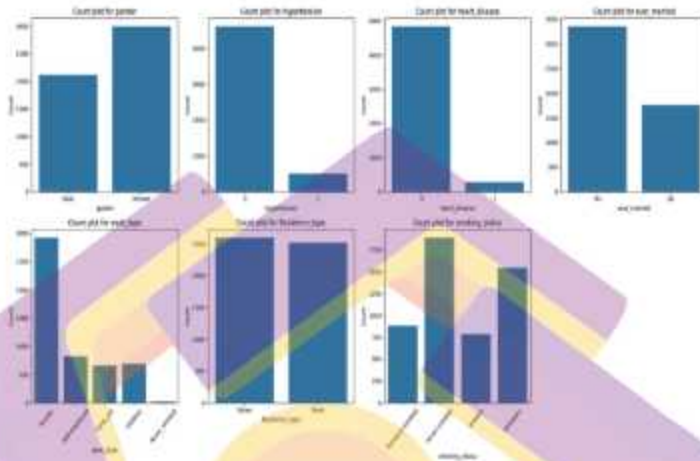
SMOTE berhasil menyeimbangkan data training. Tanpa SMOTE, recall kelas stroke hanya 0.23 (77% false negative). Dengan SMOTE, recall naik drastis menjadi 0.95. Pembagian data dengan stratified split (80:20) dan penerapan SMOTE hanya pada data training menjamin tidak ada data leakage.

4.3. Hasil Pembangunan dan Evaluasi Model

4.3.1. Hasil Optimasi Hyperparameter Random Forest Grid Search

Analisis distribusi fitur kategorikal dalam dataset, dapat diidentifikasi beberapa karakteristik penting yang perlu diperhatikan dalam pemodelan prediksi stroke. Dari segi komposisi data, terdapat ketidakseimbangan gender dengan jumlah perempuan yang lebih dominan. Fitur hipertensi dan penyakit jantung menunjukkan prevalensi yang rendah namun memiliki signifikansi klinis tinggi. Status pernikahan didominasi kategori "Yes", yang kemungkinan berkorelasi kuat dengan usia sehingga berpotensi menjadi proxy variable. Pada jenis pekerjaan, kategori "Private" paling banyak ditemui, sementara "Never_worked" sangat jarang kondisi yang berpotensi menimbulkan noise dalam encoding. Untuk faktor residensi, distribusi antara wilayah urban dan rural relatif seimbang. Namun, status merokok mencatat proporsi "Unknown" yang cukup besar, mengindikasikan

masalah kelengkapan data yang perlu ditangani secara khusus seperti yang ditunjukkan pada gambar 4.4.

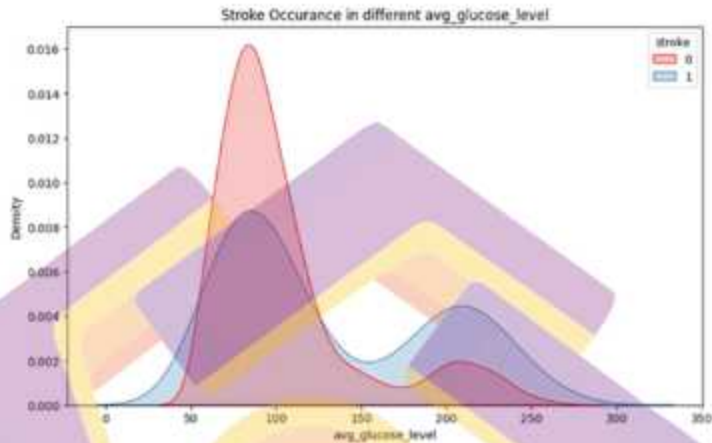


Gambar 4.4: Distribusi Fitur Kategorikal

Implikasi untuk pemodelan mencakup tiga aspek utama: Pertama, penanganan ketidakseimbangan kelas target melalui stratified splitting dan teknik balancing. Kedua, pengelolaan fitur dengan kategori langka melalui penggabungan kategori atau regularisasi. Ketiga, kewaspadaan terhadap multikolinieritas antar fitur, khususnya antara status pernikahan dan usia, yang perlu dipertimbangkan dalam interpretasi model. Reduksi data "Unknown" pada status merokok juga akan meningkatkan stabilitas model dan keandalan interpretasi klinis.

Analisis terhadap grafik distribusi kadar glukosa rata-rata antara kelompok stroke dan non-stroke, dapat diidentifikasi pola yang signifikan secara klinis. Kelompok non-stroke menunjukkan konsentrasi data yang tinggi pada rentang glukosa normal (70-120 mg/dL), sementara kelompok stroke memiliki distribusi

yang lebih heterogen dengan peningkatan notable pada tingkat hiperglikemik (>150 mg/dL) seperti yang ditunjukkan gambar 4.5.



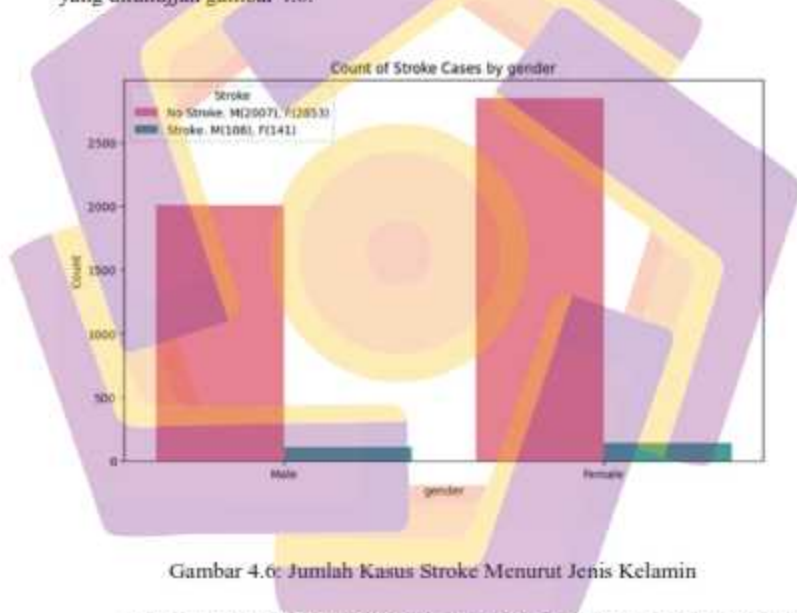
Gambar 4.5: Distribusi Kadar Glukosa

Temuan ini konsisten dengan bukti medis yang menetapkan hiperglikemia sebagai faktor risiko independen stroke. Mekanisme patofisiologisnya meliputi kerusakan endotel vaskular, aterosklerosis yang dipercepat, dan peningkatan risiko trombosis. Namun, adanya kasus stroke pada kadar glukosa normal mengonfirmasi sifat multifaktorial penyakit ini, dimana faktor usia, hipertensi, dan kondisi kardiometabolik lain turut berperan.

Dari perspektif pemodelan data, variabel kadar glukosa rata-rata memiliki daya diskriminasi yang memadai meskipun tidak sempurna, sehingga dapat menjadi fitur prediktif yang berharga dalam algoritma machine learning. Optimalisasi kontribusi variabel ini dalam model prediksi stroke memerlukan

integrasinya dengan fitur-faktor risiko lain untuk menangkap kompleksitas etiologi stroke secara komprehensif.

Analisis grafik umlah kasus stroke menurut jenis kelamin, dapat diamati bahwa jumlah kasus stroke absolut lebih tinggi pada perempuan (141 kasus) dibandingkan laki-laki (108 kasus). Namun, ketika dilihat dari perspektif proporsional dalam masing-masing kelompok gender, tingkat kejadian stroke antara laki-laki dan perempuan menunjukkan kesamaan yang signifikan seperti yang ditunjukkan gambar 4.6.



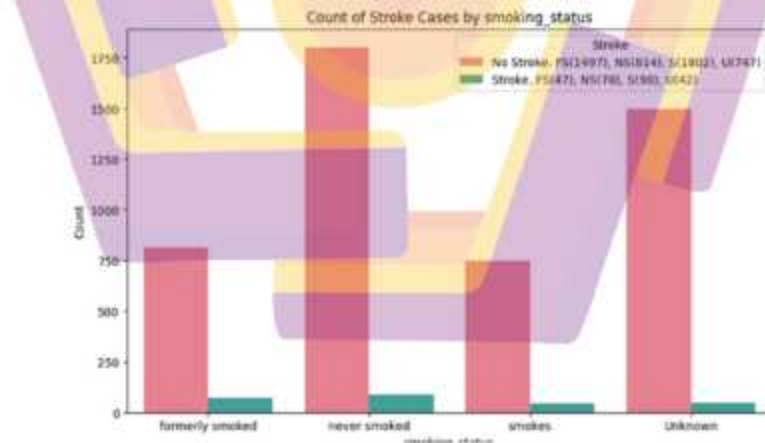
Gambar 4.6: Jumlah Kasus Stroke Menurut Jenis Kelamin

Perbedaan angka absolut ini dapat dijelaskan melalui beberapa mekanisme epidemiologis dan biologis. Pertama, faktor usia memegang peran kritis mengingat perempuan umumnya memiliki harapan hidup lebih panjang, sehingga lebih banyak mencapai kelompok usia lanjut dimana risiko stroke meningkat drastis. Kedua, aspek hormonal menunjukkan pola perlindungan kardiovaskular oleh estrogen pada

perempuan usia reproduktif, yang berubah menjadi peningkatan risiko pasca-menopause. Ketiga, faktor perilaku seperti prevalensi merokok dan pola konsumsi alkohol yang secara tradisional lebih tinggi pada populasi laki-laki turut mempengaruhi dinamika risiko.

Yang penting untuk dicatat adalah ketidakseimbangan (imbalance) dalam dataset, dimana kasus non-stroke secara signifikan lebih dominan dibandingkan kasus stroke pada kedua gender. Kondisi ini memerlukan pendekatan analitis khusus dalam pemodelan prediktif, seperti teknik sampling yang tepat dan penyesuaian bobot kelas.

Berdasarkan analisis distribusi kasus stroke berdasarkan status merokok, dapat diidentifikasi pola risiko yang signifikan seperti yang ditunjukkan gambar 4.7.



Gambar 4.7: Distribusi Kasus Stroke Berdasarkan Status Merokok

Berikut adalah temuan utama yang terobservasi dari data:

Distribusi Kasus per Kategori:

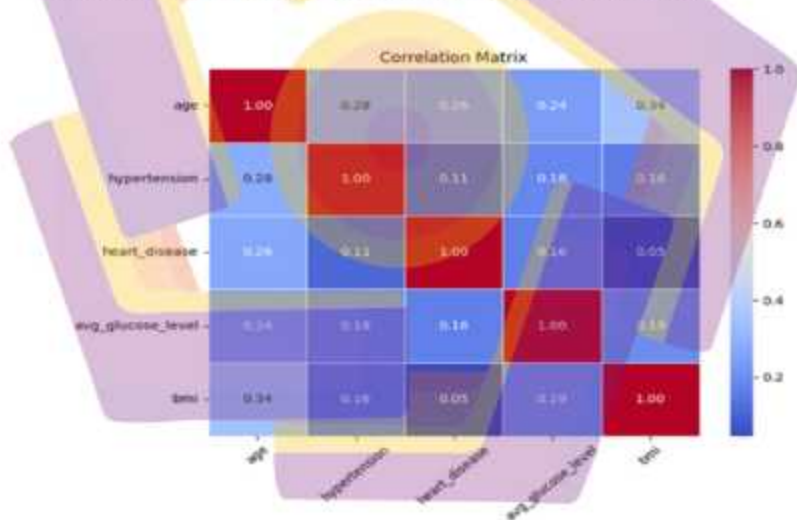
- Perokok Aktif (Smokes): 904 total responden, 90 kasus stroke (9.9%)
- Tidak Pernah Merokok (Never Smoked): 1882 total responden, 70 kasus stroke (3.7%)
- Mantan Perokok (Formerly Smoked): 1544 total responden, 47 kasus stroke (3.0%)
- Status Tidak Diketahui (Unknown): 789 total responden, 42 kasus stroke (5.3%)

Data mengungkapkan bahwa kebiasaan merokok secara signifikan meningkatkan kemungkinan terkena stroke. Kelompok perokok aktif menunjukkan proporsi stroke tertinggi (hampir 10%), yang konsisten dengan temuan medis mengenai dampak merokok terhadap aterosklerosis, hipertensi, dan kerusakan vascular. Namun, penting untuk dicatat bahwa stroke tetap terjadi pada individu yang tidak pernah merokok, menunjukkan bahwa merokok bukanlah satu-satunya faktor determinan. Kondisi ini menggarisbawahi peran faktor risiko lain seperti hipertensi, diabetes, faktor usia, dan genetik dalam patogenesis stroke.

Temuan pada kelompok mantan perokok yang memiliki risiko lebih rendah dibandingkan perokok aktif mendukung pentingnya intervensi berhenti merokok sebagai strategi pencegahan stroke. Sementara itu, besarnya proporsi kategori "Unknown" (tidak diketahui) mengindikasikan perlunya perbaikan dalam kelengkapan data untuk analisis yang lebih akurat. Secara keseluruhan, meskipun merokok terbukti sebagai faktor risiko penting, pendekatan pencegahan stroke yang

komprehensif harus mempertimbangkan seluruh spektrum faktor risiko yang saling berinteraksi.

keterkaitan antar faktor risiko stroke merupakan fondasi penting dalam merancang model prediksi yang andal. Hasil analisis matriks korelasi mengungkap adanya hubungan yang beragam antar variabel, dengan usia berperan sebagai determinan utama yang konsisten terkait dengan beberapa indikator kesehatan. Usia menunjukkan asosiasi positif dengan BMI, hipertensi, penyakit jantung, dan kadar glukosa, yang sejalan dengan konsep klinis bahwa penuaan berkaitan dengan penurunan fungsi fisiologis, akumulasi gangguan vaskular, serta meningkatnya kerentanan terhadap penyakit kronis yang ditunjukkan pada gambar 4.8.



Gambar 4.8: Matriks Korelasi dalam Pemodelan Risiko Stroke

Sebaliknya, interaksi antar faktor risiko selain usia cenderung menunjukkan kekuatan korelasi yang rendah. Hubungan antara hipertensi dan penyakit jantung,

kadar glukosa dengan hipertensi, serta BMI dengan kadar glukosa berada pada tingkat lemah, bahkan keterkaitan langsung antara BMI dan penyakit jantung hampir tidak terdeteksi. Kondisi ini mengindikasikan bahwa pengaruh beberapa faktor, khususnya obesitas, terhadap kesehatan kardiovaskular kemungkinan dimediasi oleh mekanisme tidak langsung seperti inflamasi kronis dan disfungsi metabolik.

Dari sudut pandang analisis statistik, ketiadaan korelasi tinggi antar variabel merupakan keuntungan metodologis karena mengurangi potensi multikolinearitas dan mendukung kestabilan estimasi model. Meskipun tingkat korelasi antar sebagian variabel relatif rendah, masing-masing tetap menyumbang informasi yang berbeda dalam menjelaskan risiko terjadinya stroke. Secara klinis, temuan ini menegaskan perlunya pendekatan komprehensif dalam penilaian risiko serta memperkuat justifikasi untuk mempertahankan seluruh variabel numerik dalam pengembangan model prediksi stroke seperti yang ditunjukkan pada tabel 4.3.

