

**PERBANDINGAN PERFORMA ALGORITMA MACHINE LEARNING
DALAM MENDETEKSI PENYAKIT DIABETES**

SKRIPSI

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



disusun oleh

SITI QALAM LATUCONSINA

21.11.4405

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**PERBANDINGAN PERFORMA ALGORITMA MACHINE LEARNING
UNTUK MENDETEKSI PENYAKIT DIABETES**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

SITI QALAM LATUCONSINA

21.11.4405

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

PERBANDINGAN PERFORMA ALGORITMA MACHINE LEARNING UNTUK MENDETEKSI PENYAKIT DIABETES

yang disusun dan diajukan oleh

Siti Qalam Latuconsina

21.11.4405

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 21 Februari 2025

Dosen Pembimbing,


Robert Marco, S.T., M.T., Ph.D
NIK. 190302228

HALAMAN PENGESAHAN

SKRIPSI

PERBANDINGAN PERFORMA ALGORITMA MACHINE LEARNING UNTUK MENDETEKSI PENYAKIT DIABETES

yang disusun dan diajukan oleh

Siti Qalam Latuconsina

21.11.4405

Telah dipertahankan di depan Dewan Penguji
pada tanggal 21 Februari 2025

Susunan Dewan Penguji

Nama Penguji

Anna Baita, S.Kom., M.Kom.
NIK. 190302290

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

Robert Marco, S.T., M.T., Ph.D.
NIK. 190302228

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 21 Februari 2025

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Siti Qalam Latuconsina
NIM : 21.11.4405

Menyatakan bahwa Skripsi dengan judul berikut:

Perbandingan Performa Algoritma Machine Learning untuk Mendeteksi Penyakit Diabetes

Dosen Pembimbing : Robert Marco, S.T., M.T., Ph.D

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, 21 Februari 2025

Yang Menyatakan,



10000
METERAI
TEMPER
E2895AMX197618488

Siti Qalam Latuconsina

MOTTO

“Boleh jadi kamu membenci sesuatu, padahal ia amat baik bagimu, dan boleh jadi pula kamu menyukai sesuatu, padahal dia amat buruk bagimu. Allah lebih mengetahui sedangkan kamu tidak mengetahui”

-QS. Al-Baqarah: 216

“Orang tua di rumah menanti kepulanganmu dengan hasil yang membanggakan, jangan kecewakan mereka. Simpan keluhmu, sebab letihmu tak sebanding dengan perjuangan mereka menghadapimu”

