

**PENERAPAN MACHINE LEARNING PADA PREDIKSI
PENYAKIT STROKE**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

Aldi Prasetya Sanjaya

20.11.3515

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**PENERAPAN MACHINE LEARNING PADA PREDIKSI
PENYAKIT STROKE**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

Aldi Prasetya Sanjaya

20.11.3515

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2024**

HALAMAN PERSETUJUAN

SKRIPSI

**PENERAPAN MACHINE LEARNING PADA PREDIKSI PENYAKIT
STROKE**

yang disusun dan diajukan oleh

Aldi Prasetya Sanjaya

20.11.3515

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 23 Agustus 2024

Dosen Pembimbing,



Majid Rahardi, S.Kom., M.Eng

NIK. 190302393

HALAMAN PENGESAHAN

SKRIPSI

**PENERAPAN MACHINE LEARNING PADA PREDIKSI PENYAKIT
STROKE**

yang disusun dan diajukan oleh

Aldi Prasetya Sanjaya

20.11.3515

Telah dipertahankan di depan Dewan Penguji
pada tanggal 23 Agustus 2024

Susunan Dewan Penguji

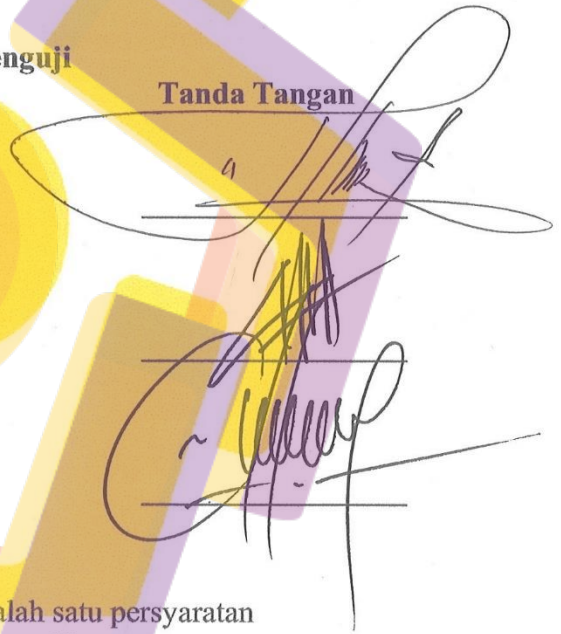
Nama Penguji

Tanda Tangan

Haryoko, S. Kom. M. Cs
NIK. 190302286

Nuri Cahyono, M.Kom
NIK. 190302278

Majid Rahardi, S.Kom., M.Eng
NIK. 190302393



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 23 Agustus 2024

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 : Aldi Prasetya Sanjaya

NIM : 20.11.3515

Menyatakan bahwa Skripsi dengan judul berikut:

**PENERAPAN MACHINE LEARNING PADA PREDIKSI PENYAKIT
STROKE**

Dosen Pembimbing : Majid Rahardi, 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, 23 Agustus 2024

Yang Menyatakan,



Aldi Prasetya Sanjaya

HALAMAN PERSEMBAHAN

- Ucapan terima kasih kepada Allah SWT yang atas karunia yang diberikan dan tidak lupa penulis selalu bersyukur atas pemberian yang telah diberikan oleh-Nya.
- Ucapan terima kasih kepada Bp Ragil Prasetyo dan Ibu Erni Ekowati yang selalu mendidik penulis sampai bisa dalam proses ini, dan juga kepada keluarga yang selalu men-suport, mendoakan, memberikan semangat sehingga kepenulisan skripsi dapat diselesaikan tepat pada waktunya.
- Ucapan terima kasih kepada Bp Majid Rahardi dan Mas Bima selaku Pembina dalam membimbing dan memberikan arahan, sehingga penulis dapat menyelesaikan penulisan naskah ini tepat waktu.
- Ucapan terima kasih kepada semua pihak yang membantu, men-suport, memberikan saran, masukan, semangat, emosional yang baik sehingga penulis dapat menyelesaikan naskah Skripsi yang telah dibuat

KATA PENGANTAR

Puji syukur kehadiran Allah SWT yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “PENERAPAN MACHINE LEARNING PADA PREDIKSI PENYAKIT STROKE”. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer di Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.