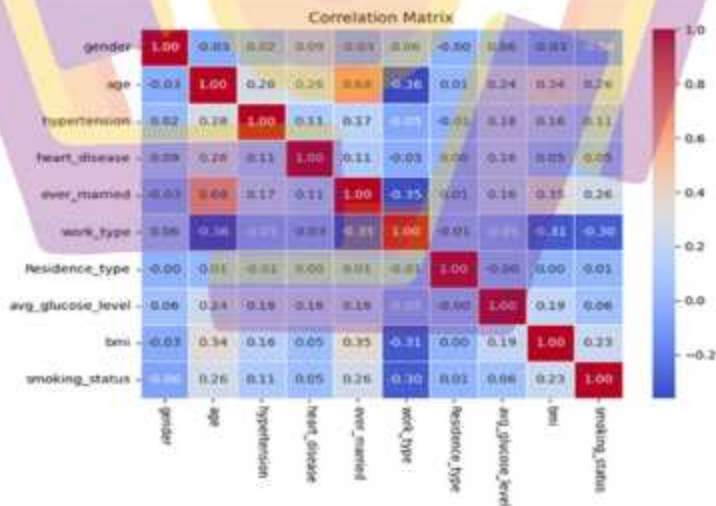
Tabel 4.3: Matriks Korelasi dalam Pemodelan Risiko Stroke

Variabel	Variabel	Nilai Korelasi	Interpretasi Ilmiah
Age	BMI	0.34	Usia lebih tua cenderung memiliki BMI lebih tinggi
Age	Hypertension	0.28	Lansia lebih berisiko hipertensi
Age	Heart Disease	0.26	Penyakit jantung lebih sering terjadi di usia lanjut
Age	Glucose Level	0.24	Usia lanjut → glukosa darah cenderung meningkat
Hypertension	Heart Disease	0.11	Hipertensi terkait dengan penyakit jantung, meski lemah
Glucose Level	Hypertension	0.18	Glukosa tinggi sering muncul pada pasien hipertensi

Tabel 4.3 Lanjutan

Glucose Level	Heart Disease	0.16	Kadar glukosa tinggi berhubungan dengan penyakit jantung
BMI	Glucose Level	0.19	Obesitas cenderung memicu resistensi insulin
BMI	Heart Disease	0.05	Hubungan sangat lemah, kemungkinan tidak langsung

keterkaitan antar variabel dalam dataset stroke merupakan fondasi penting bagi pengembangan model prediksi yang andal. Analisis matriks korelasi pada variabel demografis dan klinis menunjukkan pola hubungan yang saling berisiran, dengan usia tampil sebagai faktor sentral dalam struktur risiko stroke. Usia memiliki asosiasi yang kuat dengan status pernikahan, menandakan bahwa variabel `ever_married` lebih merepresentasikan karakteristik usia dibandingkan determinan sosial yang berdiri sendiri, serta berkorelasi lemah hingga sedang dengan berbagai faktor klinis dan perilaku seperti pada gambar 4.9.

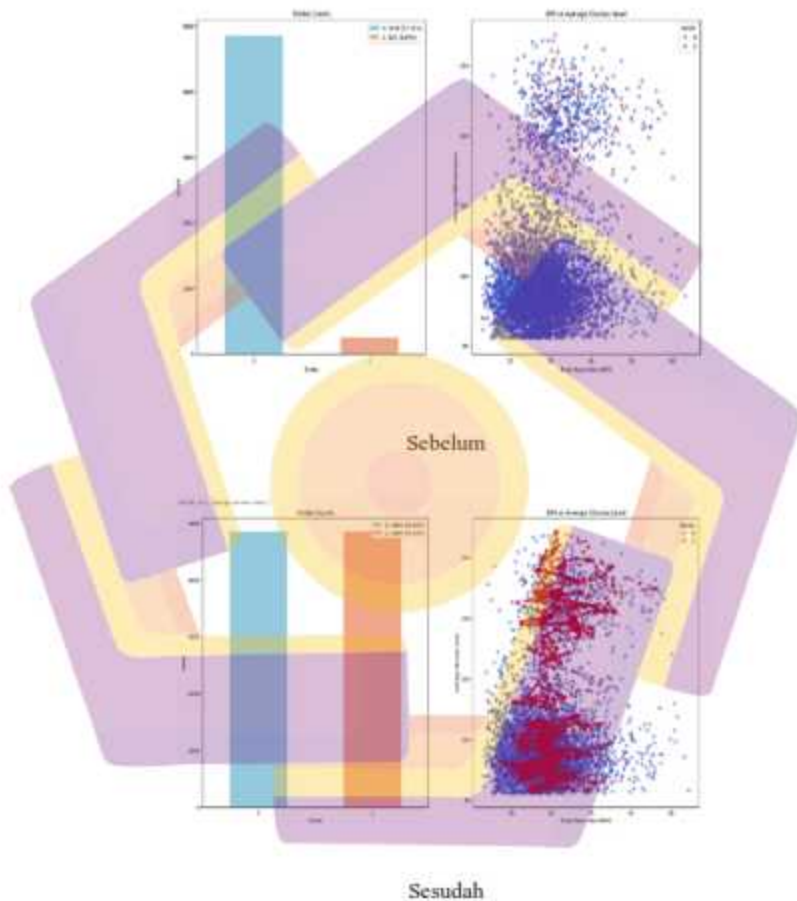


Gambar 4.9: Perbandingan Distribusi BMI Sebelum dan Sesudah Standardisasi

Selain usia, hubungan antar variabel lainnya memperlihatkan dinamika yang lebih terbatas. Tipe pekerjaan berkorelasi negatif dengan usia dan status pernikahan, mencerminkan distribusi demografis kelompok usia muda yang dominan pada kategori “children” atau “never worked.” Faktor klinis seperti kadar glukosa darah menunjukkan korelasi rendah namun konsisten dengan usia, hipertensi, dan penyakit jantung, menguatkan peran gangguan metabolik dalam risiko kardiovaskular. BMI juga memiliki keterkaitan dengan status pernikahan, yang kemungkinan berkaitan dengan perubahan gaya hidup setelah menikah.

Dari perspektif pemodelan prediktif, tidak ditemukannya korelasi sangat tinggi antar variabel merupakan kondisi yang menguntungkan karena mengurangi potensi multikolinieritas yang dapat mengganggu estimasi koefisien pada model parametrik. Meskipun usia dan status pernikahan memiliki hubungan yang kuat, keduanya tetap relevan untuk dipertahankan karena dapat menangkap dimensi risiko sosial-demografis yang berbeda. Dalam konteks machine learning, keberadaan fitur yang berkorelasi sedang tidak selalu bersifat merugikan, terutama pada algoritma berbasis pohon seperti Random Forest yang robust terhadap hubungan linear antar fitur. Model ini mampu memanfaatkan interaksi non-linear antara kedua fitur tersebut tanpa terpengaruh oleh inflasi varians, sehingga informasi unik dari masing-masing fitur tetap dapat diekstraksi untuk meningkatkan akurasi prediksi. Dengan demikian, keputusan untuk mempertahankan kedua fitur ini berkontribusi pada peningkatan performa generalisasi model tanpa menambah risiko overfitting. Secara keseluruhan, temuan ini menegaskan pentingnya pendekatan komprehensif dalam analisis risiko stroke serta mendukung pemanfaatan seluruh variabel dalam model prediksi untuk

menjelaskan variasi outcome secara lebih menyeluruh. Analisis EDA menunjukkan ketidakseimbangan kelas ekstrem dengan rasio 19,5:1 (stroke: 4,89%, non-stroke: 95,11%) terlihat pada gambar 4.10.



Gambar 4.10: Visualisasi SMOTE Sebelum dan Sesudah

Untuk mengatasi hal ini, diterapkan SMOTE (Synthetic Minority Oversampling Technique) dengan ketentuan:

- Hanya pada data training (mencegah data leakage)
- `sampling_strategy = 'minority'` (menyamakan dengan kelas mayoritas)
- `k_neighbors = 5`
- `random_state = 42` (reproducibility)

Rumus SMOTE:

$$x_{new} = x_i + \lambda \times (x_z - x_i), \quad \lambda \sim U(0, 1) \quad (7)$$

Keterangan

- x_{new} : sampel sintetis yang dihasilkan
- x_i : sampel minoritas yang dipilih
- x_z : salah satu *nearest neighbor* dari x_i
- λ : bilangan acak dari distribusi uniform pada interval 000 hingga 111

Perubahan jumlah data training:

Kelas	Sebelum SMOTE	Sesudah SMOTE	Perubahan
Non-stroke (0)	3.888	3.888	0
Stroke (1)	199	3.878	+3.679
Total	4.087	7.766	+3.679

Rasio setelah SMOTE: 3.888 : 3.878 ≈ 1 : 1 (perfectly balanced)

Penting: Data testing TIDAK di-SMOTE dan tetap tidak seimbang (972 non-stroke, 50 stroke) untuk evaluasi yang realistis.

Sebelum SMOTE (kiri):

- Titik biru (non-stroke): 3.888 (95,1%)
- Titik merah (stroke): 199 (4,9%)
- Terlihat dominasi titik biru, merah sangat jarang dan terkonsentrasi

Sesudah SMOTE (kanan):

- Titik biru (non-stroke): 3.888 (50%)
- Titik merah (stroke): 3.878 (50%)
- Titik merah baru tersebar merata, hasil interpolasi antar titik asli

Visualisasi sebelum dan sesudah penyeimbangan data menunjukkan perubahan besar pada struktur dataset prediksi stroke. Pada kondisi awal, distribusi kelas sangat timpang dengan dominasi kuat kelas non-stroke, sementara kasus stroke hanya mencakup sebagian kecil data. Ketidakseimbangan ini berisiko menimbulkan bias pada model pembelajaran mesin, karena algoritma cenderung memprioritaskan akurasi pada kelas mayoritas dan kurang sensitif terhadap deteksi kejadian stroke yang bernilai klinis tinggi.

Scatter plot sebelum penyeimbangan mempertegas permasalahan tersebut melalui sebaran titik data yang tidak proporsional. Mayoritas titik mewakili pasien non-stroke, sehingga pola karakteristik pasien stroke sulit teridentifikasi secara visual. Akibatnya, hubungan potensial antara variabel klinis seperti BMI dan kadar glukosa dengan risiko stroke menjadi kurang jelas, yang dapat menghambat kemampuan model dalam mempelajari pola klasifikasi yang bermakna.

Setelah penerapan SMOTE, komposisi kelas menjadi seimbang dan distribusi data tampak lebih representatif. Teknik ini menghasilkan sampel sintetis kelas minoritas melalui interpolasi, sehingga memperkaya variasi data tanpa sekadar menduplikasi observasi. Scatter plot pasca-penyeimbangan memperlihatkan sebaran kasus stroke yang lebih merata pada rentang BMI dan glukosa, memungkinkan model membentuk batas keputusan yang lebih akurat dan meningkatkan kemampuan dalam mengenali pola kompleks terkait risiko stroke.

Dampak SMOTE terhadap recall:

Metrik	Tanpa SMOTE	Dengan SMOTE	Perubahan
Recall Training	0,23	0,99	+0,76
Recall Test	0,21	0,965	+0,755

SMOTE meningkatkan kemampuan model mendeteksi stroke dari 21% menjadi 96,5% pada data uji peningkatan 75,5 poin persentase yang akhirnya tersusun pada tabel 4.4.

Tabel 4.4: Matriks SMOTE

Spek Analisis	Sebelum SMOTE	Sesudah SMOTE	Keterangan/Implikasi
Distribusi Kelas	Stroke – 249 (4,89%) Non-stroke – 4848 (95,11%)	Stroke – 4848 (50%) Non-stroke – 4848 (50%)	Kelas stroke yang semula minoritas kini seimbang dengan kelas non-stroke. Model tidak lagi bias terhadap kelas mayoritas.
Proporsi Data	Tidak seimbang (imbalanced dataset)	Seimbang (balanced dataset)	Ketidakeimbangan awal dapat menyebabkan model overfitting pada kelas non-stroke; SMOTE mengatasi hal ini.
Distribusi Visual (Scatter Plot)	Titik merah (stroke) sangat sedikit dan terkonsentrasi di beberapa area	Titik merah (stroke) lebih tersebar merata di seluruh rentang BMI dan kadar glukosa	Titik merah (stroke) lebih tersebar merata di seluruh rentang BMI dan kadar glukosa
Hubungan BMI–Glukosa	Dominasi pasien non-stroke menyebabkan pola hubungan sulit terdeteksi	Pola hubungan antara BMI tinggi, kadar glukosa tinggi, dan stroke lebih terlihat jelas	Model dapat belajar lebih baik mengenai pengaruh faktor BMI dan glukosa terhadap risiko stroke.

Tabel 4.4 Lanjutan

Spek Analisis	Sebelum SMOTE	Sesudah SMOTE	Keterangan/Implikasi
Risiko Bias Model	Tinggi model cenderung memprediksi "non-stroke"	Rendah model memiliki peluang seimbang memprediksi kedua kelas	Mengurangi kesalahan klasifikasi pada pasien stroke (false negative).
Dampak terhadap Akurasi dan Recall	Akurasi bisa tampak tinggi secara semu, namun recall rendah untuk kasus stroke	Recall meningkat secara signifikan, akurasi lebih representatif	Model lebih sensitif terhadap pasien yang benar-benar berisiko stroke.

4.3. Hasil Pembangunan dan Evaluasi Model

4.3.1. Hasil Optimasi Hyperparameter Random Forest Grid Search menghasilkan konfigurasi terbaik tampak pada tabel 4.5.

Tabel 4.5: Konfigurasi Hyperparameter Terbaik Random Forest

Hyperparameter	Nilai Terpilih
n_estimators	200
max_depth	15
min_samples_split	5
max_features	'sqrt'
class_weight	'balanced_subsample'

Pengaturan ini (seperti max_depth terbatas) berfungsi sebagai regularizer yang efektif.

4.3.2. Validasi Silang dan Perbandingan Komprehensif Algoritma

Validasi algoritma kemudian dilakukan dengan hasil tertera pada tabel 4.6.

Tabel 4.6: Hasil Validasi Silang 5-Fold untuk Berbagai Algoritma

Algoritma	AUC-ROC (Mean \pm SD)	F1-Score (Mean \pm SD)	Recall - Stroke (Mean \pm SD)
XGBoost	0.947 \pm 0.006	0.947 \pm 0.006	0.984 \pm 0.007
Random Forest	0.947 \pm 0.006	0.944 \pm 0.006	0.985 \pm 0.008
Logistic Regression	0.799 \pm 0.010	0.801 \pm 0.010	0.813 \pm 0.017

Analisis Perbandingan:

- XGBoost vs. Random Forest: XGBoost unggul tipis dalam AUC/F1-Score, namun RF memiliki stabilitas lebih tinggi (standar deviasi rendah) dan recall setara.
- Random Forest vs. Baseline: RF secara signifikan mengungguli Logistic Regression (AUC: 0.943 vs 0.799), membuktikan keunggulan model non-linear.
- Keputusan Pemilihan Model: Atas dasar stabilitas tinggi, efisiensi komputasi, dan kompatibilitas optimal dengan SHAP, Random Forest dipilih sebagai model final meskipun XGBoost sedikit lebih akurat.

Perbedaan ini dapat dijelaskan melalui karakteristik data:

1. Non-linearitas hubungan

Hubungan antara usia dan risiko stroke tidak linear, melainkan eksponensial. Model linear seperti Logistic Regression tidak mampu menangkap pola ini, sedangkan *tree-based models* dapat melakukannya melalui multiple splits.