HALAMAN PERSEMBAHAN

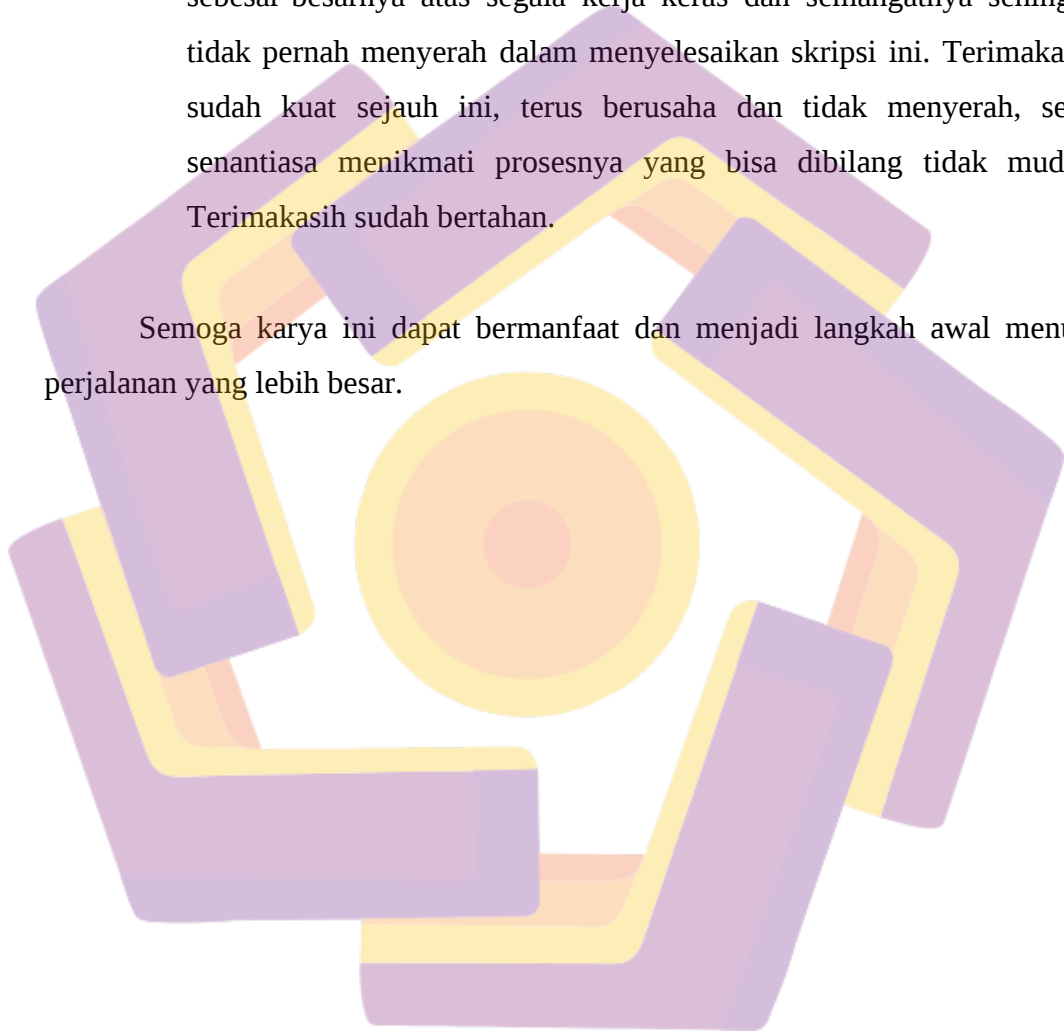
Dengan penuh rasa syukur dan kerendahan hati, karya ini saya persembahkan kepada:

1. Terimakasih kepada kedua orang tua tercinta yakni Ayahanda Fauzi Latuconsina dan Ibunda Boki Karepesina. Terimakasih atas setiap tetes keringat dalam setiap langkah pengorbanan dan kerja keras yang dilakukan untuk memberikan yang terbaik kepada penulis, mengusahakan segala kebutuhan penulis, mendidik, membimbing, dan selalu memberikan kasih sayang yang tulus, motivasi, serta dukungan dan mendoakan penulis dalam keadaan apapun agar penulis mampu bertahan untuk melangkah sedikit demi sedikit dalam meraih mimpi di masa depan. Terimakasih untuk selalu berada disisi penulis dan menjadi alasan bagi penulis dalam menyelesaikan penulisan skripsi ini.
2. Terimakasih kepada kakak dan adik saya Abd Rahman Bada Latuconsina, Abdul Qadri Latuconsina, Ramdhani Abbas Latuconsina, Hana Nazira Latuconsina, Risqi Al Fatah Latuconsina, yang selalu mendoakan, memberikan semangat serta berkontribusi dalam proses karya tulis baik tenaga maupun waktu untuk penulis. Terimakasih telah mendukung, menghibur, mendengar keluh kesah, dan mengajarkan arti kesabaran.
3. Dosen pembimbing saya, bapak Robert Marco, S.T., M.T., Ph.D., Terimakasih atas bimbingan, semangat dan kesabarannya yang luar biasa sehingga skripsi ini dapat terselesaikan.
4. Sahabat-sahabat saya yang tak kalah penting kehadirannya yaitu acit, fira, pida, gina, titin, nata, anggi, ady, yang selalu ada dalam suka maupun duka, berbagi motivasi, canda, tawa, serta semangat dan waktu yang telah diluangkan untuk menemani proses penulis hingga skripsi ini terselesaikan.
5. Teman-teman seperjuangan saya selama menempuh pendidikan dan merantau yaitu fira, tiara, rida, syahira, ayu, dyan, titi, vika, putri, elsa,

rara, sinta, cika, nurul, naya, divka, eci dan masih banyak lagi yang tidak cukup disebutkan satu persatu. Terimakasih sudah setia menemani dan mendukung penulis dari dari awal maba hingga semester akhir, baik saat senang maupun susah. Semoga kita semua menjadi orang yang sukses.

