

**SKRIPSI**

**ANALISIS HASIL LABORATORIUM PENGIDAP PENYAKIT  
DIABETES DENGAN ALGORITMA KLASIFIKASI PADA  
PUSKESMAS KECAMATAN CIRACAS JAKARTA TIMUR**



**Oleh**

**Daffy Fayyadhya Ramzy**

**NIM. 1910511044**

**PROGRAM STUDI INFORMATIKA, PROGRAM SARJANA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN**

**Jakarta**

**2023**

# PERNYATAAN ORISINALITAS

## PERNYATAAN ORISINALITAS

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

Nama : Daffy Fayyadhya Ramzy

NIM : 1910511044

Tanggal : 19 Januari 2024

Judul Skripsi : **Analisis Hasil Laboratorium Pengidap Penyakit Diabetes Dengan Algoritma Klasifikasi Pada Puskesmas Kecamatan Ciracas Jakarta Timur**

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

Jakarta, 19 Januari 2024

Yang Menyatakan,



Daffy Fayyadhya Ramzy

# PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

## PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademika Universitas Pembangunan Nasional Veteran Jakarta, Saya yang bertanda tangan di bawah ini:

Nama : Daffy Fayyadhya Ramzy  
NIM : 1910511044  
Fakultas : Ilmu Komputer  
Program Studi : Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan karya ilmiah Saya kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non-Eksklusif (*Non-Exchange Royalty Free Right*) untuk dipublikasikan dengan judul:

**Analisis Hasil Laboratorium Pengidap Penyakit Diabetes Dengan Algoritma Klasifikasi Pada Puskesmas Kecamatan Ciracas Jakarta Timur.**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media atau memformatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan skripsi Saya selama tetap mencantumkan nama Saya sebagai penulis atau pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini Saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada Tanggal : 25 Oktober 2023

Yang Menyatakan,



Daffy Fayyadhya Ramzy

# LEMBAR PENGESAHAN

## LEMBAR PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama : Daffy Fayyadhya Ramzy

NIM : 1910511044

Program Studi : S1 Informatika

Judul Tugas Akhir : Analisis Hasil Laboratorium Pengidap Penyakit Diabetes Dengan Algoritma Klasifikasi Pada Puskesmas Kecamatan Ciracas Jakarta Timur

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



Iin Ernawati, S.Kom., M.Si.

Penguji I



Musthofa Galih Pradana, S.Kom., M.Kom.

Penguji II




Bayu Hananto, S.Kom., M.Kom.

Pembimbing



Prof. R. Ir. Supriyanto, ST., M.Sc., IPM

Dekan



Dr. Widya Cholil, M.T.

Kepala Program Studi

Ditetapkan di : Jakarta

Tanggal Ujian : 11 Januari 2024

## ABSTRAK

Penyakit diabetes adalah penyakit kronis yang berupa gangguan metabolisme yang ditandai dengan kadar gula darah yang meningkat lebih dari batas normal. Jumlah pasien diabetes selal yang Setelah dilakukan pelatihan dan pengujian data dengan model klasifikasi *Random Forest* dengan beberapa rasio *split*, model yang memiliki nilai akurasi tertinggi adalah model yang menggunakan rasio *split* 70:30 dengan nilai akurasi 72,89%. Waktu yang digunakan untuk melatih dan menguji dengan model ini juga relatif rendah dengan waktu latih 21,23 detik dan waktu uji 0,76 detik. Karena dataset yang digunakan memiliki data label 0 (tidak terdiagnosis) yang jauh lebih banyak dibandingkan data label 1 (terdiagnosis), hasil yang diperoleh model-model tersebut tidak seakurat mungkin.

**Kata Kunci:** Diabetes, Data Mining, Klasifikasi

## **ABSTRACT**

Diabetes is a chronic disease in the form of metabolic disorders that are indicated by the increase of glucose in a patient's blood to an abnormal level. The amount of diabetes patients continues to increase all around the world. The diagnoses of diabetes on a patient can be predicted with the Random Forest classification algorithm. After training and testing data with the Random Forest classification model with different split ratios, the model with the highest accuracy is the model that uses the split ratio of 70:30 with 72,89% accuracy. The time taken to learn and test with this model is also relatively short with 21,23 seconds of training time and 0,76 seconds of testing time. The result may not be as accurate as possible due to the fact that the amount of data with the label 0 (undiagnosed) is far higher than the amount of data with the label 1 (diagnosed).

**Keywords:** Diabetes, Data Mining, Classification

## **KATA PENGANTAR**

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala karunia-Nya agar penulis dapat menyelesaikan skripsi ini. Skripsi berjudul “Analisis Hasil Laboratorium Pengidap Penyakit Diabetes Dengan Algoritma Klasifikasi Pada Puskesmas Kecamatan Ciracas Jakarta Timur” yang disusun sejak September 2022 ini dilaksanakan dengan bantuan dari beberapa pihak-pihak lain. Karena ini, penulis ingin menyampaikan ucapan terima kasih sebesar-besarnya kepada pihak berikut:

1. Ibu Dr. Ermatita, M.Kom selaku Dekan Fakultas Ilmu Komputer.
2. Ibu Dr. Widya Cholil, M.I.T. selaku Kepala Program Studi Informatika.
3. Bapak Bayu Hananto, S.Kom., M.Kom. selaku dosen pembimbing skripsi yang telah membantu penulis dengan memberi saran-saran untuk penyusunan skripsi.
4. Keluarga penulis yang telah memberi doa dan dukungan untuk menyelesaikan skripsi ini.
5. Teman-teman dekat penulis yang juga telah memberi doa, dukungan, dan juga saran yang membantu penulis dalam pembuatan skripsi ini.

