

**OPTIMASI PERFORMA RANDOM FOREST DENGAN RANDOM
OVERSAMPLING DAN SMOTE PADA DATASET DIABETES**

JALUR SCIENTIST

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



disusun oleh

HASBI

20.11.3778

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**OPTIMASI PERFORMA RANDOM FOREST DENGAN RANDOM
OVERSAMPLING DAN SMOTE PADA DATASET DIABETES**

JALUR SCIENTIST

untuk memenuhi salah satu syarat mencapai derajat
Sarjana Program Studi S1 Informatika



disusun oleh

HASBI

20.11.3778

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

JALUR SCIENTIST

**OPTOMASI PERFORMA RANDOM FOREST DENGAN RANDOM
OVERSAMPLING DAN SMOTE PADA DATASET DIABETES**

yang disusun dan diajukan

Hasbi

20.11.3778

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 19 Februari 2025

Dosen Pembimbing,



Theopilus Bayu Sasongko, S.Kom, M.Eng.
NIK. 190302375

HALAMAN PENGESAHAN

JALUR SCIENTIST
OPTIMASI PERFORMA RANDOM FOREST DENGAN RANDOM OVERSAMPLING
DAN SMOTE PADA DATASET DIABETES

yang disusun dan diajukan oleh

Hasbi

20.11.3778

Telah dipertahankan di depan Dewan Penguji pada tanggal 19 febuari 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Arifvanto Hadinegoro, S. Kom., M.T.
NIK.190302289

Lukman, S. Kom., M.Kom.
NIK.190302151

Theophilus Bayu Sasongko, S. Kom., M.Eng.
NIK.190302375

Skripsi ini telah diterima sebagai salah satu persyaratan untuk memperoleh gelar Sarjana

Komputer

Tanggal 19 Februari 2025

DEKAN FAKULTAS ILMU KOMPUTER



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

NIK. 190302096

HALAMAN PERNYATAAN KONTRIBUSI KARYA

Yang bertandatangan di bawah ini,

Nama mahasiswa : Hasbi

NIM : 20.11.3778

Menyatakan bahwa karya dengan judul berikut:

OPTIMASI PERFORMA RANDOM FOREST DENGAN RANDOM OVERSAMPLING DAN SMOTE PADA DATASET DIABETES

Dosen Pembimbing : Theopilus Bayu Sasongko, 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, 19 Februari 2025

Yang Menyatakan,



Hasbi

HALAMAN PERSEMBAHAN

Dengan menyebut Nama Allah Yang Maha Pengasih Lagi Maha Penyayang. Segala Puji Bagi Allah Tuhan Semesta Alam yang telah memberikan saya rahmat dan segala kemudahan dalam berbagai hal. Sebagai ungkapan rasa syukur dan terima kasih, laporan non-skripsi ini saya persembahkan kepada seluruh pihak yang telah mendukung saya, yaitu:

1. Kepada kedua orang tua saya yaitu ibunda Haidah dan almarhum ayahanda Jainudin yang senantiasa mengingatkan dan mendoakan untuk kemudahan Dan senantiasa mendukung dalam berbagai hal yang positif.
2. Kepada Bapak Theopilus Bayu Sasongko, S. Kom., M. Eng selaku dosen pembimbing yang telah memberikan arahan yang baik kepada saya.
3. Kepada teman saya yang telah menjadi bagian dari tempat untuk tukar pikiran dalam menyelesaikan suatu hal.
4. Serta untuk diri saya sendiri yang telah bersabar untuk keluar dari kesulitan dan mau terus untuk disiplin.

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah SWT yang telah memberikan Rahmat dan Karunia-Nya sehingga penulis mampu menyelesaikan laporan skripsi yang berjudul **“Optimasi Performa Random Forest dengan Random Oversampling dan SMOTE pada Dataset Diabetes”**. Laporan skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer Universitas Amikom Yogyakarta. Selama proses penyusunan tugas akhir ini, penulis menyadari bahwa terdapat berbagai kendala, tetapi dengan adanya dorongan, bantuan, bimbingan, saran, dan doa dari berbagai pihak, penulis mampu menyelesaikan laporan skripsi ini. Oleh karena itu, penulis menyampaikan terima kasih kepada:

1. Allah SWT yang telah memberikan Rahmat, Hidayah, serta Petunjuk-Nya sehingga penulis bisa melewati segala rintangan dan halangan.
2. Almarhum Bapak Zainudin dan Ibu Haidah selaku kedua orang tua penulis yang telah mendidik dan memfasilitasi segala keperluan yang diperlukan oleh penulis.
3. Almarhum Sandri, dan Ariadi selaku kakak-kakak penulis yang memberikan dukungan penuh kepada penulis serta membantu membiayai kebutuhan penulis.
4. Ibu sari sekeluarga selaku kerabat dekat penulis yang selalu memberikan motivasi serta doa kepada penulis.
5. Seluruh keluarga besar penulis yang tidak dapat disebutkan satu persatu yang selalu memberikan motivasi dan doa kepada penulis.
6. Windha Mega Pradnya Duhita, S.Kom., M.Kom. selaku dosen wali yang telah memberikan pengarahan kepada penulis sejak awal hingga akhir masa perkuliahan.
7. Bapak Theopilus Bayu Sasongko, S.Kom, M.Eng. selaku dosen pembimbing yang telah memberikan ilmu dan bimbingan serta solusi kepada penulis

