

**ANALISIS KINERJA ALGORITMA REGRESI LOGISTIK,
RANDOM FOREST, DAN NAIVE BAYES UNTUK KLASIFIKASI
DATA MEDIS PENYAKIT STROKE**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Informatika



disusun oleh

DIAS SETYA NUGRAHA

20.11.3615

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**ANALISIS KINERJA ALGORITMA REGRESI LOGISTIK,
RANDOM FOREST, DAN NAIVE BAYES UNTUK KLASIFIKASI
DATA MEDIS PENYAKIT STROKE**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Informatika



disusun oleh

DIAS SETYA NUGRAHA

20.11.3615

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS KINERJA ALGORITMA REGRESI LOGISTIK, RANDOM
FOREST, DAN NAIVE BAYES UNTUK KLASIFIKASI DATA MEDIS
PENYAKIT STROKE**

yang disusun dan diajukan oleh

Dias Setya Nugraha

20.11.3615

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 31 Januari 2025

Dosen Pembimbing,


Mulia Sulistiyono, S.Kom., M.Kom.
NIK. 190302248

HALAMAN PENGESAHAN
SKRIPSI
ANALISIS KINERJA ALGORITMA REGRESI LOGISTIK, RANDOM
FOREST, DAN NAIVE BAYES UNTUK KLASIFIKASI DATA MEDIS
PENYAKIT STROKE

yang disusun dan diajukan oleh

Dias Setya Nugraha

20.11.3615

Telah dipertahankan di depan Dewan Penguji
pada tanggal 31 Januari 2025

Susunan Dewan Penguji

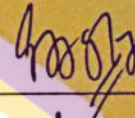
Nama Penguji

Dr. Emigawaty, S.Kom., M.Kom.
NIK. 190302226

Norhikmah, M.Kom
NIK. 190302245

Mulia Sulistiyono, S.Kom., M.Kom.
NIK. 190302248

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 31 januari 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusrini, S.Kom., M.Kom
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Dias Setya Nugraha
NIM : 20.11.3615

Menyatakan bahwa Skripsi dengan judul berikut:

ANALISIS KINERJA ALGORITMA REGRESI LOGISTIK, RANDOM FOREST, DAN NAIVE BAYES UNTUK KLASIFIKASI DATA MEDIS PENYAKIT STROKE

Dosen Pembimbing : Mulia Sulistiyono, S.Kom., 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, 31 Januari 2025

Yang Menyatakan,



Dias Setya Nugraha

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur kepada Tuhan Yang Maha Esa, serta dengan segala kerendahan hati, saya mempersembahkan skripsi ini sebagai bentuk penghormatan dan ungkapan terima kasih kepada orang-orang terdekat yang selalu memberikan dukungan, doa, dan semangat dalam perjalanan akademik saya. Dengan penuh rasa syukur dan hormat, skripsi ini saya persembahkan kepada:

1. Kedua orang tuaku tercinta Samidi, SE dan Dra. Suratmi, yang dengan segala kasih sayang, doa, dan pengorbanannya telah menjadi pilar utama dalam setiap langkah hidupku. Meskipun mereka tidak memiliki kesempatan untuk menempuh pendidikan tinggi, kebijaksanaan, semangat, dan dukungan mereka yang tanpa batas menjadi cahaya penerang dalam perjalanan akademisku hingga meraih gelar sarjana.
2. Bapak Mulia Sulistiyono, S.Kom., M.Kom., selaku dosen pembimbing, yang dengan sabar dan penuh dedikasi telah memberikan bimbingan, ilmu, serta arahan yang sangat berharga dalam penyelesaian tugas akhir ini. Terima kasih atas segala masukan, motivasi, dan waktu yang telah diberikan.
3. Kakak tercinta Samsu Adi Nugroho, ST., yang selalu menjadi tempat berbagi cerita, dukungan, serta inspirasi dalam setiap perjalanan hidup saya. Kehadiran dan semangatnya telah memberikan warna tersendiri dalam menyelesaikan perjalanan studi ini.
4. Dias Setya Nugraha, sebagai bukti perjuangan, ketekunan, dan dedikasi dalam melewati setiap tantangan. Perjalanan ini adalah cerminan dari usaha dan keyakinan bahwa setiap langkah yang ditempuh membawa makna bagi masa depan.
5. Pasangan tercinta Rida Nurjanah, yang senantiasa memberikan kehangatan, dukungan, dan dorongan dalam setiap langkah. Terima kasih telah menjadi bagian dari perjalanan ini, memahami setiap tantangan, serta selalu hadir dengan semangat dan doa yang menguatkan.

