



**KLASIFIKASI DIAGNOSIS PENYAKIT *STROKE* DENGAN  
MENGUNAKAN METODE *RANDOM FOREST***

**SKRIPSI**

**NUR ALIFFIYANTI ISKANDAR**

**1810511002**

**UNIVERSITAS PEMBANGUNAN VETERAN JAKARTA**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI SARJANA INFORMATIKA**

**2022**



**KLASIFIKASI DIAGNOSIS PENYAKIT *STROKE* DENGAN  
MENGUNAKAN METODE *RANDOM FOREST***

**SKRIPSI**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar  
Sarjana Komputer**

**NUR ALIFFIYANTI ISKANDAR**

**1810511002**

**UNIVERSITAS PEMBANGUNAN VETERAN JAKARTA  
FAKULTAS ILMU KOMPUTER  
PROGRAM STUDI SARJANA INFORMATIKA  
2022**

## PERNYATAAN ORISINALITAS

Tugas Skripsi ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun dirujuk saya nyatakan dengan benar.

Nama : Nur Aliffiyanti Iskandar

NIM : 1810511002

Tanggal : 11 Juli 2022

Bilamana dikemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 11 Juli 2022

Yang Menyatakan,



(Nur Aliffiyanti Iskandar)

## PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan dibawah ini:

Nama : Nur Aliffiyanti Iskandar

NIM : 1810511002

Fakultas : Ilmu Komputer

Program Studi : Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non eksklusif (*Non-Exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

### **Klasifikasi Diagnosis Penyakit *Stroke* Dengan Menggunakan Metode *Random Forest*.**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti di Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian Pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada Tanggal : 11 Juli 2022

Yang Menyatakan,



(Nur Aliffiyanti Iskandar)

## LEMBAR PENGESAHAN

Dengan ini dinyatakan bahwa Tugas Akhir berikut:


Nama : Nur Aliffiyanti Iskandar


NIM : 1810511002


Program Studi : S1 Informatika 2018


Judul : Klasifikasi Diagnosis Penyakit *Stroke* Dengan Menggunakan Metode *Random Forest*

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.


  
Jayanta, S.Kom., M.Si  
Penguji 1

  
Nurul Chamidah, S.Kom., M.Kom.  
Penguji 2

  
Lin Ernawati, S.Kom., M.Si  
Pembimbing 1

  
Yuni Widiastiwi, S.Kom., M.Si  
Pembimbing 2

  
Dr. Ermatita, M.Kom.  
Dekan

  
Desta Sandya Prasvita, S.Kom., M.Kom.  
Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal pengesahan : 11 Juli 2022



# **KLASIFIKASI DIAGNOSIS PENYAKIT *STROKE* DENGAN MENGUNAKAN METODE *RANDOM FOREST***

**Nur Aliffiyanti Iskandar**

## **ABSTRAK**

Penyakit *stroke* menyebabkan kematian kedua dan kecacatan ketiga di dunia, dimana 70% penderita penyakit *stroke* terjadi pada negara berpenghasilan rendah dan menengah. Sementara itu, kematian dan kecacatan yang disebabkan oleh penyakit *stroke* menyumbang 87%. Penyakit *stroke* dan TIA (*Transient Ischemic Attack*) termasuk ke dalam kasus emergensi. Namun gejala dini penyakit *stroke* sulit untuk diketahui. Data mining dapat dimanfaatkan untuk mendiagnosis penyakit. Tujuan penelitian adalah mendapatkan model terbaik dengan *Random Forest* untuk penyakit *stroke*. Model dengan penggunaan jumlah pohon 90 menghasilkan nilai yang optimal, dimana nilai *accuracy* yang dihasilkan sebesar 95.2%, *sensitivity* sebesar 4.1%, *specificity* sebesar 99.8%, *precision* sebesar 66.7%, dan *F-measure* sebesar 7.6%. Serta *ROC Curve* sebesar 0.8048 yang menandakan bahwa model termasuk ke dalam *Good Classification*.

**Kata Kunci:** *Klasifikasi, Random Forest, Stroke*

# **CLASSIFICATION OF STROKE DIAGNOSIS USING RANDOM FOREST METHOD**

**Nur Aliffiyanti Iskandar**

## **ABSTRACT**

Stroke is the second leading cause of death and third disability in the world, where 70% of stroke patients occur in low- and middle-income countries. Meanwhile, death and disability caused by stroke accounted for 87%. Stroke and TIA (Transient Ischemic Attack) are included in emergency cases. However, early symptoms of stroke are difficult to know. Data mining can be used to diagnose diseases. The aim of this research is to get the best model using Random Forest for stroke. The model using 90 trees produces the optimal value, where the accuracy value is 95.2%, sensitivity is 4.1%, specificity is 99.8%, precision is 66.7%, and F-measure is 7.6%. And ROC Curve of 0.8048 which indicates that the model is included in the Good Classification.