Penelitian ini bertujuan untuk membandingkan Mengaplikasikan bidang ilmu machine learning agar bermanfaat bagi ilmu pengetahuan dan pengembangan penelitian. Penelitian ini diharapkan dapat memberikan manfaat bagi tenaga medis supaya lebih cepat dalam pengambilan keputusan dan mendiagnosis lebih dini penyakit stroke.

Penulis menyadari bahwa penulis skripsi ini tidak lepas dari bantuan dan dukungan dari berbagai pihak. Penulis juga menyadari bahwa skripsi ini masih jauh dari sempurna dan memiliki banyak kekurangan. Oleh karena itu, penulis mengharapkan kritik dan saran membangun dari pembaca untuk perbaikan di masa depan. Semoga skripsi ini dapat bermanfaat bagi penulis dan khususnya bagi pembaca pada umumnya.

Yogyakarta, 10 Agustus 2024

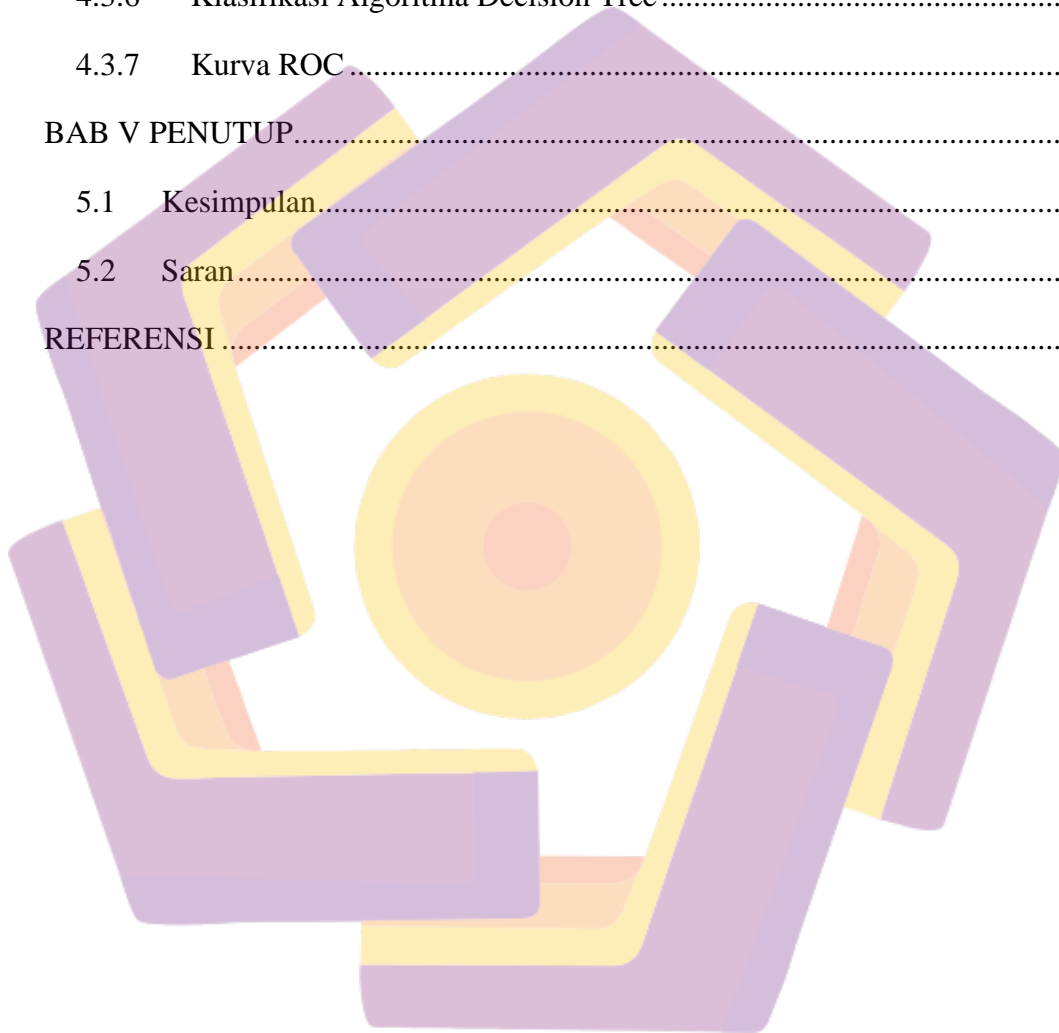
Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
INTISARI.....	xiii
ABSTRACT.....	xiv
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	4
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.....	13
2.2.1 Machine Learning	13
2.2.2 Algoritma Random Forest.....	13
2.2.3 Algoritma Naive Bayes	14
2.2.4 Algoritma Logistic Regression	15

