

Improving the Accuracy of Obesity Classification Using a Stacking Classifier on Imbalanced Data with SMOTE

Sifa Ayu Rosita sari^{1*}, M.Arief Soeleman^{2*}, Mamay Maida^{3*}, Hestiana Putri Novitasari^{4*}

* Teknik Informatika, Universitas Dian Nuswantoro

111202214204@mhs.dinus.ac.id¹, m.rief.soleman@dsn.dinus.ac.id², 111202214216@mhs.dinus.ac.id³,

111202214243@mhs.dinus.ac.id⁴

Article Info

Article history:

Received 2025-12-03

Revised 2026-01-08

Accepted 2026-01-13

Keyword:

*Obesity Classification,
Machine Learning,
Stacking Classifier,
SMOTE,
Tuning Hyperparameter.*

ABSTRACT

Overweight continues to be a prevalent public health problem related to lifestyle behavior, eating behaviour and physical activity. The aim of this work is to develop a generalized and robust machine learning model having a high accuracy for categorizing obesity-level. The study applies to the Obesity Dataset with 1610 members and some preprocessing methods such selected data cleaning, categorical attributes transformation, train/test data set split and class imbalance under utilization of SMOTE approach. The modeling process is based on two base learners namely an optimized Random Forest and Gaussian Naïve Bayes that are fused by Stacking Classifier while using Logistic Regression as the meta-model. Experimental results show that the performance of stacking is the best where it obtains an accuracy rate of 86.34%, outperforming each single model. The analysis also reveals enhancements of various classification measures: stacking can indeed model complex non-linear dependencies between instances as well as simple linear ones. In general, the results serve to demonstrate that stacking-based ensemble learning is a strong solution for predicting obesity level and holds promise against early risk detection in preventive health care systems.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Isu obesitas telah menjadi masalah kesehatan masyarakat global dengan tingkat prevalensi yang terus mengalami peningkatan [4]. Kondisi kelebihan berat badan ini bukan hanya berdampak pada penurunan kualitas hidup, tetapi juga menjadi faktor risiko utama pemicu penyakit kronis seperti hipertensi, diabetes, dan gangguan pada sistem kardiovaskular [4]. Berdasarkan berbagai penelitian, terdapat bukti kuat bahwa pola hidup memiliki kontribusi yang signifikan terhadap risiko obesitas, termasuk di antaranya adalah kebiasaan makan, tingkat aktivitas fisik, kebiasaan merokok, hingga aspek psikologis [4]. Oleh karena itu, kebutuhan akan sistem prediksi kesehatan yang memiliki ketepatan dan kecepatan tinggi semakin mendesak, menjadikan pemanfaatan *machine learning* sebagai solusi yang sangat relevan untuk mendukung program deteksi dini.

Upaya klasifikasi tingkat obesitas telah banyak dilakukan melalui penelitian menggunakan berbagai algoritma

supervised learning. Sebagai contoh, Prakoso et al. [1] melakukan studi komparasi antara *Naïve Bayes* dan *Random Forest*, yang menunjukkan keunggulan *Random Forest* dalam memprediksi kategori obesitas pada dataset berbasis faktor gaya hidup. Penelitian lain oleh Dwi et al. [2] mengonfirmasi efektivitas *Random Forest* pada data populasi Meksiko. Selain itu, *Naïve Bayes* juga dimanfaatkan oleh Maryani & Irmayansyah [3] sebagai pendekatan probabilistik untuk penentuan diagnosa obesitas pada peserta sosialisasi kesehatan. Meskipun penelitian-penelitian tersebut berhasil memberikan akurasi yang baik, sebagian besar masih berfokus pada penggunaan model tunggal tanpa mengeksplorasi potensi penggabungan model (*ensemble*) untuk menangani data yang lebih kompleks.

Walaupun demikian, mayoritas penelitian terdahulu cenderung membatasi diri pada penggunaan algoritma tunggal atau perbandingan langsung tanpa mengimplementasikan strategi optimasi model yang lebih komprehensif. Padahal, penggunaan metode *ensemble*

learning terbukti dapat meningkatkan stabilitas, akurasi, dan sensitivitas model di berbagai sektor. Untuk mengatasi masalah ketidakseimbangan kelas, teknik *oversampling* seperti SMOTE telah menunjukkan kemampuan untuk meningkatkan performa model klasifikasi secara signifikan dengan menciptakan sampel sintetis [11], [12]. Tanpa penanganan *imbalance* dan *hyperparameter tuning* yang tepat, model berisiko tinggi mengalami *overfitting* [14].

