Perbandingan Model Individual Logistic Regression dan Random Forest Dengan Model Stacking Pada Prediksi Penyakit Stroke

Muhammad Naufal Rafy Ghaly[1], Faisal Muttaqin[2], Ardhon Rakhmadi[3]
Informatika, Universitas Pembangunan Nasional "Veteran" Jawa Timur

e-mail: 22081010295@student.upnjatim.ac.id [1], faisalmuttaqin.if@upnjatim.ac.id [2], ardhon.rakhmadi.fasilkom@upnjatim.ac.id [3]

Kata Kunci:

Stroke
Logistic Regression
Random Forest
Stacking Classifier
Artificial Neural Network

ABSTRAK

Stroke merupakan penyebab utama kematian dan kecacatan global yang memerlukan deteksi dini berbasis *machine learning*. Namun, model klasifikasi individual sering kali terkendala oleh seberapa kompleks data klinis dan ketidakseimbangan kelas. Penelitian ini bertujuan membandingkan kinerja model individual *Logistic Regression* (LR) dan *Random Forest* (RF) dengan pendekatan *ensemble Stacking Classifier* berbasis *meta-learner Artificial Neural Network* (ANN). Integrasi data menggabungkan 5.110 rekam medis sekunder dan 4.448 data primer dari salah satu puskesmas di Surabaya. Prapemrosesan mencakup standarisasi *StandardScaler* dan penyeimbangan data menggunakan SMOTE melalui *5-Fold Cross Validation*. Model *stacking* memanfaatkan *out of fold predictions* dari LR dan RF sebagai *meta-features* untuk dipelajari oleh ANN. Hasil evaluasi menunjukkan model RF mencapai *Recall* 77%, akurasi 73%, dan ROC-AUC 73%, sedangkan LR mencatatkan *Recall* 71%, akurasi 71%, dan ROC-AUC 84%. Penggabungan kedua model dasar melalui arsitektur *stacking* berhasil meningkatkan performa dengan ROC-AUC 85%, akurasi 72%, dan *Recall* tertinggi sebesar 81%. Hasil ini membuktikan bahwa *Stacking Classifier* berbasis ANN mampu memberikan stabilitas performa yang lebih optimal dan sensitivitas tinggi dalam mendeteksi risiko stroke.

Keyword:

Stroke
Logistic Regression
Random Forest
Stacking Classifier
Artificial Neural Network

ABSTRACT

Stroke is a major cause of death and disability around the world requiring early detection and can be done using machine learning. However, single models often struggle because of complex medical data and uneven class sizes. This study compares two models of individual Logistic Regression (LR) and Random Forest (RF), with a combined method using a Stacking Classifier with an Artificial Neural Network (ANN) meta-learner. The study used 5,110 secondary medical records and 4,448 primary data points from a health center in Surabaya. Data was prepared by standardizing StandardScaler and balancing it using a technique called SMOTE with 5-fold cross-validation. The stacking model uses predictions from LR and RF as meta-features to be learned by ANN. Results showed that the RF model had a recall of 77%, accuracy of 73%, and an ROC-AUC of 73%. The LR model had a recall of 71%, accuracy of 71%, and an ROC-AUC of 84%. Combining both models with stacking improved results, achieving an ROC-AUC of 85%, accuracy of 72%, and a peak recall of 81%. This shows that using an ANN-based stacking method can give better and sensitive results in finding stroke risk.

Copyright © SENASIF 2026
ISSN: 2598-0076

I. PENDAHULUAN

Menurut laporan terbaru dari World Stroke Organization tahun 2025, stroke masih menjadi penyebab utama kematian dan kecacatan kedua di seluruh dunia, dengan jumlah lebih dari 93.8 juta orang yang mengalami stroke dan sekitar 7 juta orang meninggal setiap tahunnya. Lebih dari 80% kasus stroke sebenarnya dapat dicegah melalui pengendalian faktor risiko seperti hipertensi, merokok, obesitas, diabetes, serta gaya hidup tidak sehat. Namun, kesenjangan besar dalam akses terhadap layanan pencegahan, diagnosis, dan rehabilitasi menyebabkan negara berpenghasilan rendah dan menengah menanggung hingga 89% beban global stroke. Kondisi ini menegaskan bahwa stroke bukan hanya persoalan medis, tetapi juga krisis sosial dan ekonomi yang memerlukan strategi global berbasis pencegahan, pemerataan layanan kesehatan, dan peningkatan kesadaran masyarakat [1].