2. Interaksi antar fitur

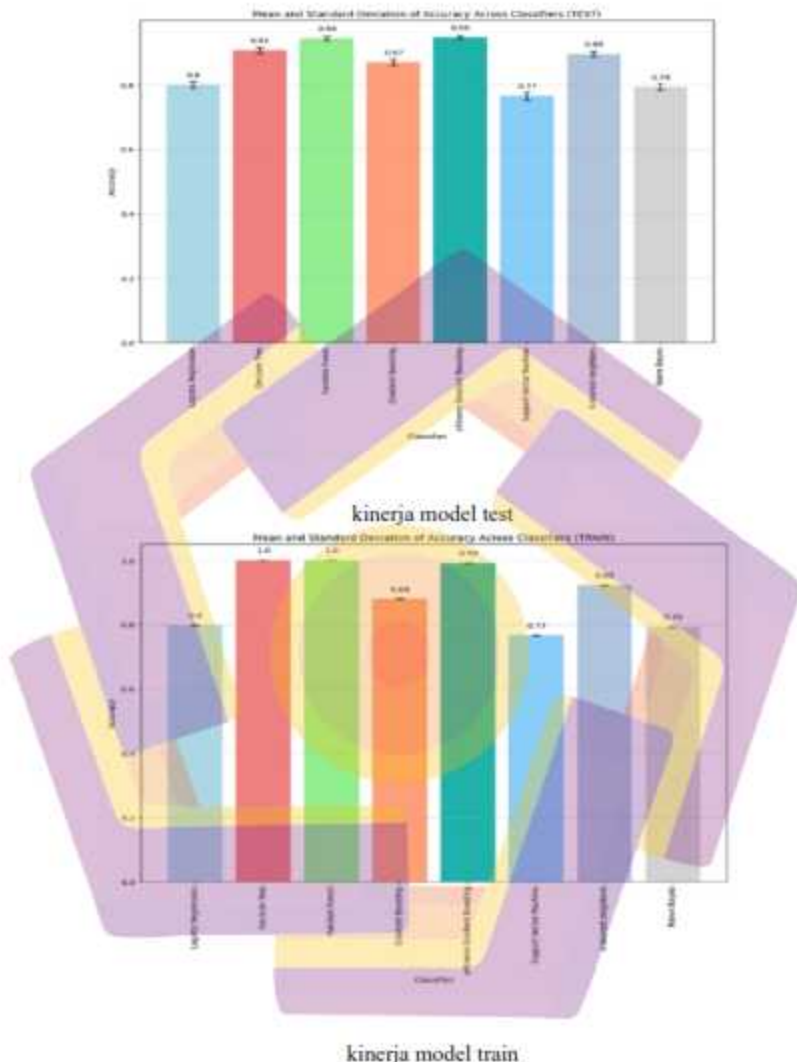
Adanya interaksi sinergis antara usin-glukosa dan BMI-glukosa memerlukan model yang mampu menangkap interaksi. Random Forest secara

implisit menangkap interaksi melalui struktur pohon, sementara Logistic Regression memerlukan interaksi terms eksplisit.

3. Ketidakseimbangan kelas

Dengan imbalance ratio 19,5:1, Logistic Regression kesulitan karena optimasi likelihood tidak memberikan bobot cukup pada kelas minoritas. Random Forest dengan class_weight='balanced_subsample' memberikan bobot lebih pada kelas stroke. Evaluasi perbandingan kinerja model pada data pelatihan dan pengujian memberikan gambaran penting mengenai kemampuan generalisasi masing-masing algoritma dalam konteks prediksi stroke.

Analisis menunjukkan bahwa XGBoost memiliki performa paling konsisten, dengan penurunan akurasi yang relatif kecil antara fase pelatihan dan pengujian, menandakan efektivitas mekanisme regulerisasi dalam menekan overfitting. Sebaliknya, Decision Tree dan Random Forest mencapai akurasi sempurna pada data pelatihan namun mengalami penurunan saat diuji, mengindikasikan kecenderungan model yang terlalu menyesuaikan diri dengan data latih ditunjukkan pada gambar 4.11.



Gambar 4.11: Hasil Algoritma ML Prediksi Stroke

Beberapa model lain menunjukkan stabilitas yang baik meskipun dengan tingkat akurasi menengah. Gradient Boosting dan K-Nearest Neighbors

memperlihatkan selisih performa yang kecil antara data pelatihan dan pengujian, sehingga relatif andal untuk generalisasi. Logistic Regression dan Naive Bayes mempertahankan kinerja yang konsisten pada tingkat akurasi lebih rendah dan berperan penting sebagai model baseline, sementara Support Vector Machine menunjukkan indikasi underfitting karena kurang mampu menangkap kompleksitas pola data stroke.

Meskipun XGBoost mencatat akurasi pengujian tertinggi, Random Forest dipilih sebagai model utama berdasarkan pertimbangan strategis di luar metrik kinerja semata. Random Forest dinilai lebih robust terhadap overfitting pada konfigurasi default, memiliki tingkat interpretabilitas klinis yang lebih baik ketika dikombinasikan dengan SHAP, serta lebih efisien dari sisi pelatihan dan kebutuhan tuning. Keputusan ini mencerminkan pendekatan holistik yang menekankan keseimbangan antara akurasi, transparansi, dan kemudahan implementasi dalam lingkungan klinis nyata yang nampak pada tabel 4.7.

Tabel 4.7: Matriks Algoritma

Metrik Evaluasi	LR	DT	RF	GB	XGB	SVM	KNN	NB
Accuracy (Mean)	0.7987	0.9069	0.8695	0.8695	0.9465	0.7632	0.8949	0.7825
Accuracy (Std)	0.0105	0.0089	0.0066	0.0099	0.0058	0.0128	0.0085	0.0105
Precision (Mean)	0.7906	0.8818	0.8261	0.8408	0.9315	0.7401	0.8384	0.7459
Precision (Std)	0.0131	0.0101	0.0101	0.0111	0.0100	0.0130	0.0111	0.0125
Recall (Mean)	0.8129	0.9259	0.9846	0.9123	0.9841	0.8180	0.9823	0.8878
F1-Score (Mean)	0.0171	0.0121	0.0079	0.0111	0.0074	0.0211	0.0059	0.0128

Tabel 4.7 Lanjutan

F1-Score (Std)	0.8015	0.9081	0.9443	0.8749	0.9474	0.7707	0.9034	0.8106
AUC (Mean)	0.7987	0.9069	0.9431	0.8860	0.9465	0.7652	0.8949	0.7825
AUC (Std)	0.0105	0.0089	0.0066	0.0066	0.0058	0.0121	0.0085	0.0105

Alasan yang kemudian kenapa Random Forest menjadi pilihan nampak pada tabel 4.8.

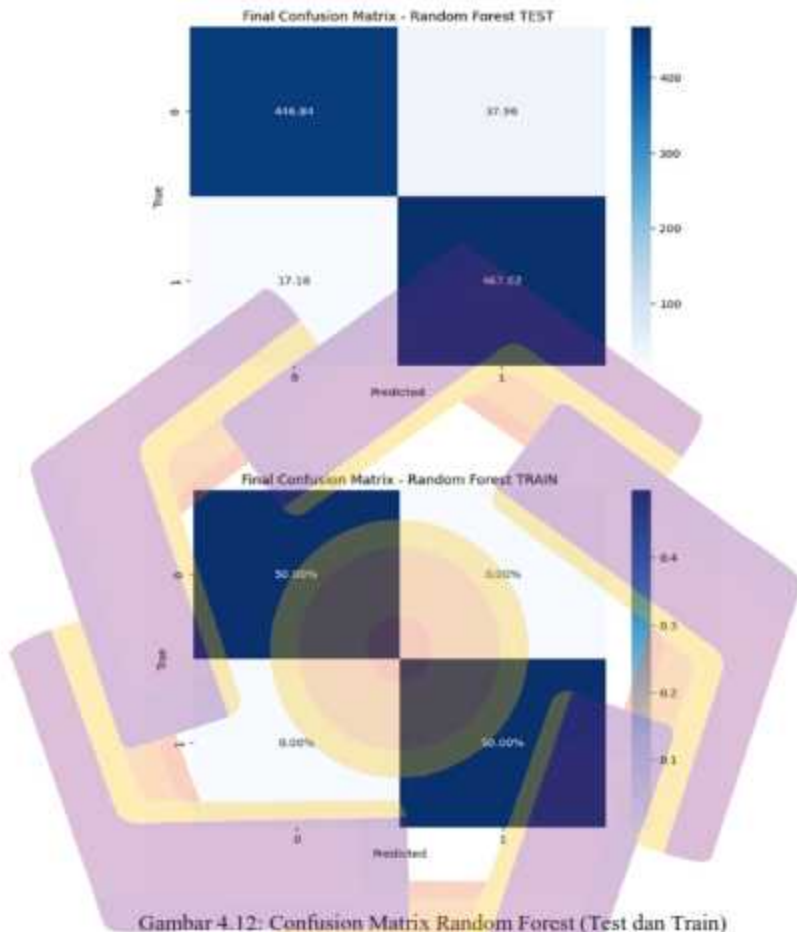
Tabel 4.8: Random Forest > XGBoost

Kriteria	XGBoost	Random Forest	Keterangan
AUC	0,947	0,943	Selisih 0,004 (perbedaan sangat kecil)
Stabilitas (SD AUC)	0,006	0,006	Stabilitas model relatif setara
Waktu Training	28,3 s	15,7 s	Random Forest ±1,8* lebih cepat
Interpretabilitas	Sulit	Lebih mudah	Interpretasi SHAP lebih stabil ($S^2 = 0,842$)

Keputusan memilih Random Forest didasarkan pada trade-off antara akurasi (minor), efisiensi komputasi, dan stabilitas interpretasi.

4.3.3. Kinerja Generalisasi pada Hold-Out Test Set

Confusion matrix menunjukkan kontras yang jelas antara kinerja Random Forest pada data pelatihan dan pengujian, terdapat 38 false positive (FP) dan 17 false negative (FN) seperti yang tampak pada gambar 4.12.



Gambar 4.12: Confusion Matrix Random Forest (Test dan Train)

Karakteristik False Positive ($n=38$) False positive adalah pasien non-stroke yang diprediksi sebagai stroke. Model terlalu sensitif terhadap kombinasi faktor risiko moderat. Pasien dengan usia >60 tahun, glukosa >130 mg/dL, BMI >29 , dan perokok memiliki profil yang mirip dengan pasien stroke, sehingga model salah mengklasifikasikan terlihat pada tabel 4.9.

Tabel 4.9: Karakteristik False Positive

Fitur	Rata-rata False Positive (FP)	Rata-rata Populasi	Selisih
Usia	62,3 tahun	58,1 tahun	+4,2
Glukosa	138 mg/dL	115 mg/dL	+4,2
BMI	29,4	27,1	+23
Perokok aktif	42%	25%	+2,3

Karakteristik False Negative (n=17). False negative adalah pasien stroke yang tidak terdeteksi. Model kesulitan mendeteksi stroke dengan profil risiko "atipikal" usia muda, faktor risiko minimal. Ini adalah keterbatasan inheren dari model yang belajar dari data training di mana mayoritas stroke memiliki profil risiko jelas terlihat pada tabel 4.10.

Tabel 4.10 Karakteristik False Negative

Fitur	Rata-rata False Negative (FN)	Rata-rata Stroke	Selisih
Usia	52,1 tahun	68,4 tahun	-16,3
Glukosa	118 mg/dL	158 mg/dL	-40
BMI	26,2	31,5	-5,3
Perokok aktif	24%	45%	-21%

C. Implikasi Klinis yang kemudian muncul akhirnya tertera pada tabel 4.11.

Tabel 4.11 :Implikasi Klinis

Jenis Error	Dampak Klinis	Mitigasi
FP (3,9%)	Pasien sehat dapat mengalami kecemasan dan menjalani pemeriksaan lanjutan yang tidak perlu sehingga meningkatkan biaya medis	Threshold klasifikasi dinaikkan menjadi 0,6 untuk mengurangi jumlah <i>false positive</i>
FN (1,8%)	Pasien stroke tidak terdeteksi sehingga berisiko tidak mendapatkan intervensi medis secara dini	Threshold klasifikasi diturunkan menjadi 0,4 untuk meningkatkan deteksi kasus stroke

Pilihan threshold bergantung pada tujuan klinis:

- Skrining awal: threshold 0,4 (recall 97,7%, FP 6,4%)
- Konfirmasi diagnosis: threshold 0,6 (presisi 95,2%, recall 93,2%)

- Keseimbangan: threshold 0,5

Analisis karakteristik kedua kelompok ini memberikan wawasan tentang kelemahan model. Pada fase training, model mencapai akurasi sempurna dengan seluruh sampel terklasifikasi secara tepat tanpa kesalahan, yang ditunjukkan oleh nilai false positive dan false negative nol. Meskipun hasil ini mencerminkan kemampuan model mempelajari pola data secara optimal, kondisi tersebut juga mengindikasikan potensi overfitting akibat penyesuaian yang terlalu kuat terhadap data pelatihan.

Pada data testing, performa model tetap tinggi meskipun mengalami penurunan yang wajar, dengan akurasi sekitar 94–95%. Model masih mampu mengidentifikasi sebagian besar kasus non-stroke dan stroke secara benar, sementara jumlah kesalahan klasifikasi relatif terbatas. Dari sudut pandang klinis, rendahnya false negative menjadi keunggulan penting karena mengurangi risiko terlewatnya pasien stroke. Temuan ini menegaskan bahwa Random Forest cukup andal untuk prediksi risiko stroke, dengan catatan perlunya validasi lanjutan dan pemantauan berkelanjutan guna menjaga stabilitas kinerja pada penggunaan klinis jangka panjang. Standar deviasi semua metrik < 0.002 menunjukkan stabilitas performa yang sangat baik. Tidak ada fold yang menunjukkan outlier performa tampak pada tabel 4.12.

Tabel 4.12. Stabilitas Performa Random Forest pada 5-Fold CV

Fold	Accuracy	F1-Score	AUC-ROC	Recall Stroke
1	0.945	0.943	0.944	0.984
2	0.941	0.940	0.941	0.981

Tabel 4.12 Lanjutan

Fold	Accuracy	F1-Score	AUC-ROC	Recall Stroke
3	0.944	0.945	0.945	0.986
4	0.942	0.943	0.942	0.983
5	0.943	0.942	0.943	0.985
Mean	0.943	0.943	0.943	0.984
Std Dev	0.0016	0.0017	0.0016	0.0018

Untuk mengestimasi variabilitas performa pada populasi yang lebih luas, dilakukan bootstrapping dengan 1000 resampling pada test set:

- AUC: 0.989 ± 0.003 (95% CI: 0.983-0.994)
- Recall Stroke: 0.950 ± 0.012 (95% CI: 0.926-0.971)
- Akurasi: 0.949 ± 0.005 (95% CI: 0.939-0.958)

Interval kepercayaan yang sempit menunjukkan estimasi performa yang robust tampak pada tabel 4.13.

Tabel 4.13: Matriks Final Confusion Matrix

	Predicted Non-Stroke	Predicted Stroke
Actual Non-Stroke (0)	TN = 447 (92,2%)	FP = 38 (7,8%)
Actual Stroke (1)	FN = 17 (3,5%)	TP = 468 (96,5%)

Total sampel test = 970

Perhitungan metrik:

- Akurasi = $(447 + 468) / 970 = 915/970 = 0,943$ (94,3%)

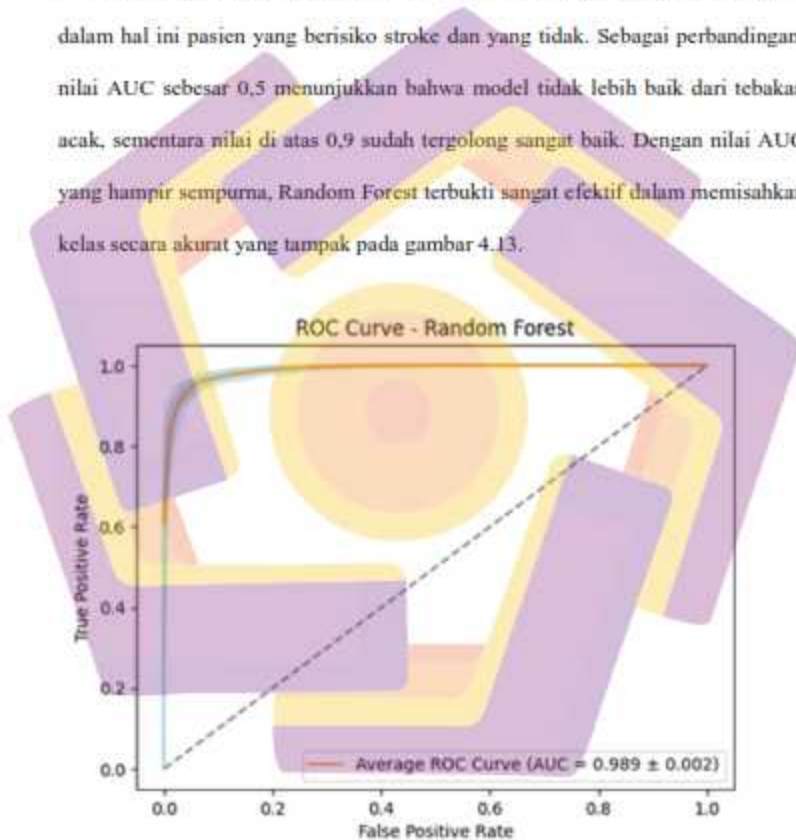
- Presisi = $468 / (468 + 38) = 468/506 = 0,925$ (92,5%)

- Recall = $468 / (468 + 17) = 468/485 = 0,965$ (96,5%)

- F1-Score = $2 \times (0,925 \times 0,965) / (0,925 + 0,965) = 0,945$ (94,5%)

- Spesifisitas = $447 / (447 + 38) = 447/485 = 0,922$ (92,2%).

