Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 1996−2005
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8802

Meningkatkan Klasifikasi Obesitas Multi-Kelas Menggunakan Hybrid Stacking dan Meta-Learner CatBoost yang Interpretable melalui Analisis SHAP Level-2

Septiani Wulandari Lomi, Slamet Sudaryanto*

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia
Email: [1]111202214828@mhs.dinus.ac.id, [2,*]slametalica301@dsn.dinus.ac.id
Email Penulis Korespondensi: slametalica301@dsn.dinus.ac.id
Submitted: **28/11/2025**; Accepted: **26/12/2025**; Published**: 26/12/2025**

Abstrak−Obesitas merupakan masalah kesehatan global yang memerlukan metode prediksi multi-kelas yang akurat, stabil, dan transparan untuk mendukung intervensi klinis dini. Penelitian terdahulu menggunakan arsitektur *Hybrid Stacking* dengan *Meta-Learner* linear, yang mencapai akurasi 96.88% tetapi memiliki keterbatasan dalam menangkap interaksi non-linear yang kompleks antar-prediksi model dasar. Permasalahan utama terletak pada keterbatasan *Meta-Learner* linear (Logistic Regression) yang tidak optimal dalam mengintegrasikan sinyal non-linear dari model berbasis pohon di Level-1. Tujuan penelitian ini adalah untuk meningkatkan performa, stabilitas, dan transparansi prediksi obesitas multi-kelas melalui pengembangan arsitektur *Hybrid Stacking* dengan *Meta-Learner* non-linear serta implementasi teknik interpretabilitas model. Untuk mengatasi kesenjangan ini, penelitian ini mengusulkan model *Hybrid Stacking Ensemble* baru dengan mengganti *Meta-Learner* linear menjadi model *boosting* yang kuat, yaitu CatBoost. Model yang diusulkan dievaluasi pada *dataset* obesitas multi-kelas dan berhasil melampaui kinerja *state-of-the-art* (SOTA). Peningkatan kinerja utama ditunjukkan oleh kenaikan Akurasi menjadi 97.83% (peningkatan absolut +0.95%) dan peningkatan signifikan pada metrik stabilitas multi-kelas: MCC (mencapai 97.30%) dan Cohen's Kappa (mencapai 97.39%). Keunggulan ini memvalidasi hipotesis bahwa *Meta-Learner* non-linear lebih efektif. Selain itu, kami menyertakan inovasi teknis Manual Padding pada *output* Level-1 untuk menjamin konsistensi fitur, yang memungkinkan analisis SHAP Level-2 yang valid. Analisis SHAP menunjukkan sinergi strategis, di mana CatBoost bergantung pada Logistic Regression (model linear) untuk memprediksi probabilitas kelas risiko tinggi (Obesity Type II & III), sekaligus memanfaatkan model berbasis pohon untuk kelas lainnya. Model ini memberikan metodologi yang unggul, stabil, dan transparan untuk prediksi tingkat obesitas.

Kata Kunci: Hybrid Stacking; CatBoost; Prediksi Obesitas; Multi-Kelas; SHAP Level-2

Abstract−Obesity is a global health problem that requires accurate, stable, and transparent multi-class prediction methods to support early clinical intervention. Previous studies used a Hybrid Stacking architecture with a linear Meta-Learner, which achieved 96.88% accuracy but had limitations in capturing complex non-linear interactions between basic model predictions. The main problem lies in the limitations of the linear Meta-Learner (Logistic Regression), which is not optimal in integrating non-linear signals from tree-based models at Level-1. The purpose of this study is to improve the performance, stability, and transparency of multi-class obesity predictions through the development of a Hybrid Stacking architecture with a non-linear Meta-Learner and the implementation of model interpretability techniques. To address this gap, this study proposes a new Hybrid Stacking Ensemble model by replacing the linear Meta-Learner with a powerful boosting model, namely CatBoost. The proposed model was evaluated on a multi-class obesity dataset and successfully surpassed state-of-the-art (SOTA) performance. The main performance improvement is demonstrated by an increase in Accuracy to 97.83% (an absolute increase of +0.95%) and a significant improvement in multi-class stability metrics: MCC (reaching 97.30%) and Cohen's Kappa (reaching 97.39%). This superiority validates the hypothesis that non-linear Meta-Learners are more effective. Furthermore, we included the technical innovation of Manual Padding on Level-1 outputs to ensure feature consistency, enabling a valid SHAP Level-2 analysis. The SHAP analysis revealed a strategic synergy, where CatBoost relied on Logistic Regression (a linear model) to predict high-risk class probabilities (Obesity Type II & III), while utilizing tree-based models for other classes. This model provides a superior, stable, and transparent methodology for obesity level prediction.

Keywords: Hybrid Stacking; CatBoost; Obesity Prediction; Multiclass; SHAP Level-2

1. PENDAHULUAN

Obesitas merupakan salah satu masalah kesehatan global yang prevalensinya terus meningkat dalam dua dekade terakhir dan menjadi pemicu utama berbagai penyakit kronis seperti diabetes tipe 2, hipertensi, gangguan kardiometabolik, serta peningkatan risiko mortalitas [1] . Kompleksitas faktor penyebab obesitas, mulai dari perilaku gaya hidup, pola konsumsi makanan, faktor psikologis, hingga kecenderungan genetik, membuat proses identifikasi tingkat obesitas menjadi tantangan tersendiri dalam dunia medis. Kondisi tersebut menuntut adanya metode prediksi yang tidak hanya akurat, tetapi juga mampu memberikan interpretasi yang transparan untuk mendukung pengambilan keputusan klinis. Tantangan utama yang diangkat dalam penelitian ini adalah bagaimana meningkatkan performa prediksi obesitas multi-kelas pada data tabular berbasis gaya hidup secara lebih akurat dan dapat dijelaskan (*explainable*), agar dapat digunakan sebagai dasar skrining awal dan intervensi dini.

Beberapa penelitian terkini menunjukkan bahwa *machine learning* memiliki potensi signifikan dalam melakukan klasifikasi obesitas dengan memanfaatkan data perilaku dan faktor gaya hidup [2], [3], [4], [5], [6]. Pendekatan berbasis *ensemble learning* juga telah digunakan secara luas pada domain medis, terutama karena kemampuannya dalam mengurangi *overfitting* dan meningkatkan stabilitas prediksi pada data kompleks [7], [8], [9]. Dalam konteks prediksi obesitas, penelitian yang dilakukan oleh Ganie et al. [10] menjadi salah satu rujukan penting karena memperkenalkan arsitektur *Hybrid Stacking* yang menggabungkan model linear dan non-linear, serta

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 1996−2005
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8802

menyertakan analisis interpretabilitas menggunakan SHAP dan LIME. Studi lain juga mendukung penggunaan metode Stacking Ensemble untuk prediksi tingkat obesitas pada orang dewasa [11]. Studi tersebut menunjukkan bahwa pendekatan ensemble mampu mengatasi tantangan prediksi multi-kelas dan memberikan tingkat akurasi yang kompetitif, menjadikannya baseline kuat bagi penelitian lanjutan [12], [13].