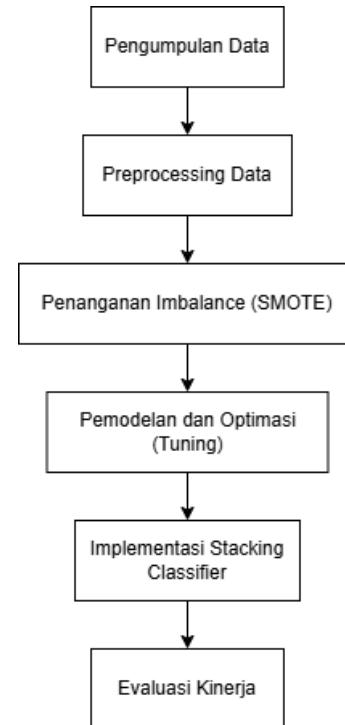
Berangkat dari kesenjangan penelitian tersebut, kami mengusulkan suatu pendekatan berjenjang dalam klasifikasi obesitas yang melibatkan penyeimbangan data menggunakan SMOTE, dilanjutkan dengan *tuning Random Forest*, dan diakhiri dengan penerapan *Stacking Classifier* sebagai model final. Pendekatan *Stacking* dirancang untuk mengintegrasikan kelebihan dari beberapa model dasar, seperti *Random Forest* yang efisien dalam menangani data tabular [1], [6], [18] dan *Naïve Bayes* yang efektif untuk pemodelan distribusi probabilistik [3], [19]. Integrasi ini diharapkan mampu mengatasi kompleksitas hubungan antarvariabel yang kerap ditemui pada data yang berhubungan dengan perilaku dan gaya hidup [5], [8].

Tujuan utama dari penelitian ini adalah merancang model klasifikasi tingkat obesitas dengan sensitivitas dan akurasi yang tinggi menggunakan dataset obesitas [24]. Penelitian ini secara spesifik bertujuan mengevaluasi efektivitas alur eksperimen yang melibatkan SMOTE, *tuning Random Forest*, dan *Stacking Classifier*, serta membandingkan hasilnya dengan model-model dasar yang telah digunakan dalam studi sebelumnya [1], [2], [3]. Kontribusi empiris dari penelitian ini diharapkan dapat memperkaya pengembangan sistem prediksi risiko obesitas yang lebih adaptif terhadap kompleksitas data kesehatan.

II. METODE

Metodologi penelitian ini dirancang sebagai sebuah kerangka kerja yang sistematis, terstruktur, dan berurutan, yang secara khusus ditujukan untuk mengonstruksi model klasifikasi obesitas dengan tingkat presisi tinggi serta mampu mengakomodasi sifat kompleks dari data faktor gaya hidup. Studi ini mengintegrasikan enam tahapan krusial yang saling berkesinambungan sesuai dengan alur pada Gambar 1, dimulai dari Pengumpulan Data primer, dilanjutkan dengan tahap *Preprocessing Data* untuk menjamin kualitas data melalui pembersihan dan transformasi teknis. Guna mengatasi tantangan distribusi data yang tidak merata, dilakukan Penanganan *Imbalance* (SMOTE) pada data latih, yang kemudian diikuti oleh fase Pemodelan dan Optimasi (Tuning) untuk memaksimalkan performa algoritma dasar melalui pencarian *hyperparameter* terbaik. Inti dari metodologi ini terletak pada Implementasi *Stacking Classifier* sebagai strategi *ensemble* yang menggabungkan kekuatan prediktif dari beberapa model berbeda, sebelum akhirnya dilakukan Evaluasi Kinerja menggunakan berbagai metrik statistik untuk menguji reliabilitas dan stabilitas model secara komprehensif. Keseluruhan rangkaian proses ini membentuk

sebuah *pipeline machine learning* yang tangguh untuk menghasilkan generalisasi prediksi yang akurat.



Gambar 1. Diagram Alir

A. Dataset

Sumber data utama riset ini adalah *Obesity Dataset*, sebuah *dataset* yang dikembangkan oleh Köklü dan Sulak [24]. Sumber data ini tersedia secara publik di *website Kaggle* berikut [Link](https://www.kaggle.com/datasets/suleymansulak/obesity): <https://www.kaggle.com/datasets/suleymansulak/obesity>-dataset. Survei online yang melibatkan 1.610 responden digunakan untuk mengidentifikasi risiko obesitas melalui analisis perilaku sosial, kebiasaan makan, dan aktivitas fisik mereka. *Dataset* ini lazim digunakan dalam studi yang menguji pemodelan risiko obesitas berbasis *machine learning* [1], [2]. Fitur yang ada mencakup berbagai variabel numerik dan kategorikal terkait langsung dengan faktor risiko obesitas. Deskripsi atribut lengkap disajikan pada Tabel I.