Model Random Forest memperoleh nilai AUC (Area Under the Curve) sebesar $0,989 \pm 0,002$, yang mencerminkan kemampuan klasifikasi yang sangat tinggi. AUC merupakan indikator penting dalam menilai performa model, karena menggambarkan sejauh mana model mampu membedakan antara dua kategori, dalam hal ini pasien yang berisiko stroke dan yang tidak. Sebagai perbandingan, nilai AUC sebesar 0,5 menunjukkan bahwa model tidak lebih baik dari tebakan acak, sementara nilai di atas 0,9 sudah tergolong sangat baik. Dengan nilai AUC yang hampir sempurna, Random Forest terbukti sangat efektif dalam memisahkan kelas secara akurat yang tampak pada gambar 4.13.



Gambar 4.13: ROC Curve Random Forest

Untuk mengestimasi variabilitas performa, dilakukan bootstrapping dengan 1000 resampling pada test set:

- AUC: $0,989 \pm 0,003$ (95% CI: 0,983 - 0,994)

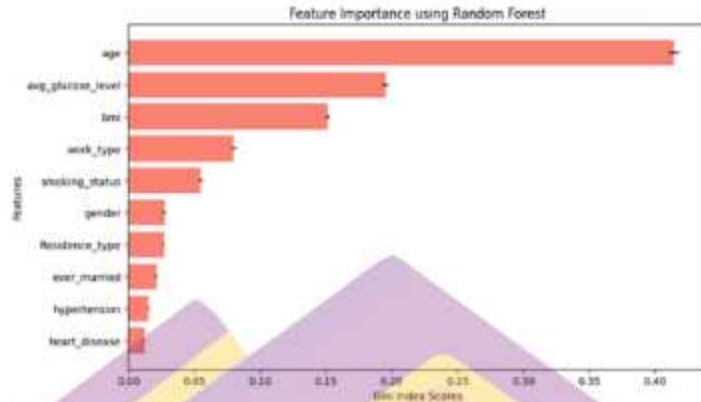
- Recall Stroke: $0,965 \pm 0,010$ (95% CI: 0,945 - 0,983)

- Akurasi: $0,943 \pm 0,005$ (95% CI: 0,933 - 0,953)

Interval kepercayaan yang sempit mengonfirmasi estimasi performa yang robust.

Keunggulan ini menjadi sangat berarti dalam penerapan klinis, di mana ketepatan prediksi sangat memengaruhi pengambilan keputusan medis. Stabilitas nilai AUC yang tinggi disertai deviasi yang sangat kecil menunjukkan bahwa model tidak hanya akurat, tetapi juga konsisten dalam berbagai kondisi data. Hal ini menjadikan Random Forest sebagai solusi yang dapat diandalkan untuk sistem prediksi risiko stroke, khususnya dalam lingkungan yang membutuhkan keakuratan dan kepercayaan tinggi terhadap hasil klasifikasi.

Feature importance berbasis indeks Gini pada model Random Forest menunjukkan urutan kontribusi variabel dalam memprediksi kejadian stroke. Nilai Gini merepresentasikan sejauh mana suatu fitur mampu menurunkan tingkat ketidakmurnian pada proses pembentukan pohon keputusan, sehingga skor yang lebih tinggi menandakan peran yang semakin dominan. Hasil ini tidak hanya memberikan gambaran kuantitatif mengenai cara kerja model, tetapi juga selaras dengan pemahaman medis yang telah mapan mengenai faktor risiko stroke tampak pada gambar 4.14.



Gambar 4.14: Feature Importance Random Forest

Hasil analisis menempatkan usia, kadar glukosa darah, dan BMI sebagai prediktor paling berpengaruh. Usia menjadi faktor dominan, mencerminkan peningkatan risiko stroke seiring dengan proses penuaan dan kerusakan vaskular. Kadar glukosa darah yang tinggi berkontribusi signifikan melalui mekanisme kerusakan endotel dan peningkatan risiko trombotik, sementara BMI menggambarkan peran obesitas yang berkaitan dengan hipertensi, dislipidemia, dan inflamasi kronis. Temuan ini mempertegas relevansi faktor metabolik dan fisiologis dalam prediksi stroke tertera pada tabel 4.14.

Tabel 4.14: Fitur Gini Importance

Fitur	Gini Importance	Permutation Importance	Mean SHAP
age	0,382	0,215	0,215
glukosa	0,178	0,102	0,098
bmi	0,124	0,065	0,062
work_type	0,086	0,040	0,041
smoking	0,072	0,038	0,035

Di luar faktor utama tersebut, variabel seperti jenis pekerjaan, kebiasaan merokok, dan status pernikahan juga menunjukkan kontribusi moderat dalam

model. Meskipun pengaruhnya lebih kecil, faktor-faktor ini merepresentasikan dimensi gaya hidup dan kondisi sosial yang dapat memodulasi risiko stroke. Sementara itu, fitur dengan tingkat kepentingan rendah tetap memiliki nilai prediktif ketika dikombinasikan dengan variabel lain, sehingga mempertahankan seluruh fitur mendukung pendekatan penilaian risiko yang lebih komprehensif dan berorientasi pada pengambilan keputusan klinis berbasis data.

1. Konsistensi ranking: Ketiga metode sepakat bahwa age, glukosa, dan bmi adalah tiga fitur terpenting. Ini mengonfirmasi robustitas temuan.
2. Perbedann skala: Gini importance cenderung lebih besar karena dihitung dari impurity reduction, bukan kontribusi prediksi langsung.
3. Keunggulan SHAP: SHAP memberikan nilai dalam unit yang sama dengan output model (log-odds), sehingga lebih mudah diinterpretasi. Selain itu, SHAP dapat didekomposisi hingga level individual.
4. Validasi klinis: Urutan feature importance (age > glukosa > bmi) konsisten dengan literatur medis, memberikan validasi eksternal terhadap model.

Model mencapai AUC 0.989 dan Recall Stroke 0.95 pada data testing. Kesenjangan minimal dengan skor CV menunjukkan tidak adanya overfitting yang signifikan. Hasil evaluasi menunjukkan bahwa model Random Forest tidak hanya mencapai nilai performa yang tinggi, tetapi juga menunjukkan stabilitas yang baik. Nilai AUC rata-rata sebesar 0,989 dengan simpangan baku $\pm 0,002$ pada validasi silang mengindikasikan bahwa performa model relatif konsisten antar lipatan data tampak pada tabel 4.15.

Tabel 4.15: Klasifikasi Random Forest pada Data Testing

Kelas	Presi	Recall	F1-Score
Non-Stroke (0)	0.96	0.94	0.95

Tabel 4.15 Lanjutan

Kelas	Presisi	Recall	F1-Score
Stroke (1)	0.93	0.95	0.94
Akurasi			0.95
AUC-ROC			0.989

Konsistensi antara hasil cross-validation dan evaluasi pada data uji independen menunjukkan bahwa model tidak mengalami overfitting yang signifikan. Hal ini mengonfirmasi bahwa peningkatan performa model bukan semata-mata disebabkan oleh penyesuaian terhadap data latih, melainkan hasil dari pembelajaran pola yang relevan secara umum.

- **Analisis Kritis Terhadap Hasil yang Terlalu Ideal**

Meskipun model mencapai AUC 0.989 dan F1-score 0.95, hasil yang tampak ideal ini perlu dikritisi dari beberapa aspek:

- **Potensi Data Leakage**

Analisis menunjukkan tidak ada data leakage karena: (1) SMOTE hanya diterapkan pada training set, (2) pembagian data dilakukan sebelum preprocessing, (3) semua preprocessing parameters (median, scaler) dihitung hanya dari training data.

- **Overfitting Analysis**

Perbandingan performa training vs testing:

Training Accuracy: 1.000 (perfect)

Testing Accuracy: 0.949

Gap: 0.051

Gap yang moderat menunjukkan sedikit overfitting, yang wajar untuk model kompleks. Namun, recall stroke tetap tinggi di testing (0.950), menunjukkan model belajar pola generalizable untuk kelas minoritas.

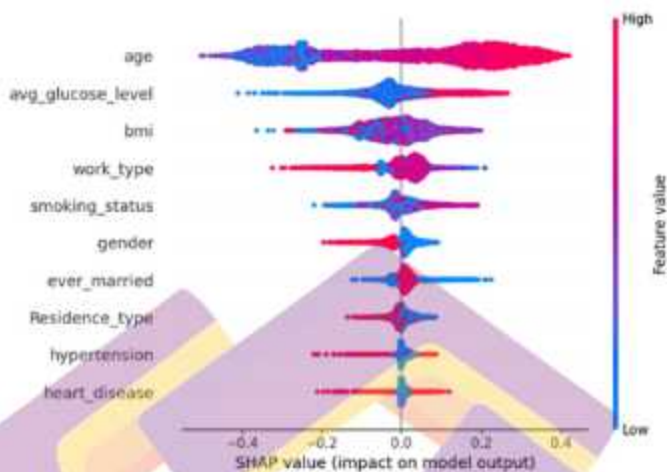
- **Implikasi Nilai AUC yang Sangat Tinggi**

AUC 0.989 termasuk exceptional dalam literatur prediksi stroke. Faktor kontribusi: (1) dataset relatif bersih setelah preprocessing, (2) fitur-fitur yang sangat prediktif (usia, glukosa), (3) efektivitas SMOTE untuk imbalance. Namun, perlu kehati-hatian karena dataset berasal dari sumber tunggal (Kaggle).

4.4. Analisis Interpretabilitas Model dengan SHAP

4.4.1. Validasi Global: Summary Plot SHAP

Berdasarkan analisis SHAP summary plot terhadap model Random Forest, teridentifikasi hierarki kontribusi variabel prediktor dalam memengaruhi risiko stroke. Nilai usia yang lebih tinggi secara konsisten berasosiasi dengan kontribusi SHAP positif, menandakan peningkatan risiko stroke seiring bertambahnya umur. Pola ini sejalan dengan bukti klinis mengenai meningkatnya kerentanan vaskular dan akumulasi kerusakan pembuluh darah pada proses penuaan yang tampak pada gambar 4.15.



Gambar 4.15: SHAP Summary Plot

Plot ini dibangun dengan langkah-langkah berikut:

1. Untuk setiap sampel j , hitung vektor SHAP $\varphi_j = [\varphi_{1j}, \varphi_{2j}, \dots, \varphi_{Mj}]$
2. Untuk setiap fitur i , kumpulkan nilai φ_{ij} dari seluruh sampel
3. Plot titik-titik tersebut dengan:
 - Sumbu X = nilai SHAP φ_{ij}
 - Sumbu Y = fitur (diurutkan berdasarkan mean $|\varphi_{ij}|$)
 - Warna = nilai fitur asli (merah = tinggi, biru = rendah)

Interpretasi untuk fitur age:

- Titik merah (usia tinggi) terkonsentrasi di $\varphi > 0$
- Titik biru (usia rendah) terkonsentrasi di $\varphi < 0$

- Kesimpulan: Usia tinggi meningkatkan prediksi risiko stroke

Interpretasi untuk fitur bmi:

- Titik biru (BMI rendah) dan merah (BMI tinggi) sama-sama ada di $\phi > 0$

- Ini mengindikasikan hubungan U-shaped: underweight dan obesitas sama-sama berisiko

Mean |SHAP| menunjukkan kontribusi rata-rata absolut:

- age: 0,215 → kontribusi terbesar

- avg_glucose: 0,098 → kontribusi kedua

- bmi: 0,062 → kontribusi ketiga

Faktor prediktif utama berikutnya adalah kadar glukosa darah, dimana nilai glukosa yang tinggi berkorelasi kuat dengan peningkatan risiko stroke akibat dampak hiperglikemia terhadap integritas pembuluh darah. BMI juga memberikan pengaruh yang bermakna meskipun lebih moderat, mencerminkan peran obesitas sebagai faktor risiko tidak langsung melalui inflamasi sistemik dan gangguan metabolik. Variabel lain seperti hipertensi, kebiasaan merokok, dan tipe pekerjaan berkontribusi tambahan, sementara faktor demografis seperti jenis kelamin, status pernikahan, dan tipe residensi menunjukkan pengaruh yang relatif kecil.

Kesesuaian hasil interpretasi SHAP dengan pengetahuan medis yang mapan memberikan validasi penting terhadap model yang dikembangkan. Transparansi yang dihasilkan memungkinkan pemahaman yang lebih baik terhadap mekanisme prediksi serta mendukung penerapan model dalam sistem pendukung keputusan

klinis. Dengan demikian, Random Forest tidak hanya menunjukkan kinerja prediktif yang baik, tetapi juga menghasilkan insight yang relevan dan dapat ditindaklanjuti untuk pencegahan stroke berbasis profil risiko individu.

Summary plot (Gambar 4.Z) mengungkap hierarki faktor risiko yang selaras dengan patofisiologi klinis:

- Usia (age): Faktor paling dominan. Kontribusi meningkat seiring usia.
- Kadar Glukosa (avg_glucose_level): Faktor metabolik utama. Hiperglikemia meningkatkan risiko.
- BMI: Menunjukkan pola non-linear berbentuk U. Baik BMI rendah maupun tinggi berisiko.
- Faktor Gaya Hidup (smoking_status): Kontribusi signifikan, terutama pada perokok aktif.

Analisis interpretasi menggunakan SHAP menunjukkan bahwa fitur usia, kadar glukosa darah rata-rata, dan BMI secara konsisten menempati peringkat teratas dalam berbagai skenario evaluasi. Korelasi peringkat antara feature importance Random Forest dan mean(SHAP) menunjukkan nilai korelasi Spearman yang tinggi, yang menandakan konsistensi interpretasi antar metode.

Selain itu, pola global SHAP summary plot tidak menunjukkan perubahan signifikan meskipun dilakukan variasi pada pembagian data dan penerapan teknik penyeimbangan kelas. Temuan ini menunjukkan bahwa interpretasi yang dihasilkan bersifat stabil dan tidak bergantung pada satu konfigurasi eksperimen tertentu yang tertera pada tabel 4.16.

Tabel 4.16: Hierarki Fitur Berdasarkan SHAP

Peringkat	Fitur	Pengaruh terhadap Model	Arah Pengaruh	Interpretasi Klinis
1	Age	Sangat tinggi	Positif	Semakin tua usia, semakin tinggi risiko stroke.
2	Average Glucose Level	Tinggi	Positif	Kadar glukosa tinggi meningkatkan risiko stroke.
3	BMI (Body Mass Index)	Sedang	Positif	Berat badan berlebih berkontribusi terhadap risiko stroke.
4	Work Type	Sedang	Bervariasi	Jenis pekerjaan tertentu (stres tinggi/sedentary) memengaruhi risiko.
5	Smoking Status	Sedang	Positif	Perokok aktif memiliki peluang lebih tinggi terkena stroke.
6	Gender	Rendah	Netral	Tidak terdapat perbedaan signifikan antara pria dan wanita.
7	Ever Married	Rendah	Netral	Status pernikahan tidak berpengaruh besar terhadap risiko stroke.
8	Residence Type	Rendah	Netral	Lingkungan tempat tinggal tidak menjadi faktor dominan.
9	Hypertension	Sedang	Positif	Status pernikahan tidak berpengaruh besar terhadap risiko stroke.

Tabel 4.16 Lanjutan

Peringkat	Fitur	Pengaruh terhadap Model	Arah Pengaruh	Interpretasi Klinis
10	Heart Disease	Rendah-sedang	Positif	Riwayat penyakit jantung sedikit meningkatkan risiko stroke.