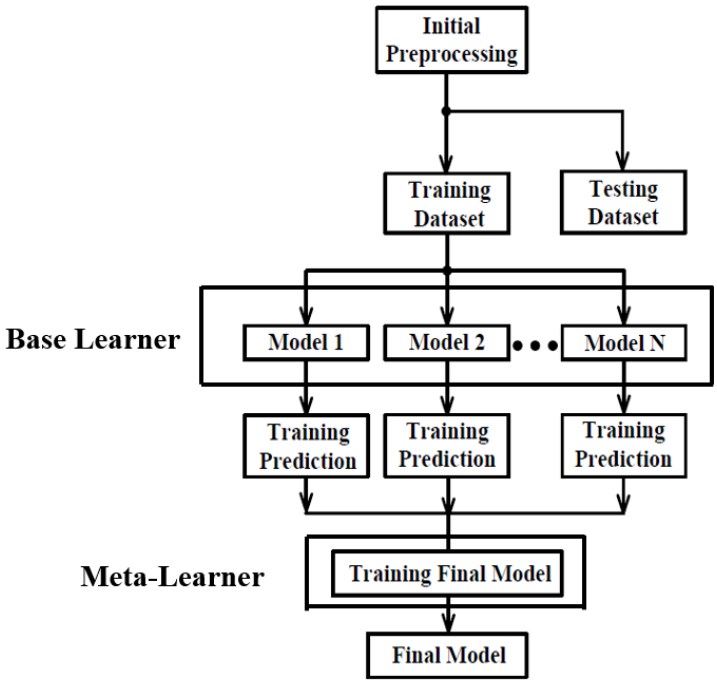
Secara jenis, stroke terbagi menjadi stroke iskemik yang mencakup 60% hingga 70% kasus global, serta perdarahan intraserebral (*Intracerebral Haemorrhage*) atau stroke hemoragik yang frekuensinya jauh lebih tinggi di negara berkembang. Perbedaan mendasar terletak pada bukti infark pada jaringan otak. *Acute Ischemic Stroke* (AIS) ditandai dengan disfungsi saraf mendadak disertai bukti infark pada pencitraan, sementara *Transient Ischemic Attack* (TIA) merupakan episode sementara tanpa bukti infark permanen. Deteksi dini sangat krusial karena pasien yang mengalami TIA memiliki risiko stroke berulang sebesar 7,5% hingga 17,4% dalam tiga bulan berikutnya jika tidak segera ditangani secara medis [2]. Kondisi ini menunjukkan perlunya penerapan prediksi berbasis *machine learning*, yang merupakan proses sistematis untuk mempelajari pola dari data historis pasien guna melakukan pengambilan keputusan klinis dengan tujuan untuk membantu tenaga medis tanpa aturan yang diprogram secara eksplisit. Analisis prediksi ini melibatkan ekstraksi fitur dari atribut klinis seperti usia, riwayat hipertensi, riwayat penyakit jantung, kadar gula darah, jenis kelamin, status nikah, jenis pekerjaan, jenis tempat tiggal, BMI, dan status merokok untuk memodelkan probabilitas terjadinya stroke pada individu secara objektif [3].

Penelitian ini akan menggunakan dua algoritma individual yang nanti hasil evaluasinya akan dibandingkan satu sama lain, model individual ini antara lain *Logistic Regression* dan *Random Forest*. *Logistic Regression* (LR) adalah model klasifikasi yang memprediksi probabilitas suatu kejadian dengan cara menghitung kombinasi linear dari fitur-fitur input seperti skor, lalu mengubah skor itu menjadi nilai 0–1. Karena outputnya berupa probabilitas, LR biasanya memodelkan logit, yaitu log perbandingan peluang (*odds*) kejadian terhadap tidak kejadian, sebagai fungsi linear dari fitur-fitur input. Model ini bersifat interpretable karena koefisien regresi dapat menjelaskan arah dan kekuatan pengaruh fitur. Alasan koefisien LR sangat interpretatif karena tanda koefisien menunjukkan apakah fitur menaikkan/menurunkan peluang kejadian, dan jika dieksponensialkan akan menghasilkan perbandingan rasio. Model dilatih dengan mencari koefisien yang memaksimalkan kecocokan data, dan untuk menghindari overfitting bisa ditambah dengan

Copyright © SENASIF 2026
ISSN: 2598-0076

regularisasi. Meski sederhana, LR kuat untuk data tabular dan baseline yang solid, asal hubungan fitur terhadap logit cukup masuk akal dan multikolinearitas tidak terlalu parah [4]. Dalam prediksi stroke, *Logistic Regression* digunakan untuk memodelkan probabilitas seseorang mengalami stroke (kelas 1) atau tidak mengalami stroke (kelas 0) berdasarkan atribut pasien seperti usia, hipertensi, kadar glukosa, BMI, riwayat penyakit jantung, dan kebiasaan merokok. Model ini sangat umum digunakan sebagai baseline model dalam penelitian ML medis karena sifatnya yang stabil, sederhana, dan mudah diinterpretasikan [5]. Random Forest merupakan algoritma ensemble bagging yang menggabungkan sejumlah decision tree atau pohon keputusan untuk meningkatkan stabilitas dan akurasi prediksi. Setiap pohon dibangun menggunakan subset data dan fitur secara acak untuk mengurangi overfitting [6]. Meskipun memiliki akurasi yang tinggi, Random Forest termasuk model black-box karena sulit menjelaskan mekanisme keputusan yang dihasilkan oleh puluhan hingga ratusan pohon [7]

Namun, muncul permasalahan di mana algoritma individual kadang dinilai kurang efektif karena terkendala oleh seberapa kompleks dataset klinis yang digunakan dan juga kelas yang tidak seimbang, di mana jumlah pasien sehat jauh lebih mendominasi dibandingkan pasien yang mengalami stroke. Hal ini dapat menyebabkan model menjadi bias dan gagal mengenali kasus positif secara optimal. Menjawab tantangan tersebut, penelitian ini mengusulkan solusi melalui pendekatan *Stacking Classifier*. *Stacking Classifier* adalah teknik ensemble learning yang bekerja dengan prinsip hierarkis yang menggabungkan beberapa algoritma dalam dua tingkat pembelajaran [8].



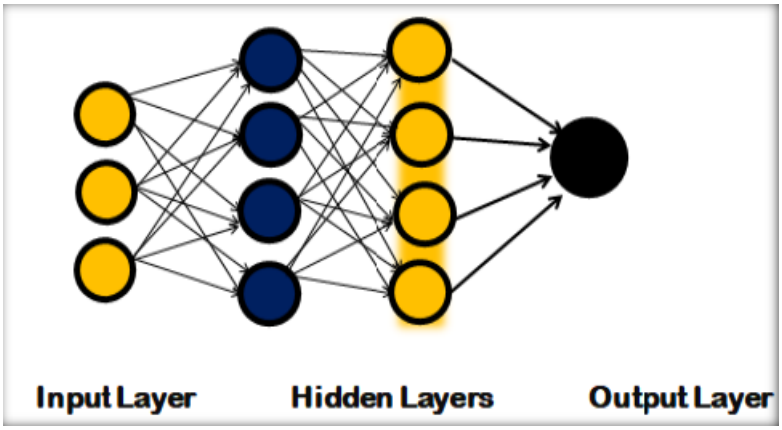
Gambar 1. Matriks Konfusi (Gupta et al., 2022)