TABEL I
DESKRIPSI ATRIBUT DATASET OBESITAS

Nama Atribut	Tipe Data
Sex	Kategorikal
Age	Numerik (int)
Height	Numerik (cm)
Overweight/Obese Families	Kategorikal
Consumption of Fast Food	Kategorikal
Frequency of consuming vegetables	Kategorikal
Number of Main Meals Daily	Kategorikal
Food Intake Between Meals	Kategorikal
Smoking	Kategorikal

Liquid Intake Daily	Kategorikal
Calculation of Calorie Intake	Kategorikal
Physical Exercise	Kategorikal
Schedule Dedicated to Technology	Kategorikal
Type of Transportation Used	Kategorikal

TABEL II
DISTRIBUSI KELAS TARGET

Kelas Target	Jumlah Sampel (n=1610)	Sampel Latih (Awal)	Sampel Latih (Setelah SMOTE)
Underweight	73 (4.53%)	58	526
Normal	658 (40.87%)	526	526
Overweight	592 (36.77%)	474	526
Obesity	287 (17.83%)	230	526
Total	1610	1288	2104

Distribusi kelas yang tidak seimbang, ditandai dengan rendahnya proporsi kategori *Underweight* (4,53%) dan *Obesity* (17,83%) dari total sampel seperti yang ditunjukkan pada Tabel II, maka diperlukan penggunaan teknik SMOTE untuk penyeimbangan data pada tahap pelatihan

B. Pre-Processing

Tahap *pre-processing* bertujuan untuk memastikan bahwa data berada dalam kondisi terbaik sebelum masuk ke proses pelatihan model. Pada tahap ini, langkah awal yang dilakukan yaitu melakukan pembersihan data dengan menghilangkan nilai yang kosong serta entri yang tidak valid sehingga model tidak menerima data yang dapat menyebabkan bias. Selain itu, atribut yang dianggap kurang berpengaruh terhadap proses prediksi turut dihapus, misalnya *Height*, *Liquid Intake Daily*, *Schedule Dedicated to Technology*, dan *Type of Transportation Used*. Keputusan tersebut didasarkan pada hasil analisis awal yang menunjukkan bahwa atribut-atribut tersebut memiliki kontribusi yang rendah terhadap model.

Usai proses pembersihan data, semua fitur kategorikal diubah ke format numerik melalui metode Label Encoding sehingga dapat diolah oleh algoritma *machine learning* yang diterapkan dalam studi ini. Metode ini dipilih untuk menjaga efisiensi komputasi mengingat model utama yang digunakan adalah algoritma berbasis pohon (*Random Forest*) yang dapat menangani data ordinal maupun nominal secara efektif. Data selanjutnya dipisahkan menjadi data latih dan data uji melalui metode *train-test split* dengan rasio pembagian 80:20, yang menghasilkan 1.288 sampel untuk data latih dan 322 sampel untuk data uji. Pembagian tersebut menerapkan parameter *stratify* pada label target sehingga distribusi setiap kelas tetap terjaga dan mencerminkan kondisi *dataset* secara keseluruhan [13]. Keseluruhan tahapan pra-pemodelan ini memiliki peran yang krusial untuk menjamin bahwa data yang diolah telah memiliki kualitas yang memadai dan mampu menunjang peningkatan performa model pada proses analisis selanjutnya [12].

C. SMOTE (*Synthetic Minority Over-sampling Technique*)

Distribusi label pada dataset memperlihatkan adanya ketimpangan jumlah data antar kelas, terutama pada kategori *Underweight* dan *Obesity* yang jauh lebih sedikit dibandingkan kelas lainnya. Situasi tersebut dapat menimbulkan kecenderungan model untuk lebih fokus mempelajari kelas mayoritas sehingga prediksi yang dihasilkan menjadi kurang akurat dan berpotensi mengarah pada bias terhadap kelas tertentu.

