

**OPTIMALISASI KLASIFIKASI PENYAKIT STROKE
MENGUNAKAN ALGORITMA GRADIENT BOOSTING
DAN RANDOM FOREST DENGAN TEKNIK SMOTE**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi SI-Informatika



disusun oleh

VIKA NUR SAFITRI

21.11.3918

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**OPTIMALISASI KLASIFIKASI PENYAKIT STROKE
MENGUNAKAN ALGORITMA GRADIENTING BOOST
DAN RANDOM FOREST DENGAN TEKNIK SMOTE**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi SI-Informatika



disusun oleh

VIKA NUR SAFITRI

21.11.3918

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**OPTIMALISASI KLASIFIKASI PENYAKIT STROKE
MENGUNAKAN ALGORITMA GRADIENT BOOSTING
DAN RANDOM FOREST DENGAN TEKNIK SMOTE**

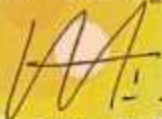
yang disusun dan diajukan oleh

VIKA NUR SAFITRI

21.11.3918

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 24 Februari 2025

Dosen Pembimbing,



Kusnawi, S.Kom, M. Eng.

NIK. 190302112

HALAMAN PENGESAHAN

SKRIPSI

**OPTIMALISASI KLASIFIKASI PENYAKIT STROKE
MENGUNAKAN ALGORITMA GRADIENT BOOSTING
DAN RANDOM FOREST DENGAN TEKNIK SMOTE**

yang disusun dan diajukan oleh

VIKA NUR SAFITRI

21.11.3918

Telah dipertahankan di depan Dewan Penguji
pada tanggal 24 Februari 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Rizqi Sukma Kharisma, S.Kom., M.Kom.
NIK. 190302215

Ike Verawati, S.Kom., M.Kom.
NIK. 190302237

Kusnawi, S.Kom., M.Eng.
NIK. 190302112



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 24 Februari 2025

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Vika Nur Safitri
NIM : 21.11.3918

Menyatakan bahwa Skripsi dengan judul berikut:

OPTIMALISASI KLASIFIKASI PENYAKIT STROKE MENGGUNAKAN ALGORITMA GRADIENT BOOSTING DAN RANDOM FOREST DENGAN TEKNIK SMOTE

Dosen Pembimbing : Kusnawi, S.Kom, M. Eng.

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, 24 Februari 2025

Yang Menyatakan,



Vika Nur Safitri

HALAMAN PERSEMBAHAN

Puji syukur penulis panjatkan ke hadirat Allah SWT yang telah melimpahkan kesehatan, rahmat, dan hidayah-Nya, sehingga penulis diberikan kekuatan untuk menyelesaikan laporan skripsi ini sebagai salah satu syarat untuk memperoleh gelar sarjana. Meskipun skripsi ini masih jauh dari kesempurnaan, penulis merasa bangga dapat mencapai tahap ini. Dengan penuh rasa syukur dan penghargaan, skripsi ini penulis persembahkan dengan tulus kepada:

1. Kedua Orang Tua Tercinta, terutama mama saya yang selalu memberikan cinta, doa, dukungan, dan pengorbanan tanpa henti. Terima kasih atas kesabaran, semangat, dan motivasi yang tiada pernah surut. Semua pencapaian ini tak lepas dari kasih sayang dan dukungan yang luar biasa.
2. Keluarga Besar yang selalu memberikan dukungan moral dan semangat selama proses pendidikan hingga skripsi ini terselesaikan.
3. Dosen Pembimbing, yang dengan sabar dan penuh dedikasi memberikan bimbingan, arahan, dan ilmu yang sangat berarti selama masa studi hingga penulisan skripsi ini.
4. Teman - teman yang telah memberikan dukungan, semangat, dan kebersamaan, baik dalam suka maupun duka selama menjalani masa perkuliahan.
5. Terima kasih kepada Universitas Amikom yang telah memberikan kesempatan dan fasilitas yang luar biasa untuk melanjutkan studi dan menyelesaikan penelitian ini.

