

**DETEKSI PENYAKIT DIABETES MENGGUNAKAN  
SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE  
DAN METODE RANDOM FOREST**

**SKRIPSI**

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



disusun oleh

**YOPIEZHAR ZHABIMA**

**21.11.4454**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2024**

**DETEKSI PENYAKIT DIABETES MENGGUNAKAN  
SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE  
DAN METODE RANDOM FOREST**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**YOPIEZHAR ZHABIMA**

**21.11.4454**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2024**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**DETEKSI PENYAKIT DIABETES MENGGUNAKAN SYNTHETIC  
MINORITY OVERSAMPLING TECHNIQUE DAN METODE RANDOM  
FOREST**

yang disusun dan diajukan oleh

**Yopiezhhar Zhabima**

**21.11.4454**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 9 Desember 2024

Dosen Pembimbing,



**Melany Mustika Dewi, S.Kom., M.Kom.**

**NIK. 190302455**

**HALAMAN PENGESAHAN**

**SKRIPSI**

**DETEKSI PENYAKIT DIABETES MENGGUNAKAN SYNTHETIC  
MINORITY OVERSAMPLING TECHNIQUE DAN METODE RANDOM  
FOREST**

yang disusun dan diajukan oleh

**Yopiezhar Zhabima**

21.11.4454

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 23 Desember 2024

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

Supriatin., S.Kom., M.Kom.  
NIK. 190302239

Yuli Astuti., S.Kom., M.Kom.  
NIK. 1903022146

Melany Mustika Dewi., Skom., M.Kom.  
NIK. 190302455


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

**DEKAN FAKULTAS ILMU KOMPUTER**



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.  
NIK. 190302096

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Nama mahasiswa : Yopiezhar Zhabima  
NIM : 21.11.4454

Menyatakan bahwa Skripsi dengan judul berikut:

### DETEKSI PENYAKIT DIABETES MENGGUNAKAN SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE DAN METODE RANDOM FOREST

Dosen Pembimbing: Melany Mustika Dewi, 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, 01 Januari 2025

Yang Menyatakan,



Yopiezhar Zhabima

## HALAMAN PERSEMBAHAN

Alhamdulillah rabbil 'aalamiin, puji serta Syukur penulis panjatkan kepada Allah SWT atas limpahan nikmat dan karunia yang telah Engkau berikan sehingga skripsi ini dapat diselesaikan dengan baik dan masih jauh dari kesempurnaan.

Skripsi ini, saya persembahkan kepada:

1. Pencapaian ini adalah sedikit persembahan kecil untuk kedua orang tua penulis, bapak Waharman dan ibu Vivi Sastri Heryanti yang telah melalui banyak perjuangan dan rasa sakit. Penulis selalu berusaha untuk tidak akan membiarkan semua itu sia-sia. Penulis selalu ingin melakukan yang terbaik untuk setiap kepercayaan yang telah diberikan. Penulis akan tumbuh dan menjadi yang terbaik, diantara yang paling baik.
2. Untuk kedua mendiang kakek dan nenek penulis yang memberikan pondasi hidup dan banyak mengajarkan arti dari kehidupan didunia ini.
3. Untuk bunga Rini, om Ari, bungsu Nova, dan om Teddy yang selalu memberikan semangat untuk menjadi orang yang kuat dan bertahan sampai sejauh ini. Serta adik-adik penulis Erfin Fadjri, Wawa, Naina, Zhazha dan Zhezhe yang membuat semangat penulis tidak pernah habis, dan semangat dari kalian yang selalu di jadikan bahan bakar oleh penulis.
4. Untuk semua orang orang yang telah mendukung dan memberikan semangat kepada penulis, baik secara langsung maupun tidak langsung.

## KATA PENGANTAR

Puji dan Syukur selalu dipanjatkan atas kehadiran ALLAH SWT yang selalu memberikan Rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “**DETEKSI PENYAKIT DIABETES MENGGUNAKAN SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE DAN METODE RANDOM**”.