Sebagai solusi, diterapkan teknik SMOTE pada data pelatihan. Metode ini bekerja dengan membuat sampel sintetis baru pada kelas minoritas berdasarkan kedekatan antar data menggunakan pendekatan *k-nearest neighbors* [11]. Dengan proses tersebut, proporsi antar kelas menjadi lebih merata sehingga model dapat meningkatkan kemampuannya dalam mengidentifikasi kelas minoritas secara lebih akurat [12]. Penerapan SMOTE dilakukan hanya pada data pelatihan (train set) setelah proses *split* dilakukan. Hal ini sangat penting untuk mencegah terjadinya *data leakage* (kebocoran data), di mana model seolah-olah memiliki akurasi tinggi namun sebenarnya hanya mengenali data sintetis yang mirip dengan data uji.

D. Tuning Random Forest

Random Forest dipilih sebagai salah satu model utama pada penelitian ini karena memiliki ketahanan yang baik terhadap *noise*, mampu mempelajari hubungan non-linear, serta telah terbukti memberikan hasil yang kuat pada analisis data perilaku dalam penelitian terdahulu [1], [6], [18]. Untuk menghasilkan performa model yang lebih optimal, dilakukan proses penyesuaian *hyperparameter* menggunakan teknik Grid Search dengan validasi silang (*cross-validation*) untuk menentukan kombinasi parameter yang paling efektif [14]. Parameter yang diuji mencakup jumlah pohon dalam model (*n_estimators*), batas maksimal kedalaman pohon keputusan (*max_depth*), serta jumlah minimum sampel pada setiap daun keputusan (*min_samples_leaf*).

Melalui proses pencarian tersebut, diperoleh konfigurasi terbaik yang digunakan dalam eksperimen ini, yaitu *n_estimators* bernilai 150, *max_depth* sebesar 20, dan *min_samples_leaf* bernilai 1. Pemilihan kombinasi parameter ini bertujuan untuk meningkatkan kemampuan model dalam melakukan generalisasi dan sekaligus mengurangi kemungkinan terjadinya *overfitting*, selaras dengan pendekatan yang disarankan pada penelitian-penelitian terdahulu [14]. Rincian parameter hasil optimasi disajikan pada Tabel III.

TABEL III
PARAMETER RANDOM FOREST SETELAH TUNING

Hyperparameter	Rentang Pencarian (Grid)	Nilai Terbaik (Optimasi)
<i>n_estimators</i>	[50, 100, 150, 200]	150
<i>max_depth</i>	[None, 10, 20, 30]	20
<i>min_samples_leaf</i>	[1, 2, 4]	1

E. Stacking Classifier

Stacking Classifier digunakan sebagai metode *ensemble learning* untuk menggabungkan kelebihan dari beberapa model dasar sehingga mampu menghasilkan prediksi yang lebih konsisten dan akurat. Pada penelitian ini, algoritma dasar yang diterapkan adalah *Random Forest* yang telah dioptimasi melalui *hyperparameter tuning* serta *Gaussian Naïve Bayes*. Proses integrasi model dilakukan dengan menggumpangkan output dari kedua *base learners* dalam bentuk nilai probabilitas kelas (*class probabilities*) kepada *meta-learner*. Penggunaan probabilitas memberikan informasi yang lebih kaya bagi *meta-learner* dibandingkan sekadar label prediksi, karena mencakup tingkat keyakinan model terhadap setiap kategori tingkat obesitas.

Output probabilitas tersebut kemudian diproses oleh *meta-learner*, yaitu *Logistic Regression*, yang berfungsi sebagai pengambil keputusan linier untuk menghasilkan prediksi akhir. Pendekatan *stacking* ini mengacu pada prinsip bahwa penggabungan beberapa algoritma dapat meningkatkan performa model secara keseluruhan, mengurangi variansi, serta memberikan stabilitas yang lebih baik dibandingkan dengan penggunaan satu model saja, sebagaimana telah disampaikan dalam berbagai penelitian mengenai *ensemble learning* [5], [8].

Untuk menjamin stabilitas model dan meminimalkan risiko bias akibat pembagian data tunggal (*single split*), penelitian ini menerapkan 5-Fold Cross Validation dalam proses pelatihan *stacking*. Melalui metode ini, dataset pelatihan dibagi menjadi lima bagian (*folds*) yang berbeda, di mana setiap bagian digunakan sebagai data validasi secara bergantian sementara empat bagian lainnya digunakan untuk pelatihan. Pendekatan ini memastikan bahwa setiap sampel data memiliki kesempatan untuk menjadi bagian dari data uji internal, sehingga memberikan estimasi performa yang lebih reliabel dan objektif dibandingkan evaluasi tanpa validasi silang.