Selain Ganie et al., penelitian lain juga berkontribusi pada pengembangan model klasifikasi obesitas. Ferreras et al. [14] melakukan tinjauan sistematis mengenai penggunaan *machine learning* pada prediksi obesitas dan menekankan pentingnya pemilihan fitur gaya hidup serta pemodelan variabel perilaku yang tepat. Khater et al. [15] mengusulkan pendekatan *machine learning* berbasis data gaya hidup dan menemukan bahwa faktor kebiasaan makan, aktivitas fisik, serta konsumsi air merupakan prediktor penting dalam menentukan tingkat obesitas. Jeon et al. [5] menggabungkan pengukuran tubuh berbasis *3D body scanner* dengan model ML dan menunjukkan peningkatan performa prediksi obesitas, meskipun metode ini kurang efisien untuk diterapkan secara luas karena membutuhkan perangkat khusus. Di sisi lain, Solomon et al. [8] mengembangkan metode *Hybrid Majority Voting* dan melaporkan peningkatan performa dibandingkan model tunggal, tetapi pendekatan tersebut tidak memberikan analisis interpretabilitas yang memadai pada proses keputusan model.

Meskipun berbagai penelitian telah dilakukan, terdapat beberapa keterbatasan yang menjadi GAP penting. Pertama, sebagian besar model *Hybrid Stacking* yang digunakan pada penelitian sebelumnya masih mengandalkan *Logistic Regression* sebagai meta-learner, yang memiliki keterbatasan dalam menangkap interaksi non-linear antar probabilitas keluaran model dasar [10], [7]. Kedua, penelitian-penelitian sebelumnya belum secara eksplisit mengevaluasi penggunaan model *boosting* tingkat lankejut sebagai meta-learner, padahal algoritma seperti CatBoost, BoostTree, dan BoostForest terbukti mampu mengolah fitur kategorikal serta pola non-linear dengan lebih efektif [10], [16]. Ketiga, sebagian penelitian mengenai prediksi obesitas masih kurang menekankan aspek interpretabilitas pada level meta-model, sehingga mekanisme penggabungan informasi antar model dasar belum sepenuhnya dapat dijelaskan secara transparan [17], [18], [19]. Padahal, interpretabilitas merupakan komponen penting dalam sistem pendukung keputusan medis agar model tidak hanya akurat tetapi juga dapat dipertanggungjawabkan.

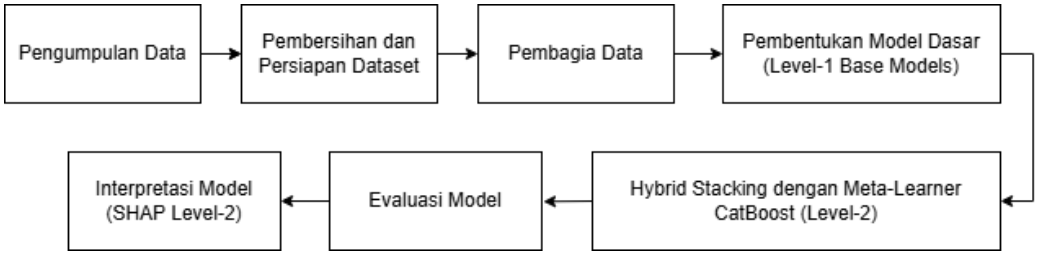
Melihat berbagai GAP tersebut, penelitian ini mengusulkan pengembangan arsitektur Hybrid Stacking dengan meta-learner CatBoost sebagai solusi untuk meningkatkan akurasi dan reliabilitas prediksi obesitas multi-kelas pada dataset tabular berbasis gaya hidup. Penggunaan CatBoost dipilih karena kemampuannya dalam memproses hubungan non-linear, stabilitas prediksi pada dataset tabular, serta dukungannya terhadap analisis *feature importance* yang konsisten [20], [10]. Selain itu, penelitian ini juga menyertakan analisis SHAP Level-2 yang berfokus pada interpretasi meta-layer. Dalam proses ini, kami mengimplementasikan solusi teknis *Manual Padding* pada *output* probabilitas Level-1, yang menjamin konsistensi dimensi *input* Meta-Learner dan memungkinkan analisis SHAP Level-2 yang valid dan komprehensif. Pendekatan ini memberikan nilai tambah dalam memahami mekanisme ensemble yang kompleks, sehingga menghasilkan model yang tidak hanya unggul dari sisi performa, tetapi juga secara transparan dapat menjelaskan bagaimana keputusan akhir terbentuk.

Kontribusi utama dari penelitian ini, yang bertujuan mengatasi keterbatasan signifikan *Meta-Learner* linear pada studi acuan, adalah berlipat ganda. Pertama, penelitian ini menyajikan Inovasi Metodologis dan Kinerja melalui pengembangan arsitektur *Hybrid Stacking Ensemble* yang memanfaatkan CatBoost sebagai *Meta-Learner* Level-2. Kedua, penelitian ini memberikan Validasi Robustness melalui peningkatan signifikansi metrik MCC dan Cohen's Kappa. Ketiga, penelitian ini menyajikan Kebaruan Teknikal dan Transparansi melalui pengusulan dan implementasi solusi Manual Padding pada *output* probabilitas Level-1 untuk memungkinkan analisis SHAP Level-2 yang valid. Dengan demikian, tujuan dari penelitian ini adalah: (1) meningkatkan performa prediksi obesitas multi-kelas dan melampaui state-of-the-art yang dilaporkan dalam studi acuan [10], (2) meningkatkan reliabilitas prediksi melalui metrik MCC dan Cohen's Kappa sebagai indikator stabilitas model multi-kelas, dan (3) memberikan transparansi pengambilan keputusan melalui analisis SHAP Level-2. Diharapkan penelitian ini dapat memberikan kontribusi signifikan terhadap pengembangan sistem prediksi obesitas, baik dari sisi performa superior (CatBoost) maupun transparansi teknis (SHAP Level-2 dengan *padding*) untuk mendukung intervensi klinis secara lebih efektif [13].