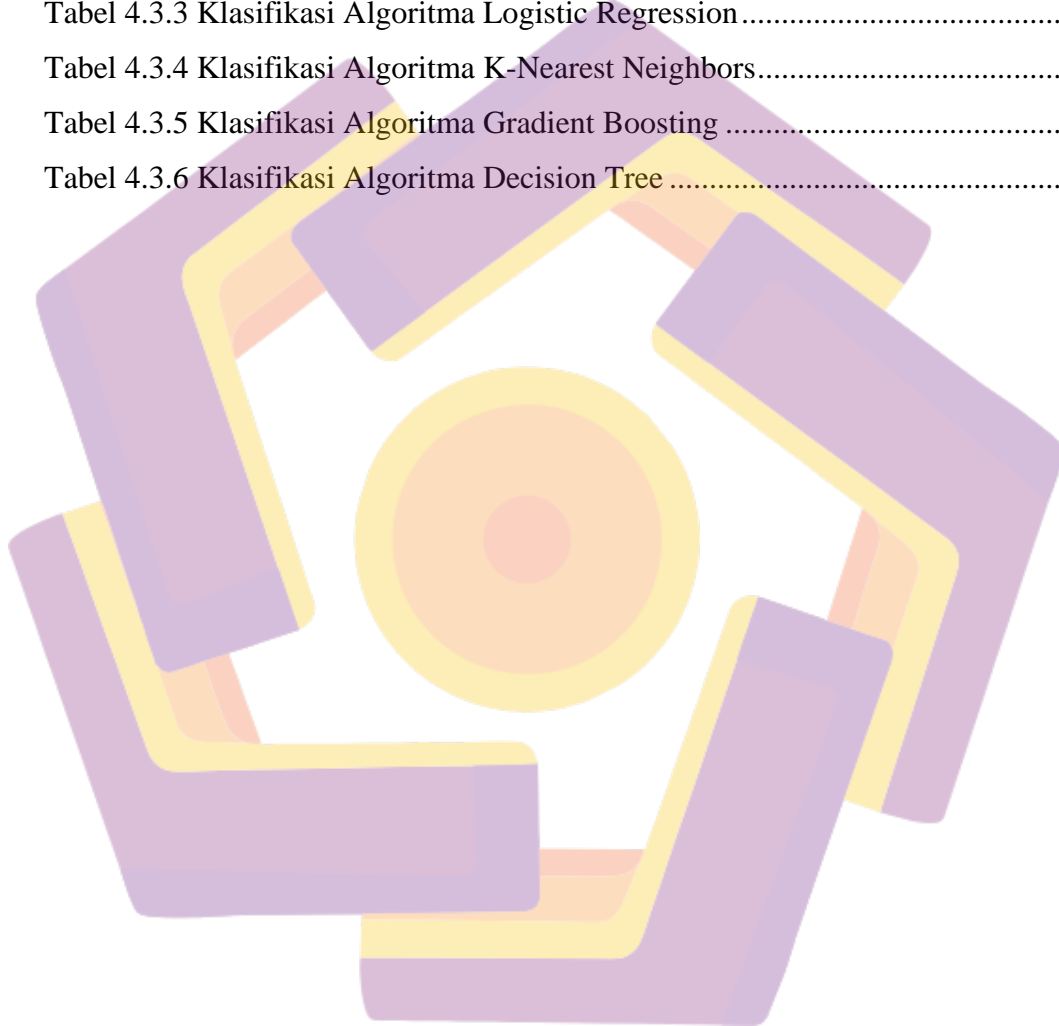
2.2.5	Algoritma K-Nearest Neighbors	15
2.2.6	Algoritma Gradient Boosting.....	16
2.2.7	Algoritma Decision Tree.....	16
2.2.8	Exploratory Data Analysis	17
2.2.9	Pre-Processing Data	17
2.2.10	Evaluasi Performa	22
BAB III METODE PENELITIAN		25
3.1	Objek Penelitian.....	25
3.2	Alur Penelitian	26
3.2.1	Pengambilan Data	27
3.2.2	Exploratory Data Analysis (EDA).....	28
3.2.3	Data Preprocessing	28
3.2.4	Analysis/Data Modeling	31
3.2.5	Evaluasi Kinerja.....	32
BAB IV HASIL DAN PEMBAHASAN		34
4.1	Exploratory Data Analysis (EDA).....	34
4.2	Pre-Processing.....	37
4.2.1	Handling Missing Value.....	37
4.2.2	Handling Duplicate Value	38
4.2.3	SMOTE.....	38
4.2.4	Encoding	39
4.2.5	Scaling	40
4.2.6	Feature Selection	40
4.3	Analysis.....	41
4.3.1	Klasifikasi Algoritma Naïve Bayes	41