TABEL IV
KONFIGURASI STACKING CLASSIFIER

Komponen	Model
Base Learner 1	Random Forest (tuned)
Base Learner 2	Gaussian Naïve Bayes
Input Meta-Learner	Probabilitas Kelas (Soft-Stacking)
Meta-Learner	Logistic Regression
Cross-Validation	5-Fold

F. Evaluasi Model

Untuk menilai kinerja model secara menyeluruh, evaluasi dilakukan menggunakan berbagai metrik seperti akurasi, precision, recall, F1-score, dan confusion matrix. Pemanfaatan beragam metrik ini bertujuan memberikan gambaran yang lebih mendalam mengenai kemampuan model dalam menyelesaikan klasifikasi multikelas, sekaligus menilai akurasi prediksi pada masing-masing kategori tingkat obesitas secara lebih spesifik. Evaluasi secara khusus

menitikberatkan pada nilai *recall* untuk kelas *Obesity*, karena kesalahan dalam mengklasifikasikan kategori tersebut dapat berimplikasi pada kurang optimalnya deteksi dini terhadap potensi risiko kesehatan.

Proses evaluasi dilakukan menggunakan *data uji* sebesar 20% dari keseluruhan *dataset*, yang diperoleh melalui metode *train-test split* sebagaimana disarankan pada studi yang membahas strategi pembagian data [13]. Hasil evaluasi ini menjadi dasar dalam menilai tingkat efektivitas keseluruhan *pipeline* pemodelan, termasuk pengaruh penerapan SMOTE, optimasi *Random Forest*, serta teknik *stacking* dalam meningkatkan performa klasifikasi.

III. HASIL DAN PEMBAHASAN

Bagian ini memaparkan hasil evaluasi terhadap *Stacking Classifier* yang dibangun melalui penggabungan model *Random Forest* dan *Gaussian Naïve Bayes*, dengan *Logistic Regression* berperan sebagai *meta-learner*. Pengujian dilakukan pada 20% data uji yang sebelumnya telah melewati tahapan *pre-processing*, penyeimbangan kelas menggunakan SMOTE, serta optimasi parameter pada *Random Forest*. Kinerja model kemudian dianalisis menggunakan metrik *akurasi*, *precision*, *recall*, *F1-score*, serta *confusion matrix* untuk memberikan penilaian yang lebih komprehensif terhadap performa klasifikasi *multikelas*.

Ringkasan performa model secara menyeluruh disajikan pada Tabel V. Model mencatatkan akurasi sebesar 86.34%, yang menunjukkan kemampuan prediksi tingkat obesitas dengan tingkat ketepatan yang cukup baik. Selain itu, digunakan pula metrik *macro precision*, *macro recall*, dan *macro F1-score* untuk mengevaluasi performa model secara seimbang pada setiap kelas sehingga tidak hanya bergantung pada kelas dengan jumlah data terbesar.

TABEL V
RINGKASAN PERFORMA STACKING

Metrik	Nilai
Akurasi (Data Uji)	86.34%
Rata-rata Akurasi (5-Fold CV)	91.06%
Standar Deviasi CV	0.94%
<i>Macro Precision</i>	84%
<i>Macro Recall</i>	82%
<i>Macro F1-Score</i>	83%

Evaluasi per kelas dilakukan untuk mengidentifikasi kemampuan model dalam membedakan masing-masing kategori target, yaitu *Underweight*, *Normal*, *Overweight*, dan *Obesity*. Analisis ini menjadi krusial karena dataset menunjukkan ketidakseimbangan distribusi antar kelas, yang berpotensi menurunkan kemampuan model dalam mengenali kelas minoritas. Detail capaian performa pada masing-masing kelas dapat dilihat pada Tabel VI.

TABEL VI
PERFORMA STACKING PER KELAS

Kelas	Precision	Recall	F1-Score
Underweight	0.77	0.67	0.71
Normal	0.92	0.91	0.91
Overweight	0.83	0.84	0.84
Obesity	0.83	0.86	0.84

Berdasarkan hasil tersebut, model menunjukkan akurasi tinggi pada kelas Normal serta performa yang cukup konsisten pada kelas *Overweight* dan *Obesity*. Performa lebih rendah terjadi pada kelas *Underweight* karena keterbatasan jumlah sampel pada kelas ini sebelum dilakukan SMOTE. Meskipun demikian, nilai *F1-score* 0.71 masih menunjukkan tingkat prediksi yang dapat diterima.