Penulisan skripsi ini tujuan sebagai syarat utama bagi penulis untuk menyelesaikan kuliah dan mendapatkan kelulusan program sarjana jurusan Informatika Universitas Amikom Yogyakarta. Selesaiannya penulisan pada skripsi ini tidak lepas dari dukungan banyak pihak yang memberikan dukungan, fasilitas, bantuan, dan spiritualitas. Oleh karena itu penulis ingin mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. M. Suyanto., MM. selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al-Fatta., S.Kom., M.Kom, Ph.D. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Ibu Windha Mega Pradnya Duhita., S.Kom., M.kom. selaku dosen wali penulis selama masa perkuliahan.
4. Ibu Melany Mustika Dewi., S.Kom., M.Kom. selaku dosen pembimbing skripsi penulis.
5. Bapak Waharman dan Ibu Vivi Sastri Heryanti selaku pemberi dana terbesar untuk penulis.

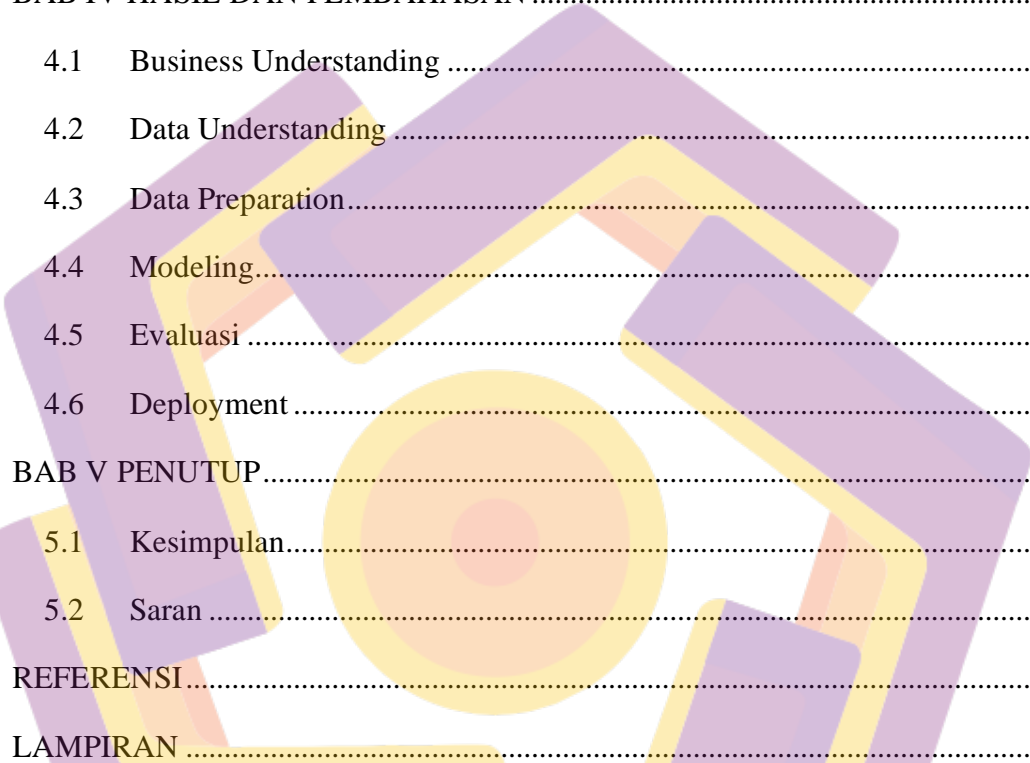
Yogyakarta, 10 Desember 2024

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 LAMPIRAN.....	xii
DAFTAR LAMBANG DAN SINGKATAN .....	xiii
DAFTAR ISTILAH .....	xiv
INTISARI .....	xv
<i>ABSTRACT</i> .....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah .....	2
1.4 Tujuan Penelitian .....	2
1.5 Manfaat Penelitian .....	3
1.6 Sistematika Penulisan .....	3
BAB II TINJAUAN PUSTAKA.....	4
2.1 Studi Literatur .....	4