Terlihat pada Gambar 1 pada tingkat awal, beberapa model dasar dilatih secara paralel untuk memproses data input. Selanjutnya, pada tingkat kedua, sebuah model tambahan yang disebut *meta-learner* akan mengambil output atau prediksi dari model-model dasar tersebut sebagai fitur input

Copyright © SENASIF 2026
ISSN: 2598-0076

baru (*meta-features*). Tugas utama dari meta-learner ini adalah mempelajari bagaimana cara terbaik untuk menggabungkan prediksi-prediksi tersebut guna menghasilkan klasifikasi akhir yang paling akurat [9]. Pada penelitian ini, pada tingkat pertama (*base-learners*), beberapa algoritma dengan karakteristik pembelajaran yang berbeda, yaitu *Logistic Regression*, dan *Random Forest*. Selanjutnya, hasil prediksi dari tingkat pertama ini tidak langsung menjadi hasil akhir, melainkan digunakan sebagai input baru untuk tingkat kedua, yaitu meta-learner yang berbasis *Artificial Neural Network* (ANN).

Artificial Neural Network (ANN) adalah kerangka arsitektur untuk komputasi yang terinspirasi dari cara sistem saraf manusia memproses informasi. ANN mensimulasikan cara neuron otak menginterpretasikan data untuk mendeteksi korelasi data yang kompleks [10].



Gambar 2. Struktural ANN (Qamar & Zardari, 2023)

Secara struktural, sistem saraf ini didefinisikan sebagai sekumpulan lapisan neuron yang saling terhubung seperti pada Gambar 2, di mana setiap neuron dapat direpresentasikan sebagai fungsi matematis yang bertugas mengumpulkan dan menganalisis informasi. Arsitektur ANN terdiri dari tiga kategori lapisan utama [11]. Pertama adalah lapisan input (Input Layer) berfungsi untuk mendefinisikan pola masukan/input. Lapisan input jaringan saraf tiruan terdiri dari kumpulan neuron input buatan. Neuron-neuron ini mengirimkan informasi dari lapisan neuron awal ke sistem untuk diproses [11]. Lapisan kedua adalah lapisan tersembunyi (Hidden Layer) terdiri dari lapisan input dan output dimana memiliki neuron buatan yang dibobot dari jumlah input [11]. Lapisan terakhir atau ketiga adalah lapisan output (Output Layer). Lapisan output bertugas memetakan informasi dari lapisan input ke dalam kategori klasifikasi yang dituju. Output prediksi akhir didapat dengan mengalikan fungsi aktivasi sigmoid dengan output yang didapat pada lapisan hidden layer [11]. Kelebihan utama penggunaan ANN dalam analisis prediksi adalah bisa melakukan pembelajaran adaptif, yaitu kempampuan untuk mempelajari tugas berdasarkan data latihan atau pengalaman terdahulu. Selain itu, ANN memiliki sifat organisir sendiri dan toleransi terhadap kesalahan, di mana kerusakan sebagian rangkaian tidak akan menghentikan keseluruhan fungsinya secara total [10].

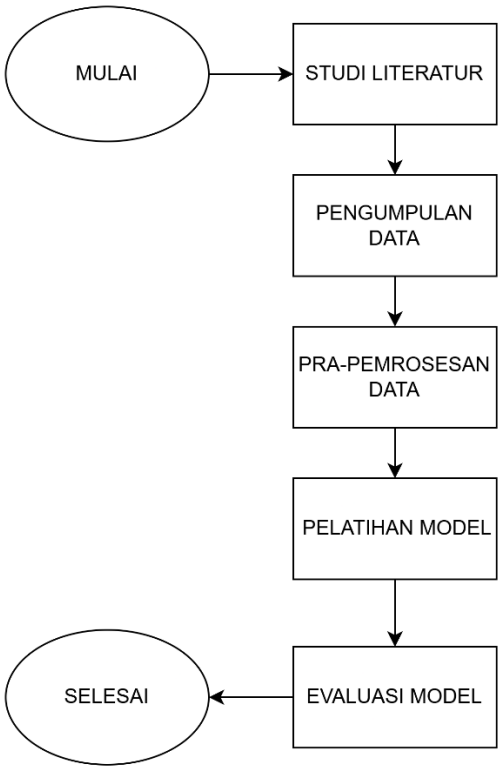
Rumusan masalah pertama dalam penelitian ini adalah bagaimana membangun dan mengoptimalkan model *Stacking Classifier* menggunakan *base learners* (LR, RF) dengan *meta-*

Copyright © SENASIF 2026
ISSN: 2598-0076

learner ANN untuk prediksi penyakit stroke. Selanjutnya . Untuk menjawab permasalahan tersebut, tujuan penelitian ini adalah untuk membangun dan mengoptimalkan model prediksi penyakit stroke menggunakan pendekatan *Stacking Classifier* yang menggabungkan algoritma *Logistic Regression* dan *Random Forest* sebagai *base learners* dengan *Artificial Neural Network* (ANN) sebagai *meta-learner*. Setelah model dibuat akan dilakukan proses evaluasi dan membandingkan performa model *Stacking Classifier* terhadap model individual dalam memprediksi risiko stroke berdasarkan metrik evaluasi yang didapatkan seperti akurasi, *precision*, *recall*, *F1-score* dan ROC-Score.

II. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen komputasional, yaitu membangun, melatih, dan mengevaluasi model machine learning stacking untuk memprediksi penyakit stroke berdasarkan data tabular pasien. Penelitian ini juga dilaksanakan melalui alur kerja sistematis yang dirancang untuk membangun model ensemble learning. Agar penelitian berjalan lancar dan terurut, dan membantu penulis dalam menyelesaikan penelitian ini, dibuat sebuah rancangan penelitian berbentuk diagram alir sebagai berikut pada Gambar 3.



