

**Klasifikasi Penyakit Diabetes pada Lansia  
Menggunakan Model *Random Forest***



**SKRIPSI**

**Diajukan Sebagai Syarat Untuk Menyelesaikan  
Pendidikan Program Strata - 1 Pada  
Program Studi Teknik Informatika**

**Oleh :**

**William Oktafian**

**2020.11.0025**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER DAN SAINS  
UNIVERSITAS INDO GLOBAL MANDIRI**

**2024**



**Klasifikasi Penyakit Diabetes pada Lansia  
Menggunakan Model *Random Forest***

**SKRIPSI**

**Diajukan Sebagai Syarat Untuk Menyelesaikan  
Pendidikan Program Strata - 1 Pada  
Program Studi Teknik Informatika**

**Oleh :**

**William Oktafian**

**2020.11.0025**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER DAN SAINS  
UNIVERSITAS INDO GLOBAL MANDIRI**

**2024**

**LEMBAR PENGESAHAN SKRIPSI**

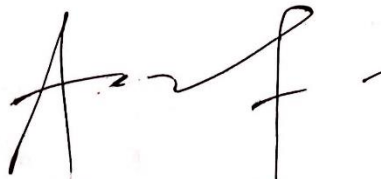
**Klasifikasi Penyakit Diabetes pada  
Lansia Menggunakan Model  
*Random Forest***

Oleh

**William Oktafian  
NPM : 2020.11.0025**

**Palembang , 05 Februari 2024**

**Pembimbing I**



**Lastri Widya Astuti, M.Kom  
NIK : 2003.01.0063**

**Pembimbing II**



**Ir Mustafa Ramadhan, M.T  
NIK : 2002.03.0172**

**Mengetahui,**

**Dekan Fakultas Ilmu Komputer dan Sains**

FAKULTAS ILMU KOMPUTER DAN SAINS



**Rudi Heriansyah, S.T., M.Eng. Ph.D  
NIK:2022.01.9315**

## LEMBAR PERSETUJUAN DEWAN PENGUJI

Pada hari Kamis tanggal 25 Januari 2024 telah dilaksanakan ujian sidang skripsi :

Nama : William Oktafian

NPM : 2020.11.0025

Judul : Klasifikasi Penyakit Diabetes pada Lansia Menggunakan

Model *Random Forest*

Oleh Prodi Teknik Informatika Fakultas Ilmu Komputer dan Sains Universitas Indo  
Global Mandiri Palembang

Palembang, 2 Februari 2024

Penguji 1,



Zaid Romegar Mair, S.T., M.Cs

NIK: 2021.01.0307

Penguji 2,



Dr Gasim, S.Kom., M.Si

NIK: 2023.01.0340

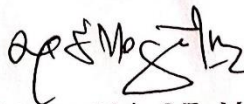
Penguji 3,



Ir Mustafa Ramadhan, M.T

NIK: 2002.03.0172

Menyetujui,  
Ka. Prodi Teknik Informatika



Zaid Romegar Mair, S.T., M.Cs

NIK: 2021.01.0307



Scanned with CamScanner



SURAT KETERANGAN REVISI SKRIPSI  
PROGRAM STUDI TEKNIK INFORMATIKA (S1)  
FASILKOM DAN SAINS  
UNIVERSITAS INDO GLOBAL MANDIRI

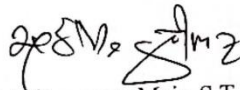
Kami yang bertanda tangan dibawah ini, menerangkan bahwa :

Nama : William Oktafian  
NPM : 2020.11.0025  
Judul : Klasifikasi Penyakit Diabetes pada Lansia Menggunakan  
Model *Random Forest*


Mahasiswa yang namanya tercantum diatas, telah selesai merevisi penulisan SKRIPSI

