

**ANALISIS PREDIKSI RISIKO OSTEOPOROSIS BERBASIS
MACHINE LEARNING MENGGUNAKAN ALGORITMA
GRADIENT BOOSTING**

LAPORAN NON-REGULER

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



Disusun oleh :

EDWIN IBRAHIM SALIM

22.11.5315

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**ANALISIS PREDIKSI RISIKO OSTEOPOROSIS BERBASIS
MACHINE LEARNING MENGGUNAKAN ALGORITMA
GRADIENT BOOSTING**

LAPORAN NON-REGULER

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika



Disusun oleh :

EDWIN IBRAHIM SALIM

22.11.5315

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

JALUR NON-REGULER

**ANALISIS PREDIKSI RISIKO OSTEOPOROSIS BERBASIS
MACHINE LEARNING MENGGUNAKAN ALGORITMA
GRADIENT BOOSTING**

yang disusun dan diajukan oleh

Edwin Ibrahim Salim

22.11.5315

telah disetujui oleh Dosen Pembimbing
pada tanggal 17 Desember 2025

Dosen Pembimbing,



Majid Rahardi, S.kom., M.Eng
NIK. 190302393

HALAMAN PENGESAHAN

JALUR NON-REGULER

**ANALISIS PREDIKSI RISIKO OSTEOPOROSIS BERBASIS
MACHINE LEARNING MENGGUNAKAN ALGORITMA
GRADIENT BOOSTING**

yang disusun dan diajukan oleh

Edwin Ibrahim Salim
22.11.5315

Telah dipertabankan di depan Dewan Penguji
pada tanggal 17 Desember 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Norhikmah, S.Kom., M.Kom.
NIK. 190302245



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



Majid Rahardi, S.Kom., M.Eng.
NIK. 190302393



Laporan ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 17 Desember 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, S.Kom., M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN KARYA

Yang bertandatangan di bawah ini,

Nama mahasiswa : Edwin Ibrahim Salim

NIM : 22.11.5315

Menyatakan bahwa Laporan dengan judul berikut:

ANALISIS PREDIKSI RISIKO OSTEOPOROSIS BERBASIS MACHINE LEARNING MENGGUNAKAN ALGORITMA GRADIENT BOOSTING

Dosen Pembimbing : Majid Rahardi, 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 kegiatan 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 ketidak-benaran 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, 17 Desember 2025

Yang Menyatakan,



Edwin Ibrahim Salim

HALAMAN PERSEMBAHAN

Dengan rasa syukur yang mendalam, dengan telah di selesaikan laporan non-reguler ini penulis mempersembahkannya kepada:

1. Kedua orang tua terkasih, yang senantiasa menjadi sumber kekuatan dan inspirasi dalam setiap langkah kehidupan penulis. Doa yang tidak pernah putus, kasih sayang yang tidak terukur, serta dukungan moril maupun materiil yang diberikan merupakan dorongan terbesar bagi penulis untuk terus berusaha dan menyelesaikan laporan ini dengan sebaik-baiknya.
2. Seluruh civitas akademika kampus Universitas Amikom Yogyakarta, yang telah memberikan kesempatan serta menyediakan lingkungan belajar yang kondusif, fasilitas pendukung yang memadai, dan suasana akademik yang membangun. Tanpa bimbingan dan pelayanan yang baik dari berbagai pihak di kampus, perjalanan studi ini tidak akan berjalan dengan lancar.
3. Bapak Majid Rahardi, selaku dosen pembimbing yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan arahan, bimbingan, serta ilmu yang sangat berarti dalam penyusunan laporan ini. Nasihat, motivasi, dan perhatian beliau menjadi dorongan berharga bagi penulis untuk terus berupaya menghasilkan karya terbaik.
4. Pemilik NIM 227190, yang selalu hadir memberikan semangat tanpa henti, dukungan moral yang tulus, serta motivasi yang kuat dalam setiap tantangan yang dihadapi selama proses pengerjaan laporan ini. Kehadiran dan dukungan tersebut menjadi bagian penting dalam perjalanan akademik penulis.
5. Teman-teman seperjuangan, yang tidak hanya memberikan bantuan teknis dan pemikiran, tetapi juga kebersamaan, canda tawa, serta pengalaman berharga selama masa perkuliahan hingga terselesaikannya laporan ini. Setiap kenangan dan dukungan yang diberikan akan selalu menjadi bagian yang tak terlupakan.