4.3.2	Klasifikasi Algoritma Random Forest	42
4.3.3	Klasifikasi Logistic Regression	43
4.3.4	Klasifikasi Algoritma K-Nearest Neighbors (KNN)	44
4.3.5	Klasifikasi Algoritma Gradient Boosting	45
4.3.6	Klasifikasi Algoritma Decision Tree	46
4.3.7	Kurva ROC	47
BAB V PENUTUP		49
5.1	Kesimpulan	49
5.2	Saran	49
REFERENSI		51



DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian.....	9
Tabel 3. 1 Sampel Dataset.....	26
Tabel 4.3.1 Klasifikasi Algoritma Naive Bayes	41
Tabel 4.3.2 Klasifikasi Algoritma Random Forset	42
Tabel 4.3.3 Klasifikasi Algoritma Logistic Regression.....	43
Tabel 4.3.4 Klasifikasi Algoritma K-Nearest Neighbors.....	45
Tabel 4.3.5 Klasifikasi Algoritma Gradient Boosting	46
Tabel 4.3.6 Klasifikasi Algoritma Decision Tree	47



DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian	27
Gambar 4. 1 Persebaran Data.....	35
Gambar 4. 2 Korelasi Data.....	36
Gambar 4.2.1 Hasil Handling Missing Value	37
Gambar 4.2.2 Hasil Handling Duplicate Value	38
Gambar 4.2.3 Before After SMOTE.....	39
Gambar 4.2.4 Hasil Encoding	40
Gambar 4.2. 5 Hasil Scaling	40
Gambar 4.2. 6 Hasil Feature Selection	41
Gambar 4.3.1 Confussion Matrix Naive Bayes	42
Gambar 4.3.2 Confussion Matrix Random Forest	43
Gambar 4.3.3 Confussion Matrix Logistic Regression.....	44
Gambar 4.3.4 Confussion Matrix K-Nearest Neighbors.....	45
Gambar 4.3.5 Confussion Maytrix Gradient Boosting	46
Gambar 4.3.6 Confussion Matrix Decision Tree	47
Gambar 4.3.7 Perbandingan Algoritma Berdasarkan Kurva ROC	48