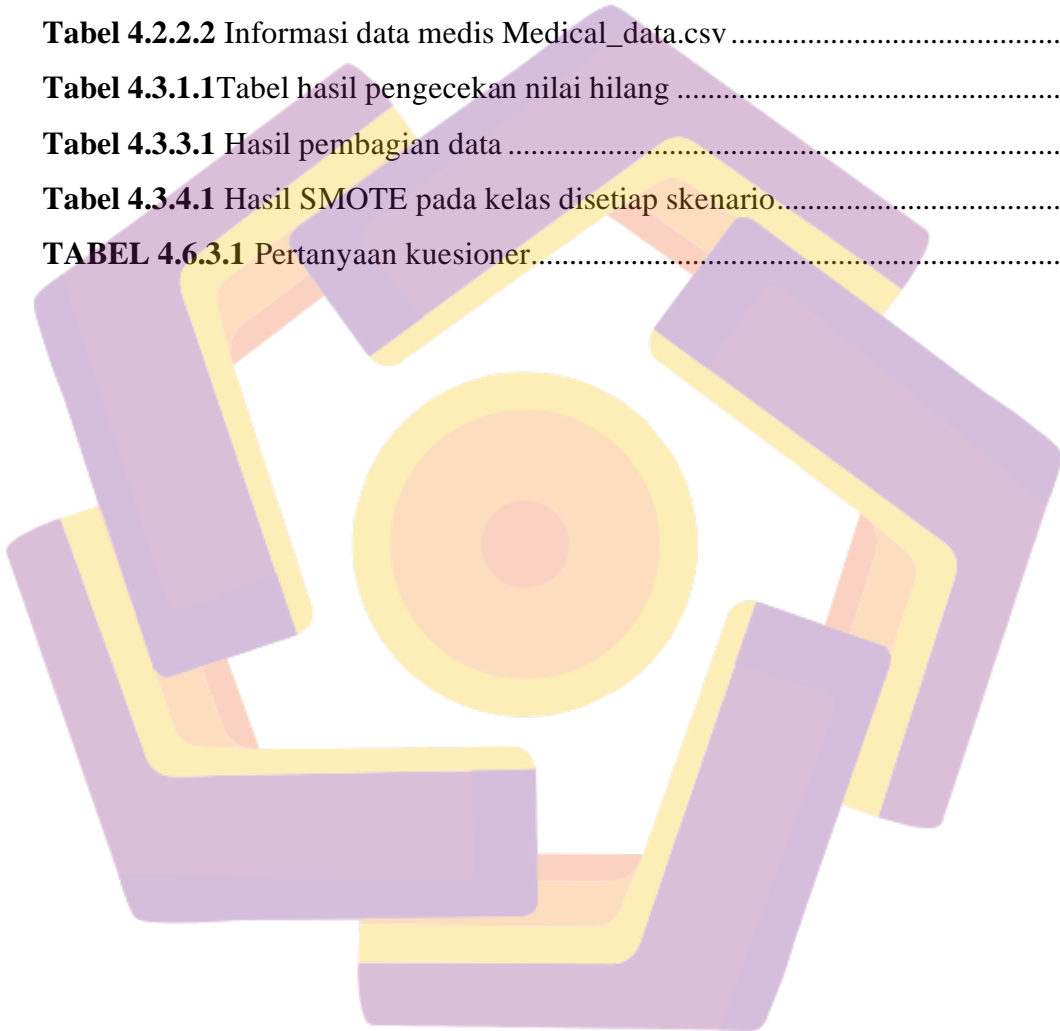




2.2	Dasar Teori .....	13
<b>BAB III METODE PENELITIAN.....</b>		<b>18</b>
3.1	Alur Penelitian .....	18
3.2	Alat dan Bahan.....	20
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>		<b>23</b>
4.1	Business Understanding .....	23
4.2	Data Understanding .....	23
4.3	Data Preparation.....	27
4.4	Modeling.....	32
4.5	Evaluasi .....	34
4.6	Deployment .....	48
<b>BAB V PENUTUP.....</b>		<b>51</b>
5.1	Kesimpulan.....	51
5.2	Saran .....	51
<b>REFERENSI.....</b>		<b>52</b>
<b>LAMPIRAN .....</b>		<b>55</b>