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas limpahan rahmat, karunia, dan hidayahNya sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Skripsi ini diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana di Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta. Pada kesempatan ini, penulis menyampaikan rasa terima kasih yang sebesar besarnya kepada:

1. Bapak Prof. Dr. Mohammad Suyanto, M.M., selaku Rektor Universitas AMIKOM Yogyakarta, yang telah memberikan kesempatan dan fasilitas untuk melaksanakan studi di universitas ini.
2. Bapak Hanif Al Fatta, S. Kom., M.Kom., Ph.D., selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta, atas dukungan dan arahnya selama proses perkuliahan.
3. Ibu Windha Mega Pradnya Duhita, S. Kom., M.Kom., selaku Ketua Program Studi S1 Informatika Universitas AMIKOM Yogyakarta, atas bimbingan dan motivasinya selama penulis menjalani masa studi.
4. Bapak Kusnawi, S. Kom, M. Eng. selaku Dosen Pembimbing yang telah sabar memberikan bimbingan, masukan, dan arahan hingga skripsi ini dapat terselesaikan.
5. Tim Dosen Penguji, atas evaluasi, saran, dan kritik yang membangun untuk menyempurnakan skripsi ini.

Akhir kata, semoga Allah SWT senantiasa melimpahkan rahmat dan karuniaNya serta membalas segala amal baik dan kebaikan dari semua pihak yang telah membantu penulis dalam menyelesaikan skripsi ini.

Yogyakarta, 6 Januari 2025

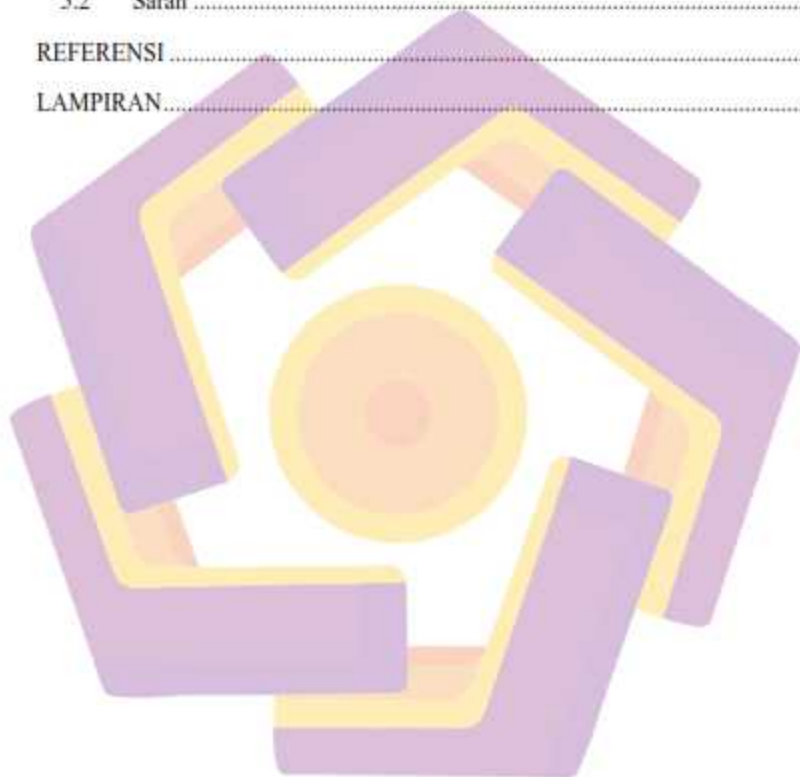
Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN.....	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xiv
INTISARI	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5