KATA PENGANTAR

Segala puji dan syukur saya panjatkan ke hadirat Allah Subhanahu wa Ta'ala atas limpahan rahmat dan hidayah-Nya, sehingga saya dapat menyelesaikan skripsi ini yang berjudul “Analisis Kinerja Algoritma Regresi Logistik, Random Forest, dan Naive Bayes untuk Klasifikasi Data Medis Penyakit Stroke”. Shalawat serta salam senantiasa tercurah kepada Nabi Muhammad Shallallahu Alaihi Wa Sallam, yang telah membawa umat manusia dari zaman kegelapan menuju era penuh ilmu dan cahaya. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Informatika di Universitas AMIKOM Yogyakarta

Sebagai penulis, saya menyadari bahwa dalam penyusunan skripsi ini masih terdapat berbagai kekurangan yang disebabkan oleh keterbatasan ilmu dan pengalaman yang saya miliki. Oleh karena itu, saya dengan terbuka menerima segala bentuk kritik dan saran yang membangun guna menyempurnakan penelitian ini di masa yang akan datang.

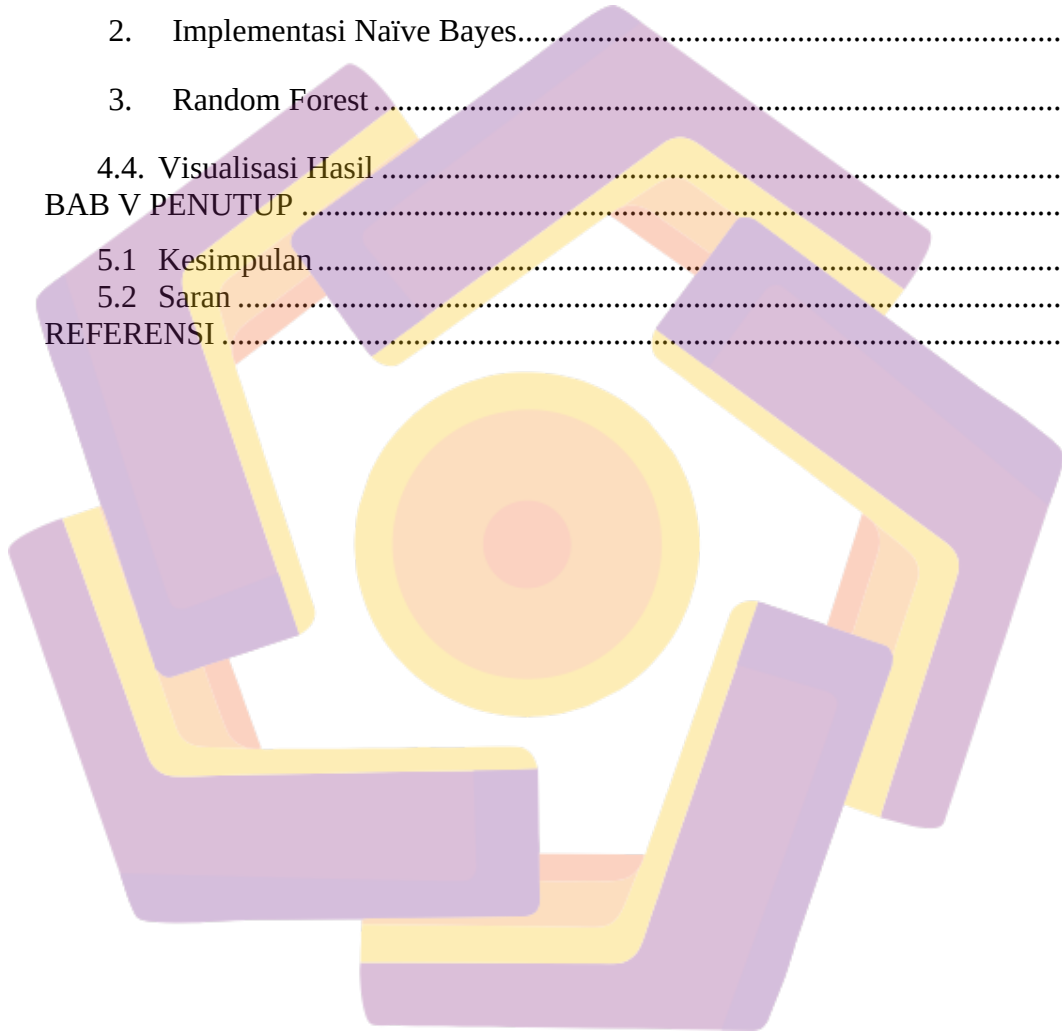
Yogyakarta, 31 Januari 2025

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.....	ix
DAFTAR GAMBAR	x
DAFTAR LAMBANG DAN SINGKATAN	xi
INTISARI	xii
ABSTRACT.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori	15
BAB III METODE PENELITIAN	22
3.1 Alur Penelitian.....	22
3.1.1 Pengumpulan Data	23
3.2 Alat dan Bahan	39
BAB IV HASIL DAN PEMBAHASAN	41
4.1. Pengumpulan Data.....	41
4.2. Preprocessing	41
4.2.1. Missing Values.....	41
4.2.2. Visualisasi Data	42