## DAFTAR TABEL

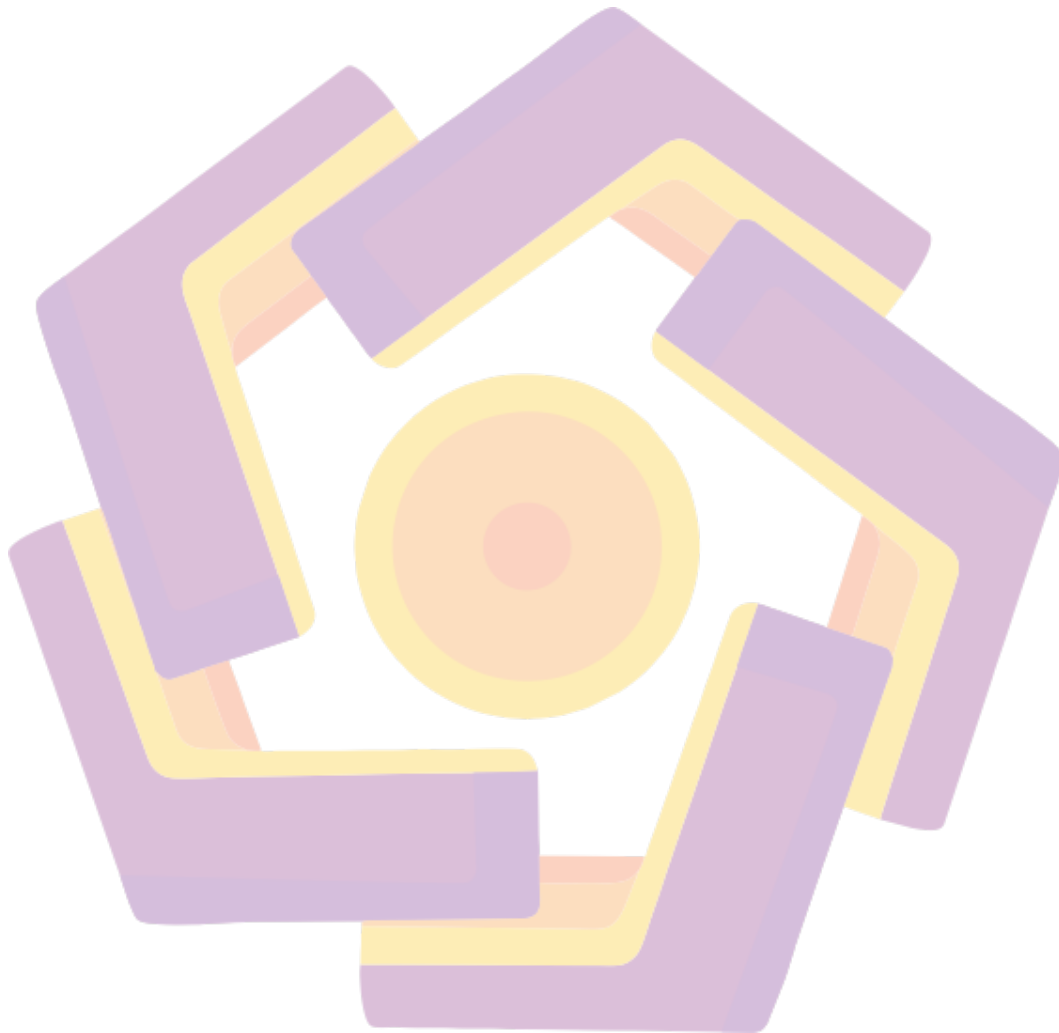
<b>Tabel 2. 1</b> Keaslian Penelitian .....	7
<b>Tabel 2.2.6.1</b> Confusion matrix .....	16
<b>Tabel 3.2.1</b> Penjelasan dataset Training.csv .....	21
<b>Tabel 3.2.2</b> Penjelasan data medis Medical_data.csv .....	22
<b>Tabel 4.2.2.1</b> Informasi data Testing.csv .....	24
<b>Tabel 4.2.2.2</b> Informasi data medis Medical_data.csv .....	24
<b>Tabel 4.3.1.1</b> Tabel hasil pengecekan nilai hilang .....	27
<b>Tabel 4.3.3.1</b> Hasil pembagian data .....	30
<b>Tabel 4.3.4.1</b> Hasil SMOTE pada kelas disetiap skenario.....	31
<b>TABEL 4.6.3.1</b> Pertanyaan kuesioner.....	50