2.2	Dasar Teori.....	11
2.2.1	Penyakit stroke.....	11
2.2.2	<i>Machine Learning</i>	11
2.2.3	<i>Random Forest</i>	12
2.2.4	<i>Gradient boosting</i>	13
2.2.5	<i>Synthetic Minority Oversampling Technique (SMOTE)</i>	13
2.2.6	<i>Hyperparameter Tuning</i>	14
BAB III METODE PENELITIAN.....		17
3.1	Objek Penelitian.....	17
3.2	Alur Penelitian.....	17
3.2.1	Dataset.....	18
3.2.2	EDA.....	19
3.2.3	Preprocessing Data.....	21
3.2.4	Split Data.....	23
3.2.5	Klasifikasi 2 Algoritma.....	23
3.2.6	<i>Hyperparameter Tuning dan K-fold</i>	23
3.2.7	Metode Evaluasi.....	25
3.3	Alat dan Bahan.....	25
BAB IV HASIL DAN PEMBAHASAN.....		27
4.1	Dataset.....	27
4.2	EDA.....	27
4.3	Preprocessing.....	32
4.4	Klasifikasi.....	37
4.5	Hyperparameter Tuning dan K fold.....	39
4.6	Evaluasi.....	41

4.8	Implementasi Website.....	44
4.7	Perbandingan Penelitian.....	45
BAB V PENUTUP		47
5.1	Kesimpulan	47
5.2	Saran	47
REFERENSI		51
LAMPIRAN.....		55



DAFTAR TABEL

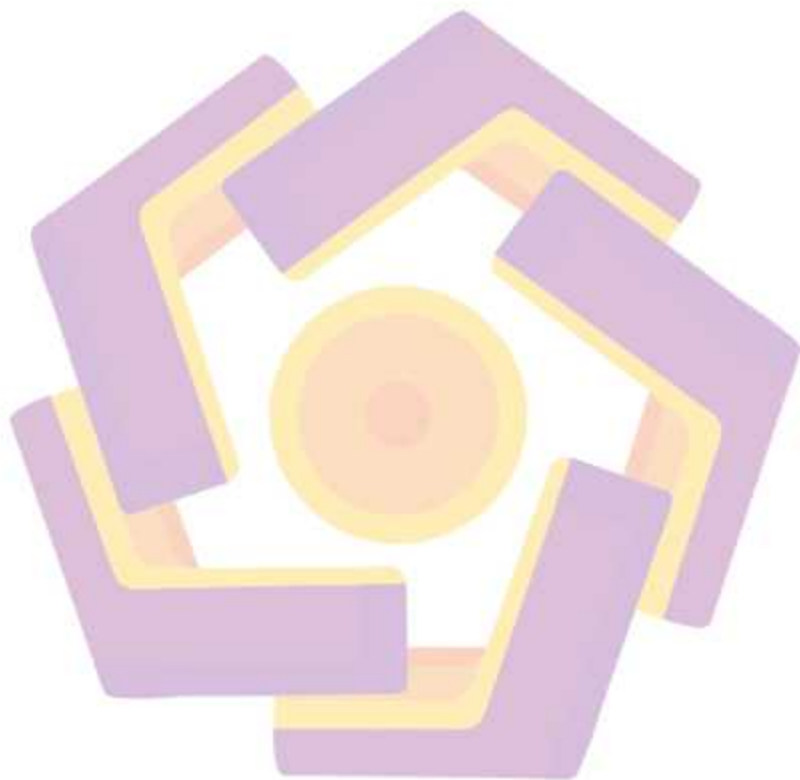
Tabel 2.1 Keaslian Penelitian	7
Tabel 2.2 Matrix.....	15
Tabel 3.1 Deskripsi Dataset	19
Tabel 3.2 Informasi Dataset	20
Tabel 3.3 Tipe Data.....	20
Tabel 3.4 Deskripsi	21
Tabel 3.5 Missing Values.....	21
Tabel 3.6 Imbalanced Data	23
Tabel 3.7 Parameter Random Forest dan Gradient Boost.....	24
Tabel 4.1 Kode Tipe data	27
Tabel 4.2 Tipe data	27
Tabel 4.3 Kode heatmap correlation matrix.....	28
Tabel 4.4 Kode Visualisasi Data.....	28
Tabel 4. 5 Kode Cek Outlier	30
Tabel 4.6 Kode Pie Chart.....	31
Tabel 4.7 Kode Data Cleaning.....	32
Tabel 4.8 Kode Winsorizing	33
Tabel 4.9 Kode Label Encoder	35
Tabel 4.10 Kode Smote.....	35
Tabel 4.11 Kode Split dan Klasifikasi	37
Tabel 4.12 Hasil Split data.....	38
Tabel 4.13 Hasil Akurasi	38
Tabel 4.14 Kode Hyperparameter dan Kfold.....	39
Tabel 4.15 Nilai parameter tuning	40
Tabel 4. 16 Hasil Parameter RandomizedSearchCV	40
Tabel 4.17 Hasil Hyperparameter Tuning dan Kfold	41
Tabel 4.18 Perbandingan Penelitian	45

