OPTIMASI MODEL KLASIFIKASI STATUS OBESITAS MENGGUNAKAN VOTING ENSEMBLE ALGORITHM BERBASIS RANDOM FOREST DAN XGBOOST

SKRIPSI

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

disusun oleh

FARHAN PURNAMA PUTRA

21.11.4397

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2025

OPTIMASI MODEL KLASIFIKASI STATUS OBESITAS MENGGUNAKAN VOTING ENSEMBLE ALGORITHM BERBASIS RANDOM FOREST DAN XGBOOST

SKRIPSI

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

disusun oleh

FARHAN PURNAMA PUTRA

21.11.439

Kepada

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025

HALAMAN PERSETUJUAN

SKRIPSI

OPTIMASI MODEL KLASIFIKASI STATUS OBESITAS MENGGUNAKAN VOTING ENSEMBLE ALGORITHM BERBASIS RANDOM FOREST DAN XGBOOST

yang disusun dan diajukan oleh

Farhan Purnama Putra

21.11.4397

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 21 Januari 2025

Dosen Pembimbing,

Ferian Fauzi Abdulloh, M. Kom.
NIK. 190302276

HALAMAN PENGESAHAN

SKRIPSI

OPTIMASI MODEL KLASIFIKASI STATUS OBESITAS MENGGUNAKAN VOTING ENSEMBLE ALGORITHM BERBASIS RANDOM FOREST DAN XGBOOST

yang disusun dan diajukan oleh

Farhan Purnama Putra

21.11.4397

Telah dipertahankan di depan Dewan Penguji
pada tanggal 21 Januari 2025

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Haryoko, S. Kom., M.Cs. NIK. 190302286	
Ahlihi Masruro, S.Kom., M.Kom. NIK. 190302148	
Ferian Fauzi Abdulloh, S.Kom., M.Kom NIK. : 190302276	

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

DEKAN FAKULTAS ILMU KOMPUTER

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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Farhan Purnama Putra
NIM : 21.11.4397

Menyatakan bahwa Skripsi dengan judul berikut:

OPTIMASI MODEL KLASIFIKASI STATUS OBESITAS MENGGUNAKAN VOTING ENSEMBLE ALGORITHM BERBASIS RANDOM FOREST DAN XGBOOST

Dosen Pembimbing : Ferian Fauzi Abdulloh, 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, 21 Januari 2025

Yang Menyatakan,

Farhan Purnama Putra

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur, karya ini dipersembahkan kepada Allah SWT yang senantiasa melimpahkan rahmat, taufik, dan hidayah-Nya selama proses penyusunan skripsi ini. Kepada kedua orang tua tercinta, yang telah memberikan kasih sayang, doa, dukungan moral, dan materi tanpa henti, saya sampaikan rasa terima kasih yang mendalam.

Ucapan terima kasih juga saya tujukan kepada dosen pembimbing yang telah mendampingi saya dengan penuh dedikasi selama pengerjaan skripsi ini. Bimbingan, arahan, serta masukan yang tegas dan mudah dipahami yang telah disampaikan sangatlah berharga dalam penyelesaian penelitian ini.

Tidak lupa, saya juga berterima kasih kepada teman sejawat yang selalu memberikan dukungan, berbagi pengalaman, dan semangat selama masa perkuliahan hingga penyusunan skripsi. Kepada Universitas Amikom Yogyakarta serta seluruh dosen pengampu mata kuliah yang telah memberikan ilmu dan wawasan selama masa studi, saya haturkan penghargaan yang setinggi-tingginya.

Semoga karya ini dapat memberikan manfaat bagi semua pihak yang membacanya dan menjadi langkah kecil menuju kontribusi yang lebih besar bagi dunia pendidikan dan penelitian

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi ini dengan judul "Optimasi Model Klasifikasi Status Obesitas Menggunakan Voting Ensemble Algorithm Berbasis Random Forest Dan XGBoost". Penyelesaian skripsi ini tidak terlepas dari dukungan dan bantuan berbagai pihak yang penulis hormati.