Gambar 3. Rancangan Penelitian

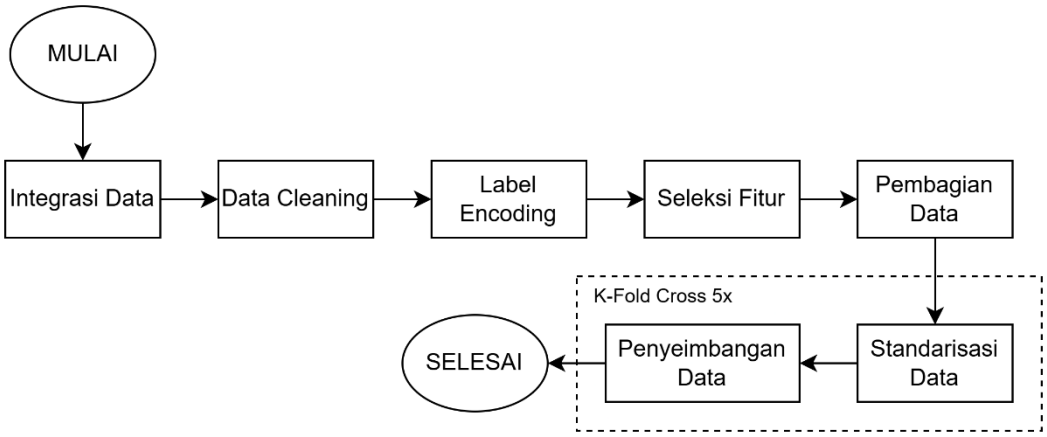
Alur penelitian yang terlihat pada Gambar 3 dimulai dengan riset pustaka untuk mencari gap penelitian yang sudah ada. Selanjutnya, dilakukan pengumpulan data untuk keperluan dataset yang nantinya akan dilakukan pra-premosesan data. Setelah itu penulis akan melatih model stacking

Copyright © SENASIF 2026
ISSN: 2598-0076

dengan menggabungkan base-learner (*Logistic Regression* dan *Random Forest*) dan *meta-learner* ANN. Setelah pelatihan selesai, pada tahap terakhir model akan diuji dan di evaluasi.

Untuk pengumpulan data yang akan dijadikan dataset pada penelitian ini, penulis memperoleh data dari dua sumber; data primer dan data sekunder. Data primer diambil dari list register pasien poliklinik yang didapat dari puskesmas di Surabaya. List berisi 4448 data tabular pasien yang berisi nama, alamat, kota asal, NIK, umur, tanggal lahir, jenis kelamin, pendidikan terakhir, pekerjaan, tipe pelayanan, status merokok, keluhan, klasifikasi penyakit, alkohol, gula darah, kolesterol, asam urat, berat badan, tinggi badan, tensi, respiratory rate, detak jantung, temperatur badan, dan keterangan administrasi lainnya.

Sebelum data dimasukkan ke dalam model untuk proses training, data mentah yang telah terkumpul dilakukan tahap pra-pemrosesan dengan tujuan untuk menyiapkan dan membersihkan data agar hasil prediksi tidak menyesatkan karena hal-hal seperti nilai yang hilang, duplikasi, atau format yang tidak seragam. Berikut pada Gambar 4 penulis telah membuat bagan alur tahap pra-pemrosesan yang akan dilakukan pada penelitian ini.



Gambar 4. Workflow Pra-pemrosesan Data

Setelah Integrasi data, data cleaning, dan label encoding data dilakukan pemilihan fitur yang bertujuan untuk menentukan atribut atau variabel prediktor yang paling relevan dan memiliki kontribusi signifikan terhadap target prediksi (kejadian stroke). Dengan mereduksi fitur yang tidak informatif, efisiensi komputasi model *Stacking Classifier* dapat ditingkatkan dan risiko *overfitting* dapat diminimalisir [12]. Pemilihan fitur dilakukan berdasarkan tinjauan pustaka medis mengenai faktor risiko utama stroke dan juga divalidasi menggunakan metode *information gain* dari nilai entropi. Nilai entropi merupakan sebuah ukuran statistik yang digunakan untuk mendeteksi tingkat ketidakpastian informasi di dalam dataset

Atribut yang menghasilkan penurunan entropi paling signifikan atau memiliki nilai gain tertinggi dianggap sebagai variabel prediktor yang paling efektif karena mampu memberikan kontribusi informasi yang besar dalam proses klasifikasi [13]. Dalam konteks penelitian ini, fitur dengan nilai IG (*Information Gain*) yang tinggi menunjukkan kekuatan diskriminatif yang kuat untuk

Copyright © SENASIF 2026
ISSN: 2598-0076

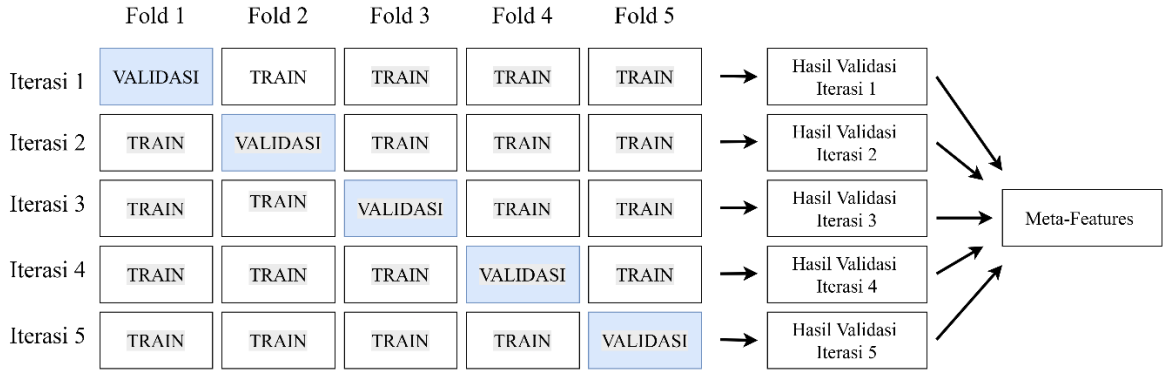
membedakan kondisi pasien secara akurat, sehingga fitur tersebut diprioritaskan sebagai input utama bagi model ML guna menjamin stabilitas akurasi dan efisiensi komputasi [12]. Jika misal ada fitur lain, misalnya status merokok dan setelah dihitung ternyata IG-nya adalah 0.001, maka status merokok bisa dibuang namun tentu saja akan ada pustaka medis yang mendukungnya.