DAFTAR GAMBAR

Gambar 2.1 K-fold	Error! Bookmark not defined.
Gambar 3.1 Alur Penelitian	18
Gambar 4.1 Heatmap correlation matrix	28
Gambar 4.2 Scater Plot Age dan Avg Glucose	29
Gambar 4.3 Scater Plot Age dan BMI	29
Gambar 4.4 Scater Plot Avg Glucose dan BMI	30
Gambar 4.5 Box Plot Outlier	31
Gambar 4.6 Pie Chart Kelas Stroke	31
Gambar 4.7 Hasil Heandling Missing Values	33
Gambar 4.8 Hasil Winsorizing	34
Gambar 4.9 Hasil Label Encoder	35
Gambar 4.10 Distribusi Kelas Stroke	36
Gambar 4.11 Distribusi Kelas Stroke Setelah Smote	37
Gambar 4.12 Confusion Matrix Gradient Boost	42
Gambar 4.13 Classification Report Gradient Boost	43
Gambar 4.14 Tampilan Implementasi Website	44
Gambar 4.15 Hasil Prediksi Stroke	44
Gambar 4.16 Hasil Prediksi Tidak Stroke	45

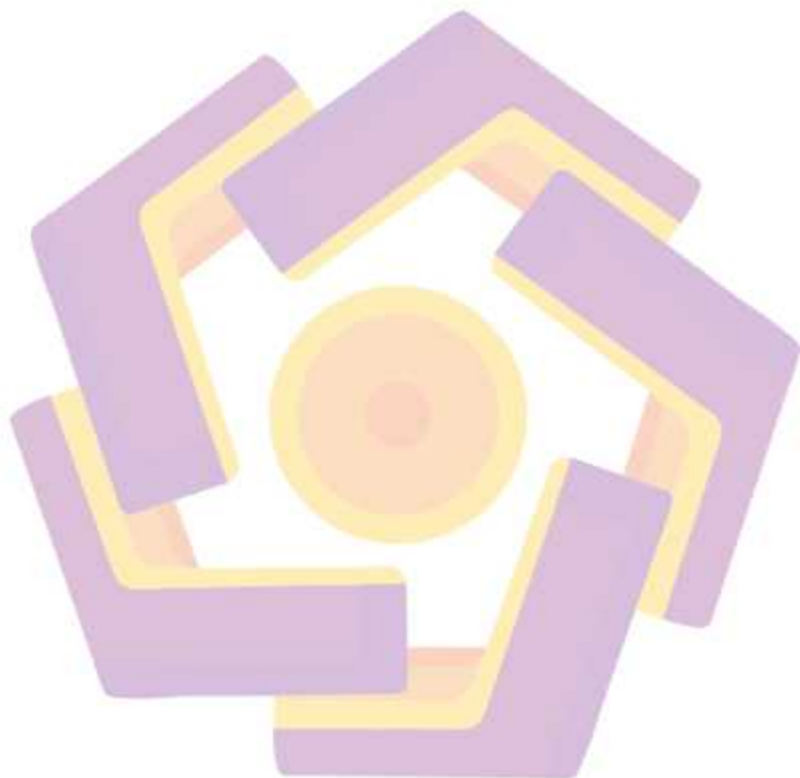
DAFTAR LAMPIRAN

Lampiran 1. Profil obyek Penelitian	10
Lampiran 2. Dokumentasi Penelitian	11



DAFTAR LAMBANG DAN SINGKATAN

PTM	Penyakit Tidak Menular
WHO	World Health Organization



DAFTAR ISTILAH

Overfitting	Model terlalu cocok dengan data latih, sehingga gagal generalisasi pada data baru.
Intervensi	Tindakan yang dilakukan untuk meningkatkan kesehatan.
Binary	Angka 0 dan 1
Ensemble	Teknik yang menggabungkan beberapa model.
Interpolasi linier	Metode untuk memperkirakan nilai di antara dua titik data.
Missing values	Nilai yang hilang.
Variabel Dependen	Variabel yang dipengaruhi variabel lain
Variabel Independen	Variabel yang mempengaruhi variabel lain



INTISARI

Stroke merupakan salah satu penyebab utama disabilitas di dunia dan menjadi penyebab kematian kedua setelah penyakit jantung. Setiap tahun, angka kejadian stroke terus meningkat, menunjukkan adanya permasalahan kesehatan masyarakat yang mendesak. Diagnosis dini stroke menjadi tantangan utama dalam penanganannya karena gejala yang bervariasi dan tidak selalu mudah dikenali. Oleh karena itu, pengembangan model prediksi yang akurat sangat diperlukan untuk mendukung deteksi dini dan pengambilan keputusan medis.