Pada kesempatan ini, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Ferian Fauzi Abdulloh, M.Kom, selaku Dosen Pembimbing yang dengan sabar memberikan bimbingan, arahan, dan motivasi yang sangat berharga dalam proses penyusunan skripsi ini.
2. Bapak/Ibu Dosen Penguji, yang telah memberikan masukan yang sangat berguna dalam pengembangan penelitian ini.
3. Orang tua tercinta, yang selalu memberikan dukungan moral dan material tanpa henti sepanjang perjalanan pendidikan penulis.
4. Semua pihak yang tidak dapat disebutkan satu per satu, yang turut membantu baik langsung maupun tidak langsung dalam penyelesaian skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna, oleh karena itu penulis mengharapkan kritik dan saran yang membangun dari pembaca. Semoga skripsi ini dapat memberikan manfaat dan kontribusi bagi perkembangan ilmu pengetahuan, khususnya di bidang Informatika.

Yogyakarta, 18 Januari 2025

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 ..... **Error! Bookmark not defined.**

DAFTAR LAMBANG DAN SINGKATAN xi

DAFTAR ISTILAH xii

INTISARI xvii

ABSTRACT xviii

BAB I PENDAHULUAN 1

1.1 Latar Belakang 1

1.2 Rumusan Masalah 2

1.3 Batasan Masalah 2

1.4 Tujuan Penelitian 3

1.5 Manfaat Penelitian 4

1.6 Sistematika Penulisan 4

BAB II TINJAUAN PUSTAKA 5

2.1 Studi Literatur 5
2.2 Dasar Teori 13
2.2.1 Binning 13
2.2.2 Feature Scaling 13
2.2.3 Data Augmentation dengan SMOTE (Synthetic Minority Over-sampling Technique) 14
2.2.4 Algoritma Random Forest 16
2.2.5 Algoritma XGBoost 17
2.2.6 Hyperparameter Tuning 17
2.2.7 Voting Ensembel Learning Random Forest dan XGBoost (Hard) 18
2.2.8 Hyperparameter tuning pada model Voting Ensembel Learning 19
BAB III METODE PENELITIAN 20
3.1 Objek Penelitian 20
3.2 Alur Penelitian 21
3.3 Alat dan Bahan 37
3.3.1 Data Penelitian 37
3.3.2 Alat/instrumen 39
BAB IV HASIL DAN PEMBAHASAN 41
4.1 Hasil Penelitian 41
BAB V PENUTUP 43
5.1 Kesimpulan 43
5.2 Saran 43
REFERENSI 45
LAMPIRAN 47

DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian.....7
Tabel 3. 1 Distribusi Data.....22

DAFTAR GAMBAR

Gambar 2. 1 Diagram Bar Sebaran Sample Kelas Pada Dataset 15
Gambar 2. 2 Struktur Dasar Algoritma Random Forest 16
Gambar 2. 3 Ilustrasi Proses Hyperparameter Tuning dengan Grid Search 18
Gambar 3. 1 Distribusi Umur.....25
Gambar 3. 2 Distribusi Tinggi Badan.....26
Gambar 3. 3 Diagram Boxplot Distribusi Data.....27
Gambar 3. 4 Diagram Distribusi Data Berdasarkan Umur (Sebelum Binning).....28
Gambar 3. 5 Diagram Distribusi Data Berdasarkan Kelompok Umur (Setelah Binning) 29
Gambar 3. 6 Diagram Distribusi Data Berdasarkan Tinggi Badan (Sebelum Binning) 30
Gambar 3. 7 Diagram Distribusi Data Berdasarkan 6 Kelompok Tinggi Badan...30
Gambar 3. 8 Diagram Distribusi Data Berdasarkan 6 Kelompok Tinggi Badan (Dengan Pembagian Frequensi Yang Sama/ Equal Frequency Binning (EFB)) ...31
Gambar 3. 9 Diagram Perbandingan Distribusi Sebelum dan Sesudah Penerapan Teknik Data Augmentation dengan SMOTE.....32
Gambar 3. 10 Diagram Perbandingan Performa Model 35
Gambar 3. 11 Confussion Matriks.....36
Gambar 4. 1 Perbandingan Performa Model Random Forest Dari Penelitian Sebelumnya Dan Model Ensembel Voting Teroptimasi.....41