Tahap selanjutnya setelah seleksi fitur adalah pembagian data. Hal ini dilakukan untuk memisahkan dataset menjadi dua bagian, yaitu data latih (*training set*) dan data uji (testing set). Pemisahan ini sangat krusial dalam penelitian machine learning untuk mensimulasikan performa model saat menghadapi data baru yang belum pernah dilihat sebelumnya (*unseen data*) [14]. Penelitian ini menggunakan proporsi 80% untuk data latih dan 20% untuk data uji. Data latih digunakan oleh model *Stacking Classifier* (*Logistic Regression, Random Forest*, dan ANN) untuk mempelajari pola fitur klinis terhadap risiko stroke. Sementara itu, data uji sebesar 20% akan disimpan dan hanya digunakan pada tahap akhir evaluasi untuk mengukur metrik akurasi, recall, dan presisi. Data uji akan dilakukan standarisasi sebelum digunakan untuk proses pengujian model stacking. Mengingat dataset penyakit stroke cenderung memiliki ketidakseimbangan kelas (*imbalanced data*), pembagian data dilakukan menggunakan teknik stratified sampling. Teknik ini menjamin bahwa proporsi label target (Stroke dan Non-Stroke) pada data latih maupun data uji tetap konsisten dengan proporsi pada dataset asli. Hal ini penting agar evaluasi model tetap objektif dan tidak bias terhadap kelas mayoritas. Selanjutnya, data latih sebesar 80% tersebut akan diolah lebih lanjut menggunakan metode *K-Fold Cross Validation* pada tahap pelatihan model stacking untuk menghasilkan prediksi *out-of-fold* bagi *meta-learner* ANN serta memastikan validitas proses pra-pemrosesan di dalam setiap lipatan (fold).

Tahap pelatihan model merupakan inti dari penelitian ini, di mana algoritma *machine learning* dilatih untuk mengenali pola-pola klinis yang mengindikasikan risiko stroke. Penelitian ini menggunakan metode *ensemble stacking*, yaitu teknik penggabungan beberapa model klasifikasi untuk meningkatkan performa prediksi dibandingkan hanya menggunakan satu model individual [15]. Arsitektur *stacking* yang dibangun terdiri dari dua lapisan utama, *base learners* dan *meta learner*. Model dasar atau base learners merupakan lapisan pertama (level 0) dalam arsitektur pembelajaran mesin berbasis stacking classifier. Pada tahap ini, beberapa algoritma klasifikasi yang berbeda digunakan untuk mempelajari pola dari dataset yang sama secara independen. Tujuan utama penggunaan berbagai base learners adalah untuk memanfaatkan keunggulan masing-masing algoritma dalam menangani variasi data, sehingga kesalahan klasifikasi yang dilakukan oleh satu model dapat dikompensasi oleh model lainnya [8]. Pada tahap ini, beberapa algoritma klasifikasi yang berbeda digunakan untuk mempelajari pola dari dataset yang sama secara independen. Penelitian ini menggunakan dua algoritma heterogen, *Logistic Regression* dan *Random Forest*.

Copyright © SENASIF 2026
ISSN: 2598-0076

Penerapan *K-Fold Cross-validation* pada lapisan *base learners* dilakukan khusus pada dataset latih (training data) dengan tujuan utama untuk menghasilkan *meta-features* yang objektif dan menghindari masalah *data leakage*. Tanpa proses ini, meta-learner (ANN) berisiko mengalami overfitting karena akan belajar dari prediksi b*ase learners* yang sudah mengenal data tersebut selama proses pelatihan [16]. Untuk workflow proses *cross-validation* yang dilakukan penulis untuk ketiga *base-learner* (*Logistic Regression*, *Random Forest*, *XgBoost*) bisa dilihat pada Gambar 5.



Gambar 5. Workflow Tahap K-Fold Cross-Validation

Dataset latih dibagi menjadi k subset (fold), dalam penelitian ini akan menjadi fold k = 5. Untuk setiap iterasi i hingga k, satu bagian fold disimpan sebagai data validasi sementara (*out-of-fold)* untuk setiap iterasi. Proses ini menggunakan 20 persen dari keseluruhan data latih. Teknik SMOTE dan scaling diterapkan secara identik di dalam setiap fold training. Kedua model dasar (*Logistic Regression* dan *Random Forest*) dilatih menggunakan fold bagian data sisanya untuk setiap iterasi, dalam penelitian ini karena menggunakan 5 subnet fold maka 4 fold digunakan training. Proses ini menggunakan 80 persen dari keseluruhan data latih. Setelah model dilatih, untuk tiap iterasi, model tersebut melaukukan prediksi pada fold validasi dengan tujuan untuk validasi data. Hasil validasi/prediksi ini disimpan sebagai bagian dari fitur baru/*meta-features*. Keseluruhan proses jika dijumlahkan maka akan ada 10 hasil validasi. Hasil prediksi dari seluruh iterasi digabungkan kembali sehingga membentuk *meta-features* yang memiliki ukuran baris yang sama dengan dataset latih asli [9]. Langkah-langkah sebelumnya diulangi lagi untuk setiap base-learner sehingga akan ada tiga meta-features yang akan digunakan untuk pelatihan ANN.