8. Seluruh bapak dan ibu dosen serta staff karyawan Universitas Amikom Yogyakarta yang telah memberikan ilmu dan bimbingan kepada penulis.
9. Rasmi selaku wanita hebat yang selalu menemani penulis dan memberikan semangat serta bantuan kepada penulis.
10. Teman-teman seperjuangan dari kelas 20 IF 08 yang tidak dapat disebutkan satu persatu yang selalu membantu dan menghibur penulis serta berbagi pengalaman dan ilmu yang bermanfaat kepada penulis.
11. Teman-teman dekat penulis yang tidak dapat disebutkan satu persatu yang telah mendukung dan mendoakan penulis.

Penulis menyadari bahwa dalam penyusunan laporan skripsi ini masih jauh dari kata sempurna. Oleh karena itu, penulis memohon maaf atas segala kekurangan dalam penulisan laporan skripsi ini. Semoga laporan skripsi ini dapat memberikan manfaat bagi penulis maupun pembaca. Akhir kata, semoga Allah SWT membalas segala kebaikan semua pihak yang telah berkontribusi dalam proses penyusunan

Yogyakarta, 19 Februari 2025

Penulis

Hasbi

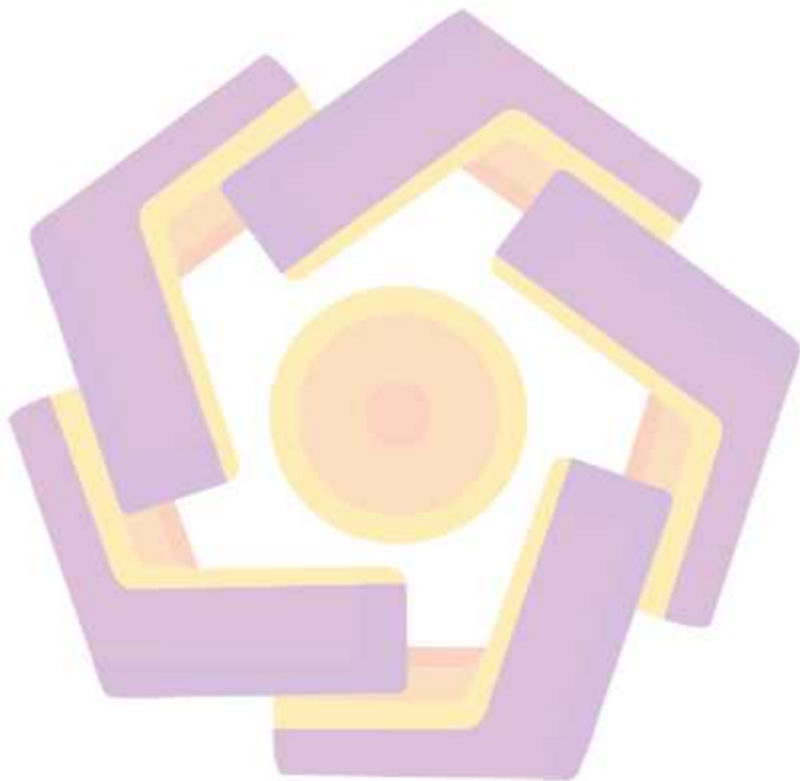
DAFTAR ISI

JALUR SCIENTIST	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KONTRIBUSI KARYA	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
INTISARI	xiii
ABSTRACT	xiv
IDENTITAS JURNAL	xv
BAB I	1
PENDAHULUAN	1
1.1 Gambaran Umum	1
1.2 Rumusan masalah	4
1.3 Batasan masalah	4
1.4 Tujuan	5
BAB II	6
TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Landasan Teori	7
BAB III	9
METODE PENELITIAN	9
3.1 Tahapan Penelitian	9
3.2 Akuisisi Data	10
3.3 Preprocessing data	10
3.4 Balancing data	11
3.5 Normalisasi data	11
3.6 Klasifikasi	11

3.7 Evaluasi	12
BAB IV	14
HASIL DAN PEMBAHASAN.....	14
4.1 Dataset.....	14
4.2 Preprocessing data	14
4.3 Balancing data	16
4.4 Normalisasi data	17
4.5 Split data.....	18
4.6 Klasifikasi dan Pengaturan Parameter.....	19
4.7 Evaluasi.....	19
BAB V.....	24
KESIMPULAN	24
5.1 Kesimpulan.....	24
5.2 Saran.....	24
REFERENCES.....	25
LAMPIRAN DAN BUKTI PENDUKUNG.....	27
A. Lampiran Lembaran review.....	27
B. Bukti Terbit/Terindex.....	34
C. Bukti Pembayaran	37