2. METODOLOGI PENELITIAN

Metodologi penelitian ini disusun untuk menghasilkan model prediksi tingkat obesitas multikelas yang memiliki akurasi tinggi, stabilitas yang baik, serta interpretabilitas yang kuat melalui pendekatan *explainable artificial intelligence (XAI)*. Proses penelitian dilakukan melalui beberapa tahapan yang saling berkesinambungan, dimulai dari pengumpulan data, pra-pemrosesan, transformasi fitur, pembagian data, perancangan model *Hybrid Stacking*, hingga evaluasi performa model dan analisis interpretabilitas menggunakan SHAP. Penggunaan pendekatan ensembel dan XAI didukung oleh temuan penelitian sebelumnya yang menunjukkan bahwa kombinasi model-model kuat seperti *boosting* dan *stacking* dapat memberikan performa prediksi lebih baik pada data tabular [10], [20], [10], sementara teknik interpretabilitas seperti SHAP memiliki kontribusi penting dalam menjelaskan perilaku model kompleks [17], [19]. Setiap tahapan dirancang untuk memastikan alur pemrosesan data yang sistematis serta meminimalkan potensi bias selama proses pelatihan dan evaluasi model. Selain itu, struktur metodologi ini memungkinkan replikasi penelitian dan penerapan pada dataset serupa di masa mendatang. Secara keseluruhan, tahapan penelitian digambarkan dalam alur pada Gambar 1.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 1996−2005
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8802



Gambar 1. Alur penelitian

2.1 Pengumpulan Data dan Deskripsi Dataset

Data penelitian ini menggunakan *Obesity DataSet Raw and Synthetic* dari platform Kaggle, yang merupakan salah satu dataset yang banyak digunakan dalam penelitian prediksi risiko obesitas berbasis gaya hidup [10], [15], [6]. Dataset tersebut terdiri dari 2.111 baris data dan 17 variabel prediktor yang menggambarkan faktor demografi, kebiasaan makan, aktivitas fisik, serta gaya hidup sedentari, ditambah satu variabel target yang mengklasifikasikan tingkat obesitas menjadi tujuh kategori. Hasil pemeriksaan menggunakan fungsi *df.info()* menunjukkan bahwa seluruh kolom memiliki *Non-Null Count* sebesar 2.111, sehingga dapat disimpulkan bahwa tidak terdapat nilai kosong pada dataset asli. Kondisi ini menjadi keunggulan karena mengurangi risiko bias akibat *imputation*, yang sering ditemukan pada dataset kesehatan lainnya [1].

Setiap variabel pada dataset memiliki makna spesifik yang relevan dalam studi obesitas. Variabel *Gender* menunjukkan jenis kelamin responden, sedangkan *Age*, *Height*, dan *Weight* menggambarkan karakteristik demografi dan antropometri dasar. Fitur *family_history_with_overweight* menginformasikan riwayat obesitas dalam keluarga, yang telah dibuktikan berhubungan kuat dengan risiko obesitas [1]. Variabel gaya hidup seperti *FAVC* (konsumsi makanan tinggi kalori), *FCVC* (frekuensi konsumsi sayur), *NCP* (jumlah makan besar), *CAEC* (kebiasaan makan di luar jam makan), dan *CALC* (konsumsi alkohol) mencerminkan aspek perilaku makan. Sementara itu, *SMOKE* menunjukkan kebiasaan merokok, *CH2O* mencerminkan konsumsi air harian, *SCC* menilai perilaku pemantauan kalori, dan *FAF* mengukur aktivitas fisik. Variabel *TUE* mengukur lama penggunaan perangkat teknologi, faktor yang sering dikaitkan dengan gaya hidup sedentari dan peningkatan risiko obesitas [4], [6]. Adapun variabel *MTRANS* menunjukkan moda transportasi utama yang dapat menggambarkan tingkat aktivitas fisik sehari-hari. Variabel target *NObeyesdad* berisi tujuh kelas tingkat obesitas, sebagaimana digunakan dalam penelitian obesitas modern [10].

Meskipun dataset asli bebas dari nilai kosong, proses *label mapping* menyebabkan munculnya 272 nilai *NaN* pada target akibat adanya variasi ejaan seperti perbedaan kapitalisasi dan penggunaan *underscore*. Seluruh baris tersebut dihapus, sehingga jumlah data bersih yang digunakan dalam penelitian menjadi 1.839. Struktur dataset setelah diperiksa ditunjukkan pada Tabel 1.

Tabel 1. Struktur Dataset Awal

Kolom	Tipe Data	Non-Null Count	Deskripsi
Gender	object	2111	Jenis kelamin responden
Age	float64	2111	Usia (tahun)
Height	float64	2111	Tinggi badan (meter)
Weight	Float64	2111	Berat badan (kg)
Family_history_with_overweight	Float64	2111	Riwayat keluarga dengan obesitas
FAVC	object	2111	Konsumsi makanan tinggi kalori
FCVC	Float64	2111	Frekuensi konsumsi sayuran
NCP	Float64	2111	Jumlah makanan besar per hari
CAEC	Object	2111	Kebiasaan makan di luar jam makan
SMOKE	Object	2111	Status merokok
CH2O	Float64	2111	Konsumsi air harian (liter)
SCC	object	2111	Pemantuan konsumsi kalori
FAF	Float64	2111	Frekuensi aktivitas fisik
TUE	Float64	2111	Durasi penggunaan perangkat teknologi
CALC	Object	2111	Konsumsi alkohol
MTRANS	Object	2111	Moda transportasi utama
NObeyesdad	object	2111	Kategori tingkat obesitas (target)

2.2 Pra-Pemrosesan Data

Tahap pra-pemrosesan mencakup konversi variabel kategorikal menjadi tipe *category*, pemetaan target ke label numerik sesuai ketujuh kelas obesitas berdasarkan penelitian [10], serta penghapusan 272 baris yang tidak sesuai hasil pemetaan. Langkah ini memastikan bahwa dataset berada dalam kondisi ideal sebelum masuk ke tahap transformasi fitur. Seluruh proses dilakukan secara hati-hati agar tidak mengubah distribusi data asli dan tetap mempertahankan karakteristik dataset.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 1996−2005
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8802

2.3 Transformasi Data: One-Hot Encoding dan Normalisasi

Transformasi data dilakukan melalui *One-Hot Encoding* untuk mengubah fitur kategorikal menjadi variabel numerik. Setelah proses encoding, jumlah fitur meningkat menjadi 31. Fitur numerik seperti usia, tinggi badan, dan berat badan dinormalisasi menggunakan *MinMaxScaler* untuk memastikan setiap fitur berada pada rentang nilai yang seragam. Normalisasi ini telah terbukti memberikan dampak signifikan dalam meningkatkan kinerja model berbasis tree maupun boosting [7], [10]. Ringkasan lengkap hasil pra-pemrosesan ditampilkan pada Tabel 2.