4.2.3.	Normalisasi	43
4.2.4.	Encoding	46
4.2.5.	SMOTE	47
4.3.	Implementasi Algoritma	48
1.	Implementasi Regresi Logistik	49
2.	Implementasi Naïve Bayes.....	58
3.	Random Forest	66
4.4.	Visualisasi Hasil	78
BAB V	PENUTUP	83
5.1	Kesimpulan	83
5.2	Saran	85
REFERENSI		87

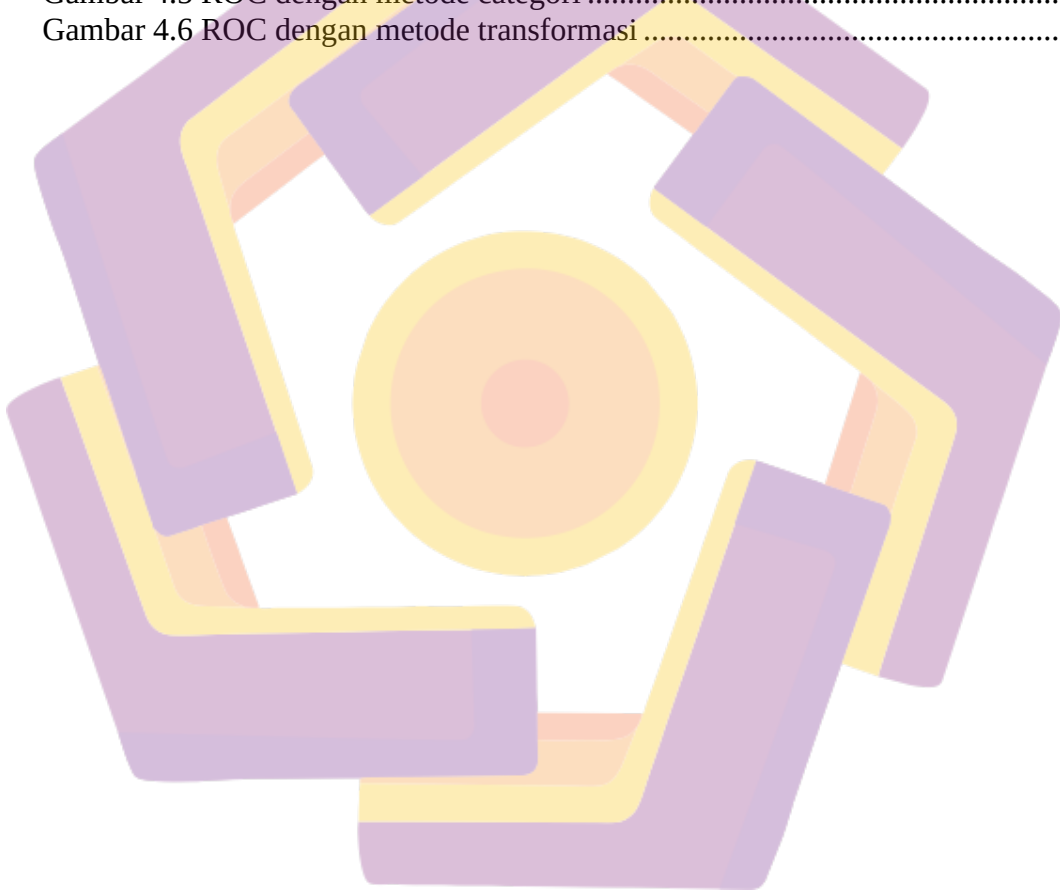


DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian	8
Tabel 2. 2 Confussion Matrix	20
Tabel 3. 1 Tabel dataset	23
Tabel 3. 2 Missing Values	26
Tabel 3. 3 Normaliasi.....	32
Tabel 3. 4 Encoding	33
Tabel 3. 5 Spesifikasi Hardware	40
Tabel 4. 1 missing values sebelum dibersihkan.....	42
Tabel 4. 2 missing values setelah dibersihkan	42
Tabel 4. 3 Normaliasasi data menjadi categorical	45
Tabel 4. 4 Normaliasasi data dengan mengubah jenis data	45
Tabel 4. 5 Encoding	47
Tabel 4. 6 Perbandingan jumlah data.....	48
Tabel 4. 7 Hasil klasifiaksi data uji 90:10 Regresi Logistik.....	50
Tabel 4. 8 Hasil klasifiaksi data uji 80:20 Regresi Logistik.....	50
Tabel 4. 9 Hasil klasifiaksi data uji 70:30 Regresi Logistik.....	50
Tabel 4. 10 Hasil klasifiaksi data uji 60:40 Regresi Logistik.....	51
Tabel 4. 11 Hasil perbandingan Regresi Logistik + tranformasi dan Regresi Logistik + categorical	56
Tabel 4. 12 Hasil klasifiaksi data uji 90:10 Naïve Bayes	59
Tabel 4. 13 Hasil klasifiaksi data uji 80:20 Naïve Bayes	59
Tabel 4. 14 Hasil klasifiaksi data uji 70:30 Naïve Bayes	59
Tabel 4. 15 Hasil klasifiaksi data uji 60:40 Naïve Bayes	60
Tabel 4. 16 Hasil perbandingan Naïve Bayes + tranformasi dan Naïve Bayes + categorical.....	65
Tabel 4. 17 Hasil klasifiaksi data uji 90:10 Random Forest.....	68
Tabel 4. 18 Hasil klasifiaksi data uji 90:10 Random Forest.....	68
Tabel 4. 19 Hasil klasifiaksi data uji 80:20 Random Forest.....	68
Tabel 4. 20 Hasil klasifiaksi data uji 70:30 Random Forest.....	68
Tabel 4. 21 Hasil klasifiaksi data uji 60:40 Random Forest.....	68
Tabel 4. 22 Hasil perbandingan Random Forest + tranformasi dan Random Forest + categorical	74
Tabel 4. 23 Perbandingan hasil klasifikasi	76