DAFTAR TABEL

<u>Tabel 1.1. Identitas Jurnal</u>	xv
<u>Tabel 1.2 Confusion Matrix</u>	12
<u>Tabel 1.3 Split data</u>	19
<u>Tabel 3. Hasil Pengaturan Parameter Metode Random Forest</u>	19



DAFTAR GAMBAR

<u>Gambar 1. Alur Penelitian</u>	9
<u>Gambar 2. Dataset</u>	14
<u>Gambar 3. Missing value</u>	15
<u>Gambar 4. Pengubahan Missing Value</u>	15
<u>Gambar 5. Hasil Pengubahan Missing Value</u>	16
<u>Gambar 6. Hasil Balancing data menggunakan SMOTE</u>	16
<u>Gambar 7. Hasil Balancing data menggunakan Random Oversampling</u>	17
<u>Gambar 8. Normalisasi data pada dataset setelah dilakukan Smote</u>	18
<u>Gambar 9. Normalisasi data pada dataset setelah dilakukan Random Oversampling</u>	18
<u>Gambar 10. Hasil confusion matrix pada dataset smote</u>	20
<u>Gambar 11. Hasil confusion matrix pada dataset Random Oversampling</u>	21
<u>Gambar 12. Nilai Tuning Parameter</u>	22
<u>Gambar 13. Hasil akurasi</u>	23

DAFTAR LAMPIRAN

<u>Lampiran 1. Lembaran Review</u>	27
<u>Lampiran 2. Catatan Reviewer A</u>	30
<u>Lampiran 3. Catatan Revisi Reviewer B</u>	33
<u>Lampiran 4. Lembaran Review</u>	34
<u>Lampiran 5. Bukti Karya Ilmiah Terbit di JMIB</u>	35
<u>Lampiran 6. MIB terindeks Sinta 3</u>	36
<u>Lampiran 7. MIB terindek Google Scholar</u>	36
<u>Lampiran 8. MIB terindeks Garuda Kemdikbud</u>	36



INTISARI

Diabetes adalah penyakit kronis serius yang menempati peringkat ketiga penyebab kematian di Indonesia setelah stroke dan penyakit jantung, dengan jumlah penderita sekitar 10 juta orang yang diperkirakan akan meningkat dua hingga tiga kali lipat dalam 10 tahun mendatang. Penyakit ini ditandai dengan kadar gula darah tinggi, memiliki gejala seperti sering buang air kecil, rasa haus berlebihan, kelelahan, dan luka yang sulit sembuh. Komplikasi serius meliputi Neuropati, Retinopati, penyakit jantung, dan gangguan pencernaan. Penelitian ini menggunakan teknologi Machine Learning untuk meningkatkan akurasi diagnosis diabetes, memanfaatkan *Pima Indians Diabetes Dataset* yang memiliki masalah ketidakseimbangan data (imbalance class). Untuk mengatasi masalah ini, diterapkan metode **SMOTE** dan **Random Oversampling**. Algoritma **Random Forest** dipilih untuk klasifikasi karena kemampuannya dalam mengurangi overfitting dan menghasilkan prediksi yang lebih akurat. Penelitian sebelumnya menunjukkan akurasi terbaik sebesar 95,96% menggunakan algoritma Decision Tree, namun belum menangani masalah ketidakseimbangan data atau skala data yang berbeda. Penelitian ini bertujuan untuk mengatasi kekurangan tersebut dan menghasilkan model klasifikasi diabetes yang lebih akurat. Hasil akhir dari penelitian ini diharapkan mampu memberikan kontribusi dalam pengembangan metode diagnosis diabetes yang lebih efektif, mendukung layanan kesehatan, dan mencegah komplikasi berbahaya.

Kata Kunci: Random Forest, Random Oversampling, SMOTE, Diabetes.

ABSTRACT

Diabetes is a serious chronic disease that ranks third as the cause of death in Indonesia after stroke and heart disease, with the number of sufferers around 10 million people which is expected to increase two to three times in the next 10 years. This disease is characterized by high blood sugar levels, has symptoms such as frequent urination, excessive thirst, fatigue, and wounds that are difficult to heal. Serious complications include Neuropathy, Retinopathy, heart disease, and digestive disorders. This study uses Machine Learning technology to improve the accuracy of diabetes diagnosis, utilizing the Pima Indians Diabetes Dataset which has a data imbalance problem (imbalance class). To overcome this problem, the SMOTE and Random Oversampling methods are applied. The Random Forest algorithm was chosen for classification because of its ability to reduce overfitting and produce more accurate predictions. Previous studies have shown the best accuracy of 95.96% using the Decision Tree algorithm, but have not addressed the problem of data imbalance or different data scales. This study aims to overcome these shortcomings and produce a more accurate diabetes classification model. The final results of this study are expected to be able to contribute to the development of more effective diabetes diagnosis methods, support health services, and prevent dangerous complications. Keywords: Random Forest, Random Oversampling, SMOTE, Diabetes

Kata Kunci: Random Forest, Random Oversampling, SMOTE, Diabetes.