DAFTAR LAMBANG DAN SINGKATAN

WHO	World Health Organization (Organisasi Kesehatan Dunia)
BMI	Body Mass Index (Indeks Massa Tubuh)
ANN	Artificial Neural Network (Jaringan Syaraf Tiruan)
KNN	K-Nearest Neighbors (Tetangga Terdekat-K)
RF	Random Forest
SVM	Support Vector Machine
XGBoost	Extreme Gradient Boosting
EDA	Exploratory Data Analysis
DA	Data Augmentation
ERF-XGB	Ensembel Random Forest - XGBoost
ML	Machine Learning
std	Standar Deviasi Fitur
SMOTE	Synthetic Minority Over-sampling Technique
λ	lambda
L1	Lasso Regression
L2	Ridge Regression
EFB	Equal Frequency Binning
EWB	Equal Width Binning
CPU	Central Processing Unit (Unit Pemrosesan Pusat)
GHz	Gigahertz (Satuan frekuensi, 1 GHz = 1 miliar siklus per detik)
RAM	Random Access Memory (Memori Akses Acak)
GB	Gigabyte (Satuan penyimpanan data, 1 GB = 1 miliar byte)

DAFTAR ISTILAH

Binning	Pengelompokan variabel dengan sebaran nilai yang luas
Feature Scaling	Teknik untuk menyelaraskan skala fitur-fitur dalam dataset agar berada dalam rentang yang seragam.
Kuantil (Quantile)	ukuran statistik yang membagi sekumpulan data yang diurutkan menjadi bagian yang sama besar.
Data Augmentation	Teknik memperbanyak data pelatihan secara sintetik untuk meningkatkan performa model.
Hyperparameter Tunning	Pengoptimalan parameter model untuk meningkatkan performa.
GridSearchCV	Metode pencarian kombinasi parameter terbaik melalui pencarian grid.
Voting Ensemble Model	Model ensembel yang menggabungkan hasil prediksi beberapa model dengan cara memberikan suara (voting) untuk menentukan prediksi akhir.
Standardization	Salah satu teknik feature scaling yaitu mengubah skala data sehingga memiliki rata-rata (mean) 0 dan standar deviasi 1.
Obesitas	Kondisi kelebihan lemak tubuh yang berlebihan dan dapat membahayakan kesehatan.
Underweight	Kondisi di mana seseorang memiliki berat badan lebih rendah dari berat ideal berdasarkan IMT, yang dapat mengindikasikan kurangnya nutrisi atau masalah kesehatan lainnya.
Overweight	Kondisi di mana seseorang memiliki berat badan yang lebih tinggi dari berat ideal berdasarkan indeks massa tubuh (IMT/BMI), tetapi belum mencapai kategori obesitas.
Diskretisasi	Teknik mengonversi data kontinu menjadi data kategori atau diskrit, seperti dalam proses binning.
Noise Data	Variasi acak atau gangguan dalam data yang tidak mencerminkan pola sebenarnya.
Outlier	Nilai yang sangat berbeda dari sebagian besar data lainnya dalam dataset.
Stabilitas Model	Kemampuan model untuk memberikan hasil yang konsisten dan tidak terlalu terpengaruh oleh variasi kecil dalam data pelatihan.
Equal Frequency Binning (EFB)	Metode binning yang membagi data menjadi bin dengan jumlah data yang sama
Equal Width Binning (EWB)	Metode binning yang membagi data menjadi bin dengan jarak data yang sama
Interval	Rentang nilai dalam binning yang digunakan untuk mengelompokkan data kontinu.