Penulis menyadari bahwa laporan ini masih jauh dari kesempurnaan. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun demi perbaikan di masa mendatang. Semoga laporan ini dapat memberikan manfaat dan menjadi referensi bagi pihak yang membutuhkan.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat, karunia, dan kemudahan yang diberikan sehingga laporan non-reguler ini dapat diselesaikan dengan baik. Penyusunan laporan ini merupakan bagian dari pemenuhan persyaratan akademik yang harus ditempuh oleh penulis.

Dalam kesempatan ini, penulis menyampaikan penghargaan dan ucapan terima kasih kepada Bapak Majid Rahardi selaku Dosen Pembimbing, atas bimbingan, arahan, dan dukungan yang diberikan selama proses penelitian dan penyusunan laporan ini berlangsung. Penulis juga mengucapkan terima kasih kepada Tim Dosen Penguji yang telah memberikan masukan berharga guna penyempurnaan laporan ini.

Ucapan terima kasih penulis tujukan pula kepada kedua orang tua, yang selalu memberikan doa, dukungan, dan dorongan moral dalam setiap langkah yang penulis tempuh. Tidak lupa, penulis menyampaikan terima kasih kepada pihak kampus dan seluruh civitas akademika yang telah menyediakan fasilitas, lingkungan belajar, serta kesempatan bagi penulis untuk berkembang.

Akhirnya, penulis menyampaikan terima kasih kepada seluruh pihak yang telah membantu, baik secara langsung maupun tidak langsung, dalam proses penyelesaian laporan ini. Semoga laporan ini dapat memberikan manfaat bagi pihak yang membutuhkan dan menjadi bagian dari kontribusi penulis dalam bidang ilmu pengetahuan.