**Keyword:** Classification, Random Forest, Stroke

## KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadiran Allah SWT atas segala karunia-Nya sehingga tugas akhir ini dapat disusun dan dapat diselesaikan, dengan judul “Klasifikasi Diagnosis Penyakit *Stroke* Dengan Menggunakan Metode *Random Forest*”. Penulis juga mengucapkan terima kasih kepada:

1. Kedua orang tua yang memberikan dukungan, mendoakan, selalu menemani di masa-masa sulit, dan memberikan kepercayaan bahwa penulis mampu menyelesaikan tugas akhir ini.
2. Ibu Dr. Ermatita, M.Kom sebagai Dekan Fakultas Ilmu Komputer.
3. Bapak Desta Sandya Prasvita, S.Kom., M.Kom sebagai Ketua Program Studi S1 Informatika.
4. Ibu Iin Ernawati, S.Kom., M.Si dan Ibu Yuni Widiastiwi, S.Kom., M.Si sebagai dosen pembimbing yang telah memberikan arahan dan saran yang sangat bermanfaat.
5. Ibu dan Bapak dosen Fakultas Ilmu Komputer yang sudah mengajarkan dan memberikan ilmu-ilmu bermanfaat.
6. Teman-teman S1 Informatika 2018, terutama Luthfiyah, Ein, Erisa, dan Haiqal yang selalu sabar dan memberikan dukungan kepada penulis.
7. Teman terdekat penulis, Della, Ara, Dinda, Penta, Ditya, dan Firman yang selalu menemani dan memberikan dukungan kepada penulis.

Penulis menyadari bahwa penulisan ini masih memiliki banyak kekurangan. Penulis berharap dapat menerima kritik dan saran yang membangun agar dapat melakukan penelitian yang lebih baik lagi.

Jakarta, 11 Juli 2022

Penulis



## DAFTAR ISI

HALAMAN SAMPUL .....	i
HALAMAN JUDUL .....	ii
PERNYATAAN ORISINALITAS .....	iii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS .....	iv
LEMBAR PENGESAHAN .....	v
ABSTRAK .....	vi
ABSTRACT .....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI .....	ix
DAFTAR TABEL .....	xi
DAFTAR GAMBAR .....	xii
DAFTAR SIMBOL .....	xiii
DAFTAR LAMPIRAN .....	xiv
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	2
1.3 Ruang Lingkup .....	2
1.4 Tujuan Penelitian .....	2
1.5 Manfaat Penelitian .....	3
1.6 Luaran Yang Diharapkan .....	3
1.7 Sistematika Penulisan .....	3
BAB II TINJAUAN PUSTAKA .....	4
2.1 <i>Stroke</i> .....	4
2.1.1 Penyebab <i>Stroke</i> .....	4
2.2 <i>Data Mining</i> .....	6
2.2.1 Pengelompokan <i>Data Mining</i> .....	7
2.2.2 Tahapan <i>Data Mining</i> .....	7
2.3 Klasifikasi .....	9
2.4 <i>Decision Tree</i> .....	10
2.5 <i>Random Forest</i> .....	10

2.6	<i>Confusion Matrix</i> .....	12
2.7	<i>ROC Curve</i> .....	14
2.8	Penelitian Relevan .....	15
BAB III METODOLOGI PENELITIAN .....		16
3.1	Kerangka Pikir .....	16
3.1.1	Identifikasi Masalah .....	16
3.1.2	Studi Literatur.....	16
3.1.3	Pengumpulan Data.....	17
3.1.4	Pra-proses Data.....	18
3.1.5	Membagi Data <i>Train</i> dan Data <i>Test</i> .....	18
3.1.6	Pelatihan Model.....	18
3.1.7	Pengujian Model.....	19
3.1.8	Evaluasi Model .....	19
3.2.0	Analisis .....	19
3.2	Alat Bantu Penelitian .....	20
3.3	Jadwal Penelitian .....	20
BAB IV HASIL DAN PEMBAHASAN .....		22
4.1	Pengumpulan Data .....	22
4.2	Pra-proses Data .....	23
4.2.1	<i>Data Cleaning</i> .....	24
4.2.2	<i>Data Transformasi</i> .....	29
4.3	Membagi Data <i>Train</i> dan Data <i>Test</i> .....	30
4.4	Implementasi <i>Random Forest</i> .....	31
4.5	Evaluasi.....	32
4.6	Analisis .....	42
BAB V PENUTUP .....		44
5.1	Kesimpulan .....	44
5.2	Saran .....	44
DAFTAR PUSTAKA.....		45
RIWAYAT HIDUP .....		47
LAMPIRAN .....		48