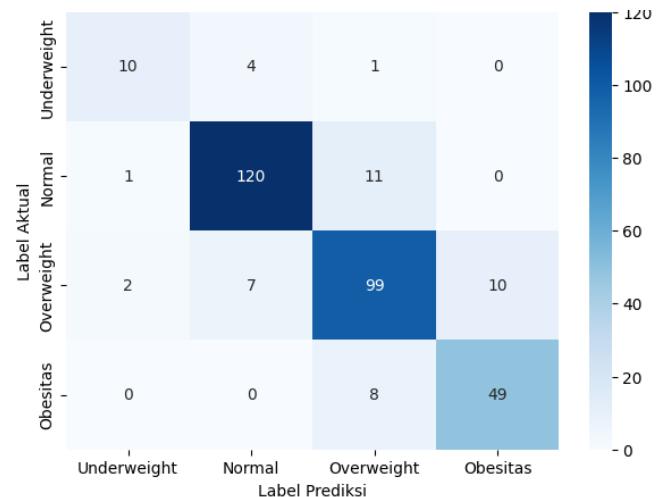
Selain metrik skor, *confusion matrix* memberikan informasi terkait pola kesalahan model dalam melakukan klasifikasi pada tiap kategori. Ringkasan *confusion matrix* ditampilkan pada Tabel VII.

TABEL VII
CONFUSION MATRIX STACKING CLASSIFIER

Aktual → Prediksi	1	2	3	4
1 (Underweight)	10	4	1	0
2 (Normal)	1	120	11	0
3 (Overweight)	2	7	99	10
4 (Obesity)	0	0	8	49

Visualisasi *confusion matrix* dalam bentuk *heatmap* diterapkan untuk memberikan representasi visual yang lebih intuitif dan komprehensif terhadap pola kesalahan klasifikasi yang dilakukan oleh model. Melalui skema gradasi warna yang merepresentasikan frekuensi prediksi, *heatmap* ini memungkinkan identifikasi cepat terhadap sel-sel diagonal yang menunjukkan keberhasilan prediksi benar (*True Positive*), sekaligus menyoroti area di luar diagonal yang mengindikasikan adanya misklasifikasi. Penggunaan intensitas warna yang proporsional dengan jumlah sampel mempermudah peneliti dalam mendeteksi kelas mana yang memiliki tingkat ambiguitas tinggi serta menilai sejauh mana model mampu mempertahankan separabilitas antar label tingkat obesitas.

Lebih lanjut, visualisasi ini menjadi instrumen penting dalam mengevaluasi efektivitas *decision boundary* yang dibentuk oleh *meta-learner*. Dengan melihat distribusi kesalahan pada *heatmap*, peneliti dapat membedakan apakah kesalahan klasifikasi bersifat acak atau memiliki pola sistematis, seperti tumpang tindih antara kategori yang berdampingan secara klinis. Oleh karena itu, penerapan *heatmap* tidak sekadar melengkapi analisis kuantitatif berbasis angka, tetapi juga berfungsi sebagai alat evaluasi kualitatif yang mendalam untuk memandu strategi peningkatan performa model, terutama dalam memitigasi kesalahan pada kelas-kelas yang memiliki dampak klinis signifikan.



Gambar 2. Visualisasi *Confusion Matrix* Model *Stacking Classifier*

Berdasarkan Tabel VII dan Gambar 2, ditemukan pola kesalahan klasifikasi yang signifikan antara kelas *Normal* dan *Overweight*. Sebanyak 11 sampel kategori *Normal* salah diprediksi sebagai *Overweight*. Hal ini menunjukkan adanya tumpang tindih (*overlap*) fitur gaya hidup pada area *borderline* antara kedua kategori tersebut. Secara klinis, individu dengan pola makan dan aktivitas fisik di batas atas kategori normal sering kali memiliki kemiripan data dengan individu di kategori awal *Overweight*, sehingga model memerlukan fitur antropometri yang lebih spesifik untuk membedakannya.