Yogyakarta, 17 Desember 2025

Penulis



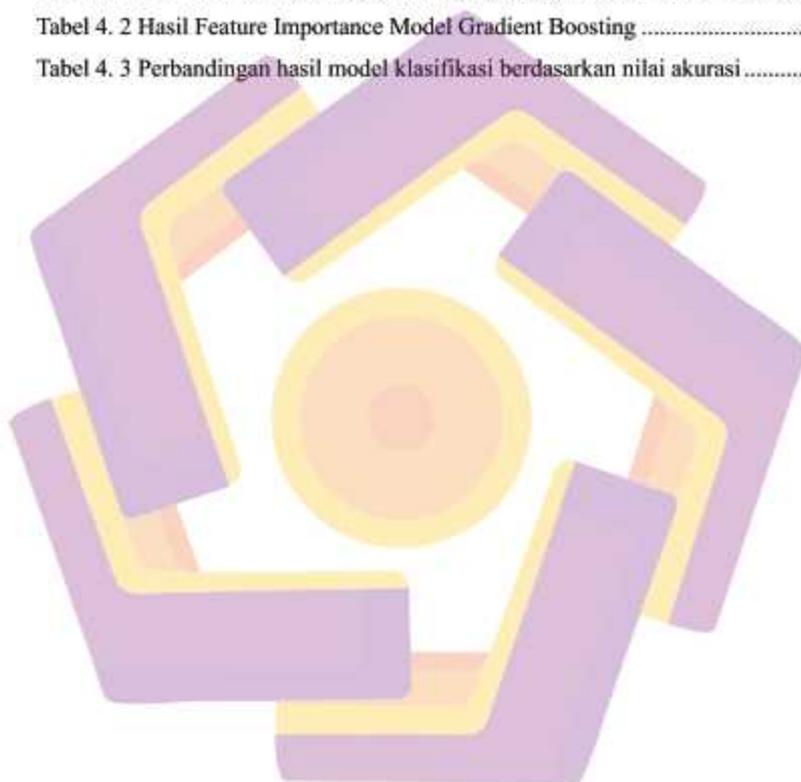
DAFTAR ISI

Halaman Judul.....	i
Halaman Persetujuan.....	ii
Halaman Pengesahan.....	iii
Halaman Pernyataan Keaslian Karya.....	iv
Halaman Persembahan.....	v
Kata Pengantar.....	vi
Daftar Isi.....	vii
Daftar Tabel.....	ix
Daftar Gambar.....	x
Daftar Lampiran.....	xi
Daftar Lambang dan Singkatan.....	xii
Daftar Istilah.....	xiv
Intisari.....	xvi
<i>Abstract</i>	xvii
Bab I Pendahuluan.....	1
1.1. Latar belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Batasan Masalah.....	4
1.4. Tujuan.....	4
Bab II Tinjauan Pustaka.....	6
2.1. Studi Literatur.....	6
2.2. Landasan Teori.....	7
BAB III Metode Penelitian.....	9
3.1. Akuisisi Data.....	10
3.2. Pra-Pemrosesan Data.....	13
3.3. Pembagian Data.....	14
3.4. Algoritma Gradient Boosting.....	16
3.5. Penyetelan Parameter (Hyperparameter Tuning).....	18
3.6. Evaluasi Model.....	20
BAB IV Pembahasan.....	22
4.1. Deskripsi Dataset dan Karakteristik Data.....	22

4.2. Hasil Pra-Pemrosesan Data	24
4.3. Hasil Pelatihan Model Gradient Boosting.....	26
4.4. Analisis Feature Importance	28
4.5. Perbandingan Model Klasifikasi	30
4.6. Implikasi Praktis dan Keterbatasan.....	31
4.7. Ringkasan Hasil	32
BAB V Kesimpulan	34
5.1. Kesimpulan	34
5.2. Saran.....	35
Referensi	37
Curriculum Vitae	39
Lampiran dan Bukti Pendukung.....	41
a. Letter of Acceptance (LOA).....	41
b. Lembar Review	42
c. Bukti Terbit/Terindex	51
d. Sertifikat sebagai Penyaji.....	51
e. Bukti pembayaran	52

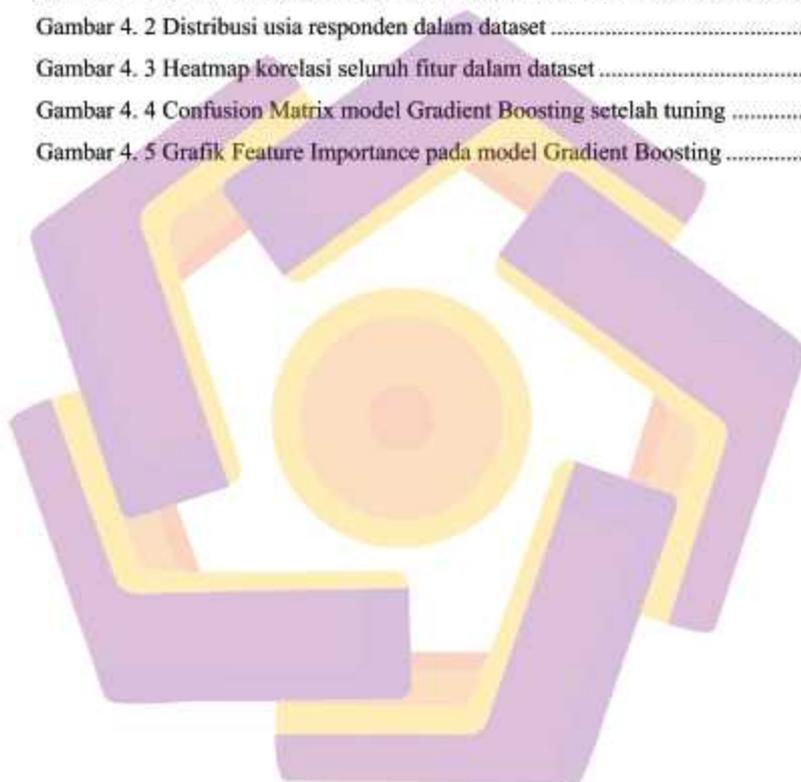
DAFTAR TABEL

Tabel 3. 1 Dataset.....	10
Tabel 3. 2 Sampel Dataset Osteoporosis Risk Prediction	11
Tabel 3. 3 Parameter Gradient Boosting	19
Tabel 4. 1 Hasil evaluasi model Gradient Boosting sebelum dan setelah tuning .	26
Tabel 4. 2 Hasil Feature Importance Model Gradient Boosting	28
Tabel 4. 3 Perbandingan hasil model klasifikasi berdasarkan nilai akurasi	30