Bin Width	Panjang setiap interval dalam metode equal-width binning.
Mean	Rata-rata
SMOTE (Synthetic Minority Over-sampling Technique)	Teknik data augmentation khusus untuk menangani imbalanced data dengan membuat sampel sintetis dari kelas minoritas melalui interpolasi.
Imbalanced Data	Kondisi di mana jumlah sampel dalam satu kelas jauh lebih sedikit dibandingkan kelas lain dalam dataset.
Random Forest	Algoritma ensemble learning berbasis pohon keputusan (decision trees) yang bekerja secara kolektif untuk meningkatkan akurasi prediksi.
XGBoost (Extreme Gradient Boosting)	Algoritma gradient boosting yang menggabungkan beberapa model prediksi lemah (weak learners) untuk membentuk model yang lebih kuat.
Gradient Boosting	Teknik machine learning yang membangun model secara bertahap, dengan setiap model baru difokuskan untuk memperbaiki kesalahan dari model sebelumnya.
Weak Learners	Model sederhana dengan performa di atas rata-rata yang digunakan dalam ensemble learning untuk membentuk model yang lebih kuat.
Regularisasi (L1 & L2)	Teknik untuk mencegah overfitting dengan menambahkan penalti terhadap parameter model yang besar.
Dataset	Kumpulan data yang terstruktur dalam bentuk tabel, terdiri dari baris dan kolom.
Kaggle	Platform daring yang menyediakan berbagai dataset publik serta alat untuk analisis data dan pembelajaran mesin (machine learning).
EDA (Exploratory Data Analysis)	Proses eksplorasi data untuk memahami distribusi, pola, dan anomali sebelum pemodelan lebih lanjut.
Variabel	Atribut atau kolom dalam dataset yang mewakili karakteristik tertentu dari objek penelitian.
Diagram Boxplot	Diagram yang menunjukkan distribusi data berdasarkan lima nilai utama (minimum, kuartil pertama, median, kuartil ketiga, dan maksimum), sering digunakan untuk mendeteksi outlier.
Preprocessing Data	Tahap persiapan data sebelum digunakan dalam analisis atau pemodelan, seperti penanganan outlier dan binning.
n_estimators	Parameter yang digunakan pada tuning untuk jumlah pohon
max_depth	Parameter yang digunakan pada tuning untuk kedalaman maksimum
LabelEncoder	Teknik encoding dalam scikit-learn yang mengubah nilai kategori menjadi angka (label numerik) agar dapat digunakan dalam algoritma machine learning.
bootstrap	