## DAFTAR GAMBAR

<b>Gambar 2.2.4.1</b> Random Forest [4] .....	15
<b>Gambar 3.1</b> Alur Penelitian .....	18
<b>Gambar 4.2.3.1</b> Distribusi kelas pada dataset.....	25
<b>Gambar 4.2.3.2</b> Distribusi kelas pada dataset.....	26
<b>Gambar 4.3.2.1</b> Pair plot fitur dan target pada dataset.....	29
<b>Gambar 4.3.2.2</b> Pair plot fitur dan targer pada data medis .....	29
<b>Gambar 4.3.4.1</b> Code cell menyeimbangkan data menggunakan SMOTE .....	31
<b>Gambar 4.4.1.1</b> Code cell untuk inialisai RandomForestClassifier.....	32
<b>Gambar 4.4.1.2</b> Code cell untuk inialisai RandomForestClassifier.....	32
<b>Gambar 4.4.1.3</b> Code cell untuk inialisai RandomForestClassifier.....	32
<b>Gambar 4.5.1.1</b> Code cell prediksi .....	35
<b>Gambar 4.5.1.2</b> Code cell prediksi .....	35
<b>Gambar 4.5.1.3</b> Code cell prediksi .....	35
<b>Gambar 4.5.2.1</b> Evaluasi skenario 1 (90:10).....	36
<b>Gambar 4.5.2.2</b> Evaluasi skenario 2 (80:20).....	37
<b>Gambar 4.5.2.3</b> Evaluasi skenario 3 (70:30).....	37
<b>Gambar 4.5.2.4</b> Evaluasi skenario 4 (60:40).....	38
<b>Gambar 4.5.2.5</b> Evaluasi skenario 5 (50:50).....	38
<b>Gambar 4.5.3.1</b> Evaluasi skenario 1 (90:10) tanpa SMOTE .....	39
<b>Gambar 4.5.3.2</b> Evaluasi skenario 2 (80:20) tanpa SMOTE .....	40
<b>Gambar 4.5.3.3</b> Evaluasi skenario 3 (70:30) tanpa SMOTE .....	40
<b>Gambar 4.5.3.4</b> Evaluasi skenario 4 (60:40) tanpa SMOTE .....	41
<b>Gambar 4.5.3.5</b> Evaluasi skenario 5 (50:50) tanpa SMOTE .....	41
<b>Gambar 4.5.4.1</b> Evaluasi data medis menggunakan skenario 1 (90:10) .....	42
<b>Gambar 4.5.4.2</b> Evaluasi data medis menggunakan skenario 2 (80:20) .....	43
<b>Gambar 4.5.4.3</b> Evaluasi data medis menggunakan skenario 3 (70:30) .....	43
<b>Gambar 4.5.4.4</b> Evaluasi data medis menggunakan skenario 4 (60:40) .....	44
<b>Gambar 4.5.4.5</b> Evaluasi data medis menggunakan skenario 5 (50:50) .....	44
<b>Gambar 4.5.5.2</b> Untuk hasil pasien tidak berisiko terkena diabetes.....	47
<b>Gambar 4.5.5.3</b> Untuk hasil pasien berisiko terkena diabetes .....	48

**Gambar 4.6.1.1** Code cell untuk menyimpan model.....48  
**Gambar 4.6.2.1** Struktur website .....49  
**Gambar 4.6.3.1** Tampilan awal website.....49