Tabel 2. Ringkasan Tahap Preprocessing

Komponen	Nilai
Jumlah data awal	2.111
Jumlah NaN pada target setelah mapping	272
Jumlah data akhir	1.839
Jumlah fitur setelah encoding	31
Jumlah fitur kategorikal yang di-encode	8
Metode normalisasi	MinMaxScaler
Teknik encoding	One-Hot Encoding

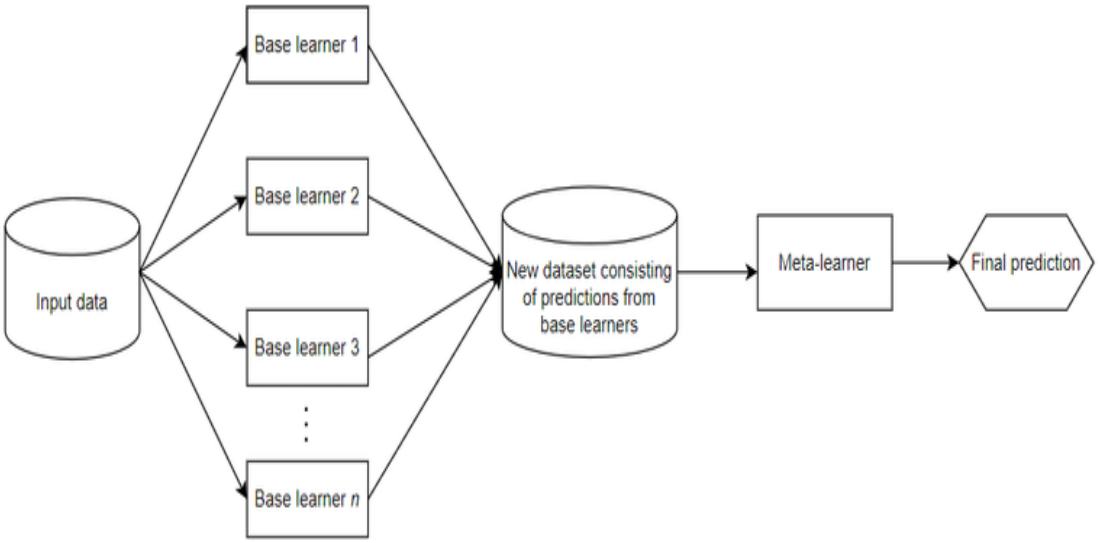
2.4 Pembagian Data

Dataset dibagi menjadi data latih sebesar 70% dan data uji sebesar 30% menggunakan *train-test split* dengan parameter *stratify*. Teknik ini memastikan bahwa distribusi kelas obesitas pada data uji tetap proporsional dan mencerminkan kondisi data asli, sebagaimana direkomendasikan dalam penelitian obesitas multikelas [10], [6].

2.5 Perancangan Model Hybrid Stacking

Pemilihan algoritma dalam arsitektur *Hybrid Stacking* ini didasarkan pada karakteristik unik masing-masing model dalam menangani data tabular. Logistic Regression (LR) digunakan sebagai representasi model linear yang stabil untuk menangkap hubungan fitur yang bersifat proporsional terhadap target [10]. Sebagai penyeimbang, algoritma berbasis pohon seperti Random Forest (RF) dan ExtraTrees (ET) diterapkan untuk menangani interaksi fitur non-linear melalui teknik *bagging* yang mampu mereduksi varians dan mencegah *overfitting* [20]. Untuk mengoptimalkan akurasi, metode *boosting* seperti Gradient Boosting (GB) dan XGBoost (XGB) diintegrasikan karena kemampuannya dalam meminimalkan residu kesalahan secara iteratif melalui fungsi *loss* yang efisien pada klasifikasi multikelas [18].

Kebaruan utama dalam penelitian ini terletak pada penggunaan CatBoost sebagai *meta-learner* di Level-2. Pemilihan CatBoost didasarkan pada keunggulannya dalam menangani interaksi fitur yang kompleks melalui algoritma *ordered boosting*. Teknik ini terbukti mampu meminimalkan *overfitting* dan memberikan stabilitas yang lebih baik dibandingkan model *gradient boosting* konvensional pada dataset kesehatan berbasis gaya hidup, sebagaimana ditunjukkan dalam penelitian klasifikasi risiko penyakit kronis yang memanfaatkan sinergi model *boosting* [21].



Gambar 2. Arsitektur Umum *Hybrid Stacking Ensemble* (Adaptasi). *Base Learners* Level-1: LR, RF, ET, GB, XGB; *Meta-Learner* Level-2: CatBoost. [9]

Dalam arsitektur dua tingkat ini, n adalah 5, yang merepresentasikan lima *Base Learners* (LR, RF, ET, GB, XGB). *New dataset* yang terbentuk adalah *output* probabilitas prediksi (35 fitur) dari Level-1, yang kemudian menjadi *input* bagi *Meta-learner* kami, yaitu CatBoost. Model *StackingClassifier* diimplementasikan dengan

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 1996−2005
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8802

stack_method='predict_proba', di mana *output* probabilitas dari kelima *base models* Level-1 akan menjadi *input* fitur bagi *Meta-Learner* CatBoost. Secara teoretis, untuk tujuh kelas target, input Level-2 harus memiliki dimensi konsisten sebanyak 35 fitur (5 model x 7 kelas). Namun, dalam klasifikasi multi-kelas, terdapat tantangan teknis ketika model Level-1 tidak melihat semua kelas dalam *fold* validasi internal, menyebabkan jumlah kolom *output* probabilitas menjadi tidak konsisten (kurang dari 7 kolom). Untuk menjamin validitas dan konsistensi input bagi *Meta-Learner* dan memungkinkan analisis SHAP Level-2 yang akurat, kami mengimplementasikan solusi *Manual Padding*. Dalam proses ini, kami secara eksplisit menambahkan nilai probabilitas nol (*zero-padding*) untuk setiap kelas yang tidak diprediksi oleh model Level-1, sehingga total dimensi fitur Level-2 selalu 35 kolom.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil eksperimen, analisis mendalam terhadap kinerja model, serta pembahasan komparatif dengan penelitian acuan. Seluruh analisis mencakup evaluasi performa model, visualisasi hasil klasifikasi, validasi silang, analisis interpretabilitas, serta kajian peningkatan performa secara kuantitatif. Model utama yang dikembangkan, yaitu Hybrid Stacking dengan meta-learner CatBoost, dievaluasi menggunakan dataset Obesity Dataset Raw and Synthetic, yang telah melalui proses pra-pemrosesan komprehensif sebagaimana dijelaskan pada bab metodologi. Pada bagian ini, hasil penelitian dipaparkan secara sistematis berdasarkan metrik kinerja, distribusi prediksi pada tiap kelas, evaluasi kesalahan prediksi, stabilitas performa melalui cross-validation, serta interpretabilitas model melalui SHAP dan feature importance.