min_samples_leaf	
min_samples_split	
Iteratif	Berulang atau dilakukan secara berulang kali dalam suatu proses hingga mencapai hasil yang diinginkan.
Objek Penelitian	Subjek atau entitas yang menjadi fokus dalam penelitian.
Klasifikasi	Metode dalam pembelajaran mesin atau statistik untuk mengelompokkan data ke dalam kategori tertentu.
StandardScaler	Metode yang digunakan dalam Feature Scaling untuk menstandarisasi fitur dengan menghilangkan rata-rata dan menskalakan ke variansi satuan.
Cross-validation	Teknik untuk mengevaluasi kinerja model machine learning dengan membagi data menjadi beberapa set pelatihan dan pengujian secara berulang dan merata, untuk memastikan model dapat menggeneralisasi dengan baik.
Precision	Metrik kinerja yang mengukur akurasi dari prediksi positif, dihitung sebagai jumlah true positive dibagi dengan jumlah true positive dan false positive.
Recall	Metrik kinerja yang mengukur kemampuan model untuk mengidentifikasi semua instance positif, dihitung sebagai jumlah true positive dibagi dengan jumlah true positive dan false negative.
F1-score	Metrik kinerja yang menggabungkan precision dan recall menjadi satu nilai tunggal, dihitung sebagai rata-rata harmonis antara precision dan recall.
Confusion Matrix	Tabel yang digunakan untuk mengevaluasi kinerja model klasifikasi dengan menampilkan jumlah true positive, false positive, true negative, dan false negative.
Underweight, Normal, Overweight, Obese	Kategori status berat badan yang digunakan dalam tugas klasifikasi terkait prediksi obesitas.
Integers	Angka bulat, yang digunakan untuk variabel seperti usia dan tinggi badan.
Binary value	Nilai yang hanya terdiri dari dua kemungkinan, biasanya 0 atau 1, untuk menunjukkan kategori yang berbeda (misalnya, jenis kelamin atau kebiasaan makan).
Fast Food	Makanan yang cepat saji, sering kali tinggi kalori dan rendah gizi.
Vegetables	Sayuran, kategori makanan yang dikonsumsi oleh responden yang sering dikonsumsi dalam penelitian ini.
Main Meals	Makanan utama dalam sehari, yang dikategorikan berdasarkan frekuensinya (1-2 kali, 3 kali, atau lebih).
Snack	Makanan ringan yang dikonsumsi di antara makan utama.
Smoking	Kebiasaan merokok, dikategorikan dalam dataset dengan nilai biner (Ya atau Tidak).
Liquid Intake	Jumlah konsumsi cairan harian, dibagi menjadi tiga kategori berdasarkan volume.

Calorie Intake	Penghitungan kalori makanan yang dikonsumsi oleh responden, yang dikategorikan dalam bentuk biner (Ya atau Tidak).
Physical Exercise	Aktivitas fisik atau latihan yang dilakukan oleh responden dalam seminggu.
Technology Usage	Waktu yang dihabiskan untuk menggunakan teknologi/gawai dalam sehari.
Transportation	Jenis kendaraan yang digunakan oleh responden untuk mobilitas sehari-hari.
Range	Rentang
Class (Body Status)	Kategori status tubuh responden, seperti Underweight, Normal, Overweight, dan Obesity.
Interpretasi	Proses memahami, menjelaskan, atau memberikan makna terhadap sesuatu berdasarkan analisis atau sudut pandang
.xlsx	Format file yang digunakan untuk menyimpan dataset, yang memungkinkan penyimpanan data dalam bentuk tabel.
Windows 10 Pro	Sistem operasi yang digunakan untuk menjalankan perangkat keras dalam penelitian ini.
Intel Core i7	Jenis prosesor yang digunakan dalam perangkat keras penelitian untuk mendukung pemrosesan data.
RAM 10 GB	Memori akses acak (RAM) pada perangkat keras penelitian yang mendukung eksekusi analisis data dan model pembelajaran mesin.
Google Colab	Platform berbasis cloud yang memungkinkan eksekusi kode Python tanpa memerlukan perangkat keras lokal yang kuat.
Python	Bahasa pemrograman utama yang digunakan dalam penelitian ini untuk pemrosesan data dan pengembangan model.
Pandas	Pustaka Python untuk manipulasi data dalam bentuk tabel (DataFrame).
Scikit-learn	Pustaka Python untuk machine learning, yang digunakan untuk evaluasi model dalam penelitian ini.
Matplotlib dan Seaborn	Pustaka Python untuk visualisasi data dalam bentuk grafik.
XGBoost dan Random Forest	Algoritma untuk pengembangan model klasifikasi dalam penelitian ini.
Optimized Voting Classifier	Model ensemble yang menggabungkan berbagai algoritma (seperti Random Forest dan XGBoost) dan dioptimalkan untuk meningkatkan akurasi prediksi.
Stacking	Metode ensemble learning di mana prediksi dari beberapa model digabungkan menggunakan meta-model.
Bagging	Teknik machine learning yang menggabungkan prediksi dari beberapa model yang dilatih pada subset data yang