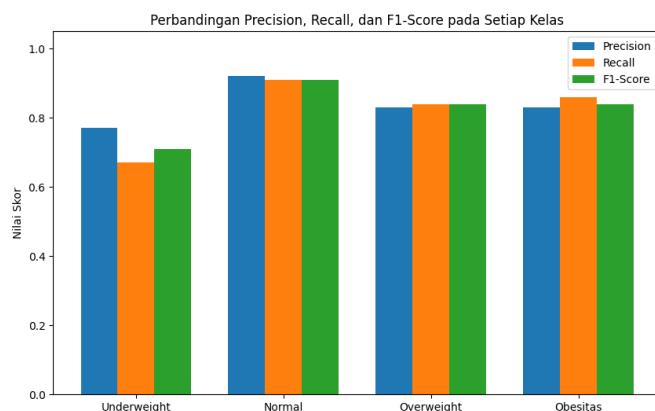
Selain analisis teknis, temuan pada Tabel VII memiliki implikasi klinis yang signifikan. Kesalahan klasifikasi pada kategori *Normal* yang diprediksi sebagai *Overweight* (11 sampel) dikategorikan sebagai *false positive*. Dalam konteks kesehatan masyarakat, hal ini lebih dapat diterima karena individu akan menerima edukasi preventif lebih awal. Sebaliknya, performa tinggi pada kelas *Obesity* dengan nilai recall 0.86 menunjukkan bahwa model sangat andal dalam mengidentifikasi individu berisiko tinggi. Hal ini meminimalkan risiko 'misdiagnosis' atau *false negative* pada pasien obesitas, sehingga intervensi medis seperti pencegahan diabetes dan hipertensi dapat dilakukan secara tepat sasaran sebelum kondisi pasien memburuk [4].

A. Analisis Kinerja Model

Hasil evaluasi yang ditampilkan pada subbab sebelumnya menunjukkan bahwa Stacking Classifier memiliki performa yang baik dalam mengklasifikasikan tingkat obesitas pada data uji. Pada bagian ini, dibahas interpretasi hasil masing-masing metrik serta perbandingan performa model sebelum dan sesudah penerapan teknik stacking.

- 1) *Analisis Per Kelas:* Skor precision dan recall pada kelas *Normal* tercatat sangat unggul, yakni masing-masing 0.92 dan 0.91. Hasil tersebut menunjukkan bahwa model mampu mengenali pola pada kelas tersebut dengan tingkat ketepatan dan konsistensi yang sangat tinggi. Jumlah sampel

yang besar pada kelas Normal memberikan model kesempatan belajar yang lebih optimal. Pada kelas *Overweight* dan *Obesity*, nilai F1-score berada pada kisaran 0.84 yang menegaskan bahwa model cukup konsisten dalam membedakan kedua kelas ini. Berbeda dengan kelas lainnya, performa kelas *Underweight* cenderung lebih rendah, terutama pada nilai recall yang hanya mencapai 0.67. Kondisi ini diduga dipengaruhi oleh ketidakseimbangan distribusi data pada kelas *Underweight* sebelum penerapan SMOTE. Meskipun teknik SMOTE telah membantu menyeimbangkan data, karakteristik kelas minoritas masih menjadi tantangan dalam proses pembelajaran model. Untuk memperjelas performa per kelas, visualisasi metrik *precision*, *recall*, dan F1-score ditampilkan dalam bentuk grafik batang pada Gambar 3.



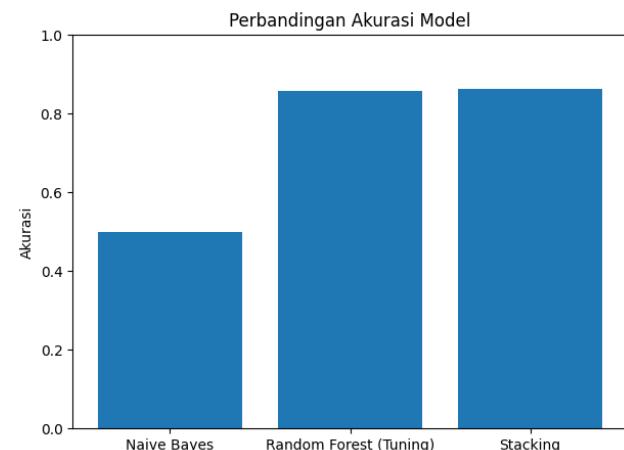
Gambar 3. Grafik *Precision*, *Recall*, dan F1-Score per Kelas

2) *Perbandingan Model Dasar dengan Stacking*: Implementasi *Stacking Classifier* mencapai akurasi 86,34%, mengungguli seluruh model dasar secara konsisten. Secara kuantitatif, *Stacking* memberikan peningkatan 0,63% dibandingkan *Random Forest (Tuned)* (85,71%) dan peningkatan signifikan sebesar 36,34% dibandingkan *Gaussian Naïve Bayes* (50,00%). Meskipun kenaikan terhadap *Random Forest* terlihat marginal, dalam klasifikasi medis, setiap peningkatan fraksional sangat krusial untuk meminimalkan risiko kesalahan diagnosis. Sebagai penutup, keberhasilan ini membuktikan bahwa sinergi antar-algoritma melalui arsitektur *stacking* mampu mengompensasi kelemahan model tunggal dan menghasilkan generalisasi yang lebih tangguh dalam memprediksi kompleksitas data risiko obesitas.