6. Siti Qalam Latuconsina, last but no least! Diri saya sendiri. Apresiasi sebesar-besarnya atas segala kerja keras dan semangatnya sehingga tidak pernah menyerah dalam menyelesaikan skripsi ini. Terimakasih sudah kuat sejauh ini, terus berusaha dan tidak menyerah, serta senantiasa menikmati prosesnya yang bisa dibilang tidak mudah. Terimakasih sudah bertahan.

Semoga karya ini dapat bermanfaat dan menjadi langkah awal menuju perjalanan yang lebih besar.



KATA PENGANTAR

Puji syukur kepada Allah SWT yang telah melimpahkan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan proposal penelitian skripsi ini yang berjudul **“Perbandingan Performa Algoritma Machine Learning dalam Mendeteksi Penyakit Diabetes”**. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Informatika, di Universitas AMIKOM Yogyakarta. Penyusunan skripsi ini tidak terlepas dari dukungan, bimbingan, serta bantuan dari berbagai pihak. Oleh karena itu, dengan segala hormat penulis mengucapkan terima kasih kepada :

1. Bapak Hanif Al Fatta, S.Kom., M.Kom., Ph.D., selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
2. Bapak Robert Marco, S.T., M.T., Ph.D., selaku dosen pembimbing yang telah memberikan bimbingan, arahan, dan motivasi selama proses penelitian dan penulisan skripsi ini.
3. Tim Dosen Penguji, yang telah memberikan masukan serta saran yang sangat berarti dalam penyempurnaan skripsi ini.
4. Seluruh Dosen dan Staf Universitas AMIKOM Yogyakarta, yang telah memberikan ilmu dan pengalaman berharga selama masa perkuliahan.
5. Orang tua dan keluarga tercinta, yang selalu memberikan doa, dukungan, serta semangat yang tiada henti dalam menyelesaikan studi ini.

Yogyakarta, 21 Februari 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
MOTTO	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xiv
INTISARI	xv
ABSTRACT.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6

2.2	Dasar Teori.....	13
2.2.1	Diabetes.....	13
2.2.2	Machine Learning	13
2.2.3	SMOTE	15
2.2.4	Hyperparameter Tuning	16
2.2.5	K-Fold Cross Validation	16
2.2.6	Confusion Matrix	17
BAB III	METODE PENELITIAN	20
3.1	Objek Penelitian.....	20
3.2	Alur Penelitian	20
3.3	Alat dan Bahan.....	26
3.3.1	Alat.....	26
BAB IV	HASIL DAN PEMBAHASAN	28
4.1	Proses Pengumpulan Data.....	28
4.2	Eksplorasi Data (EDA)	28
4.3	Pre-processing.....	31
4.4	Modeling	36
4.5	Evaluasi.....	41
BAB V	PENUTUP	42
5.1	Kesimpulan	42
5.2	Saran	43
REFERENSI		44
LAMPIRAN		48

DAFTAR TABEL

Tabel 2.1. Keaslian Penelitian	10
Tabel 2.2. Confusion Matrix	17
Tabel 3.1. Deskripsi Data Diabetes	21
Tabel 3.2. Dataset Sebelum Pre-processing	22
Tabel 3.3. Dataset Setelah Pre-processing	23
Tabel 3.4. Parameter Algoritma yang digunakan	24
Tabel 4.1. Perbandingan Performa Model dengan K-Fold Cross-Validation	37
Tabel 4.2. Perbandingan Performa Model dengan Confusion Matrix	39