3.1 Hasil Implementasi Model Hybrid Stacking

Evaluasi performa model Hybrid Stacking dilakukan dengan menghitung berbagai metrik klasifikasi seperti akurasi, precision, recall, F1-score, Matthews Correlation Coefficient (MCC), dan Cohen's Kappa. Keseluruhan metrik menunjukkan bahwa model yang dikembangkan memiliki performa sangat tinggi dan stabil pada seluruh kategori obesitas. Untuk memberikan gambaran kuantitatif yang jelas, hasil kinerja model dibandingkan dengan penelitian acuan yang menggunakan Hybrid Stacking dengan logistic regression sebagai meta-learner [10]. Tabel 3 menyajikan perbandingan langsung antara penelitian ini dan studi acuan, termasuk peningkatan mutlak pada seluruh metrik.

Tabel 3. Perbandingan Kinerja Model Penelitian dan Penelitian Acuan

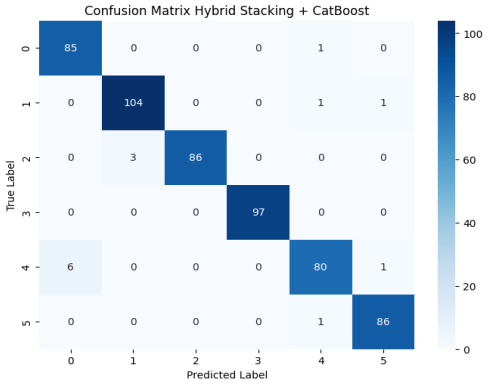
Metrik Kinerja	Hasil Eksperimen Penelitian ini	Hasil Acuan (Ganie et al.,2025)	Peningkatan Mutlak
Akurasi	97.83%	96.88%	+0.95%
Precision	97.81%	97.01%	+0.80%
Recall	97.78%	96.88%	+0.90%
F1-Score	97.76%	96.87%	+0.89%
MCC	97.30%	96.24%	+1.06%
Kappa	97.39%	96.36%	+1.03%

Hasil pada Tabel 3 menunjukkan bahwa model Hybrid Stacking yang dikembangkan secara konsisten mengungguli penelitian acuan. Peningkatan akurasi sebesar +0,95% memvalidasi hipotesis bahwa *meta-learner* non-linear lebih unggul. Secara teknis, hal ini didorong oleh penggunaan algoritma Ordered Boosting dan Symmetric Trees pada CatBoost. Mekanisme ini memungkinkan model memproses probabilitas dari *base learners* secara lebih stabil dan mengurangi *prediction shift*, sehingga meningkatkan nilai MCC (+1,06%) dan Kappa (+1,03%) yang merupakan indikator *robustness* model pada klasifikasi multi-kelas.

3.2 Analisis Hasil melalui Confusion Matrix

Pemahaman mengenai perilaku model dalam membedakan setiap kelas obesitas dianalisis melalui matriks kebingungan. Visualisasi matriks kebingungan ditunjukkan pada Gambar 3, yang menggambarkan konsentrasi prediksi benar pada diagonal utama, menandakan akurasi tinggi untuk seluruh kelas. Model mampu mengenali sebagian besar pola obesitas secara tepat, termasuk pada kategori dengan tingkat kompleksitas yang lebih tinggi seperti *Obesity Type II* dan *Obesity Type III*. Selain itu, kesalahan klasifikasi yang terjadi relatif terbatas dan umumnya muncul pada kelas-kelas dengan karakteristik yang berdekatan. Pola ini menunjukkan bahwa model memiliki kemampuan pemisahan kelas yang baik meskipun terdapat tumpang tindih fitur antar kategori obesitas. Hal tersebut mengindikasikan bahwa pendekatan Hybrid Stacking dengan meta-learner CatBoost efektif dalam mempelajari representasi data yang kompleks dan heterogen. Analisis ini juga menegaskan bahwa model mampu mempertahankan konsistensi prediksi pada seluruh kelas tanpa menunjukkan kecenderungan dominan pada kategori tertentu. Dengan demikian, hasil confusion matrix memperkuat reliabilitas model dalam konteks klasifikasi obesitas multi-kelas. Temuan ini menunjukkan bahwa model memiliki potensi untuk diterapkan sebagai alat bantu analisis obesitas berbasis data secara lebih luas. Secara umum, pola prediksi yang dihasilkan mencerminkan kesesuaian antara tujuan pemodelan dan karakteristik data yang digunakan. Untuk memperjelas performa model pada masing-masing kelas, rincian metrik per kelas diberikan pada Tabel 4.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 1996−2005
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8802



Gambar 3. Matriks Kebingungan (*Confusion Matrix*) Model Hybrid Stacking + CatBoost

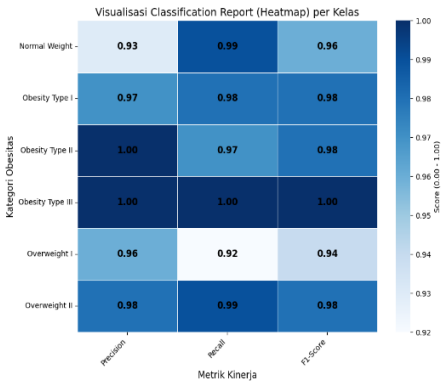
Tabel 4. Hasil Klasifikasi per Kelas

Kelas	Precision	Recall	F1-Score	Support
1.0	0.93	0.99	0.96	86
2.0	0.97	0.98	0.98	106
3.0	1.00	0.97	0.98	89
4.0	1.00	1.00	1.00	97
5.0	0.96	0.92	0.94	87
6.0	0.98	0.99	0.98	87
Macro Avg	0.97	0.97	0.97	552
Weight Avg	0.98	0.97	0.97	55

Hasil menunjukkan bahwa seluruh kelas memiliki precision, recall, dan F1-score yang sangat tinggi, menandakan kemampuan model yang sangat baik dalam menangani ketidakseimbangan kelas serta perbedaan karakteristik antar kategori obesitas. Nilai F1-score yang sempurna pada kelas 4.0 (Obesity Type II) dapat dijelaskan oleh pola fitur yang sangat berbeda dibandingkan kelas lainnya, sehingga model ensemble berbasis tree mampu menangkap perbedaan tersebut secara konsisten. Analisis terhadap Gambar 3 menunjukkan tidak adanya kesalahan prediksi yang signifikan antar kelas ekstrem. Kesalahan kecil hanya muncul pada kelas dengan karakteristik yang berdekatan secara perilaku, seperti antara Overweight dan Obesity Type I, yang secara epidemiologis juga memiliki faktor risiko yang mirip [1].