Meta-learner merupakan lapisan kedua (level 1) dalam arsitektur *stacking classifier* yang bertugas untuk mengombinasikan hasil prediksi dari berbagai *base learners*. Pada penelitian ini, *Artificial Neural Network* (ANN) dipilih sebagai meta-learner karena kemampuannya dalam mempelajari hubungan non-linear yang kompleks di antara prediksi-prediksi yang dihasilkan pada level 0 [10]. Proses pada tahap ini tidak lagi menggunakan fitur medis mentah, melainkan menggunakan *meta-features* yang telah dibentuk melalui mekanisme *K-Fold Cross-validation* pada tahap sebelumnya [9].

Setelah tahap pelatihan model selesai, tahap terakhir akan dilakukan tahapan pengujian dan evaluasi model yang bertujuan untuk menguji hasil output prediksi dan mengukur kinerja model yang

Copyright © SENASIF 2026
ISSN: 2598-0076

telah dilatih. Setiap pengujian model, base maupun meta, akan menggunakan data testing/validasi sebesar 20% yang telah dipisahkan sebelumnya untuk menguji output dari model yang telah dibangun. Data testing digunakan agar model diuji menggunakan data yang belum tersentuh sama sekali sehingga output pengujian yang dihasilkan tidak sama dengan output pada saat pembuatan model (training). Pada level *base-learner* yang melalui proses k-*fold Cross Validation*, pengujian model dilakukan pada output pelatihan model base menggunakan data 4 fold (80% dari data training) yang telah dilakukan standarisasi / *Standard Scaler* dan penyeimbangan data. 1 Fold validasi (20% dari data training) digunakan untuk menguji hasil output dari model base-learners tersebut. Proses pengujian ini dilakukan pada tiap iterasi (5 iterasi). Pada level meta-learner, pengujian model ANN dilakukan menggunakan *meta-features* dari model base-learners fit yang merupakan model yang dibuat dari keseluruhan model data training (tanpa proses *k-fold Cross Validation*) dan diuji menggunakan data testing yang sudah terstandarisasi. Hasil prediksi model menggunakan threshold (ambang batas prediksi) 50%, ini berarti jika angka kepercayaan model diatas ambang batas tersebut maka model memprediksi stroke (1) jika data dimasukkan.

III. HASIL DAN PEMBAHASAN

Sebelum data diolah dilakukan integrasi data untuk menggabungkan kedua dataset (*primer* dan *sekunder)*. Dataset yang telah digabung memiliki jumlah 8518 baris. Pada penitian ini dilakukan tahapan pra-preprocessing pada dataset yang sudah didapatkan dan digabung sebelum pelatihan model dan evaluasi model dimana dilakukan agar data yang dipakai menjadi bersih dari data kosong, ganda, dan sebagainya. Baris yang memliki atribut ganda dicek menggunakan library pandas dan dihapus. Total data setelah pembersihan adalah 8316 baris. Setelah pembersihan data, jumlah data baru berjumlah 8105 baris.

Setelah itu dilakukan label encoding yang termasuk dalam transformasi data kategorikal agar data bisa dimasukkan kedalam algoritma *machine learning*. Sebelum pembagian data, dilakukan seleksi fitur menggunakan teknik *information gain*. Didapatkan hasil IG untuk setiap atribut pada Tabel 2 berikut.

Tabel 1. Score Information Gain Atribut Sebelum Seleksi

Atribut	Information Gain Score
age	0.030109
hearth_disease	0.014561
ever_married	0.012443
hypertension	0.009883
BMI	0.008507
work_type	0.007335
smoking_status	0.005967

Copyright © SENASIF 2026
ISSN: 2598-0076

avg_glucose_level	0.004124
residence_type	0.001558
gender	0.000000

Terlihat pada Tabel 2 skor *information gain* atribut gender bernilai 0, ini memiliki arti bahwa atribut gender tidak berpengaruh sama sekali terhadap prediksi stroke, sehingga agar meningkatkan efisiensi komputasi maka atribut ini akan dihapus. Tabel 3 berikut adalah score *information gain* setelah atribut gender dihapus.

Tabel 2. Score Information Gain Atribut Setelah Seleksi

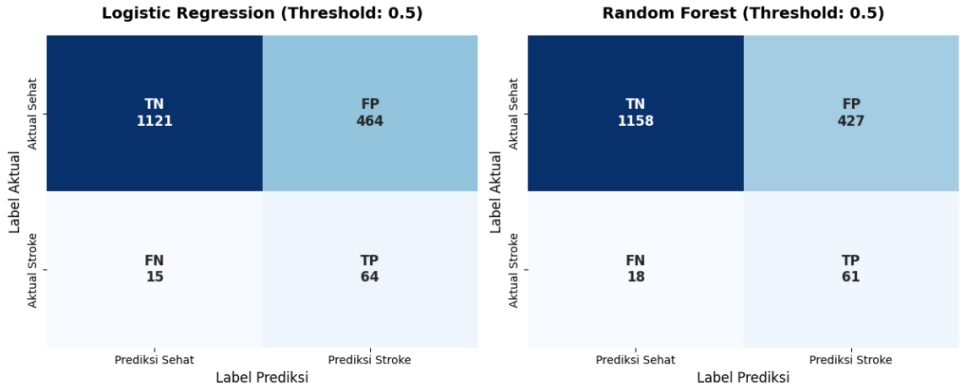
Atribut	**Information Gain Score**
age	0.028505
hearth_disease	0.016757
ever_married	0.012064
work_type	0.009479
BMI	0.009233
hypertension	0.006880
residence_type	0.004687
avg_glucose_level	0.004069
smoking_status	0.001241

Setelah itu dilakukan pembagian data dimana 80% dijadikan data training (6653 baris) dan 20% dijadikan data testing (1663 baris). Selanjutnya telah dilakukan *K-Fold Cross Validation* sebanyak 5 kali untuk dataset *training*. Setelah itu telah dilakukan SMOTE dan standarisasi pada setiap fold yang telah dibuat. Setiap fold dibangun model *base-learner logistic regression* dan *random* f*orest* yang hasil akhirnya berupa penggabungan setiap hasil validasi dalam bentuk *meta-features training*. Setelah itu dibuat lagi model baru individual (*final fit*) tanpa proses SMOTE yang digunakan untuk proses evaluasi model individual dan dijadikan *meta-features* untuk *testing*. Model *final fit* ini menggunakan data t*raining* yang sama dan diujikan menggunakan data uji. Menggunakan *meta-features training* dan *testing* tersebut, dibangun model stacking dengan *meta-learner Artificial Neural Network* yang dimana menggabungkan *meta features Logistic Regresion* dan *Random Forest*. Selama proses tuning penulis menyadari bahwa proses dataset yang diggunakan tidak seimbang meski sudah diberi SMOTE, sehingga diterapkan *class weight* (beban kelas) untuk memperbagus hasil model.