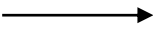
## DAFTAR TABEL

Tabel 1. Deskripsi Atribut <i>Dataset Stroke</i> .....	17
Tabel 2. Jadwal Penelitian.....	20
Tabel 3. <i>Dataset Stroke</i> .....	22
Tabel 4. <i>Dataset Stroke</i> Setelah <i>LabelEncoder</i> .....	29
Tabel 5. Keterangan <i>Dataset Stroke</i> Setelah <i>LabelEncoder</i> .....	29
Tabel 6. Tipe Data Pada <i>Dataset Stroke</i> .....	30
Tabel 7. Pembagian Data <i>Train</i> dan Data <i>Test</i> .....	31
Tabel 8. Perbandingan Model <i>Random Forest</i> Dengan Jumlah Pohon Berbeda .....	32
Tabel 9. <i>Confusion Matrix</i> .....	33
Tabel 10. <i>Feature Importance</i> .....	35
Tabel 11. Data Sampel <i>Stroke</i> .....	36
Tabel 12. Sampel Data <i>Stroke</i> Hipertensi <i>Yes</i> .....	39
Tabel 13. Sampel Data <i>Stroke</i> Hipertensi <i>No</i> .....	41

## DAFTAR GAMBAR

Gambar 1. <i>Data Mining</i> Dalam Proses <i>Knowledge Discovery from Data</i> (KDD)....	8
Gambar 2. <i>Confusion Matrix</i> .....	12
Gambar 3. <i>ROC Curve</i> .....	14
Gambar 4. Kerangka Pikir.....	16
Gambar 5. <i>Pie Chart Pasien Stroke</i> .....	23
Gambar 6. Atribut Mengandung <i>Missing Value</i> .....	24
Gambar 7. <i>Bar Chart</i> Atribut <i>Gender</i> Sebelum Penanganan <i>Missing Value</i> .....	25
Gambar 8. <i>Bar Chart</i> Atribut <i>Gender</i> Setelah Penanganan <i>Missing Value</i> .....	25
Gambar 9. <i>Bar Chart</i> Atribut <i>Smoking_Status</i> Sebelum Penanganan <i>Missing Value</i> .....	26
Gambar 10. <i>Bar Chart</i> Atribut <i>Smoking_Status</i> Setelah Penanganan <i>Missing Value</i> .....	26
Gambar 11. Setelah Penanganan <i>Missing Value</i> .....	27
Gambar 12. <i>Boxplot</i> Atribut BMI.....	28
Gambar 13. <i>Boxplot</i> Atribut BMI Setelah Penanganan <i>Outlier</i> .....	28
Gambar 14. <i>Bar Chart</i> Model <i>Random Forest</i> Dengan Jumlah Pohon Berbeda .....	33
Gambar 15. Grafik Probabilitas Model <i>Random Forest</i> .....	34
Gambar 16. <i>Bar Chart Feature Importance</i> .....	36
Gambar 17. <i>Root</i> Pertama Dengan <i>Hypertension</i> .....	39
Gambar 18. <i>Root</i> Kedua Dengan <i>Ever Married</i> .....	40
Gambar 19. Hasil <i>Tree</i> Dengan Data Sampel .....	42

## DAFTAR SIMBOL

No	Simbol	Nama	Keterangan
1.		Terminator	Simbol <i>start</i> (permulaan) atau <i>stop</i> (akhir) dari sebuah proses.
2.		Proses	Simbol pengolahan data atau perhitungan data pada program.
3.		<i>Input-Output</i>	Simbol proses <i>input</i> ataupun <i>output</i>
4.		Garis alir ( <i>flow line</i> )	Arah aliran program.

## DAFTAR LAMPIRAN

Lampiran 1. <i>Dataset Stroke</i> .....	49
Lampiran 2. <i>Dataset Stroke</i> Setelah Pra-proses .....	50
Lampiran 3. Hasil Turnitin.....	51