Palembang, 2 Februari 2024

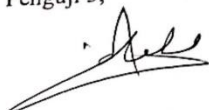
Penguji 1,

  
Zaid Romegar Mair, S.T.,M.Cs  
NIK: 2021.01.0307

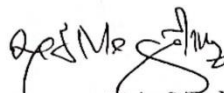
Penguji 2,

  
Dr Gasim, S.Kom.,M.Si  
NIK: 2023.01.0340

Penguji 3,

  
Ir Mustafa Ramadhan, M.T  
NIK: 2002.03.0172

Menyetujui,  
Ka. Prodi Teknik Informatika

  
Zaid Romegar Mair, S.T., M.Cs  
NIK: 2021.01.0307

# Klasifikasi Penyakit *Diabetes* Pada Lansia Menggunakan Model *Random Forest*

## ABSTRAK

Bertambahnya umur akan diiringi dengan penurunan fungsi tubuh, timbulnya berbagai penyakit, keseimbangan tubuh dan risiko jatuh. Status kesehatan lansia yang semakin menurun seiring bertambahnya umur ini berlawanan dengan keinginan para lansia itu sendiri agar tetap sehat. Penyakit terbanyak pada lanjut usia adalah penyakit tidak menular antara lain hipertensi, artritis, stroke, penyakit paru obstruktif kronis, dan *Diabetes mellitus*. Faktor yang dapat mempengaruhi kualitas hidup dan kesehatan lansia adalah keterbatasan akses lansia terhadap pelayanan kesehatan. Teknologi *machine learning* dapat digunakan untuk mengklasifikasi suatu penyakit yang dialami oleh pasien lansia. Berdasarkan penelitian sebelumnya yang berjudul Perbandingan Model *Decision Tree*, *Naive Bayes* dan *Random Forest* untuk Prediksi Klasifikasi Penyakit Jantung, telah digunakan model *random forest* yang menghasilkan tingkat akurasi terbesar dari dua model *machine learning* yang lainnya, akurasi yang didapatkan oleh model *random forest* sebesar 75%, sedangkan pada penelitian yang berjudul Prediksi Kemungkinan *Diabetes* pada Tahap Awal Menggunakan Algoritma Klasifikasi *Random Forest* menghasilkan akurasi sebesar 97,88%. Maka dari itu pada penelitian ini akan berfokus pada penggunaan model *random forest*. Pada kali ini model *random forest* dan dataset yang diambil dari puskesmas merdeka berhasil mendapatkan tingkat akurasi sebesar 96%, nilai tersebut dapat dihasilkan dari rata-rata percobaan model *random forest* menggunakan *k-fold cross validation* sebanyak 5-fold pengujian.

Kata Kunci : lansia, *machine learning*, penyakit, *random forest*.

## ***Classifying Diabetes in the Elderly Using a Random Forest Model***

### ***ABSTRACT***

*Increasing age is accompanied by a decline in bodily functions, the onset of various diseases, body balance issues, and the risk of falls. The deteriorating health status of the elderly as they age contradicts their desire to remain healthy. The most common diseases among the elderly are non-communicable diseases including hypertension, arthritis, stroke, chronic obstructive pulmonary disease, and diabetes mellitus. Factors that can affect the quality of life and health of the elderly include limited access to healthcare services. Machine learning technology can be used to classify a disease experienced by elderly patients. Based on previous research entitled "Comparison of Decision Tree, Naive Bayes, and Random Forest Models for Heart Disease Classification Prediction", the random forest model has been used, which yielded the highest accuracy rate compared to the other two machine learning models. The accuracy obtained by the random forest model was 75%, whereas in the study titled "Prediction of Early Stage Diabetes Using Random Forest Classification Algorithm", it yielded an accuracy of 97.88%. Therefore, this research will focus on the use of the random forest model. This time, the random forest model and dataset taken from Merdeka community health center managed to achieve an accuracy rate of 96%. This value was obtained from the average of the random forest model experiments using 5-fold cross-validation testing.*

*Keywords: disease, elderly, machine learning, random forest.*

## KATA PENGANTAR

Puji dan syukur saya ucapkan atas kehadiran Allah Subhanahu Wata'ala berkat rahmat dan hidayahnya akhirnya penulis dapat menyelesaikan penelitian ini dengan baik tepat pada waktunya, tidak lupa shalawat serta salam selalu dilimpahkan kepada junjungan kita Nabi besar Nabi Muhammad SAW beserta keluarga sahabat para pengikut dan insyaallah kita semua hingga akhir zaman.

Proposal Skripsi yang penulis buat dengan judul “ *Klasifikasi Penyakit Diabetes Pada Lansia Menggunakan Model Random Forest* ” disusun guna memenuhi syarat kelulusan dalam memperoleh gelar Sarjana (S1) pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Indo Global Mandiri (UIGM) Palembang.