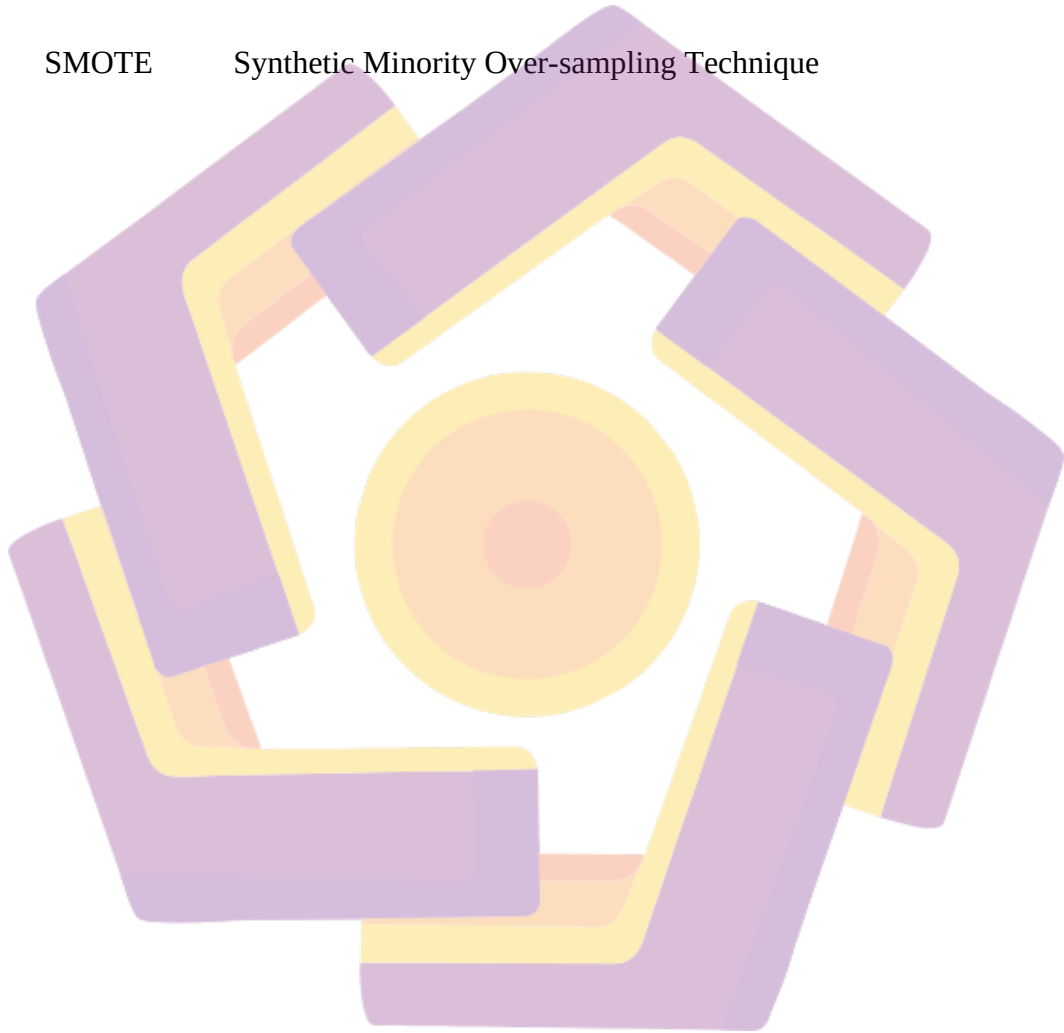
DAFTAR GAMBAR

Gambar 2.1 Scatter Plot Klasifikasi	16
Gambar 3.1 Alur Penelitian	22
Gambar 3.2 Visualisasi Data.....	31
Gambar 3.3 Visualisasi Variabel Stroke	34
Gambar 4.1 Distribusi Kejadian Stroke	43
Gambar 4.2 Distribusi Stroke sebelum dan setelah menggunakan SMOTE	48
Gambar 4.3 Diagram Batang dengan metode kategori	79
Gambar 4.4 Diagram Batang dengan metode transformasi	80
Gambar 4.5 ROC dengan metode kategori	81
Gambar 4.6 ROC dengan metode transformasi	82



DAFTAR LAMBANG DAN SINGKATAN

LR	Regresi Logistic
NB	Naive Bayes
RF	Random Forest
SMOTE	Synthetic Minority Over-sampling Technique



INTISARI

Penelitian ini bertujuan untuk menganalisis kinerja tiga algoritma klasifikasi, yaitu Regresi Logistik, Random Forest, dan Naive Bayes, dalam mengklasifikasikan data medis. Dalam dunia medis, proses klasifikasi data sangat penting untuk meningkatkan akurasi diagnosis penyakit, namun pemilihan algoritma yang tepat sering kali menjadi tantangan karena adanya variasi dalam kompleksitas dan karakteristik data. Metode yang digunakan dalam penelitian ini adalah evaluasi kinerja ketiga algoritma menggunakan beberapa parameter, seperti akurasi, recall, precision, dan F1-Score, dengan menggunakan dataset medis yang berisi informasi tentang pasien yang berisiko terkena penyakit. Hasil penelitian menunjukkan bahwa setiap algoritma memiliki keunggulan dan kelemahan dalam hal kinerja pada dataset tertentu. Regresi Logistik memberikan interpretasi yang mudah, namun Random Forest lebih unggul dalam menangani data dengan hubungan non-linear, sementara Naive Bayes bekerja dengan baik pada dataset yang relatif sederhana. Hasil akhir dari penelitian ini memberikan gambaran yang jelas tentang algoritma mana yang