INTISARI

Penelitian ini bertujuan untuk menganalisis hubungan antara faktor risiko hipertensi (merokok, aktivitas fisik dan konsumsi garam), prevalensi hipertensi dan prevalensi diabetes melitus dengan perbedaan prevalensi stroke di berbagai provinsi di Indonesia. Riset Kesehatan Dasar (Riskesdas) 2007 di 33 provinsi menunjukkan bahwa prevalensi stroke adalah 8,2 per 1000 penduduk, dan prevalensi tertinggi berasal dari provinsi Aceh (16,6/‰). Terdapat 11 rumah sakit yang terlibat dalam penelitian epidemiologi ini. Hasil: 1807 pasien stroke yang dikumpulkan dari Oktober 2012 - April 2013, stroke iskemik merupakan kasus terbanyak (67,1%) dan hemoragik 32,9%. dan hipertensi merupakan faktor risiko yang paling banyak ditemukan baik untuk perdarahan (71,2%) maupun stroke iskemik (63,4%), disusul diabetes melitus dan dislipidemia. Kematian tercatat 20,3% kematian setelah 48 jam, 18,3% \leq 48 jam pada stroke perdarahan, dibandingkan dengan 8,3% kematian pada stroke iskemik setelah 48 jam, dan 3,5% \leq 48 jam.

Penelitian ini melakukan prediksi penyakit stroke dengan menerapkan beberapa algoritma Machine Learning, seperti Naïve Bayes, Random Forest, dan Logistic Regression. Dataset yang digunakan dalam penelitian ini berisi faktor-faktor yang dapat menyebabkan penyakit stroke. Penelitian ini menggunakan beberapa metrik evaluasi seperti akurasi, presisi, recall, f1-score dan Area Under Curve (AUC) untuk memilih model terbaik dari ketiga algoritma tersebut. Penelitian membuktikan bahwa Naïve Bayes mencapai akurasi sebesar 71,34%, Random Forest sebesar 95%, Logistic Regression sebesar 72,78%, K-Nearest Neighbors sebesar 88,42%, Gradient Boosting sebesar 82,55%, dan Decision Tree sebesar 92,23%. Penelitian ini memberikan gambaran mengenai manfaat penerapan Machine Learning dalam melakukan prediksi penyakit stroke. Penerapan Machine Learning dalam melakukan prediksi penyakit stroke dapat membantu dunia medis mendiagnosa penyakit dengan lebih awal.

Kata kunci: Penyakit Stroke, Machine Learning, Prediksi

ABSTRACT

This study aimed to analyze the relationship between hypertension risk factors (smoking, physical activity and salt consumption), prevalence of hypertension and prevalence of diabetes mellitus with differences in stroke prevalence in various provinces in Indonesia. Basic Health Research (Riskesdas) 2007 in 33 provinces showed that the prevalence of stroke was 8.2 per 1000 population, and the highest prevalence was from Aceh province (16.6/‰). There were 11 hospitals involved in this epidemiologic study. Results: 1807 stroke patients collected from October 2012 - April 2013, ischemic stroke was the most common case (67.1%) and hemorrhagic 32.9%. and hypertension was the most common risk factor for both hemorrhage (71.2%) and ischemic stroke (63.4%), followed by diabetes mellitus and dyslipidemia. Mortality was 20.3% after 48 hours, 18.3% \leq 48 hours in bleeding stroke, compared with 8.3% after 48 hours and 3.5% \leq 48 hours in ischemic stroke.

This research predicts stroke disease by applying several Machine Learning algorithms, such as Random Forest, Naïve Bayes, and Logistic Regression. The dataset used in this research contains factors that can cause stroke disease. This study uses several evaluation metrics such as accuracy, precision, recall, f1-score and Area Under Curve (AUC) to select the best model from the three algorithms. The research proved that Naïve Bayes achieved an accuracy of 71.34%, Random Forest of 95%, Logistic Regression of 72.78%, K-Nearest Neighbors of 88.42%, Gradient Boosting of 82.55%, and Decision Tree of 92.23%. This research provides an overview of the benefits of applying Machine Learning in predicting stroke disease. The application of Machine Learning in predicting stroke disease can help the medical world diagnose diseases earlier..

Keyword: Stroke Disease, Machine Learning, Prediction