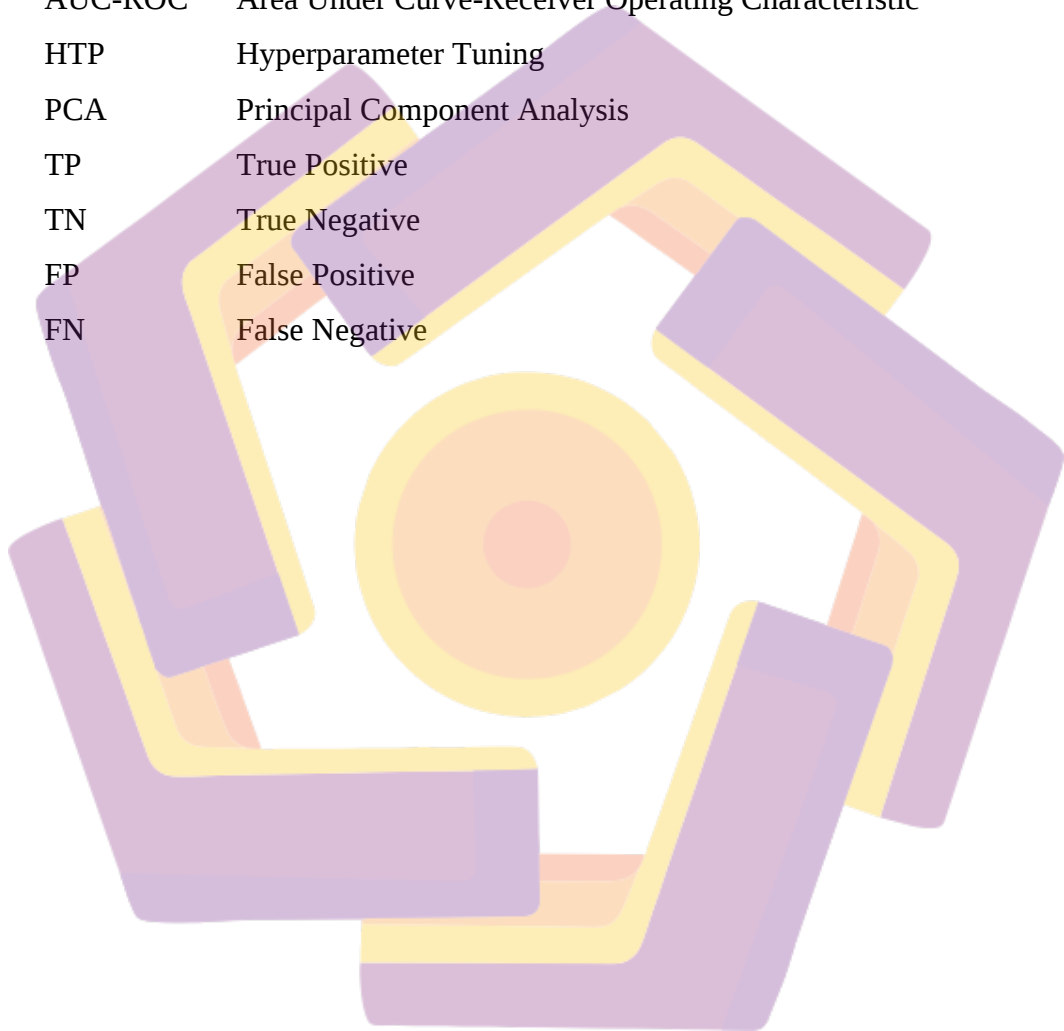


DAFTAR GAMBAR

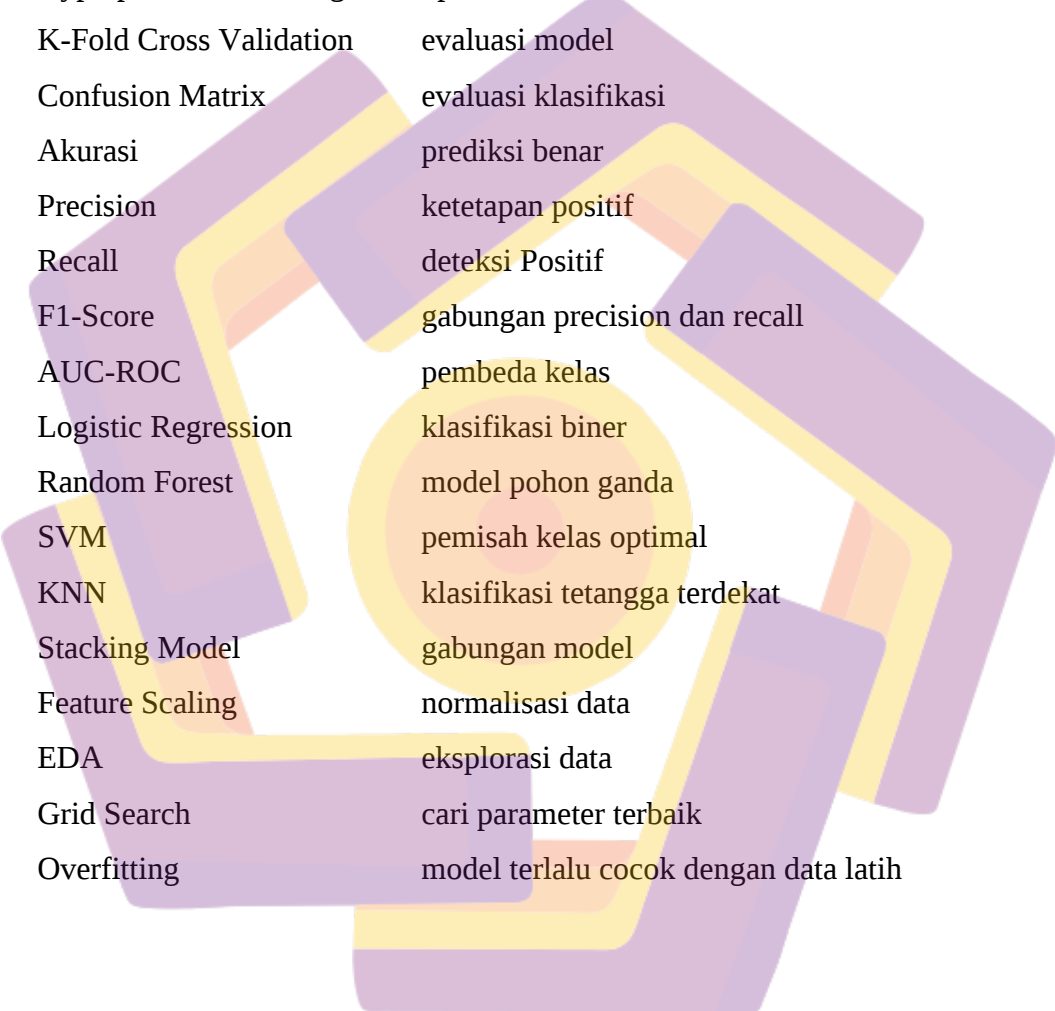
Gambar 3.1. Alur Penelitian	20
Gambar 4.1. Deskripsi Dataset	28
Gambar 4.2. Visualisasi Data menggunakan Histogram	29
Gambar 4.3. Analisis Boxplot	30
Gambar 4.4. Heatmap Korelasi Fitur	31
Gambar 4.5. Tampilan Output Dataset	32
Gambar 4.6. Handling Missing Value	33
Gambar 4.7. Handling Duplicated	33
Gambar 4.8. Handling Outlier	34
Gambar 4.9. Normalisasi Data	34
Gambar 4.10. Sebelum Penanganan Ketidakseimbangan Data	35
Gambar 4.11. Sesudah Penanganan Ketidakseimbangan Data	36
Gambar 4.12. Splitting Data Train dan Data Test	36
Gambar 4.13. Data diubah menjadi Array Numpy	36
Gambar 4.14. Penyetelan Hyperparameter Tuning	37
Gambar 4.15. Confusion Matrix Logistic Regression	38
Gambar 4.16. Confusion Matrix Random Forest	38
Gambar 4.17. Confusion Matrix SVM	39
Gambar 4.18. Confusion Matrix KNN	39
Gambar 4.19. Confusion Matrix Stacking Model	39
Gambar 4.20. Test Accuracy Comparison Between Algorithms	41