## DAFTAR LAMPIRAN

<b>Lampiran 1</b> Kode app.py.....	55
<b>Lampiran 2</b> Kode requirements.txt.....	56
<b>Lampiran 3</b> Tampilan jika tidak menginput data .....	56
<b>Lampiran 4</b> Tampilan jika mendapatkan hasil berisiko terkena diabetes .....	57
<b>Lampiran 5</b> Tampilan jika mendapatkan hasil tidak berisiko terkena diabetes....	57
<b>Lampiran 6</b> Tampilan keterangan untuk input data .....	58
<b>Lampiran 7</b> Asal instansi responden .....	58
<b>Lampiran 8</b> Presentase hasil pertanyaan pertama.....	58
<b>Lampiran 9</b> Presentase hasil pertanyaan kedua .....	59
<b>Lampiran 10</b> presentase hasil pertanyaan ketiga.....	59



## DAFTAR LAMBANG DAN SINGKATAN

SMOTE	Synthetic Minority Oversampling Technique
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
mm Hg	Milimeter Air Raksa
$\mu\text{U/mL}$	Mikro Unit per Milimeter
$\text{kg/m}^2$	Kilo Meter per Meter Persegi



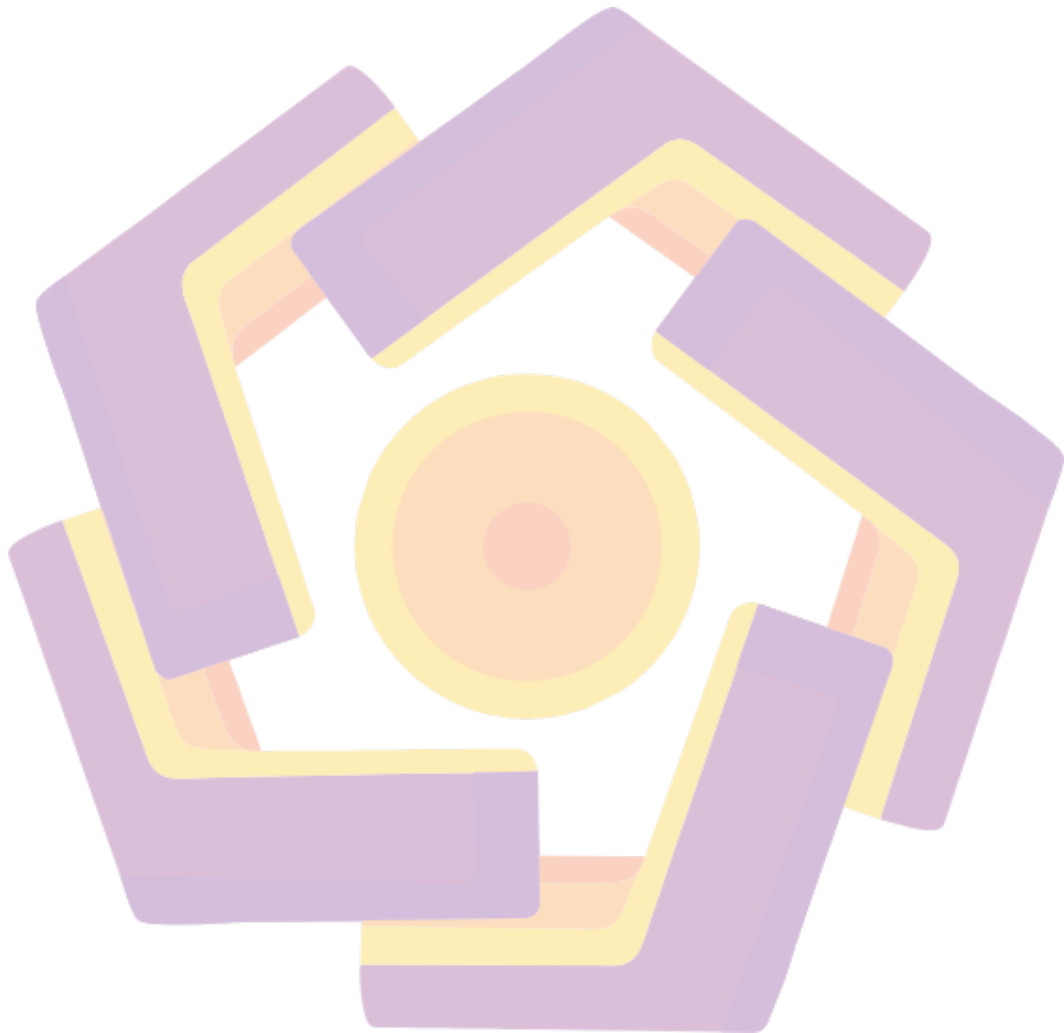
## DAFTAR ISTILAH

Decision Tree

Pohon keputusan

Machine Learning

Mesin Pembelajaran



## INTISARI

Diabetes adalah kondisi kronis yang ditandai tingginya kadar gula akibat gangguan dalam produksi atau fungsi insulin. Komplikasi serius mulai dari penyakit jantung, gangguan ginjal, kerusakan saraf, dan bahkan sampai kematian. Tercatat pada tahun 2021 penderita diabetes sebanyak 19,5 juta jiwa dan diprediksi akan menyentuh angka 28,6 juta jiwa pada tahun 2045. Meskipun angka ini tinggi setengah dari penderita tidak menyadarinya, maka dilakukan penelitian ini untuk membuat alternatif yang memudahkan dalam pendeteksi penyakit diabetes dan diharapkan dapat memberikan peringatan dini supaya Masyarakat bisa segera memperbaiki pola hidup agar terhindar dari penyakit diabetes.

Random forest merupakan salah satu metode untuk melakukan klasifikasi. Pada penelitian ini digunakan metode random forest yang dilatih menggunakan 10 skenario yaitu 5 skenario di testing menggunakan data train dan 5 skenario testing menggunakan data medis rumah, pembagian data (90:10, 80:20, 70:30, 60:40, 50:50) dan data train sudah di seimbangkan dengan SMOTE. SMOTE sendiri merupakan metode untuk mengatasi ke tidak seimbangan data. Untuk melakukan analisis data peneliti menggunakan Teknik CRISP-DM yang memiliki tahapan Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment.