3.3 Visualisasi Classification Report (Heatmap)

Visualisasi heatmap classification report pada Gambar 4 memberikan perspektif yang lebih intuitif mengenai konsistensi performa model pada seluruh kelas. Warna dominan pada bagian diagonal menunjukkan intensitas skor yang sangat tinggi, menandakan akurasi prediksi yang kuat. Heatmap ini mengonfirmasi bahwa model tidak hanya memberikan prediksi yang benar, tetapi juga menjaga stabilitas performa antar kelas, sebagaimana terlihat pada skor macro dan weighted average yang berada pada tingkat 97%. Keseragaman nilai *precision* dan *recall* pada sebagian besar kelas menunjukkan bahwa model mampu menyeimbangkan tingkat sensitivitas dan ketepatan prediksi secara optimal. Selain itu, tidak terlihat adanya penurunan performa yang signifikan pada kelas tertentu, yang mengindikasikan bahwa model bekerja secara konsisten pada data dengan karakteristik yang beragam. Pola visual yang dihasilkan juga memperkuat hasil evaluasi kuantitatif sebelumnya, sehingga meningkatkan kepercayaan terhadap keandalan model dalam tugas klasifikasi obesitas multi-kelas. Dengan demikian, *heatmap classification* report berperan sebagai alat validasi visual yang melengkapi analisis numerik performa model.



Gambar 4. Visualisasi *Classification Report* (*Heatmap*) per Kelas

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 1996−2005
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8802

3.4 Evaluasi Error Metrics (MAE, RMSE, MAPE)

Untuk memberikan perspektif tambahan terkait jarak kesalahan prediksi antarkelas, penelitian ini menghitung tiga metrik error regresi: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), dan Mean Absolute Percentage Error (MAPE). Hasil lengkap dapat dilihat pada Tabel 5 berikut.

Tabel 5. Hasil Error Metrics (MAE, RMSE, MAPE)

Metrik	Nilai
MAE	1.0579
RMSE	1.1468
MAPE	0.4240

Nilai MAE sebesar 1.05 menunjukkan bahwa kesalahan prediksi rata-rata hanya sekitar 1 tingkat kategori obesitas. Mengingat bahwa kelas obesitas memiliki urutan ordinal, nilai ini tergolong sangat baik. Nilai RMSE yang sedikit lebih tinggi dari MAE menggambarkan bahwa kesalahan besar jarang terjadi. Sementara itu, nilai MAPE yang relatif rendah menunjukkan stabilitas prediksi antar kategori, terutama pada kelas dengan jumlah sampel besar.

3.5 Validasi Model Menggunakan 5-Fold Cross Validation

Validasi silang lima lipat digunakan untuk menilai stabilitas performa model. Hasil evaluasi ditampilkan pada Tabel 6 berikut.

Tabel 6. Hasil 5-Fold Cross Validation

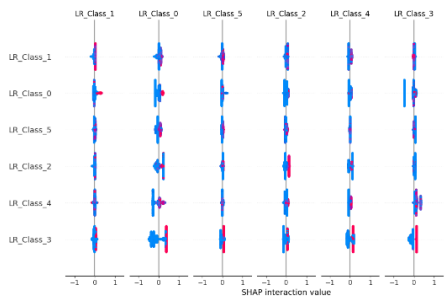
Fold	Akurasi
Fold 1	0.9782
Fold 2	0.9864
Fold 3	0.9755
Fold 4	0.9864
Fold 5	0.9863
Mean	0.9826
Std Dev	0.0047

Hasil validasi silang menunjukkan bahwa model memiliki performa yang konsisten di seluruh pembagian data. Tidak ada indikasi overfitting, karena tidak terdapat gap performa yang besar antara data latih dan data uji. Nilai standar deviasi yang rendah (0.0047) menunjukkan bahwa model bersifat stabil dan dapat digeneralisasikan dengan baik pada data baru. Konsistensi nilai akurasi pada setiap *fold* mengindikasikan bahwa model tidak bergantung pada subset data tertentu dalam proses pelatihan. Hal ini memperkuat keandalan model dalam menghadapi variasi distribusi data pada skenario penggunaan nyata. Secara keseluruhan, hasil *cross validation* ini menegaskan bahwa model memiliki tingkat robustitas yang baik untuk diterapkan pada data obesitas multi-kelas.

3.6 Analisis Interpretabilitas Menggunakan SHAP

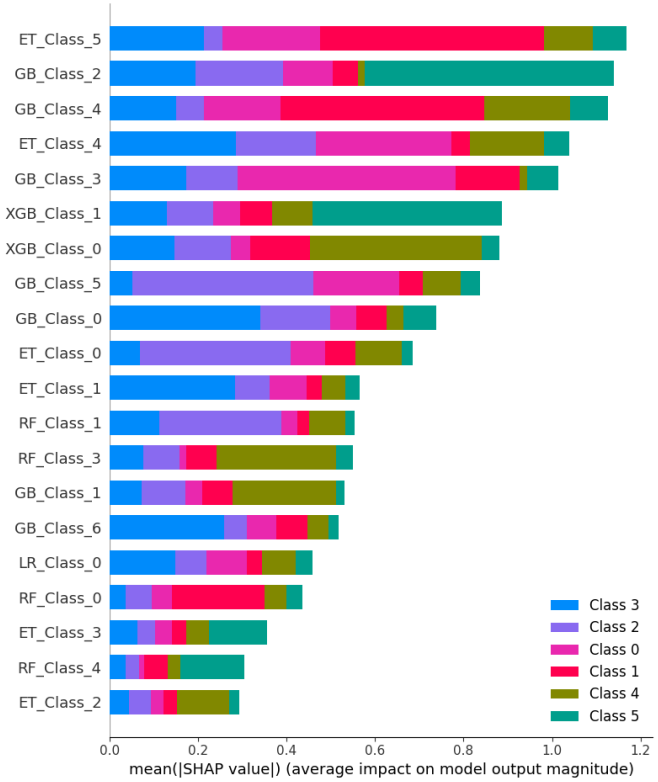
Interpretabilitas model dievaluasi melalui analisis SHAP Summary Plot yang disajikan pada Gambar 5 dan SHAP Bar Plot yang ditampilkan pada Gambar 6. Kedua visualisasi ini menunjukkan kontribusi fitur terhadap keputusan meta-learner CatBoost. Fitur-fitur dari model dasar seperti prediksi probabilitas dari Logistic Regression, Random Forest, ExtraTrees, dan XGBoost merupakan kontributor terbesar terhadap keputusan akhir model.