Tidak lupa penulis mengucapkan terimakasih atas bantuan yang diberikan selama penyusunan skripsi ini kepada

1. Dr. Marzuki Alie, SE.,MM, selaku Rektor Universitas Indo Global Mandiri Palembang.
2. Rudi Heriansyah, ST., M.Eng. Ph.D selaku Dekan Fakultas Ilmu Komputer Universitas Indo Global Mandiri.
3. Zaid Romegar Mair, S.T., M.Cs sebagai Ketua Program Studi Teknik Informatika Universitas Indo Global Mandiri dan Dosen Pembimbing Akademik.
4. Lastri Widya Astuti, M.Kom, sebagai Dosen Pembimbing I.
5. Ir Mustafa Ramadhan, M.T, sebagai Dosen Pembimbing II.
6. Bapak/Ibu Dosen Fakultas Ilmu Komputer dan Karyawan/Karyawati Universitas Indo Global Mandiri.
7. Keluarga dan Teman -teman yang telah memberikan dukungan moril maupun materil sehingga penulis bisa menyelesaikan skripsi ini.

Dengan segala kerendahan hati penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, oleh karena itu dibutuhkan kritik dan saran untuk perbaikan dan pengembangan skripsi ini sangat diharapkan. Akhir kata, semoga skripsi ini bermanfaat bagi semua pihak terima kasih.

Palembang, November 2023

Penulis William Oktafian  
2020.11.0025



## DAFTAR ISI

<b>HALAMAN JUDUL LUAR .....</b>	<b>iii</b>
<b>HALAMAN JUDUL DALAM .....</b>	<b>iii</b>
<b>LEMBAR PENGESAHAN SKRIPSI .....</b>	<b>iii</b>
<b>LEMBAR PERSETUJUAN DEWAN PENGUJI.....</b>	<b>iv</b>
<b>SURAT KETERANGAN REVISI SKRIPSI.....</b>	<b>v</b>
<b>ABSTRAK .....</b>	<b>vi</b>
<b><i>ABSTRACT</i> .....</b>	<b>vii</b>
<b>KATA PENGANTAR.....</b>	<b>viii</b>
<b>DAFTAR ISI.....</b>	<b>ixii</b>
<b>DAFTAR GAMBAR .....</b>	<b>xii</b>
<b>DAFTAR TABEL .....</b>	<b>xiv</b>
<b>DAFTAR RUMUS.....</b>	<b>xv</b>
<b>DAFTAR LAMPIRAN .....</b>	<b>xvi</b>
<b>BAB 1 PENDAHULUAN .....</b>	<b>1</b>
1.1    Latar Belakang .....	1
1.2    Perumusan Masalah.....	2
1.3    Batasan Masalah.....	2
1.4    Tujuan Dan Manfaat Penelitian.....	3
1.4.1    Tujuan Penelitian.....	3
1.4.2    Manfaat Penelitian .....	3
1.5    Sistematika Penulisan.....	3
<b>BAB 2 KAJIAN PUSTAKA DAN DASAR TEORI .....</b>	<b>5</b>
2.1 <i>Data Mining</i> .....	5
2.2 <i>Machine Learning</i> .....	6
2.2.1 <i>Unsupervised learning</i> .....	6
2.2.2 <i>Supervised learning</i> .....	7
2.2.3 <i>Random Forest</i> .....	8
2.3    Klasifikasi.....	10
2.4 <i>Diabetes</i> .....	10

2.5	<i>Confusion Matrix</i> .....	12
2.6	<i>K Fold Cross Validation</i> .....	13
2.7	Kodingan .....	14
2.7.1	<i>Python</i> .....	14
2.7.2	<i>Library Python</i> .....	14
2.7.3	<i>Flowchart</i> .....	20
2.8	Penelitian Terdahulu .....	21
<b>BAB 3 METODOLOGI PENELITIAN</b> .....		<b>24</b>
3.1	Tahapan Penelitian.....	24
3.2	Studi Literatur.....	25
3.3	Persiapan <i>Software</i> dan <i>Hardware</i> .....	25
3.3.1	<i>Software</i> (Perangkat Lunak).....	25
3.3.2	<i>Hardware</i> (Perangkat Keras).....	25
3.4	Pengumpulan data .....	26
3.5	<i>Preprocessing</i> .....	28
3.5.1	<i>Data Sample Penyakit Diabetes</i> . .....	29
3.5.2	<i>Cleaning</i> .....	32
3.5.3	<i>K-fold Cross Validation</i> .....	32
3.5.4	Simpan Data Latih dan Data Uji .....	32
3.6	<i>Design Model</i> .....	33
3.6.1	<i>Bootstrap sampling</i> .....	33
3.6.2	<i>Random Forest Classification</i> .....	33
3.6.3	Evaluasi <i>Model</i> .....	53
3.7	Pengambilan Kesimpulan.....	53
<b>BAB 4 IMPLEMENTASI DAN EVALUASI</b> .....		<b>54</b>
4.1	Pengumpulan data .....	54
4.2	<i>Preprocessing</i> .....	54
4.2.1	<i>Data Cleansing</i> .....	55
4.3	<i>Visualisasi dataset</i> .....	56
4.4	Inisialisasi Model <i>Random Forest Classifier</i> .....	63
4.5	Inisialisasi <i>K-fold Cross Validation</i> .....	64
4.6	Inisialisasi <i>List</i> .....	64