Penelitian ini bertujuan untuk mengembangkan model klasifikasi stroke yang akurat dengan menggunakan algoritma *Random Forest* dan *Gradient Boosting*, serta mengatasi ketidakseimbangan data menggunakan teknik *Synthetic Minority Over-sampling Technique (SMOTE)*. Selain itu, penelitian ini juga bertujuan untuk mengevaluasi dan mengoptimalkan pengaturan *hyperparameter* menggunakan *random search* guna meningkatkan performa model dalam mengklasifikasikan penyakit stroke.

Hasil penelitian menunjukkan bahwa setelah diterapkan *SMOTE* dan optimasi *hyperparameter*, algoritma *Random Forest* mencapai akurasi sebesar 95,06% sedangkan *Gradient Boosting* memberikan performa yang lebih tinggi dengan akurasi 97,43%. Algoritma *Gradient Boosting* lebih efektif dalam melakukan klasifikasi penyakit stroke dibandingkan *Random Forest*.

Kata kunci: Stroke, Data Mining, Random Forest, Gradient Boosting, SMOTE, Hyperparameter Tuning

ABSTRACT

Stroke is one of the the leading cause of disability in the world and is second only to heart disease, heart disease. Every year, the incidence of stroke continues to increase, indicating an urgent public health problem. Diagnosis early diagnosis of stroke is a major challenge in its management due to symptoms that varied and not always easily recognizable. Therefore, the development of accurate model is needed to support early detection and medical decision-making, medical decision making.

This research aims to develop an accurate stroke classification model using Random Forest and Gradient Boosting algorithms, as well as overcoming data imbalance. Forest and Gradient Boosting algorithms, as well as overcoming data imbalance using the Synthetic Minority Over-sampling Technique (SMMT). using Synthetic Minority Over-sampling Technique (SMOTE). In addition, this research also aims to evaluate and optimize the hyperparameter settings using random search to improve the performance of the model in classifying stroke disease.

The results showed that after applying SMOTE and hyperparameter optimization, the Random Forest algorithm achieved an accuracy of 95.06% while Gradient Boosting provided higher performance with an accuracy of 97.43%. The Gradient Boosting algorithm is more effective in classifying stroke diseases than Random Forest.

Keyword: Stroke, Data Mining, Random Forest, Gradient Boosting, SMOTE, Hyperparameter Tuning.