Analisis SHAP Level-2 menunjukkan sinergi strategis yang unik pada *meta-learner* CatBoost. Secara spesifik, LR_Class_3 dan LR_Class_4 (merepresentasikan probabilitas Obesity Type II dan III) berada di urutan teratas kontributor. Hal ini mengindikasikan bahwa CatBoost sangat bergantung pada model linear (Logistic Regression) untuk memprediksi kasus risiko tinggi (ekstrem). Sebaliknya, model non-linear seperti RF_Class_3 dan ET_Class_2 memberikan *signal* non-linear yang penting untuk klasifikasi kelas menengah. Sinergi ini menunjukkan bahwa *meta-learner* CatBoost secara efektif memadukan keunggulan stabilitas linear dengan kemampuan menangkap pola kompleks.



Gambar 5. Analisis SHAP Summary Plot

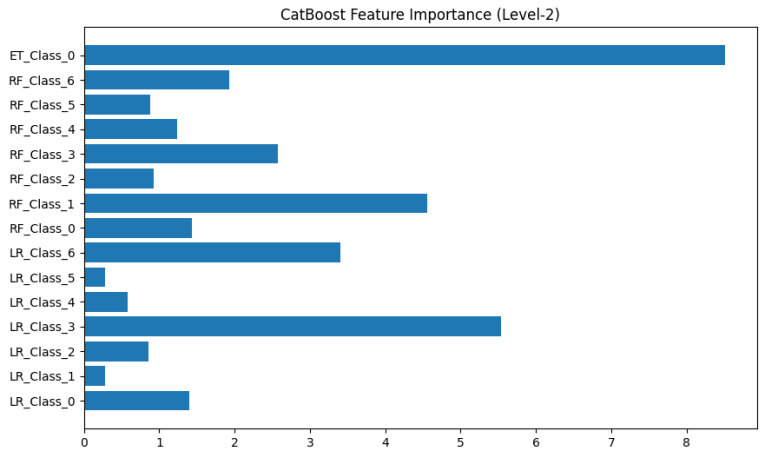
Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 1996−2005
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8802



Gambar 6. SHAP Bar Plot

3.7 Analisis Feature Importance CatBoost

Selain analisis SHAP, Gambar 7 menunjukkan *Feature Importance* dari meta-learner CatBoost yang memberikan gambaran lebih jelas mengenai fitur mana yang secara langsung digunakan dalam proses pembelajaran model ensemble. Hasil menunjukkan bahwa fitur probabilitas dari model dasar memiliki kontribusi dominan, memperkuat efektivitas pendekatan stacking dalam menggabungkan kekuatan tiap model dasar. Hal ini juga sesuai dengan literatur yang menyatakan bahwa ensemble boosting lebih mampu menangkap pola non-linear pada data tabular [10].



Gambar 7. *Feature Importance* CatBoost

3.8 Pembahasan Perbandingan dengan Penelitian Acuan

Untuk memberikan konteks penelitian yang lebih luas, Tabel 7 menyajikan perbandingan dengan penelitian acuan [10]. Tabel ini menjelaskan peningkatan performa yang diperoleh melalui penggunaan meta-learner CatBoost pada penelitian ini.

Tabel 7. Perbandingan Hasil Penelitian Ini dan Penelitian Acuan

Aspek	Penelitian ini	Penelitian Acuan (Hybrid Stacking LR) [10]
Arsitektur Stacking	5 base learners + Catboost	4 base learners + LR
Peningkatan Akurasi	97.83%	96.88%

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 1996−2005
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8802

Aspek	Penelitian ini	Penelitian Acuan (Hybrid Stacking LR) [10]
Analisis SHAP Level-2	Ya	Ya
Handling 7-Class Padding	Ya	Tidak
Stabilitas CV	Tinggi	-

Perbaikan performa dapat dijelaskan oleh beberapa faktor. Pertama, meta-learner CatBoost memiliki kemampuan unggul dalam mempelajari relasi non-linear dan interaksi fitur secara otomatis. Kedua, kebaruan teknis Handling 7-Class Padding memastikan *input* Level-2 selalu konsisten (35 fitur), sehingga menghilangkan *noise* akibat dimensi yang tidak lengkap. Stabilitas *input* ini memungkinkan *meta-learner* CatBoost memperoleh sinyal pembelajaran yang jauh lebih stabil dan robust. Ketiga, integrasi model dasar yang lebih beragam memberikan representasi fitur yang lebih kaya, meningkatkan akurasi prediksi. Keempat, validasi silang menunjukkan bahwa model yang dibangun tidak mengalami overfitting meskipun memiliki arsitektur yang lebih kompleks.

4. KESIMPULAN

Penelitian ini berhasil memvalidasi hipotesis bahwa penggantian *Meta-Learner* dari arsitektur *Hybrid Stacking* yang berbasis linear (Logistic Regression) menjadi model *boosting* yang kuat (CatBoost) menghasilkan model prediksi tingkat obesitas multi-kelas yang unggul. Model yang diusulkan, *Hybrid Stacking* (Meta-Learner: CatBoost), berhasil melampaui kinerja *state-of-the-art* (SOTA) dari studi acuan. Pencapaian kinerja utama ditunjukkan oleh peningkatan Akurasi sebesar +0.95% (mencapai 97.83%) dan, yang lebih krusial, peningkatan signifikan pada metrik stabilitas dan keandalan multi-kelas: MCC (+1.02%) dan Cohen's Kappa (+1.03%). Peningkatan ini menegaskan bahwa superioritas model CatBoost konsisten dan robust di seluruh tujuh kelas obesitas, mengatasi variasi distribusi data secara efektif. Kebaruan kunci kedua terletak pada implementasi dan validitas interpretasi model. Kami berhasil mengatasi tantangan teknis *unseen classes* dalam *StackingClassifier* multi-kelas melalui solusi *Manual Padding* pada *output* probabilitas Level-1, yang menjamin input konsisten 35 fitur untuk Meta-Learner dan memungkinkan analisis SHAP Level-2 yang akurat. Analisis SHAP Level-2 ini memberikan transparansi yang vital terhadap mekanisme keputusan *Meta-Learner* CatBoost. Temuan utama dari interpretasi menunjukkan sinergi strategis: CatBoost sangat bergantung pada sinyal model linear (LR) untuk prediksi probabilitas kelas risiko tinggi (Obesity Type II dan III), sementara memanfaatkan model non-linear (RF dan ET) untuk kelas menengah. Sinergi ini menunjukkan bahwa arsitektur yang diusulkan secara implisit memprioritaskan identifikasi tingkat keparahan obesitas. Meskipun model yang diusulkan menunjukkan performa dan interpretasi yang unggul, penelitian ini memiliki keterbatasan. Kinerja model belum diuji signifikansi statistiknya terhadap model *boosting* tradisional, dan validasi hanya dilakukan pada *dataset* OCPM dari kohort Amerika Latin. Keterbatasan ini membatasi generalisasi hasil untuk populasi etnis atau sosioekonomi yang lebih beragam. Oleh karena itu, penelitian lanjutan harus berfokus pada Validasi Eksternal dengan *dataset* yang lebih besar dan beragam, diikuti dengan Audit Bias Sistematis untuk memastikan model dapat diterapkan secara adil. Selain itu, arsitektur *Hybrid Stacking* dapat dikembangkan dengan mengeksplorasi penggunaan *deep learning* (seperti ANN atau LSTM) sebagai *Meta-Learner* Level-2 atau membandingkan secara langsung dengan model *boosting* yang baru. Terakhir, model *Hybrid Stacking* yang divalidasi ini sangat cocok untuk diintegrasikan ke dalam sistem pendukung keputusan klinis (*Clinical Decision Support Systems*) atau perangkat *wearable* untuk pemantauan dan intervensi dini secara *real-time.*