4.7	Eksekusi <i>Model</i> .....	64
4.8	Menghitung akurasi dari model.....	65
4.9	Evaluasi .....	66
<b>BAB 5 KESIMPULAN DAN SARAN .....</b>		<b>72</b>
5.1	Kesimpulan.....	72
5.2	Saran .....	72
<b>Daftar Pustaka.....</b>		<b>73</b>
<b>Lampiran .....</b>		<b>75</b>

## DAFTAR GAMBAR

<b>Gambar 2.1</b> Algoritma sederhana <i>Random Forest</i> .	9
<b>Gambar 2.2</b> <i>5-Fold Cross Validation</i> .	14
<b>Gambar 2.3</b> logo <i>pandas</i>	16
<b>Gambar 2.4</b> Logo <i>matplotlib</i>	17
<b>Gambar 2.5</b> logo <i>Scikit-Learn</i>	18
<b>Gambar 2.6</b> logo <i>numpy</i>	19
<b>Gambar 2.7</b> logo <i>seaborn</i>	20
<b>Gambar 3.1</b> Skema Alur Penelitian.	25
<b>Gambar 3.2</b> <i>Flowchart Preprocessing Dataset</i> .	28
<b>Gambar 3.3</b> Pembagian Data <i>Traning</i> dan <i>Testing K-fold Cross Validatioan</i> .	32
<b>Gambar 3.4</b> <i>Flowchart</i> proses <i>Design Model</i> .	33
<b>Gambar 3.5</b> <i>Root</i> Pohon Keputusan 1.	35
<b>Gambar 3.6</b> <i>Node 1.1</i> Pohon Keputusan 1.	37
<b>Gambar 3.7</b> <i>Node 1.1.1</i> Pohon Keputusan 1.	38
<b>Gambar 3.8</b> <i>Root</i> pohon keputusan 2.	40
<b>Gambar 3.9</b> <i>Node 1.1</i> Pohon Keputusan 2.	41
<b>Gambar 3.10</b> <i>Node 1.1.1</i> Pohon Keputusan 2.	43
<b>Gambar 3.11</b> Pohon Keputusan 2.	44
<b>Gambar 3.12</b> <i>Root</i> pohon keputusan 3.	46
<b>Gambar 3.13</b> <i>Node 1.1</i> Pohon Keputusan 2.	47
<b>Gambar 3.14</b> <i>Node 1.1.1</i> Pohon Keputusan 3.	49
<b>Gambar 3.15</b> <i>Node 1.1.1.1</i> Pohon Keputusan 3.	51
<b>Gambar 3.16</b> Pohon Keputusan 3.	52
<b>Gambar 4.1</b> Lima dataset pertama.	54
<b>Gambar 4.2</b> Lima dataset pertama yang telah dinormalisasi.	55
<b>Gambar 4.3</b> Kode program untuk melihat <i>Missing Value</i> .	55
<b>Gambar 4.4</b> Hasil dari kode program untuk melihat <i>missing value</i> .	55