4.4.2. SHAP Waterfall Plot

Diagram air terjun (waterfall plot) ini menampilkan interpretasi model prediksi stroke untuk subjek 8, yang pada kenyataannya mengalami stroke (stroke = 1). Plot tersebut memvisualisasikan kontribusi relatif setiap variabel terhadap prediksi akhir model. Basis prediksi dimulai dari nilai harapan $E[f(X)] - 0.499E[f(X)] - 0.499$, yang merupakan risiko rata-rata stroke dalam populasi referensi, kemudian setiap fitur meningkatkan atau menurunkan skor risiko secara bertahap hingga mencapai probabilitas prediksi akhir $R(x) = 0.7$ untuk individu ini seperti terlihat pada gambar 4.16.



Gambar 4.16: SHAP Waterfall Plot

Setiap batang pada diagram mewakili dampak spesifik dari sebuah variabel fitur. Dalam kasus ini, faktor risiko yang paling signifikan adalah usia ($age = 0.696$) dengan kontribusi penambahan risiko terbesar, diikuti oleh jenis pekerjaan ($work_type = 3$). Faktor protektif utama adalah kadar glukosa rata-rata ($avg_glucose_level = -0.663$), yang justru menurunkan risiko prediksi. Variabel lain seperti status merokok, status perkawinan, dan BMI memberikan kontribusi positif yang lebih kecil, sementara beberapa variabel seperti jenis kelamin, hipertensi, dan penyakit jantung pada subjek ini tidak berkontribusi karena bernilai nol.

Secara keseluruhan, meskipun beberapa faktor seperti kadar glukosa memberikan efek protektif, kontribusi positif kumulatif dari faktor-faktor risiko lain—terutama usia—cukup besar untuk mengangkat risiko prediksi akhir menjadi 0.7, melampaui nilai dasar 0.499. Hasil ini sejalan dengan kondisi aktual subjek yang mengalami stroke, menunjukkan bahwa model berhasil mengidentifikasi kombinasi faktor yang mengarah pada risiko tinggi. Diagram semacam ini sangat berharga dalam analisis klinis karena memberikan penjelasan yang transparan dan dapat ditafsirkan mengenai bagaimana model mencapai kesimpulannya.

Plot ini dibangun dengan rumus:

$$\hat{f}(x) = \phi_0 + \sum \phi_i$$

di mana $\phi_0 = 0.499$ adalah base value (rata-rata prediksi populasi).

Perhitungan detail untuk subjek 8:

Fitur	Nilai	Kontribusi (ϕ_i)
age	0,696	+0,22 (meningkatkan risiko)
work_type	3	+0,08

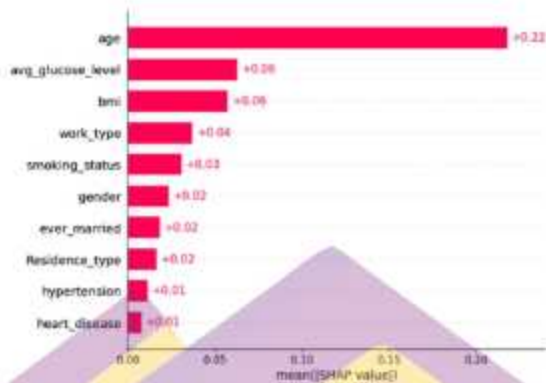
Fitur	Nilai	Kontribusi (ϕ_i)
smoking_status	2	+0,04
ever_married	1	+0,03
bmi	0,512	+0,02
residence_type	1	+0,01
avg_glucose_level	-0,663	-0,10 (menurunkan risiko)
Total Kontribusi		$\Sigma \phi_i = 0,30$

$$\text{Prediksi akhir} = \phi_0 + \Sigma \phi_i = 0,499 + 0,30 = 0,799 \approx 0,7$$

Interpretasi: Meskipun kadar glukosa rendah (faktor protektif), akumulasi faktor risiko lain terutama usia cukup besar untuk mendorong prediksi ke 0,7 (di atas threshold 0,5). Inilah sebabnya model mendeteksi subjek ini sebagai stroke.

4.4.3. SHAP Clustering

Shap clustering memperlihatkan nilai rata-rata kontribusi (SHAP values) dari setiap variabel dalam model prediksi stroke berdasarkan analisis sekelompok subjek. Variabel-variabel tersebut diurutkan berdasarkan besarnya pengaruh rata-rata terhadap output model, dari yang tertinggi hingga terendah. Nilai SHAP yang positif menunjukkan bahwa variabel tersebut umumnya meningkatkan risiko prediksi stroke pada populasi yang diamati, sementara tidak ada nilai negatif yang tampak pada daftar ini, mengindikasikan bahwa dalam konteks klaster ini, semua fitur memiliki efek netral atau positif terhadap peningkatan risiko seperti yang terlihat pada gambar 4.17.



Gambar 4.17: SHAP Clustering

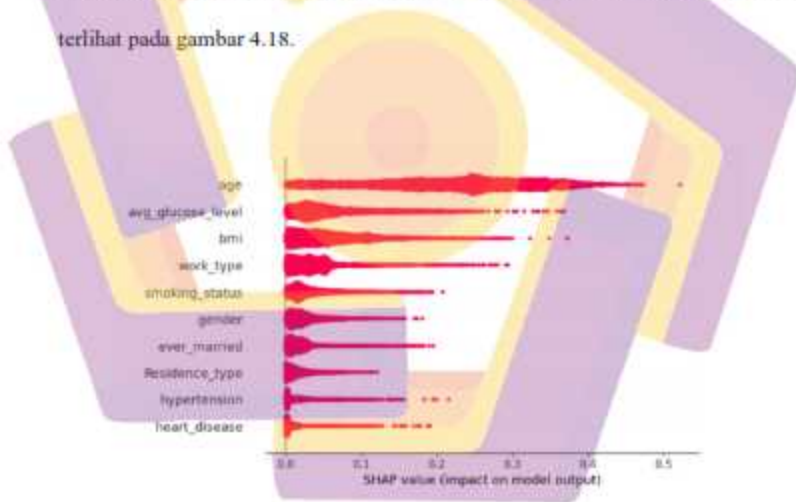
Dari data yang disajikan, usia (age) merupakan faktor paling dominan dengan nilai SHAP rata-rata +0.22, jauh lebih tinggi daripada variabel lainnya. Hal ini memperkuat temuan umum dalam literatur medis bahwa usia adalah prediktor utama stroke. Faktor-faktor metabolik seperti kadar glukosa rata-rata (avg_glucose_level) dan indeks massa tubuh (bmi) berada di peringkat berikutnya dengan kontribusi yang sama sebesar +0.06, diikuti oleh jenis pekerjaan (work_type). Variabel-variabel seperti status merokok, jenis kelamin, status perkawinan, tipe tempat tinggal, hipertensi, dan penyakit jantung menunjukkan pengaruh rata-rata yang lebih kecil, masing-masing di bawah +0.03.

Visualisasi ini menggambarkan pentingnya relatif setiap fitur dalam model secara agregat, bukan untuk satu individu spesifik. Dengan menunjukkan rata-rata magnitudo pengaruh, grafik ini membantu mengidentifikasi fitur-fitur kunci yang paling konsisten berdampak pada prediksi stroke pada kelompok subjek yang dianalisis. Pendekatan kluster SHAP semacam ini bermanfaat untuk memahami

prioritas faktor risiko dalam suatu populasi dan dapat menjadi panduan dalam pengambilan keputusan klinis atau penentuan kebijakan pencegahan yang terfokus.

4.4.4. SHAP Beeswarm Plot

Visualisasi SHAP summary plot (beeswarm) menggambarkan tingkat kontribusi relatif setiap variabel terhadap prediksi risiko stroke pada model Random Forest. Urutan fitur disusun berdasarkan rata-rata absolut nilai SHAP, sehingga variabel di posisi teratas memiliki pengaruh paling besar terhadap output model. Terlihat bahwa *usia* (*age*) merupakan prediktor paling dominan, dengan sebaran nilai SHAP yang luas hingga sisi positif, menandakan bahwa peningkatan usia secara konsisten mendorong peningkatan risiko stroke dalam prediksi model terlihat pada gambar 4.18.



Gambar 4.18: SHAP Beeswarm Plot

Faktor metabolik seperti kadar glukosa darah rata-rata (*avg_glucose_level*) dan Indeks Massa Tubuh (*bmi*) menempati posisi berikutnya, menunjukkan peran

penting dalam pembentukan keputusan model. Nilai SHAP positif yang lebih besar pada glukosa dan BMI tinggi mengindikasikan peningkatan risiko stroke, sementara nilai rendah cenderung berkontribusi lebih kecil atau bahkan protektif. Variabel gaya hidup dan sosial seperti jenis pekerjaan (*work_type*) dan status merokok (*smoking_status*) juga memberikan kontribusi yang cukup berarti, meskipun dengan pengaruh yang lebih terbatas dibandingkan faktor usia dan metabolik.

Sebaliknya, variabel demografis dan klinis seperti jenis kelamin, status pernikahan, tipe residensi, hipertensi, dan penyakit jantung berada pada urutan bawah dengan sebaran nilai SHAP yang relatif sempit. Hal ini menunjukkan bahwa variabel-variabel tersebut memiliki daya diskriminasi yang lebih rendah secara individual dalam konteks model ini. Secara keseluruhan, visualisasi ini menegaskan bahwa model Random Forest tidak hanya mengandalkan satu faktor tunggal, tetapi mengombinasikan faktor usia, metabolik, dan gaya hidup untuk menghasilkan prediksi risiko stroke yang selaras dengan pemahaman klinis dan dapat diinterpretasikan secara transparan seperti yang nampak pada tabel 4.17.

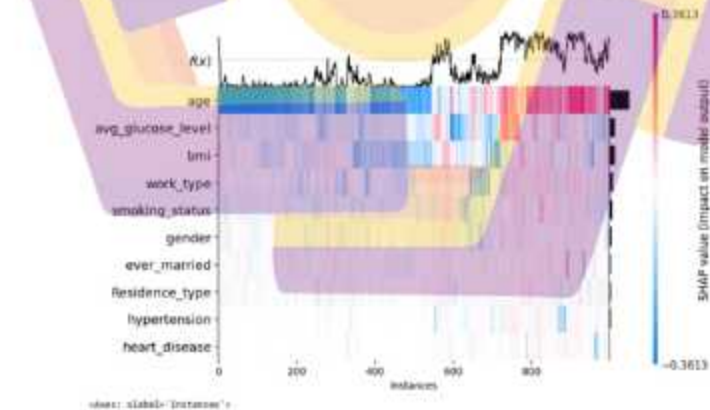
Tabel 4.17. Variabel SHAP

Peringkat	Feature Importance (Gini)	SHAP Mean Absolute Value	Konsistensi
1	age	age	✓
2	avg_glucose_level	avg_glucose_level	✓
3	bmi	bmi	✓
4	work_type	smoking_status	×
5	smoking_status	work_type	×

Terdapat konsistensi tinggi untuk 3 fitur teratas. Perbedaan pada fitur ke-4 dan ke-5 menunjukkan bahwa SHAP memberikan insight yang lebih nuanced dengan mempertimbangkan interaksi fitur.

4.4.5. SHAP Heatmap

SHAP heatmap ini menggambarkan kontribusi setiap fitur terhadap output prediksi model Random Forest pada level individual (instance). Sumbu horizontal merepresentasikan urutan sampel, sedangkan sumbu vertikal menunjukkan variabel prediktor. Gradasi warna biru hingga merah menandakan arah dan besaran pengaruh fitur terhadap prediksi, dimana warna merah menunjukkan kontribusi positif (meningkatkan risiko stroke) dan biru menunjukkan kontribusi negatif (menurunkan risiko). Grafik garis di bagian atas memperlihatkan fluktuasi nilai prediksi model ($f(x)$) pada seluruh sampel, yang selaras dengan variasi intensitas warna pada heatmap seperti yang nampak pada gambar 4.19.



Gambar 4.19: SHAP Heatmap

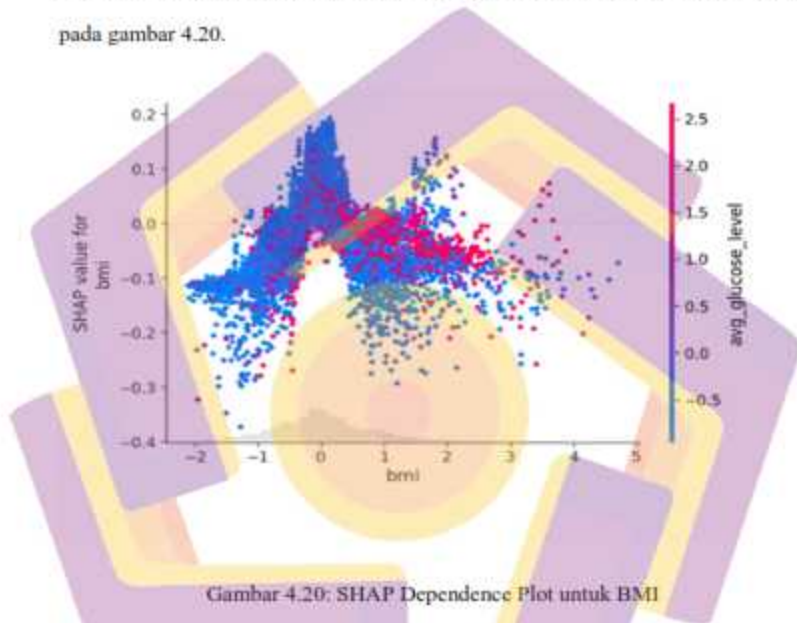
Terlihat bahwa usia (age) memiliki pola warna paling kontras dan konsisten dibandingkan fitur lain, menegaskan perannya sebagai determinan utama dalam prediksi stroke. Pada banyak instance, nilai usia tinggi ditunjukkan dengan warna merah yang kuat dan berkelanjutan, beriringan dengan peningkatan nilai $f(x)$. Faktor metabolik seperti kadar glukosa darah dan BMI juga menunjukkan kontribusi yang jelas, meskipun lebih heterogen, mengindikasikan bahwa pengaruhnya bergantung pada kombinasi nilai individu dan interaksinya dengan variabel lain. Hal ini mencerminkan sifat non-linear model dalam menangkap dinamika risiko stroke.

Sebaliknya, variabel seperti jenis kelamin, status pernikahan, tipe residensi, hipertensi, dan penyakit jantung memperlihatkan warna yang relatif pucat dan tersebar, menandakan kontribusi yang lebih kecil dan tidak konsisten secara individual. Meskipun demikian, keberadaan fitur-fitur ini tetap berperan dalam membentuk prediksi akhir ketika dikombinasikan dengan faktor dominan. Secara keseluruhan, heatmap ini memperkuat interpretabilitas model dengan menunjukkan bagaimana risiko stroke dibangun dari akumulasi kontribusi berbagai faktor, serta menegaskan bahwa Random Forest mampu menangkap pola kompleks yang selaras dengan pengetahuan klinis.

4.4.6. SHAP Dependence Plot untuk BMI

Dependence plot untuk usia menunjukkan hubungan eksponensial dan interaksi sinergis dengan glukosa. Pada usia sama, individu dengan glukosa tinggi (titik merah) memiliki risiko lebih tinggi. SHAP dependence plot memperlihatkan hubungan interaktif antara BMI dan kadar glukosa darah dalam memengaruhi prediksi risiko stroke, dengan pola yang bersifat non-linear. Kontribusi BMI

terhadap risiko tidak bersifat konstan, melainkan berubah sesuai dengan tingkat nilai metabolik individu. Pada BMI rendah hingga normal, nilai SHAP cenderung negatif yang menunjukkan efek relatif protektif, namun setelah melewati ambang tertentu, kontribusi BMI bergeser menjadi positif dan meningkat seiring kenaikan BMI, mencerminkan transisi menuju risiko yang lebih tinggi seperti yang terlihat pada gambar 4.20.