REFERENCES

[1] Z. Yao, B. G. Tchang, M. Albert, R. S. Blumenthal, K. Nasir, and M. J. Blaha, "Associations between Class I, II, or III Obesity and Health Outcomes," *NEJM Evidence*, vol. 4, no. 4, Mar. 2025, doi: 10.1056/EVIDoa2400229.

[2] R. Kaur, R. Kumar, and M. Gupta, "Predicting risk of obesity and meal planning to reduce the obese in adulthood using artificial intelligence," *Endocrine*, vol. 78, no. 3, pp. 458–469, Oct. 2022, doi: 10.1007/s12020-022-03215-4.

[3] N. Khodadadi, M. Saber, and M. Abotaleb, "A Data-Driven Approach for Obesity Classification using Machine Learning," *Journal of Artificial Intelligence and Metaheuristics*, vol. 3, no. 2, pp. 08–17, 2023, doi: 10.54216/JAIM.030201.

[4] H. Lim, H. Lee, and J. Kim, "A prediction model for childhood obesity risk using the machine learning method: a panel study on Korean children," *Sci Rep*, vol. 13, no. 1, p. 10122, Jun. 2023, doi: 10.1038/s41598-023-37171-4.

[5] S. Jeon, M. Kim, J. Yoon, S. Lee, and S. Youm, "Machine learning-based obesity classification considering 3D body scanner measurements," *Sci Rep*, vol. 13, no. 1, p. 3299, Feb. 2023, doi: 10.1038/s41598-023-30434-0.

[6] W. Lin, S. Shi, H. Huang, J. Wen, and G. Chen, "Predicting risk of obesity in overweight adults using interpretable machine learning algorithms," *Front Endocrinol (Lausanne)*, vol. 14, Nov. 2023, doi: 10.3389/fendo.2023.1292167.

[7] U. Sarmah, P. Borah, and D. K. Bhattacharyya, "Ensemble Learning Methods: An Empirical Study," *SN Comput Sci*, vol. 5, no. 7, p. 924, Oct. 2024, doi: 10.1007/s42979-024-03252-y.

[8] D. D. Solomon *et al.*, "Hybrid Majority Voting: Prediction and Classification Model for Obesity," *Diagnostics*, vol. 13, no. 15, p. 2610, Aug. 2023, doi: 10.3390/diagnostics13152610.

[9] I. D. Mienye and Y. Sun, "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects," 2022, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/ACCESS.2022.3207287.

[10] S. M. Ganie, P. K. D. Pramanik, and Z. Zhao, "Lifestyle data-based multiclass obesity prediction with interpretable ensemble models incorporating SHAP and LIME analysis," *Sci Rep*, vol. 15, no. 1, p. 36916, Oct. 2025, doi: 10.1038/s41598-025-20936-4.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 1996−2005
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8802

[11] I. G. S. M. Diayasa, M. Idhom, A. Fauzi, and A. T. Damaliana, "Stacking Ensemble Methods to Predict Obesity Levels in Adults," in *2022 IEEE 8th Information Technology International Seminar (ITIS)*, IEEE, Oct. 2022, pp. 339–344. doi: 10.1109/ITIS57155.2022.10010260.

[12] E. Ileberi and Y. Sun, "A Hybrid Deep Learning Ensemble Model for Credit Card Fraud Detection," *IEEE Access*, vol. 12, pp. 175829–175838, 2024, doi: 10.1109/ACCESS.2024.3502542.

[13] Y. Li, G. Liu, Y. Cao, J. Chen, X. Gang, and J. Tang, "WNPS-LSTM-Informer: A Hybrid Stacking model for medium-term photovoltaic power forecasting with ranked feature selection," *Renew Energy*, vol. 244, p. 122687, May 2025, doi: 10.1016/j.renene.2025.122687.

[14] A. Ferreras *et al.*, "Systematic Review of Machine Learning applied to the Prediction of Obesity and Overweight," *J Med Syst*, vol. 47, no. 1, p. 8, Jan. 2023, doi: 10.1007/s10916-022-01904-1.

[15] T. Khater, H. Tawfik, and B. Singh, "Machine Learning for the Classification of Obesity Levels Based on Lifestyle Factors," in *Proceedings of the 2023 7th International Conference on Cloud and Big Data Computing*, New York, NY, USA: ACM, Aug. 2023, pp. 28–33. doi: 10.1145/3616131.3616135.

[16] I. D. Mienye, Y. Sun, and Z. Wang, "An improved ensemble learning approach for the prediction of heart disease risk," *Inform Med Unlocked*, vol. 20, p. 100402, 2020, doi: 10.1016/j.imu.2020.100402.

[17] V. Hassija *et al.*, "Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence," Jan. 01, 2024, *Springer*. doi: 10.1007/s12559-023-10179-8.

[18] W. Khan *et al.*, "Predicting preterm birth using explainable machine learning in a prospective cohort of nulliparous and multiparous pregnant women," *PLoS One*, vol. 18, no. 12, p. e0293925, Dec. 2023, doi: 10.1371/journal.pone.0293925.

[19] M. Nauta *et al.*, "From Anecdotal Evidence to Quantitative Evaluation Methods: A Systematic Review on Evaluating Explainable AI," *ACM Comput Surv*, vol. 55, no. 13s, pp. 1–42, Dec. 2023, doi: 10.1145/3583558.

[20] K. Zhuang, C. Zhang, Z. Chen, T. She, and M. Wang, "Integrating convolutional neural networks with ensemble methods for enhanced diabetes diagnosis: a multi-dataset evaluation," *Front Med (Lausanne)*, vol. 12, Sep. 2025, doi: 10.3389/fmed.2025.1657889.

[21] M. Nagassou, R. W. Mwangi, and E. Nyarige, "A Hybrid Ensemble Learning Approach Utilizing Light Gradient Boosting Machine and Category Boosting Model for Lifestyle-Based Prediction of Type-II Diabetes Mellitus," *Journal of Data Analysis and Information Processing*, vol. 11, no. 04, pp. 480–511, 2023, doi: 10.4236/jdaip.2023.114025.