<b>Gambar 4.5</b> Kode program untuk melihat Duplicated Value. ....	56
<b>Gambar 4.6</b> Hasil kode program <i>Duplicated Value</i> .....	56
<b>Gambar 4.7</b> Kode program untuk melihat distribusi data pada setiap atribut ....	56
<b>Gambar 4.8</b> Distribusi data atribut umur. ....	57
<b>Gambar 4.9</b> Distribusi data atribut riwayat diabetes. ....	57
<b>Gambar 4.10</b> Distribusi data atribut riwayat diabetes turunan. ....	58
<b>Gambar 4.11</b> Distribusi data atribut sistol. ....	58
<b>Gambar 4.12</b> Distribusi data atribut distol. ....	59
<b>Gambar 4.13</b> Distribusi data atribut IMT. ....	59
<b>Gambar 4.14</b> Distribusi data atribut gula darah.....	59
<b>Gambar 4.15</b> Distribusi data atribut gula darah puasa. ....	60
<b>Gambar 4.16</b> Distribusi data atribut jenis kelamin. ....	60
<b>Gambar 4.17</b> Distribusi data atribut kehamilan. ....	61
<b>Gambar 4.18</b> Distribusi data atribut lingk pinggang. ....	61
<b>Gambar 4.19</b> Distribusi data atribut diagnosa. ....	62
<b>Gambar 4.20</b> Kode program untuk melihat relasi antar atribut .....	62
<b>Gambar 4.21</b> Relasi antar atribut. ....	63
<b>Gambar 4.22</b> Fitur (X) dan Label (y) . ....	63
<b>Gambar 4.23</b> Inisialisasi <i>Model Random Forest Classifier</i> . ....	64
<b>Gambar 4.24</b> Inisialisasi <i>K-Fold Cross Validation</i> . ....	64
<b>Gambar 4.25</b> Inisialisasi List. ....	64
<b>Gambar 4.26</b> Eksekusi Model. ....	65
<b>Gambar 4.27</b> kode program untuk mencari nilai rata-rata akurasi. ....	66
<b>Gambar 4.28</b> Hasil akurasi model. ....	66

## DAFTAR TABEL

<b>Table 2.1</b> <i>Confusion Matrix table</i> .....	12
<b>Table 2.2</b> Simbol- Simbol <i>Flowchart</i> .....	21
<b>Table 2.3</b> Tinjauan Penelitian Terdahulu .....	21
<b>Table 3.1</b> Format pengumpulan Dataset.....	26
<b>Table 3.2</b> Data Sample Penyakit Diabetes Mentah.....	29
<b>Table 3.3</b> Data Sample Penyakit Diabetes Mentah (Lanjut). .....	30
<b>Table 3.4</b> Data Sample Penyakit Diabetes.....	30
<b>Table 3.5</b> Data Sample Penyakit Diabetes (Lanjut).....	31
<b>Table 3.6</b> Tabel Perhitungan pada <i>Root</i> Pohon Keputusan 1.....	34
<b>Table 3.7</b> Tabel Perhitungan pada <i>Node</i> 1.1 Pohon Keputusan 1.....	36
<b>Table 3.8</b> Tabel Perhitungan pada <i>Node</i> 1.1.1 Pohon Keputusan 1.....	38
<b>Table 3.9</b> Komparasi akurasi model <i>random forest classification</i> .....	53
<b>Table 4.1</b> Tabel perhitungan akurasi model.....	67
<b>Table 4.2</b> 80 dataset penyakit diabetes. ....	68
<b>Table 4.3</b> 80 dataset penyakit diabetes (lanjutan).....	69
<b>Table 4.4</b> 80 dataset penyakit diabetes (lanjutan).....	70
<b>Table 4.5</b> 80 dataset penyakit diabetes (lanjutan).....	71

## DAFTAR RUMUS

<b>Persamaan 1</b> <i>Entropy Random Forest</i> .....	10
<b>Persamaan 2</b> <i>Information Gain Random Forest</i> .....	10
<b>Persamaan 3</b> Menghitung Nilai Akurasi.....	12
<b>Persamaan 4</b> Menghitung Nilai Presisi .....	12
<b>Persamaan 5</b> Menghitung Nilai <i>Recall</i> .....	13
<b>Persamaan 6</b> Menghitung Nilai <i>Specificity</i> .....	13
<b>Persamaan 7</b> Menghitung Nilai <i>F1 Score</i> .....	13

## DAFTAR LAMPIRAN

<b>Lampiran 1</b> Daftar Riwayat Hidup .....	75
<b>Lampiran 2</b> Kartu Bimbingan.....	76
<b>Lampiran 3</b> Surat Keterangan Tidak Plagiat .....	77
<b>Lampiran 4</b> Kode Program .....	78