Evaluasi terbaik didapat pada skenario ke 5 dari model Random Forest menggunakan SMOTE pada prediksi penyakit diabetes yang diuji menggunakan 100% data medis. Untuk Precision kelas "0" 67% dan kelas "1" 94%, untuk Recall kelas "0" 67% dan kelas "1" 94%, untuk F1-Score kelas "0" 67% dan kelas "1" 94%, serta nilai Accuracy 90%. Hasil dari model tersebut dimuat dan dijadikan sebuah website untuk prediksi diabetes.

**Kata kunci:** CRISP-DM, Prediksi, Random Forest, Smote, Website.



## **ABSTRACT**

*Diabetes is a chronic disease characterized by high blood sugar levels due to impaired insulin production or function. Serious complications that can arise include heart disease, kidney disorders, nerve damage, and even death. It was recorded that in 2021 the number of diabetes sufferers reached 19.5 million people and is predicted to reach 28.6 million people in 2045. Although this figure is quite high, half of the sufferers are not aware of it, so this research was conducted to create an alternative that makes it easier to detect diabetes and is expected to provide early warning so that people can immediately improve their lifestyle to avoid diabetes.*

*Random forests are one method for classification. In this study, the random forest method was used which was trained using 10 data distribution scenarios (90:10, 80:20, 70:30, 60:40, 50:50) and the data has been balanced with SMOTE. SMOTE itself is a method to overcome data imbalance. To conduct data analysis, researchers used the CRISP-DM technique which has stages of Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment.*

*The best evaluation was obtained in the 5th scenario of the Random Forest model using SMOTE in predicting diabetes disease which was tested using 100% medical data. For Precision class "0" 67% and class "1" 94%, for Recall class "0" 67% and class "1" 94%, for F1-Score class "0" 67% and class "1" 94%, and Accuracy value 90%. The results of the model are uploaded and made into a website for diabetes prediction.*

**Keywords:** CRISP-DM, Prediction, Random Forest, Smote, Website.