DAFTAR GAMBAR

Gambar 3. 1 Flowchart/Alur Penelitian	9
Gambar 3. 2 Skema pembagian dataset menjadi data latih (80%) dan data uji (20%) serta tahapan pelatihan dan evaluasi model.	15
Gambar 4. 1 Distribusi Kelas Osteoporosis Sebelum SMOTE.....	23
Gambar 4. 2 Distribusi usia responden dalam dataset	24
Gambar 4. 3 Heatmap korelasi seluruh fitur dalam dataset	25
Gambar 4. 4 Confusion Matrix model Gradient Boosting setelah tuning	27
Gambar 4. 5 Grafik Feature Importance pada model Gradient Boosting	29



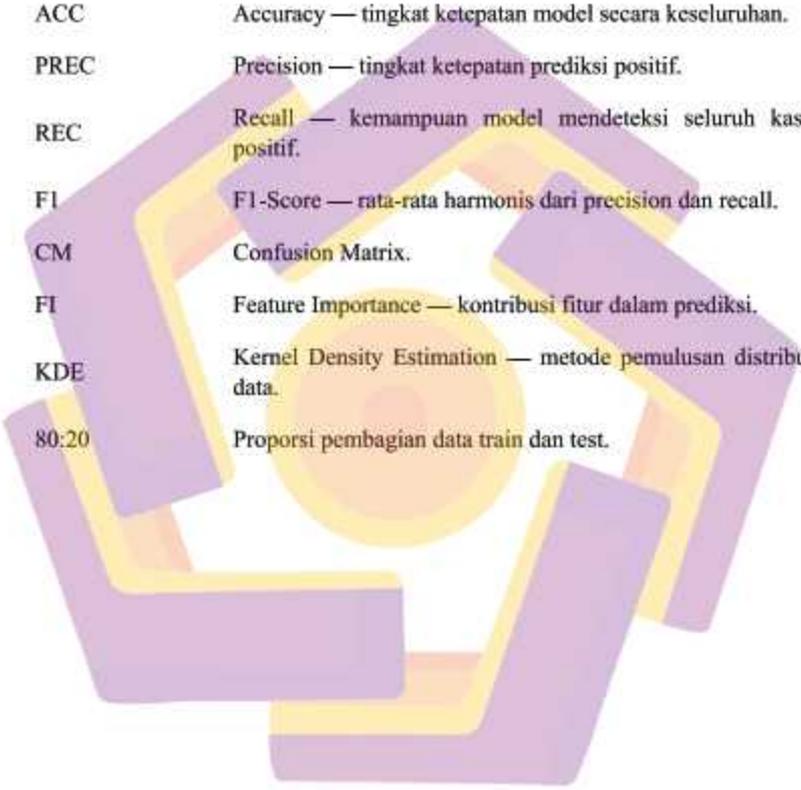
DAFTAR LAMPIRAN

Letter of Acceptance (LOA)	38
Lembar Review	39
Bukti Terbit/Terindex	43
Sertifikat sebagai Penyaji	43
Bukti Pembayaran	44