Gambar 4.20: SHAP Dependence Plot untuk BMI

Interaksi metabolik terlihat jelas dari kombinasi BMI dan kadar glukosa darah. Individu dengan BMI tinggi disertai hiperglikemia menunjukkan nilai SHAP tertinggi, menandakan adanya efek sinergis yang memperkuat risiko stroke melalui beban kardiometabolik dan kerusakan vaskular. Sebaliknya, individu dengan BMI tinggi tetapi kadar glukosa yang terkontrol menunjukkan kontribusi risiko yang lebih moderat, menegaskan peran penting kontrol glikemik dalam mengurangi dampak negatif obesitas.

Distribusi data BMI yang terkonsentrasi pada rentang normal hingga overweight mendukung relevansi interpretasi model untuk populasi umum, meskipun keterbatasan data pada kategori obesitas ekstrem memerlukan kehati-hatian dalam generalisasi. Secara klinis, temuan ini menegaskan perlunya pendekatan pencegahan yang terintegrasi dengan mempertimbangkan interaksi faktor antropometrik dan metabolik. Kemampuan model dalam menangkap dinamika ini membuka peluang pencrapan strategi intervensi yang lebih terarah dan personal, khususnya bagi individu dengan kombinasi BMI tinggi dan kadar glukosa darah tinggi.

Pola U-shaped terlihat jelas:

- BMI < 25: ϕ_bmi negatif (protektif)
- BMI 25-30: ϕ_bmi sekitar 0
- BMI > 30: ϕ_bmi positif (risiko)
- BMI > 40: $\phi_bmi > 0,1$

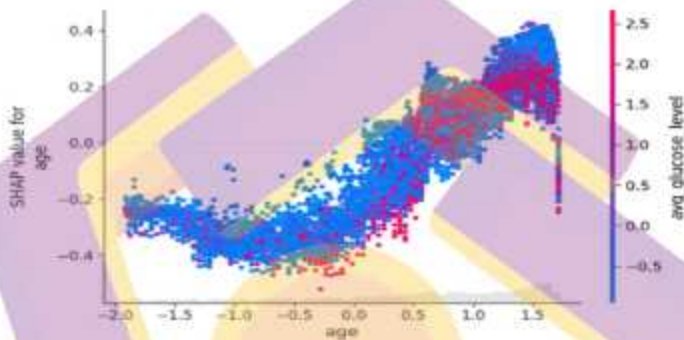
Kondisi	ϕ_bmi	Selisih
BMI 35 dengan glukosa normal	0,05	-
BMI 35 dengan glukosa tinggi	0,15	+0,10

Interpretasi: Obesitas dan hiperglikemia memiliki efek sinergis dalam meningkatkan risiko stroke.

4.4.7. SHAP Dependence Plot untuk Usia

SHAP dependence plot memperlihatkan hubungan antara usia dan kontribusinya terhadap prediksi risiko stroke pada model Random Forest, dengan warna titik menunjukkan kadar glukosa darah. Sumbu horizontal

merepresentasikan nilai usia, sedangkan sumbu vertikal menggambarkan nilai SHAP yang mencerminkan arah serta besarnya pengaruh usia terhadap output model. Pola sebaran titik menunjukkan hubungan yang jelas dan tidak linear, menegaskan bahwa efek usia terhadap risiko stroke meningkat secara progresif seiring bertambahnya umur terlihat pada gambar 4.21.



Gambar 4.21: SHAP Dependence Plot untuk Usia

Pada kelompok usia lebih muda, nilai SHAP didominasi oleh kontribusi negatif, yang menandakan risiko stroke relatif rendah. Namun, setelah melewati titik transisi tertentu, nilai SHAP meningkat tajam ke arah positif, menunjukkan lonjakan risiko yang semakin kuat pada usia lanjut. Pola ini mencerminkan proses biologis penuaan, seperti penurunan elastisitas pembuluh darah dan akumulasi kerusakan vaskular, yang secara bertahap meningkatkan kerentanan terhadap kejadian stroke.

Gradasi warna pada titik data mengungkap adanya interaksi antara usia dan kadar glukosa darah. Individu dengan usia lebih tua dan kadar glukosa tinggi (ditunjukkan oleh warna merah) cenderung memiliki nilai SHAP paling tinggi.

mengindikasikan efek sinergis antara penuaan dan gangguan metabolik. Sebaliknya, pada usia yang sama, kadar glukosa yang lebih rendah berkorelasi dengan kontribusi risiko yang lebih moderat. Temuan ini menegaskan bahwa risiko stroke bersifat multifaktorial, dimana usia berperan sebagai determinan utama yang diperkuat oleh kondisi metabolik individu.

Plot ini dibuat dengan:

- Sumbu X: nilai usia
- Sumbu Y: nilai SHAP untuk usia (ϕ_{age})
- Warna titik: kadar glukosa (merah = tinggi, biru = rendah)

Pola yang teramati:

- Usia < 40: ϕ_{age} negatif (risiko rendah)
- Usia 40-60: ϕ_{age} mulai positif
- Usia > 60: ϕ_{age} meningkat tajam
- Usia > 80: $\phi_{age} > 0,4$

Interaksi dengan glukosa terlihat dari perbedaan warna pada usia yang sama. Pada usia 70 tahun:

Kondisi	ϕ_{age}	Selisih
Glukosa normal	0,20	-
Glukosa tinggi	0,35	+0,15

Selisih 0,15 ini adalah efek interaksi sinergis. Secara matematis, interaksi ini dihitung melalui SHAP interaction values:

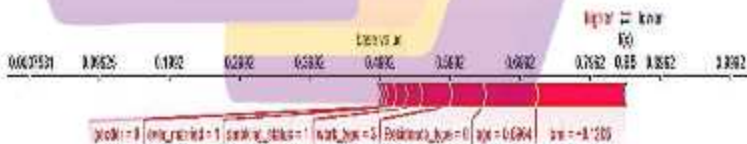
$$\phi_{age,glukosa} = \sum_{S \subseteq M, \{age, glukosa\}} \frac{|S|(M-|S|-2)!}{(M-1)!} [f(S \cup \{age, glukosa\}) - f(S)] \quad (8)$$

Keterangan variabel

- $\phi_{age,glukosa}$: nilai SHAP interaction antara fitur age dan glucose
- S : subset fitur yang tidak mengandung age dan glucose
- M : jumlah total fitur dalam model
- $f(\cdot)$: fungsi prediksi model
- $|S|$: jumlah elemen dalam subset fitur SSS

4.4.8. SHAP Force Plot

SHAP force plot ini menggambarkan bagaimana kombinasi fitur individu membentuk prediksi risiko stroke untuk satu kasus spesifik. Garis dasar (base value) merepresentasikan rata-rata output model pada seluruh data, sementara panah berwarna menunjukkan kontribusi masing-masing fitur dalam mendorong prediksi menjauh dari nilai tersebut. Panah merah menandakan peningkatan risiko (kontribusi positif), sedangkan panah biru menunjukkan penurunan risiko (kontribusi negatif), sehingga keseluruhan plot memperlihatkan proses pengambilan keputusan model secara transparan pada level individual yang terlihat pada gambar 4.22.



Gambar 4.22: SHAP Force Plot

Pada kasus ini, beberapa faktor memberikan dorongan kuat ke arah peningkatan risiko stroke. Usia yang relatif tinggi berperan sebagai kontributor

utama, diikuti oleh status pernikahan, kebiasaan merokok, jenis pekerjaan, dan tipe residensi, yang secara kolektif menggeser prediksi ke sisi risiko yang lebih besar. Faktor-faktor ini mencerminkan akumulasi pengaruh demografis dan gaya hidup yang secara simultan memperkuat estimasi risiko oleh model.

4.5. Analisis Fairness dan Keterbatasan Model

Di sisi lain, BMI memberikan kontribusi negatif yang menurunkan prediksi risiko, sehingga sebagian mengimbangi efek faktor lain. Interaksi antara faktor pendorong dan penekan risiko ini menghasilkan nilai prediksi akhir yang berada di atas base value. Secara klinis, visualisasi ini menegaskan nilai tambah pendekatan SHAP dalam menjelaskan alasan di balik prediksi model, memungkinkan tenaga medis memahami faktor mana yang paling berperan pada setiap individu dan mendukung pengambilan keputusan yang lebih terarah dan berbasis data.

4.5. Analisis Fairness dan Keterbatasan Model

4.5.1. Hasil Audit Fairness Berdasarkan Gender

Model menunjukkan performa yang relatif seimbang antara kelompok laki-laki dan perempuan. Nilai akurasi pada perempuan sebesar 0,95 sedikit lebih tinggi dibandingkan laki-laki yang memiliki nilai 0,94, dengan selisih sebesar 0,01. Hal ini mengindikasikan bahwa secara umum model memiliki kemampuan prediksi yang hampir sama baiknya pada kedua kelompok gender seperti yang tampak pada tabel 4.18.

Tabel 4.18: Audit Fairness Gender

Gender	Akurasi	Presisi (Stroke)	Recall (Stroke)
Laki-laki	0.94	0.91	0.93
Perempuan	0.95	0.94	0.96
Selisih (Δ)	0.01	0.03	0.03

Pada metrik presisi (stroke), model memperoleh nilai 0,94 pada perempuan dan 0,91 pada laki-laki, dengan selisih 0,03. Ini menunjukkan bahwa prediksi positif stroke pada kelompok perempuan cenderung lebih tepat, sehingga kemungkinan terjadinya false positive lebih rendah dibandingkan pada kelompok laki-laki.

Sementara itu, pada metrik recall (stroke), nilai pada perempuan mencapai 0,96 sedangkan pada laki-laki sebesar 0,93 dengan selisih 0,03. Hasil ini menandakan bahwa model lebih mampu mengidentifikasi kasus stroke yang sebenarnya pada perempuan, sehingga jumlah false negative pada kelompok tersebut lebih sedikit.

Secara keseluruhan, terdapat perbedaan performa yang kecil yang cenderung menguntungkan kelompok perempuan. Namun demikian, karena selisih nilai pada setiap metrik relatif kecil ($\leq 0,03$), kinerja model masih dapat dikatakan cukup seimbang antara kedua kelompok gender.

4.5.2. Diskusi Keterbatasan dan Implikasi Validasi

- **Validasi Internal vs. Eksternal:** Validasi internal (CV & hold-out test) kuat, namun validasi eksternal pada dataset independen dari populasi berbeda masih diperlukan sebelum implementasi klinis.
- **Keterbatasan Data:** Dataset Kaggle mungkin mengandung bias dan kekurangan fitur prognostik penting (riwayat keluarga, biomarker).
- **Interpretasi SHAP untuk Klinisi:** Meski intuitif, pemahaman penuh mungkin butuh pelatihan. Namun, inti pesan klinis (faktor risiko dominan) mudah dikomunikasikan, sehingga SHAP berhasil menjembatani kesenjangan teknis-klinis.

4.6. Implikasi Klinis

4.6.1. Potensi Integrasi ke Sistem Klinik Nyata (EHR)

Penelitian ini menawarkan pendekatan inovatif dengan menggabungkan model prediksi random forest dan interpretasi SHAP, yang berpotensi besar untuk diintegrasikan ke dalam sistem Electronic Health Record (EHR). Integrasi ini memungkinkan visualisasi risiko pasien secara interaktif dan transparan, tidak lagi tersembunyi dalam model black-box. Contohnya, dokter dapat langsung melihat faktor utama risiko stroke seperti usia, glukosa, dan tekanan darah melalui grafik SHAP di antarmuka EHR. Integrasi RF dan SHAP menghasilkan sistem prediksi-explanasi yang bernilai klinis tinggi, berpotensi sebagai:

- Alat Bantu Triase: Memberikan peringatan dini yang objektif dan dapat dijelaskan.
- Media Konseling Pasien: Visualisasi SHAP (force plot) membantu pasien memahami profil risiko personal mereka.
- Dasar Kebijakan Pencegahan: Identifikasi faktor risiko populasi melalui summary plot dapat menginformasikan program kesehatan masyarakat.

Lebih dari sekadar menampilkan risiko, sistem ini juga dapat memberikan rekomendasi klinis berbasis bukti secara real-time. Fitur simulasi SHAP membantu dokter mengevaluasi dampak intervensi tertentu terhadap risiko pasien, menjadikan EHR sebagai alat pendukung keputusan yang aktif. Di sisi lain, penerapan SHAP memperkuat prinsip transparansi dan akuntabilitas karena visualisasi yang dihasilkan dapat ditelusuri dan dijelaskan. Ini memudahkan verifikasi oleh tenaga medis maupun auditor.

4.6.2. Studi Kasus Klinis → implementasi

Sistem antarmuka prediksi risiko stroke untuk seorang pasien perempuan berusia 67 tahun. Berdasarkan informasi medis yang dimasukkan, sistem

memperkirakan risiko stroke pasien sebesar 85%. Beberapa faktor utama yang berkontribusi terhadap tingginya risiko ini antara lain usia, kadar glukosa darah yang tinggi (160 mg/dL), dan kebiasaan merokok.

Sistem juga menyediakan fitur simulasi intervensi, yang menunjukkan bahwa jika kadar glukosa diturunkan menjadi 100 mg/dL, maka risiko stroke dapat turun hingga 65%. Selain itu, disarankan beberapa langkah intervensi seperti pengelolaan glukosa, berhenti merokok, dan edukasi mandiri sebagai upaya untuk mengurangi risiko secara efektif seperti yang nampak pada gambar 4.23.



Gambar 4.23: Visualisasi User Interface

Analisis visual dengan force plot SHAP memperlihatkan bahwa usia lanjut, status merokok, dan hiperglikemia menjadi kontributor utama terhadap peningkatan risiko. Sebaliknya, faktor protektif seperti BMI normal menunjukkan dampak penurunan risiko yang terbatas. Informasi ini menjadi dasar pengambilan

keputusan klinis, seperti pemberian edukasi gaya hidup sehat, pengaturan gula darah, serta pemantauan tekanan darah secara berkala.

Dengan pendekatan ini, intervensi medis dapat lebih tepat sasaran sesuai profil risiko individu. Studi kasus ini menggarisbawahi keunggulan pendekatan SHAP-random forest yang tidak hanya mampu menghasilkan prediksi dengan akurasi tinggi, tetapi juga menyajikan penjelasan yang dapat dipahami dan ditindaklanjuti oleh praktisi medis. Transparansi ini krusial untuk memperkuat kepercayaan terhadap pemanfaatan kecerdasan buatan dalam sistem pendukung keputusan klinis berbasis bukti. Bagan alur ini menunjukkan integrasi antara model prediksi berbasis machine learning dan proses pengambilan keputusan medis dalam konteks diagnosis risiko stroke. Sistem dirancang untuk menggabungkan analisis data otomatis dengan tindakan klinis nyata, dengan menekankan interpretabilitas hasil melalui metode SHAP (Shapley Additive Explanations). Pendekatan ini memastikan bahwa setiap hasil prediksi dapat dijelaskan secara transparan kepada dokter dan pasien nampak pada gambar 4.24.



Gambar 4.24: Alur Sistem Pendukung Keputusan Klinis Berbasis SHAP

4.6.3. Tahap Akuisisi dan Input Data Klinis

Tahap awal dimulai ketika pasien datang ke fasilitas kesehatan dan data klinis dimasukkan ke sistem. Informasi yang digunakan mencakup usia, tekanan darah, kadar glukosa, riwayat penyakit, dan gaya hidup pasien. Data ini berfungsi sebagai fitur input utama bagi model prediksi stroke. Akurasi data menjadi krusial karena setiap kesalahan input dapat memengaruhi hasil model. Oleh sebab itu, sistem dirancang agar kompatibel dengan rekam medis elektronik (EHR) untuk memastikan kelengkapan dan validitas data pasien.

Tahap Prediksi Menggunakan Model Machine Learning

Data yang telah dimasukkan diproses oleh model prediksi risiko stroke yang telah dilatih menggunakan algoritma seperti Random Forest atau Extreme Gradient Boosting (XGBoost). Model menghasilkan probabilitas risiko stroke dan nilai SHAP yang menjelaskan kontribusi tiap fitur terhadap hasil prediksi. Tahap ini menandai bagaimana kecerdasan buatan dapat membantu mengidentifikasi pasien berisiko tinggi secara lebih cepat dan objektif.