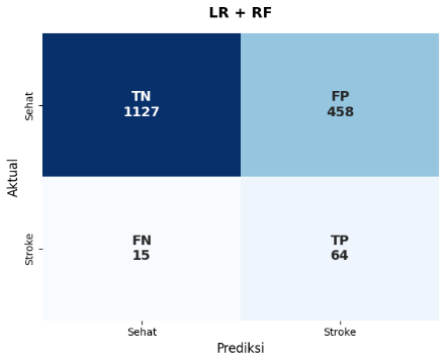
Proses terakhir adalah proses evaluasi dimana akan dilihat matriks evaluasi setiap model. Pada Tabel 4 berikut adalah proses hasil evaluasi untuk setiap model. Namun, sebelum menghitung matriks evaluasi, penulis telah membuat matriks konfusi untuk setiap model pada Gambar 8 dan

Copyright © SENASIF 2026
ISSN: 2598-0076

Gambar 6. Proses evaluasi selanjutnya dilakukan dan penulis telah membuat tabel hasil evaluasinya pada Tabel 3.



Gambar 6. Matriks Konfusi Base-Learner



Gambar 7 Matriks Konfusi Meta-Learner

Berdasarkan matriks konfusi pada Gambar 6 dan Gambar 7 tersebut diambil beberapa kesimpulan :

1. Logistic Regression (LR)
    - True Negative (TN) = 1121: Model berhasil menebak dengan benar 1.121 orang yang aslinya memang Sehat.
    - True Positive (TP) = 64: Model berhasil mendeteksi dengan benar 64 pasien yang aktualnya mengalami Stroke.
    - False Positive (FP) = 464 : Model memprediksi 464 orang terkena stroke, padahal aslinya mereka sehat.
    - False Negative (FN) = 15 : Model gagal mendeteksi 15 pasien stroke karena diprediksi sehat.
2. Random Forest (RF)
    - True Negative (TN) = 1158: Model berhasil menebak dengan benar 1.158 orang yang memang Sehat.
    - True Positive (TP) = 61: Model berhasil mendeteksi 61 pasien Stroke.
    - False Positive (FP) = 427 : Model memprediksi 427 orang terkena stroke, padahal aslinya sehat.

Copyright © SENASIF 2026
ISSN: 2598-0076

- False Negative (FN) = 18 : Model gagal mendeteksi 18 pasien stroke.

3. Stacking Artifical Neural Network (LR + RF)
    - True Negative (TN) = 1127: Model berhasil menebak dengan benar 1.127 orang yang aslinya memang Sehat.
    - True Positive (TP) = 64: Model berhasil mendeteksi dengan benar 64 pasien yang aktualnya mengalami Stroke.
    - False Positive (FP) = 458 : Model memprediksi 458 orang terkena stroke, padahal aslinya mereka sehat.
    - False Negative (FN) = 15 : Model gagal mendeteksi 15 pasien stroke karena diprediksi sehat.

Tabel 3. Hasil Evaluasi Model

Model	Akurasi Total	ROC-AUC	Presisi (0)	Recall (0)	F1-Score (0)	Presisi (1)	Recall (1)	F1-Score (1)
LR	71%	84%	99%	70%	82%	12%	81%	21%
RF	73%	83%	98%	73%	84%	12%	77%	21%
Stacking ANN (LR + RF)	72%	85%	99%	71%	83%	12%	81%	21%

Berdasarkan hasil tersebut, dapat diambil bahwa Stacking ANN Menjadi Model Paling Optimal (berdasarkan ROC-AUC). Metrik ROC-AUC menunjukkan seberapa baik model secara keseluruhan mampu membedakan antara pasien sehat (0) dan pasien stroke (1). Model Stacking ANN (LR + RF) berhasil mencapai skor ROC-AUC tertinggi, yaitu 85%. Ini membuktikan bahwa meta-learner ANN berhasil menggabungkan pola atau suara terbaik dari kedua *base-learner* (LR di 84% dan RF di 83%) untuk menghasilkan batas keputusan yang lebih bagus. Dalam kasus penyakit mematikan seperti stroke, kita ingin menekan angka *False Negative* sekecil mungkin (pasien stroke yang salah dibilang sehat). Oleh karena itu, metrik Recall (1) adalah yang paling krusial. *Logistic Regression* (LR) sangat sensitif dengan Recall (1) sebesar 81%, namun *Random Forest* (RF) turun menjadi 77%. Kehebatan *stacking* ANN di sini adalah kemampuannya menambal kelemahan RF dan mengambil kelebihan LR. Model Stacking sukses mempertahankan sensitivitas deteksi di angka maksimal yaitu Recall (1) sebesar 81%. Jika hanya melihat akurasi total, model RF terlihat paling bagus (73%). Namun, akurasi tinggi tersebut sebagian besar disumbang dari kemampuannya menebak pasien yang memang sehat (Recall 0: 73%). Sebagai gantinya, RF paling sedikit dalam mendeteksi pasien yang benar-benar stroke. Semua model menunjukkan nilai Presisi (1) yang hampir sama dan sangat rendah, yaitu 12% (diikuti F1-Score 1 sebesar 21%), artinya, dari 100 orang yang diprediksi beresiko stroke oleh sistem, faktanya hanya 12 orang yang benar-benar stroke.