paling efektif dalam konteks klasifikasi data medis dan dapat digunakan sebagai acuan dalam pemilihan algoritma untuk aplikasi di bidang kesehatan. Penelitian ini dapat dimanfaatkan oleh profesional kesehatan dan peneliti dalam mengoptimalkan penggunaan algoritma untuk analisis data medis.

Kata Kunci: Regresi Logistik, Random Forest, Naive Bayes, klasifikasi data medis, kinerja algoritma

ABSTRACT

This study aims to analyze the performance of three classification algorithms: Logistic Regression, Random Forest, and Naive Bayes, in classifying medical data. In the medical field, data classification is crucial for improving the accuracy of disease diagnosis. However, selecting the most suitable algorithm remains a challenge due to variations in data complexity and characteristics. The method employed in this research involves evaluating the performance of the three algorithms using several metrics, such as accuracy, recall, precision, and F1-Score, on a medical dataset containing information about patients at risk of diseases. The results demonstrate that each algorithm has its strengths and weaknesses depending on the dataset. Logistic Regression offers straightforward interpretation, Random Forest excels in handling non-linear relationships, while Naive Bayes performs well on relatively simple datasets. The findings provide clear insights into which algorithm is most effective for medical data classification and can serve as a reference for algorithm selection in healthcare applications. This research is beneficial for healthcare professionals and researchers in optimizing the use of algorithms for medical data analysis.

Keywords: *Logistic Regression, Random Forest, Naive Bayes, medical data classification, algorithm performance*