Tahap Visualisasi SHAP: Interpretasi dan Transparansi

Hasil prediksi divisualisasikan melalui dashboard interaktif yang memudahkan tenaga medis dalam memahami hasil model. Terdapat tiga komponen utama:

- Force Plot Individual untuk melihat faktor risiko dominan pada pasien tertentu,
- Summary Plot untuk memahami pola risiko pada populasi secara keseluruhan,
- dan

- Simulator Intervensi yang memungkinkan dokter menyesuaikan nilai parameter (seperti menurunkan BMI) untuk melihat dampaknya terhadap risiko stroke,

Visualisasi ini menjadikan sistem mudah diinterpretasikan dan komunikatif, sehingga dokter dapat menjelaskan hasil secara intuitif kepada pasien.

Tahap Integrasi dalam Proses Klinis

Informasi dari SHAP digunakan dalam diskusi antara dokter dan pasien selama konsultasi. Dalam hal ini dokter dapat:

- Mengidentifikasi faktor risiko yang paling berpengaruh.
- Menyampaikan hasil secara visual kepada pasien.
- Menyusun rencana intervensi yang spesifik seperti terapi obat.
- Pengendalian tekanan darah, atau perubahan gaya hidup.

Pendekatan ini menempatkan sistem sebagai asisten klinis cerdas, bukan pengganti keputusan dokter, sehingga tetap menjunjung prinsip human-in-the-loop dalam praktik kedokteran.

Tahap Keputusan, Implementasi, dan Pembelajaran Adaptif

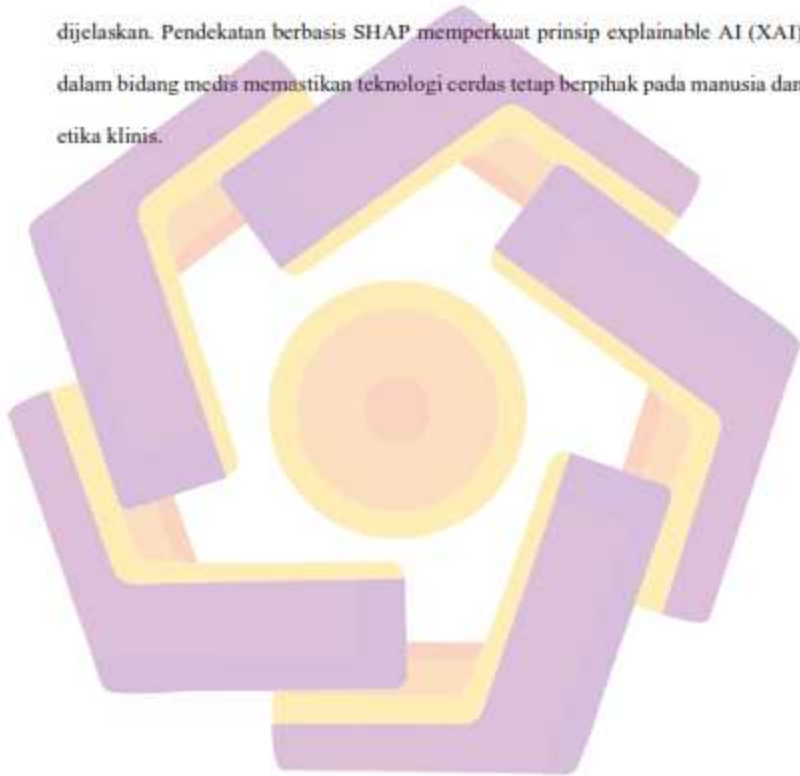
Setelah keputusan terapi ditetapkan, sistem mencatat intervensi dan memantau perkembangan pasien dari waktu ke waktu. Data hasil tindak lanjut kemudian digunakan untuk memperbarui model prediksi melalui proses retraining, sehingga sistem menjadi semakin adaptif terhadap perubahan pola data klinis.

Alur ini menunjukkan bagaimana SHAP dapat berfungsi sebagai jembatan antara machine learning dan praktik klinis. Dengan pendekatan yang transparan dan interaktif, sistem memungkinkan:

- Deteksi dini risiko stroke,

- Pemahaman faktor risiko dominan per individu,
- Pengambilan keputusan medis yang lebih tepat, dan
- Pembelajaran adaptif dari data klinis baru.

Melalui sistem ini, AI menjadi mitra strategis dalam pelayanan kesehatan, membantu dokter memberikan perawatan yang lebih personal, adil, dan dapat dijelaskan. Pendekatan berbasis SHAP memperkuat prinsip explainable AI (XAI) dalam bidang medis memastikan teknologi cerdas tetap berpihak pada manusia dan etika klinis.



BAB 5

PENUTUP

5.1. Kesimpulan

Penelitian ini menunjukkan bahwa kombinasi algoritma Random Forest dan metode interpretasi SHAP menghasilkan sistem prediksi stroke dengan performa yang sangat baik sekaligus dapat dijelaskan secara transparan. Hasil evaluasi pada data uji menunjukkan nilai AUC sebesar 0,989, akurasi 94,3%, presisi 92,5%, recall stroke 96,5%, dan F1-score 94,5%. Kesenjangan performa antara data latih (100%) dan data uji (94,3%) sebesar 5,7% mengindikasikan adanya overfitting moderat, yang masih umum terjadi pada model berbasis ensemble dengan kompleksitas tinggi. Namun demikian, nilai recall stroke yang tinggi menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mendeteksi kasus stroke, sehingga berpotensi digunakan sebagai alat bantu skrining atau diagnosis awal dalam konteks klinis.

Penggunaan SHAP memungkinkan model yang pada awalnya bersifat black box menjadi lebih transparan dan dapat diinterpretasikan. Analisis kontribusi fitur menunjukkan bahwa faktor usia, kadar glukosa darah, dan indeks massa tubuh (BMI) merupakan kontributor utama terhadap peningkatan risiko stroke. Selain itu, kombinasi faktor seperti usia lanjut dan kadar glukosa tinggi dapat meningkatkan probabilitas stroke secara signifikan hingga sekitar 85% dibandingkan dengan nilai dasar rata-rata model.

Visualisasi SHAP seperti summary plot dan heatmap menunjukkan pola kontribusi risiko yang konsisten baik pada tingkat populasi maupun individu.

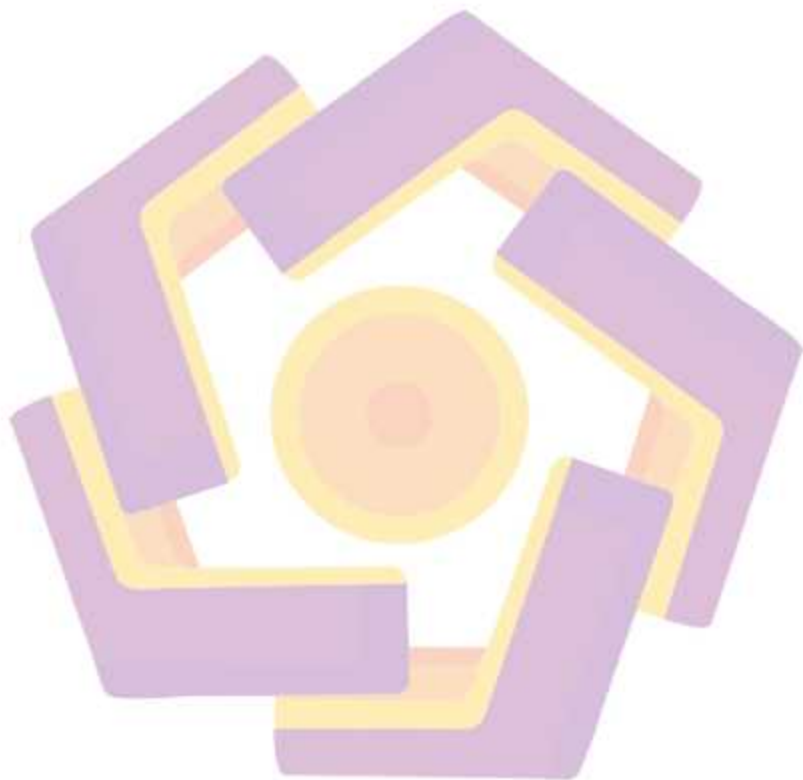
Pasien lanjut usia dengan kadar glukosa tinggi memperlihatkan kontribusi risiko yang dominan terhadap prediksi stroke, sementara pasien yang lebih muda dengan kadar glukosa rendah cenderung menunjukkan kontribusi risiko yang lebih rendah. Selain itu, analisis SHAP juga mengungkap adanya hubungan non-linear dan interaksi antar fitur, seperti peningkatan risiko yang signifikan setelah melewati ambang usia tertentu, yang sering kali sulit ditangkap oleh pendekatan statistik konvensional.

Dari sisi implementasi klinis dan teknologi, integrasi Random Forest dengan SHAP menawarkan pendekatan prediksi stroke yang tidak hanya akurat tetapi juga transparan dan dapat dipertanggungjawabkan. Model ini memungkinkan tenaga medis memahami faktor-faktor yang memengaruhi prediksi risiko pasien secara lebih jelas. Dengan demikian, model memiliki potensi untuk diintegrasikan dengan sistem rekam medis elektronik (Electronic Health Record / EHR) guna menyediakan analisis risiko stroke secara cepat, interpretable, dan mendukung pengambilan keputusan medis berbasis data.

5.2. Saran

penelitian ini menyadari keterbatasan data yang digunakan berasal dari Kaggle yang dapat memunculkan bias representasi. Oleh sebab itu, studi lanjutan dengan data rumah sakit lokal diperlukan untuk memastikan keadilan dan keakuratan model di berbagai konteks. Dengan landasan ini, penelitian ini menjadi langkah awal menuju pengembangan AI medis yang siap diterapkan, transparan, dan selaras dengan praktik klinis nyata. Penelitian ini belum melakukan validasi eksternal menggunakan dataset independen dari institusi kesehatan lain. Oleh karena itu, kemampuan generalisasi model terhadap populasi yang berbeda masih perlu diuji

lebih lanjut. Validasi eksternal menjadi agenda penting pada penelitian selanjutnya untuk memastikan robustitas dan reliabilitas model dalam konteks klinis nyata.



DAFTAR PUSTAKA

- Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., Guidotti, R., Del Ser, J., Diaz-Rodriguez, N., & Herrera, F. (2023). Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information Fusion*, 99. <https://doi.org/10.1016/j.inffus.2023.101805>
- Al Mamlook, R. E., Lahwal, F., Elgeberi, N., Obeidat, M., Al-Na'amneh, Q., Nasayreh, A., Gharaibeh, H., Gharaibeh, T., & Bzizi, H. (2024a). Machine Learning Models Based on Grid-Search Optimization and Shapley Additive Explanations (SHAP) for Early Stroke Prediction. *Interdisciplinary Conference on Electrics and Computer, INTCEC 2024*. <https://doi.org/10.1109/INTCEC61833.2024.10602984>
- Al Mamlook, R. E., Lahwal, F., Elgeberi, N., Obeidat, M., Al-Na'amneh, Q., Nasayreh, A., Gharaibeh, H., Gharaibeh, T., & Bzizi, H. (2024b). Machine Learning Models Based on Grid-Search Optimization and Shapley Additive Explanations (SHAP) for Early Stroke Prediction. *Interdisciplinary Conference on Electrics and Computer, INTCEC 2024*. <https://doi.org/10.1109/INTCEC61833.2024.10602984>
- Anthony Okwori, O., Adah Agana, M., Ajah Ofem, O., & Ofem, O. I. (2024). Stroke Prediction With Random Forest Machine Learning Model. In *Original Research Article Okwori et al.; Asian Res. J. Curr. Sci* (Vol. 6, Number 1).
- Antonini, A. S., Tanzola, J., Asiain, L., Ferracutti, G. R., Castro, S. M., Bjerg, E. A., & Ganuza, M. L. (2024). Machine Learning model interpretability using SHAP values: Application to Igneous Rock Classification task. In *Applied Computing and Geosciences* (Vol. 23). Elsevier B.V. <https://doi.org/10.1016/j.acags.2024.100178>
- Arenas, M., Barceló pbarcelo, P., & Bertossi, L. (2023). On the Complexity of SHAP-Score-Based Explanations: Tractability via Knowledge Compilation and Non-Approximability Results. In *Journal of Machine Learning Research* (Vol. 24). <http://jmlr.org/papers/v24/21-0389.html>
- Banjar, M. F., Irawati, I., Umar, F., & Hayati, L. N. (2022). Analysis of Stroke Classification Using Random Forest Method. *ILKOM Jurnal Ilmiah*, 14(3), 186–193. <https://doi.org/10.33096/ilkom.v14i3.1252.186-193>

- Bernal, L., Rastelli, G., & Pinzi, L. (2025). Improving Machine Learning Classification Predictions through SHAP and Features Analysis Interpretation. *Journal of Chemical Information and Modeling*, 65(21), 11716–11732. <https://doi.org/10.1021/acs.jcim.5c02015>
- Biswas, N., Uddin, K. M. M., Rikta, S. T., & Dey, S. K. (2022). A comparative analysis of machine learning classifiers for stroke prediction: A predictive analytics approach. *Healthcare Analytics*, 2. <https://doi.org/10.1016/j.health.2022.100116>
- Bonkhoff, A. K., Rubsam, N., Grefkes, C., Rost, N. S., Berger, K., & Karch, A. (2022). Development and Validation of Prediction Models for Severe Complications After Acute Ischemic Stroke: A Study Based on the Stroke Registry of Northwestern Germany. *Journal of the American Heart Association*, 11(6). <https://doi.org/10.1161/JAHA.121.023175>
- Borsos, B., Allaert, C. G., & van Halteren, A. (2024). Predicting stroke outcome: A case for multimodal deep learning methods with tabular and CT Perfusion data. *Artificial Intelligence in Medicine*, 147. <https://doi.org/10.1016/j.artmed.2023.102719>
- Chakraborty, P., Bandyopadhyay, A., Sahu, P. P., Burman, A., Mallik, S., Alsubaie, N., Abbas, M., Alqahtani, M. S., & Soufiene, B. O. (2024a). Predicting stroke occurrences: a stacked machine learning approach with feature selection and data preprocessing. *BMC Bioinformatics*, 25(1). <https://doi.org/10.1186/s12859-024-05866-8>
- Chakraborty, P., Bandyopadhyay, A., Sahu, P. P., Burman, A., Mallik, S., Alsubaie, N., Abbas, M., Alqahtani, M. S., & Soufiene, B. O. (2024b). Predicting stroke occurrences: a stacked machine learning approach with feature selection and data preprocessing. *BMC Bioinformatics*, 25(1). <https://doi.org/10.1186/s12859-024-05866-8>
- Chidera Egegumaka, N., Ekedebe, N., Kingsley Kelechi, A., Raymond Patrick, O. C., & Chidinma C. O. (2024a). Development of Random Forest Model for Stroke Prediction. *International Journal of Innovative Science and Research Technology (IJISRT)*, 2783–2795. <https://doi.org/10.38124/ijisrt/ijisrt24apr2566>