Copyright © SENASIF 2026
ISSN: 2598-0076

IV. SIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa penggunaan model gabungan bertingkat terbukti lebih unggul dan optimal dibandingkan dengan model prediksi yang berdiri sendiri dalam mendeteksi risiko penyakit stroke. Pendekatan metode gabungan ini berhasil menutupi kelemahan dari masing-masing model tunggal dan mengambil pola terbaiknya, sehingga menghasilkan sebuah sistem yang jauh lebih peka dalam mengenali pasien yang benar-benar mengalami stroke. Tingkat kepekaan yang tinggi ini sangat krusial dalam dunia medis guna menekan angka kegagalan diagnosis yang berisiko fatal.

Meskipun model yang dibangun terkesan sangat sensitif sehingga sering memberikan peringatan palsu kepada pasien yang sebenarnya sehat, perilaku ini dinilai wajar dan sangat dibenarkan secara klinis. Dalam konteks keselamatan nyawa pasien, jauh lebih baik model memberikan alarm peringatan agar pasien melakukan pemeriksaan lebih lanjut daripada model gagal mendeteksi dan membiarkan pasien stroke tanpa penanganan.

DAFTAR RUJUKAN

[1] V. L. Feigin *et al.*, “World Stroke Organization: Global Stroke Fact Sheet 2025,” Feb. 01, 2025, *SAGE Publications Inc.* doi: 10.1177/17474930241308142.

[2] N. A. Hilkens, B. Casolla, T. W. Leung, and F. E. de Leeuw, “Stroke,” Jun. 29, 2024, *Elsevier B.V.* doi: 10.1016/S0140-6736(24)00642-1.

[3] I. Werdiningsih *et al.*, “Analisis Prediksi Stroke Menggunakan Pendekatan Decision Tree dengan Seleksi Fitur dan Neural Network,” 2023.

[4] Kehinde Josephine Olowe, Ngozi Linda Edoh, Stephane Jean Christophe Zouo, and Jeremiah Olamijuwon, “Comprehensive review of logistic regression techniques in predicting health outcomes and trends,” *World Journal of Advanced Pharmaceutical and Life Sciences*, vol. 7, no. 2, pp. 016–026, Nov. 2024, doi: 10.53346/wjapls.2024.7.2.0039.

[5] M. A. Vk, M. Danish, and A. Nf, “ICCICT-057 STROKE PREDICTION USING EXPLAINABLE ARTIFICIAL INTELLIGENCE (XAI).”

[6] H. A. Salman, A. Kalakech, and A. Steiti, “Random Forest Algorithm Overview,” Dec. 15, 2024, *Mesopotamian Academic Press*. doi: 10.58496/BJML/2024/007.

[7] A. Murtovi, M. Schlüter, and B. Steffen, “An Efficient Compilation-Based Approach to Explaining Random Forests Through Decision Trees,” in *International Conference on Agents and Artificial Intelligence*, Science and Technology Publications, Lda, 2025, pp. 484–495. doi: 10.5220/0013188600003890.

Copyright © SENASIF 2026
ISSN: 2598-0076

[8] A. Gupta, V. Jain, and A. Singh, “Stacking Ensemble-Based Intelligent Machine Learning Model for Predicting Post-COVID-19 Complications,” *New Gener. Comput.*, vol. 40, no. 4, pp. 987–1007, Dec. 2022, doi: 10.1007/s00354-021-00144-0.

[9] M. Barton and B. Lennox, “Model stacking to improve prediction and variable importance robustness for soft sensor development,” *Digital Chemical Engineering*, vol. 3, Jun. 2022, doi: 10.1016/j.dche.2022.100034.

[10] R. Qamar and B. A. Zardari, “Artificial Neural Networks: An Overview,” Jan. 15, 2023, *Mesopotamian Academic Press*. doi: 10.58496/MJCSC/2023/015.

[11] O. A. Montesinos López, A. Montesinos López, and J. Crossa, “Fundamentals of Artificial Neural Networks and Deep Learning,” in *Multivariate Statistical Machine Learning Methods for Genomic Prediction*, Springer International Publishing, 2022, pp. 379–425. doi: 10.1007/978-3-030-89010-0_10.

[12] D. Theng and K. K. Bhoyar, “Feature selection techniques for machine learning: a survey of more than two decades of research,” Mar. 01, 2024, *Springer Science and Business Media Deutschland GmbH*. doi: 10.1007/s10115-023-02010-5.

[13] R. Lamba, T. Gulati, and A. Jain, “A Hybrid Feature Selection Approach for Parkinson’s Detection Based on Mutual Information Gain and Recursive Feature Elimination,” *Arab. J. Sci. Eng.*, vol. 47, no. 8, pp. 10263–10276, Aug. 2022, doi: 10.1007/s13369-021-06544-0.

[14] A. Jude and J. Uddin, “Explainable Software Defects Classification Using SMOTE and Machine Learning,” *Annals of Emerging Technologies in Computing*, vol. 8, no. 1, pp. 35–49, 2024, doi: 10.33166/AETiC.2024.01.00.

[15] D. Ahmed Muhtasim *et al.*, “Dense Stack Meta Ensemble Classifier (DSMEC): An Advanced Stacking Ensemble Approach for Early Cardiovascular Disease Prediction With Robust Feature Engineering,” *IEEE Access*, vol. 13, pp. 148735–148752, 2025, doi: 10.1109/ACCESS.2025.3598285.

[16] R. Bin Tareaf, A. M. Korga, S. Wefers, and K. Hanken, “Direct vs. Cross-Validated Stacking in Ensemble Learning: Evaluating the Trade-Off between Inference Time and Generalizability on Fashion-MNIST,” in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Feb. 2024, pp. 453–461. doi: 10.1145/3651671.3651764.

Copyright © SENASIF 2026
ISSN: 2598-0076