DAFTAR LAMBANG DAN SINGKATAN

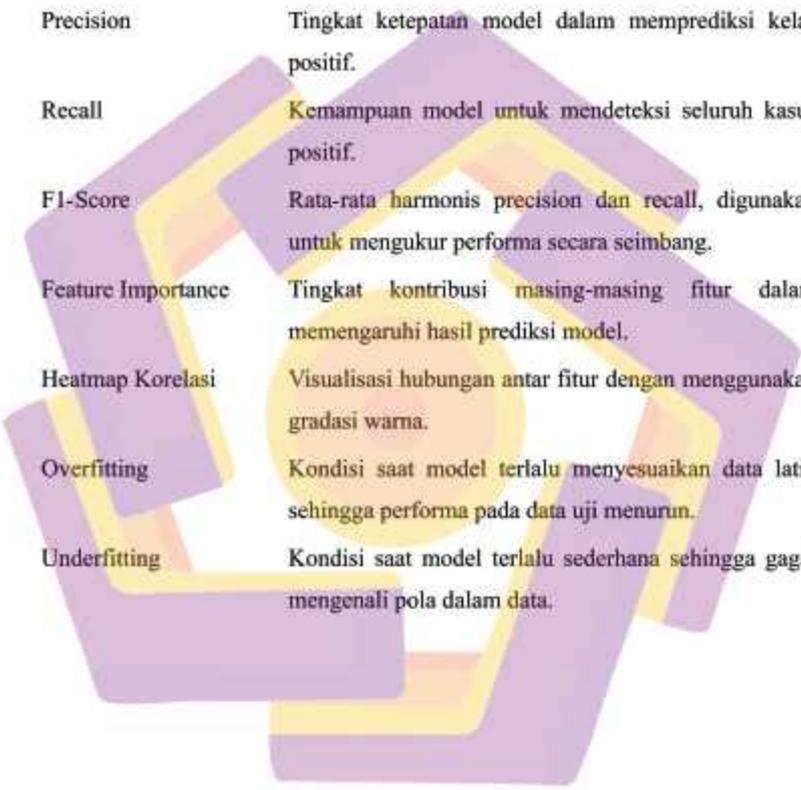
$F_0(x)$	Model awal sebelum proses boosting dimulai.
$F_m(x)$	Model pada iterasi ke- m dalam algoritma Gradient Boosting.
$h_m(x)$	Weak learner (pohon keputusan sederhana) pada iterasi ke- m .
r_{im}	Residual atau negatif gradien untuk sampel ke- i pada iterasi ke- m .
γ_m	Step size yang diperoleh melalui proses optimasi pada iterasi ke- m .
ν	Learning rate yang mengatur besar kontribusi dari setiap weak learner.
$L(y_i, \hat{y}^i)$	Fungsi loss yang digunakan dalam pembaruan model.
Arg min	Operator untuk mencari nilai yang meminimalkan suatu fungsi.
$\sum_{i=1}^n$	Simbol sigma, melambangkan penjumlahan dari $i = 1$ sampai n .
$\partial F(x_i) / \partial L$	Turunan parsial fungsi loss terhadap prediksi model.
n	Jumlah total sampel dalam dataset.
M	Jumlah iterasi atau jumlah weak learner dalam ensemble.
ML	Machine Learning.
AI	Artificial Intelligence (Kecerdasan Buatan).
GB	Gradient Boosting.
XGB	XGBoost (Extreme Gradient Boosting).
RF	Random Forest.
LR	Logistic Regression.
TP	True Positive — jumlah sampel positif yang diprediksi benar.



TN	True Negative — jumlah sampel negatif yang diprediksi benar.
FP	False Positive — sampel negatif yang salah diprediksi positif.
FN	False Negative — sampel positif yang salah diprediksi negatif.
ACC	Accuracy — tingkat ketepatan model secara keseluruhan.
PREC	Precision — tingkat ketepatan prediksi positif.
REC	Recall — kemampuan model mendeteksi seluruh kasus positif.
F1	F1-Score — rata-rata harmonis dari precision dan recall.
CM	Confusion Matrix.
FI	Feature Importance — kontribusi fitur dalam prediksi.
KDE	Kernel Density Estimation — metode pemulusan distribusi data.
80:20	Proporsi pembagian data train dan test.

DAFTAR ISTILAH

Dataset	Kumpulan data yang digunakan dalam penelitian untuk proses pelatihan dan pengujian model.
Fitur (Feature)	Variabel atau atribut dalam dataset yang digunakan sebagai input untuk model machine learning.
Label / Target	Nilai keluaran yang ingin diprediksi oleh model, yaitu status osteoporosis (1 = Ya, 0 = Tidak).
Pra-Pemrosesan Data	Proses pembersihan, normalisasi, encoding, dan persiapan data sebelum dilakukan pelatihan model.
Missing Values	Nilai kosong pada fitur tertentu yang memerlukan penanganan agar tidak mengganggu proses pemodelan.
Label Encoding	Metode mengubah data kategorikal menjadi numerik agar dapat diproses oleh algoritma machine learning.
Standardisasi	Proses menyamakan skala fitur numerik agar tidak terjadi dominasi nilai antarvariabel.
Train-Test Split	Pembagian dataset menjadi data latih dan data uji untuk mengevaluasi performa model.
Gradient Boosting	Algoritma ensemble yang membangun model secara bertahap untuk memperbaiki error dari model sebelumnya.
Ensemble Learning	Teknik menggabungkan beberapa model untuk meningkatkan akurasi prediksi.
Weak Learner	Model sederhana, seperti pohon keputusan kecil, yang digunakan dalam metode boosting.
Hyperparameter Tuning	Proses mencari kombinasi parameter terbaik agar model bekerja optimal.
RandomizedSearchCV	Metode pencarian hyperparameter secara acak dengan evaluasi menggunakan validasi silang.