DAFTAR LAMBANG DAN SINGKATAN

SMOTE	Synthetic Minority Over-sampling Technique
SVM	Support Vector Machines
KNN	K-Nearest Neighbor
AUC-ROC	Area Under Curve-Receiver Operating Characteristic
HTP	Hyperparameter Tuning
PCA	Principal Component Analysis
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative



DAFTAR ISTILAH



Diabetes Mellitus	penyakit dengan kadar gula darah tinggi
Machine Learning	AI yang memungkinkan komputer belajar dari data
SMOTE	metode untuk menangani ketidakseimbangan data
Hyperparameter Tuning	optimasi model
K-Fold Cross Validation	evaluasi model
Confusion Matrix	evaluasi klasifikasi
Akurasi	prediksi benar
Precision	ketetapan positif
Recall	deteksi Positif
F1-Score	gabungan precision dan recall
AUC-ROC	pembeda kelas
Logistic Regression	klasifikasi biner
Random Forest	model pohon ganda
SVM	pemisah kelas optimal
KNN	klasifikasi tetangga terdekat
Stacking Model	gabungan model
Feature Scaling	normalisasi data
EDA	eksplorasi data
Grid Search	cari parameter terbaik
Overfitting	model terlalu cocok dengan data latih

INTISARI

Penyakit diabetes merupakan salah satu penyakit kronik yang memiliki prevalensi tinggi di seluruh dunia. Deteksi dini penyakit ini sangat penting untuk mengurangi risiko komplikasi yang lebih parah. Dalam penelitian ini, dilakukan perbandingan performa beberapa algoritma machine learning dalam mendeteksi penyakit diabetes menggunakan dataset Pima Indian Diabetes yang tersedia di Kaggle. Algoritma yang diuji meliputi Logistic Regression, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbor (KNN) dan Stacking Model. Untuk mengatasi permasalahan ketidakseimbangan data, diterapkan metode Synthetic Minority Over-sampling Technique (SMOTE), sedangkan optimasi model dilakukan menggunakan Hyperparameter Tuning. Hasil penelitian ini menunjukkan bahwa algoritma Random Forest dan Stacking Model memberikan performa terbaik dibandingkan algoritma lainnya. Stacking Model mencapai akurasi tertinggi sebesar 82,5%, diikuti oleh Random Forest dengan akurasi 81,5%. Model ini juga memiliki nilai Area Under Curve-Receiver Operating Characteristic (AUC-ROC) tertinggi, yang menunjukkan kemampuannya dalam membedakan hasil ini, penggunaan Stacking Model dapat direkomendasikan untuk implementasi sistem deteksi dini diabetes yang lebih optimal.

Kata kunci: Machine Learning, Diabetes, SMOTE, Hyperparameter Tuning, Stacking Model.

ABSTRACT

Diabetes is a chronic disease that has a high prevalence worldwide. Early detection of this disease is very important to reduce the risk of more severe complications. In this study, we compared the performance of several machine learning algorithms in detecting diabetes using the Pima Indian Diabetes dataset available on Kaggle. The algorithms tested include Logistic Regression, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Stacking Model. To overcome the problem of data imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) method is applied, while model optimization is carried out using Hyperparameter Tuning. The results of this study show that the Random Forest algorithm and Stacking Model provide the best performance compared to other algorithms. Stacking Model achieved the highest accuracy of 82.5%, followed by Random Forest with 81.5% accuracy. This model also has the highest Area Under Curve-Receiver Operating Characteristic (AUC-ROC) value, which shows its ability to distinguish these results, the use of the Stacking Model can be recommended for the implementation of a more optimal diabetes early detection system.

Keyword: Machine Learning, Diabetes, SMOTE, Hyperparameter Tuning, Stacking Model