	berbeda untuk mengurangi variansi dan meningkatkan akurasi.
Feature Selection	Proses pemilihan fitur yang paling relevan untuk model, biasanya menggunakan metode statistik atau algoritma untuk mengurangi overfitting dan meningkatkan performa.
Cloud Computing	Penggunaan server jarak jauh yang dihosting di internet untuk menyimpan, mengelola, dan memproses data, memungkinkan kapasitas komputasi yang lebih kuat.
Web-Based Application	Aplikasi perangkat lunak yang dijalankan di server web daripada diinstal secara lokal di perangkat pengguna.
Mobile Application	Aplikasi perangkat lunak yang dirancang untuk dijalankan di perangkat mobile seperti smartphone atau tablet.

INTISARI

Obesitas adalah penyakit serius dan kronis yang dipengaruhi oleh interaksi faktor genetik dan lingkungan, termasuk kebiasaan sosial, psikologis, pola makan, serta kurangnya aktivitas fisik. Saat ini, lebih dari 2 miliar orang di dunia mengalami obesitas atau kelebihan berat badan, yang menjadi masalah kesehatan signifikan di berbagai kelompok usia.

Penelitian menunjukkan bahwa status obesitas dapat diklasifikasikan menggunakan metode kecerdasan buatan seperti Artificial Neural Network, K-Nearest Neighbors, Random Forest, dan Support Vector Machine. Namun, akurasi klasifikasi dari algoritma-algoritma ini masih dapat ditingkatkan.

Penelitian ini berfokus pada optimasi akurasi model klasifikasi status obesitas dengan menggunakan algoritma Random Forest, yang sebelumnya mencapai akurasi tertinggi (87,82%). Upaya optimasi dilakukan melalui pengelompokan variabel dengan sebaran luas (binning) seperti tinggi badan dan umur, Feature scaling dan Data Augmentation, Hyperparameter tuning menggunakan GridSearchCV untuk mencari kombinasi parameter terbaik pada Random Forest dan XGBoost, penerapan voting ensemble model dengan kombinasi Random Forest dan XGBoost, serta optimasi lebih lanjut melalui hyperparameter tuning pada model ensemble tersebut.

Melalui penerapan teknik optimasi ini terbukti meningkatkan nilai akurasi hingga angka 89,08%. Dengan hasil yang lebih akurat, penelitian ini diharapkan dapat berkontribusi dalam upaya pencegahan dan penanganan obesitas secara lebih efektif.

Kata kunci: Klasifikasi Status Obesitas, Kecerdasan Buatan Dalam Obesitas, Optimasi Random Forest, XGBoost, Model Voting Ensemble, Hyperparameter Tuning Dalam Classification, Data Augmentation Dan Feature Scaling

ABSTRACT

Obesity is a serious and chronic disease influenced by the interaction of genetic and environmental factors, including social habits, psychological aspects, dietary patterns, and lack of physical activity. Currently, more than 2 billion people worldwide suffer from obesity or being overweight, making it a significant health issue across various age groups.

Research shows that obesity status can be classified using artificial intelligence methods such as Artificial Neural Network, K-Nearest Neighbors, Random Forest, and Support Vector Machine. However, the classification accuracy of these algorithms can still be improved.

This study focuses on optimizing the accuracy of obesity status classification models using the Random Forest algorithm, which previously achieved the highest accuracy (87.82%). Optimization efforts include variable grouping with wide distribution (binning), such as height and age, feature scaling, data augmentation, hyperparameter tuning using GridSearchCV to find the best parameter combination for Random Forest and XGBoost, the application of a voting ensemble model combining Random Forest and XGBoost, and further optimization through hyperparameter tuning on the ensemble model.

The implementation of these optimization techniques successfully increased the accuracy to 89.08%. With more accurate results, this study is expected to contribute to more effective obesity prevention and management efforts.

Keyword: *Obesity Status Classification, Artificial Intelligence In Obesity, Random Forest Optimization, XGBoost, Voting Ensemble Model, Hyperparameter Tuning in Classification, Data Augmentation and Feature Scaling*