Learning Rate	Parameter yang mengatur besar kontribusi setiap weak learner saat memperbarui model.
Confusion Matrix	Matriks evaluasi yang menunjukkan jumlah prediksi benar dan salah pada model klasifikasi.
Accuracy	Persentase prediksi benar dari seluruh sampel uji.
Precision	Tingkat ketepatan model dalam memprediksi kelas positif.
Recall	Kemampuan model untuk mendeteksi seluruh kasus positif.
F1-Score	Rata-rata harmonis precision dan recall, digunakan untuk mengukur performa secara seimbang.
Feature Importance	Tingkat kontribusi masing-masing fitur dalam memengaruhi hasil prediksi model.
Heatmap Korelasi	Visualisasi hubungan antar fitur dengan menggunakan gradasi warna.
Overfitting	Kondisi saat model terlalu menyesuaikan data latih sehingga performa pada data uji menurun.
Underfitting	Kondisi saat model terlalu sederhana sehingga gagal mengenali pola dalam data.

INTISARI

Osteoporosis merupakan penyakit degeneratif yang ditandai oleh penurunan massa tulang dan meningkatnya risiko patah tulang, sehingga diperlukan upaya deteksi dini yang efektif dan mudah diakses. Penelitian ini bertujuan untuk mengembangkan model prediksi risiko osteoporosis berbasis data pola hidup sebagai alternatif metode skrining awal non-radiologis. Data penelitian menggunakan dataset sekunder berjumlah 1.958 sampel yang terdiri atas 979 kasus positif dan 979 kasus negatif, dengan variabel yang mencakup usia, jenis kelamin, perubahan hormonal, riwayat keluarga, aktivitas fisik, asupan kalsium, vitamin D, kebiasaan merokok, konsumsi alkohol, kondisi medis, serta penggunaan obat. Proses penelitian meliputi pembersihan data, pengkodean fitur kategorikal, standarisasi, dan pembagian data menggunakan metode train-test split. Model dikembangkan menggunakan algoritma Gradient Boosting dengan penyetelan parameter melalui `RandomizedSearchCV`. Hasil penelitian menunjukkan bahwa model yang dihasilkan mencapai akurasi sebesar 91%, dengan nilai precision 0.99, recall 0.82, dan F1-score 0.90. Temuan ini menunjukkan bahwa Gradient Boosting memiliki kemampuan yang baik dalam memprediksi risiko osteoporosis berbasis data pola hidup dan berpotensi dimanfaatkan sebagai alat bantu skrining dini di fasilitas layanan kesehatan dasar.

Kata kunci: osteoporosis, prediksi risiko, machine learning, gradient boosting, pola hidup.

ABSTRACT

Osteoporosis is a degenerative disease characterized by decreased bone mass and a higher risk of fractures, emphasizing the need for early and accessible screening methods. This study aims to develop a machine learning-based model for predicting osteoporosis risk using lifestyle-related data as a non-radiological early screening alternative. The dataset used in this study consists of 1,958 samples, comprising 979 positive and 979 negative cases, with features including age, gender, hormonal changes, family history, physical activity, calcium and vitamin D intake, smoking habits, alcohol consumption, medical conditions, and medication use. The research process includes data cleaning, categorical feature encoding, standardization, and data splitting using the train-test split method. The Gradient Boosting algorithm was employed as the primary classification model, with hyperparameter optimization conducted using RandomizedSearchCV. The results show that the optimized model achieved an accuracy of 91%, along with a precision of 0.99, recall of 0.82, and an F1-score of 0.90. These findings indicate that the Gradient Boosting algorithm performs effectively in predicting osteoporosis risk based on lifestyle data and has the potential to be utilized as an early screening tool in primary healthcare services.

Keywords: osteoporosis, risk prediction, machine learning, gradient boosting, lifestyle factors.