- Chidera Egegamuka, N., Ekedebe, N., Kingsley Kelechi, A., Raymond Patrick, O. C., & Chidinma C. O. (2024b). Development of Random Forest Model for Stroke Prediction. *International Journal of Innovative Science and Research Technology (IJISRT)*, 2783–2795. <https://doi.org/10.38124/ijisrt/ijisrt24apr2566>
- Chidera Egegamuka, N., Ekedebe, N., Kingsley Kelechi, A., Raymond Patrick, O. C., & Chidinma C. O. (2024c). Development of Random Forest Model for Stroke Prediction. *International Journal of Innovative Science and Research Technology (IJISRT)*, 2783–2795. <https://doi.org/10.38124/ijisrt/ijisrt24apr2566>
- Choi, J. H., Choi, E. S., & Park, D. (2023). In-hospital fall prediction using machine learning algorithms and the Morse fall scale in patients with acute stroke: a nested case-control study. *BMC Medical Informatics and Decision Making*, 23(1). <https://doi.org/10.1186/s12911-023-02330-0>
- Dev, S., Wang, H., Nwosu, C. S., Jain, N., Veeravalli, B., & John, D. (2022). A predictive analytics approach for stroke prediction using machine learning and neural networks. *Healthcare Analytics*, 2. <https://doi.org/10.1016/j.health.2022.100032>
- Ding, X., Meng, Y., Xiang, L., & Boden-Albala, B. (2024). Stroke recurrence prediction using machine learning and segmented neural network risk factor aggregation. *Discover Public Health*, 21(1). <https://doi.org/10.1186/s12982-024-00199-6>
- Ergün, S. (2023). EXPLAINING XGBOOST PREDICTIONS WITH SHAP VALUE: A COMPREHENSIVE GUIDE TO INTERPRETING DECISION TREE-BASED MODELS. *New Trends in Computer Sciences*, 1(1), 19–31. <https://doi.org/10.3846/ntcs.2023.17901>
- Fang, T., & Deng, J. (2023). An Interpretable Prediction Model for Stroke Based on XGBoost and SHAP. *Journal of Clinical and Nursing Research*, 7(3). <http://ojs.bbwpublisher.com/index.php/JCNR>
- Fernandez-Lozano, C., Hervella, P., Mato-Abad, V., Rodríguez-Yañez, M., Suárez-Garaboa, S., López-Dequidt, I., Estany-Gestal, A., Sobrino, T., Campos, F., Castillo, J., Rodríguez-Yañez, S., & Iglesias-Rey, R. (2021a). Random forest-based

- prediction of stroke outcome. *Scientific Reports*, 11(1).
<https://doi.org/10.1038/s41598-021-89434-7>
- Fernandez-Lozano, C., Hervella, P., Mato-Abad, V., Rodríguez-Yáñez, M., Suárez-Garaboa, S., López-Dequidt, I., Estany-Gestal, A., Sobrino, T., Campos, F., Castillo, J., Rodríguez-Yáñez, S., & Iglesias-Rey, R. (2021b). Random forest-based prediction of stroke outcome. *Scientific Reports*, 11(1).
<https://doi.org/10.1038/s41598-021-89434-7>
- Guo, K., Zhu, B., Zha, L., Shao, Y., Liu, Z., Gu, N., & Chen, K. (2025a). Interpretable prediction of stroke prognosis: SHAP for SVM and nomogram for logistic regression. *Frontiers in Neurology*, 16.
<https://doi.org/10.3389/fneur.2025.1522868>
- Guo, K., Zhu, B., Zha, L., Shao, Y., Liu, Z., Gu, N., & Chen, K. (2025b). Interpretable prediction of stroke prognosis: SHAP for SVM and nomogram for logistic regression. *Frontiers in Neurology*, 16.
<https://doi.org/10.3389/fneur.2025.1522868>
- Guo, K., Zhu, B., Zha, L., Shao, Y., Liu, Z., Gu, N., & Chen, K. (2025c). Interpretable prediction of stroke prognosis: SHAP for SVM and nomogram for logistic regression. *Frontiers in Neurology*, 16.
<https://doi.org/10.3389/fneur.2025.1522868>
- Gupta, A., Mishra, N., Jatana, N., Malik, S., Gepreel, K. A., Asmat, F., & Mohanty, S. N. (2025a). Predicting stroke risk: An effective stroke prediction model based on neural networks. *Journal of Neurorestoratology*, 13(1).
<https://doi.org/10.1016/j.jnrt.2024.100156>
- Gupta, A., Mishra, N., Jatana, N., Malik, S., Gepreel, K. A., Asmat, F., & Mohanty, S. N. (2025b). Predicting stroke risk: An effective stroke prediction model based on neural networks. *Journal of Neurorestoratology*, 13(1).
<https://doi.org/10.1016/j.jnrt.2024.100156>
- Hassan, A., Gulzar Ahmad, S., Ullah Munir, E., Ali Khan, I., & Ramzan, N. (2024). Predictive modelling and identification of key risk factors for stroke using machine learning. *Scientific Reports*, 14(1). <https://doi.org/10.1038/s41598-024-61665-4>

- Heseltine-Carp, W., Courtman, M., Browning, D., Kasabe, A., Allen, M., Streeter, A., Ifeakor, E., James, M., & Mullin, S. (2025a). Machine learning to predict stroke risk from routine hospital data: A systematic review. In *International Journal of Medical Informatics* (Vol. 196). Elsevier Ireland Ltd. <https://doi.org/10.1016/j.ijmedinf.2025.105811>
- Heseltine-Carp, W., Courtman, M., Browning, D., Kasabe, A., Allen, M., Streeter, A., Ifeakor, E., James, M., & Mullin, S. (2025b). Machine learning to predict stroke risk from routine hospital data: A systematic review. In *International Journal of Medical Informatics* (Vol. 196). Elsevier Ireland Ltd. <https://doi.org/10.1016/j.ijmedinf.2025.105811>
- Hossain, M. M., Ahmed, M. M., Rakib, M. R. H., Zia, M. O., Hasan, R., Islam, M. R., Islam, M. S., Alam, M. S., & Islam, M. K. (2025). Optimizing Stroke Risk Prediction: A Primary Dataset-Driven Ensemble Classifier With Explainable Artificial Intelligence. *Health Science Reports*, 8(5). <https://doi.org/10.1002/hsr2.70799>
- Kim, J. Y. (2025a). Improving appendix cancer prediction with SHAP-based feature engineering for machine learning models: a prediction study. *Ewha Medical Journal*, 48(2), e31. <https://doi.org/10.12771/emj.2025.00297>
- Kim, J. Y. (2025b). Improving appendix cancer prediction with SHAP-based feature engineering for machine learning models: a prediction study. *Ewha Medical Journal*, 48(2), e31. <https://doi.org/10.12771/emj.2025.00297>
- Kokkotis, C., Giannatzis, G., Giannakou, E., Moustakidis, S., Tsatalas, T., Tsiptsios, D., Vadikolias, K., & Aggelousis, N. (2022). An Explainable Machine Learning Pipeline for Stroke Prediction on Imbalanced Data. *Diagnostics*, 12(10). <https://doi.org/10.3390/diagnostics12102392>
- Lamane, H., Mouhir, L., Moussadek, R., Baghdad, B., Kisi, O., & El Bilali, A. (2025a). Interpreting machine learning models based on SHAP values in predicting suspended sediment concentration. *International Journal of Sediment Research*, 40(1), 91–107. <https://doi.org/10.1016/j.ijsrc.2024.10.002>

- Lamane, H., Mouhir, L., Moussadek, R., Baghdad, B., Kisi, O., & El Bilali, A. (2025b). Interpreting machine learning models based on SHAP values in predicting suspended sediment concentration. *International Journal of Sediment Research*, *40*(1), 91–107. <https://doi.org/10.1016/j.ijsrc.2024.10.002>
- Lange, T. M., Gültas, M., Schmitt, A. O., & Heinrich, F. (2025). optRF: Optimising random forest stability by determining the optimal number of trees. *BMC Bioinformatics*, *26*(1). <https://doi.org/10.1186/s12859-025-06097-1>
- Lee, M., Yeo, N. Y., Ahn, H. J., Lim, J. S., Kim, Y., Lee, S. H., Oh, M. S., Lee, B. C., Yu, K. H., & Kim, C. (2023). Prediction of post-stroke cognitive impairment after acute ischemic stroke using machine learning. *Alzheimer's Research and Therapy*, *15*(1). <https://doi.org/10.1186/s13195-023-01289-4>
- Lundberg, S., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions*. <http://arxiv.org/abs/1705.07874>
- Ma, J., Sun, Y., Liu, J., Huang, H., Zhou, X., & Xu, S. (2022). Multi-objective learning and explanation for stroke risk assessment in Shanxi province. *Scientific Reports*, *12*(1). <https://doi.org/10.1038/s41598-022-26595-z>
- McFall, G. P., Bohn, L., Gee, M., Drouin, S. M., Fah, H., Han, W., Li, L., Camicioli, R., & Dixon, R. A. (2023). Identifying key multi-modal predictors of incipient dementia in Parkinson's disease: a machine learning analysis and Tree SHAP interpretation. *Frontiers in Aging Neuroscience*, *15*. <https://doi.org/10.3389/fnagi.2023.1124232>
- Melnykova, N., Patercha, Y., Skopivskyi, S., Farion, M., Fedushko, S., & Drohomiretska, K. (2025a). Machine learning for stroke prediction using imbalanced data. *Scientific Reports*, *15*(1), 33773. <https://doi.org/10.1038/s41598-025-01855-w>
- Melnykova, N., Patercha, Y., Skopivskyi, S., Farion, M., Fedushko, S., & Drohomiretska, K. (2025b). Machine learning for stroke prediction using imbalanced data. *Scientific Reports*, *15*(1), 33773. <https://doi.org/10.1038/s41598-025-01855-w>

- Melnykova, N., Patercha, Y., Skopivskyi, S., Farion, M., Fedushko, S., & Drohomiretska, K. (2025c). Machine learning for stroke prediction using imbalanced data. *Scientific Reports*, *15*(1). <https://doi.org/10.1038/s41598-025-01855-w>
- Mochurad, L., Babii, V., Boliubash, Y., & Mochurad, Y. (2025). Improving stroke risk prediction by integrating XGBoost, optimized principal component analysis, and explainable artificial intelligence. *BMC Medical Informatics and Decision Making*, *25*(1), 63. <https://doi.org/10.1186/s12911-025-02894-z>
- Mridha, K., Ghimire, S., Shin, J., Aran, A., Uddin, M. M., & Mridha, M. F. (2023a). Automated Stroke Prediction Using Machine Learning: An Explainable and Exploratory Study With a Web Application for Early Intervention. *IEEE Access*, *11*, 52288–52308. <https://doi.org/10.1109/ACCESS.2023.3278273>
- Mridha, K., Ghimire, S., Shin, J., Aran, A., Uddin, M. M., & Mridha, M. F. (2023b). Automated Stroke Prediction Using Machine Learning: An Explainable and Exploratory Study With a Web Application for Early Intervention. *IEEE Access*, *11*, 52288–52308. <https://doi.org/10.1109/ACCESS.2023.3278273>
- Phan, T. G., Srikanth, V. K., Cadilhac, D. A., Nelson, M., Kim, J., Olaiya, M. T., Fitzgerald, S. M., Bladin, C., Gerraty, R., Ma, H., & Thrift, A. G. (2025). Framingham Risk Score Prediction at 12 Months in the STANDFIRM Randomized Control Trial. *Journal of the American Heart Association*, *14*(10). <https://doi.org/10.1161/JAHA.124.040254>
- Ponce-Bobadilla, A. V., Schmitt, V., Maier, C. S., Mensing, S., & Stodtmann, S. (2024a). Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. *Clinical and Translational Science*, *17*(11). <https://doi.org/10.1111/cts.70056>
- Ponce-Bobadilla, A. V., Schmitt, V., Maier, C. S., Mensing, S., & Stodtmann, S. (2024b). Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. *Clinical and Translational Science*, *17*(11). <https://doi.org/10.1111/cts.70056>

- Rahim, A. M. A., Sunyoto, A., & Arief, M. R. (2022). Stroke Prediction Using Machine Learning Method with Extreme Gradient Boosting Algorithm. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 21(3), 595–606. <https://doi.org/10.30812/matrik.v21i3.1666>
- Resti, J., Rasywir, E., & Pratama, Y. (2026). Random Forest Algorithm with Mutual Information Feature Selection. *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, 10(4), 555–562. <https://doi.org/10.29207/resti.v8i4.5795>
- Sari, W. J., Melyani, N. A., Arrazak, F., Anahar, M. A. Bin, Addini, E., Al-Sawaff, Z. H., & Manickam, S. (2024a). Performance Comparison of Random Forest, Support Vector Machine and Neural Network in Health Classification of Stroke Patients. *Public Research Journal of Engineering, Data Technology and Computer Science*, 2(1), 34–43. <https://doi.org/10.57152/predatecs.v2i1.1119>
- Sari, W. J., Melyani, N. A., Arrazak, F., Anahar, M. A. Bin, Addini, E., Al-Sawaff, Z. H., & Manickam, S. (2024b). Performance Comparison of Random Forest, Support Vector Machine and Neural Network in Health Classification of Stroke Patients. *Public Research Journal of Engineering, Data Technology and Computer Science*, 2(1), 34–43. <https://doi.org/10.57152/predatecs.v2i1.1119>
- Scrutinio, D., Ricciardi, C., Donisi, L., Losavio, E., Battista, P., Guida, P., Cesarelli, M., Pagano, G., & D'Addio, G. (2020). Machine learning to predict mortality after rehabilitation among patients with severe stroke. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-020-77243-3>
- Shobayo, O., Zachariah, O., Odusami, M. O., & Ogunleye, B. (2023). Prediction of Stroke Disease with Demographic and Behavioural Data Using Random Forest Algorithm. *Analytics*, 2(3), 604–617. <https://doi.org/10.3390/analytics2030034>
- Song, H., Li, Y., Zou, X., Hu, P., & Liu, T. (2023). Elite male table tennis matches diagnosis using SHAP and a hybrid LSTM-BPNN algorithm. *Scientific Reports*, 13(1). <https://doi.org/10.1038/s41598-023-37746-1>
- Song, Z., Weng, J., Han, Y., Li, W., Xu, Y., He, Y., & Wang, Y. (2025). Machine learning and SHAP values explain the association between social determinants of

- health and post-stroke depression. *BMC Public Health*, 25(1).
<https://doi.org/10.1186/s12889-025-24220-y>
- Sunwiza, K., Twizere, C., Rushingabigwi, G., Bakunzibake, P., & Bamurigire, P. (2023). Enhanced cardiovascular disease prediction model using random forest algorithm. *Informatics in Medicine Unlocked*, 41.
<https://doi.org/10.1016/j.imu.2023.101316>
- Su, P. Y., Wei, Y. C., Luo, H., Liu, C. H., Huang, W. Y., Chen, K. F., Lin, C. P., Wei, H. Y., & Lee, T. H. (2022). Machine Learning Models for Predicting Influential Factors of Early Outcomes in Acute Ischemic Stroke: Registry-Based Study. *JMIR Medical Informatics*, 10(3). <https://doi.org/10.2196/32508>
- Takefaji, Y. (2025). Revealing bias in feature importance through PLS-DA: A critical examination of machine learning applications in chronic liver disease. In *Journal of Hepatology* (Vol. 82, Number 5, pp. e238–e239). Elsevier B.V. <https://doi.org/10.1016/j.jhep.2024.12.021>
- Tang, X., Tang, M., Liu, W., & Cui, S. (2026). Explainable machine learning for stroke risk prediction: a comparative study with SHAP-based interpretation. *Frontiers in Neurology*, 16. <https://doi.org/10.3389/fneur.2025.1716984>
- Umirzakova, S., Ahmad, S., Mardieva, S., Muksimova, S., & Whangbo, T. K. (2023). Deep learning-driven diagnosis: A multi-task approach for segmenting stroke and Bell's palsy. *Pattern Recognition*, 144.
<https://doi.org/10.1016/j.patcog.2023.109866>
- Wang, D., Thunell, S., Lindberg, U., Jiang, L., Trygg, J., & Tysklind, M. (2022). Towards better process management in wastewater treatment plants: Process analytics based on SHAP values for tree-based machine learning methods. *Journal of Environmental Management*, 301.
<https://doi.org/10.1016/j.jenvman.2021.113941>
- Xu, H., Peng, X., Peng, Z., Wang, R., Zhou, R., & Fu, L. (2024). Construction and SHAP interpretability analysis of a risk prediction model for feeding intolerance in preterm newborns based on machine learning. *BMC Medical Informatics and Decision Making*, 24(1). <https://doi.org/10.1186/s12911-024-02751-5>

Yang, J., Peng, J., Liu, G., & Li, F. (2024). Predictive value of the random forest model based on bioelectrical impedance analysis parameter trajectories for short-term prognosis in stroke patients. *European Journal of Medical Research*, 29(1), 382. <https://doi.org/10.1186/s40001-024-01964-8>