Penulis mengucapkan terima kasih kepada seluruh pihak yang telah membantu pembuatan skripsi ini.

## DAFTAR ISI

<b>PERNYATAAN ORISINALITAS.....</b>	<b>ii</b>
<b>PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS .....</b>	<b>iii</b>
<b>LEMBAR PENGESAHAN .....</b>	<b>iv</b>
<b>ABSTRAK .....</b>	<b>v</b>
<b>ABSTRACT .....</b>	<b>vi</b>
<b>KATA PENGANTAR.....</b>	<b>vii</b>
<b>DAFTAR ISI.....</b>	<b>viii</b>
<b>DAFTAR GAMBAR.....</b>	<b>x</b>
<b>DAFTAR TABEL .....</b>	<b>xi</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
<b>1.1. Latar Belakang .....</b>	<b>1</b>
<b>1.2. Rumusan Masalah .....</b>	<b>2</b>
<b>1.3. Tujuan Penelitian .....</b>	<b>2</b>
<b>1.4. Batasan Penelitian .....</b>	<b>2</b>
<b>1.5. Luaran Penelitian .....</b>	<b>2</b>
<b>1.6. Sistematika Penulisan .....</b>	<b>3</b>
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>4</b>
<b>2.1. Diabetes Melitus .....</b>	<b>4</b>
<b>2.2. Data Mining .....</b>	<b>4</b>
<b>2.3. Klasifikasi.....</b>	<b>6</b>
<b>2.4. <i>Random Forest</i> .....</b>	<b>7</b>
<b>2.5. <i>Feature Importance</i> .....</b>	<b>8</b>
<b>2.6. <i>Random Over Sampling</i> .....</b>	<b>9</b>
<b>2.7. Ulasan Penelitian Terdahulu .....</b>	<b>9</b>
<b>BAB III METODOLOGI PENELITIAN .....</b>	<b>13</b>
<b>3.1. Alur Penelitian.....</b>	<b>13</b>
<b>3.1.1. Identifikasi Masalah .....</b>	<b>14</b>
<b>3.1.2. Studi Literatur.....</b>	<b>14</b>
<b>3.1.3. Pengumpulan Data.....</b>	<b>14</b>
<b>3.1.4. <i>Preprocess Data</i>.....</b>	<b>14</b>
<b>3.1.5. <i>Data Mining</i> .....</b>	<b>14</b>



3.1.6. Evaluasi .....	15
3.2. Perangkat Penelitian .....	15
3.3. Jadwal Penelitian.....	16
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>17</b>
4.1. Akuisisi Data .....	17
4.2. <i>Preprocessing Data</i> .....	19
4.2.1. <i>Cleaning Data</i> .....	20
4.2.2. Transformasi Data .....	21
4.2.3. Balancing Data .....	25
4.3. <i>Data Mining</i> .....	25
4.3.1. <i>Feature Importance</i> .....	27
4.3.2. Implementasi Algoritma <i>Random Forest</i> .....	28
4.4. Hasil .....	29
4.4.1. <i>Random Split</i> Rasio 90:10 .....	30
4.4.2. <i>Random Split</i> Rasio 80:20 .....	31
4.4.3. <i>Random Split</i> Rasio 70:30 .....	32
4.5. Pembahasan .....	32
<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>34</b>
5.1. Kesimpulan .....	34
5.2. Saran.....	34
<b>DAFTAR PUSTAKA .....</b>	<b>35</b>
<b>RIWAYAT HIDUP .....</b>	<b>37</b>
<b>LAMPIRAN.....</b>	<b>38</b>

## DAFTAR GAMBAR

Gambar 2. 1. Langkah-langkah KDD (Han, Kamber and Pei, 2011) .....	5
Gambar 2. 2. Alur Kerja Random Forest (S.Farel, 2023) .....	7
Gambar 3. 1. Alur Penelitian .....	13
Gambar 4. 1. Grafik jumlah data label 0 dan 1 .....	27
Gambar 4. 2. Grafik Feature Importance Random Forest.....	28
Gambar 4. 3. Tree Pertama Random Forest.....	29
Gambar 4. 4. Confusion Matrix Rasio 90:10 .....	30
Gambar 4. 5. Confusion Matrix Rasio 80:20 .....	31
Gambar 4. 6. Confusion Matrix Rasio 70:30 .....	32

## DAFTAR TABEL

Tabel 3. 1. Jadwal Pelaksanaan Penelitian.....	16
Tabel 4. 1. Sampel data bagian 1 .....	18
Tabel 4. 2. Sampel data bagian 2 .....	19
Tabel 4. 3. Sampel data bagian 3 .....	19
Tabel 4. 4. Jumlah Missing Value di atribut .....	20
Tabel 4. 5. Atribut Kategorik Yang Terubah dan Penjelasannya .....	21
Tabel 4. 6. Contoh Pengubahan Data Kategorik ke Numerik.....	22
Tabel 4. 7. Contoh Data Sebelum Penghapusan String yang Tidak Relevan 1 ....	22
Tabel 4. 8. Contoh Data Sebelum Penghapusan String yang Tidak Relevan 2 ....	23
Tabel 4. 9. Contoh Data Setelah Penghapusan String yang Tidak Relevan 1 .....	23
Tabel 4. 10. Contoh Data Setelah Penghapusan String yang Tidak Relevan 2 ....	23
Tabel 4. 11. Atribut-atribut dan label yang digunakan dalam penelitian.....	24
Tabel 4. 12. Contoh data sebelum normalisasi .....	24
Tabel 4. 13. Contoh data setelah normalisasi.....	25
Tabel 4. 14. Contoh data 1 .....	26
Tabel 4. 15. Contoh data 2 .....	26
Tabel 4. 16. Contoh data 3 .....	26
Tabel 4. 17. Perbandingan hasil model-model.....	33