TABEL VIII
PERBANDINGAN AKURASI

Model	Akurasi	Selisih Peningkatan (Stacking)
Gaussian Naïve Bayes	50.00%	+36.34%
Random Forest (Tuned)	85.71%	+0.63%
Stacking Classifier	86.34%	-

Pada tabel tersebut, *Stacking* memiliki nilai akurasi tertinggi. Ini menunjukkan bahwa penggabungan model mampu memanfaatkan kelebihan kedua algoritma dasar *Random Forest* yang kuat pada pola kompleks dan *Naïve Bayes* yang efisien pada pola distribusi sederhana. Untuk memperkuat analisis, grafik perbandingan akurasi model dapat ditampilkan menggunakan *bar chart*.



Gambar 4. Perbandingan Akurasi Gaussian Naïve Bayes, Random Forest, dan Stacking

TABEL IX
PERBANDINGAN F1-SCORE PER KELAS ANTAR MODEL

Kelas Target	Gaussian Naïve Bayes	Random Forest (Tuned)	Stacking Classifier
Underweight	0.27	0.65	0.71
Normal	0.51	0.90	0.91
Overweight	0.53	0.83	0.84
Obesity	0.59	0.84	0.84

Mengingat karakteristik dataset yang tidak seimbang (*imbalance*), evaluasi tidak hanya didasarkan pada akurasi global, tetapi juga pada metrik *per-class* seperti *Precision*, *Recall*, dan *F1-Score*. Berdasarkan Tabel IX, terlihat bahwa model *Stacking Classifier* memberikan peningkatan *F1-Score* yang paling signifikan pada kelas minoritas, yaitu *Underweight* (0.71) dan *Obesity* (0.84). Hal ini menunjukkan bahwa penggabungan model mampu menangkap fitur-fitur pembeda pada kelas dengan jumlah sampel sedikit lebih baik dibandingkan model tunggal. Peningkatan pada nilai *Recall* di kelas *Obesity* sangat krusial dalam konteks kesehatan masyarakat, karena meminimalisir risiko individu obesitas yang tidak terdeteksi oleh sistem

TABEL X
PERBANDINGAN AKURASI DATA LATIH DAN DATA UJI

Kondisi Data	Akurasi
Data Latih (Training Set)	99.84%
Data Uji (Test Set)	86.34%
Selisih (Gap) Akurasi	13.50%

Berdasarkan Tabel X, terlihat adanya selisih akurasi sebesar 13.50% antara tahap pelatihan dan pengujian. Tingginya akurasi latih yang mencapai 99.84% menunjukkan bahwa model cenderung 'menghafal' pola data sintetis yang dihasilkan oleh teknik SMOTE secara berlebihan⁸. Fenomena *overfitting* ini menjadi salah satu keterbatasan dalam penelitian ini, mengingat ukuran dataset yang relatif terbatas ($n=1610$) menyebabkan model sangat sensitif terhadap redundansi data hasil *oversampling*.

B. Pembahasan

Selain membandingkan *Stacking Classifier* dengan model tunggal, validasi model diperkuat melalui pengujian 5-Fold Cross Validation yang menghasilkan rata-rata akurasi sebesar 91,06% dengan standar deviasi yang rendah, yaitu 0,94%. Hasil ini secara ilmiah membuktikan bahwa model memiliki stabilitas yang sangat baik dan nilai akurasi 86,34% pada data uji merupakan hasil yang konsisten, bukan sekadar kebetulan dari satu skenario pembagian data (*single split*) saja. Hal ini sekaligus menjawab kekhawatiran mengenai reliabilitas model pada dataset dengan karakteristik yang kompleks.

Sebagai contoh, pada kelas *Underweight*, model menghasilkan *precision* yang sangat rendah (0.16), namun *recall* cukup tinggi (0.87). Pola ini mengindikasikan bahwa model sering melakukan kesalahan dalam bentuk *false positive*, yakni salah mengklasifikasikan sampel dari kelas lain sebagai *Underweight*. Rincian hasil evaluasi model dirangkum pada Tabel XI.

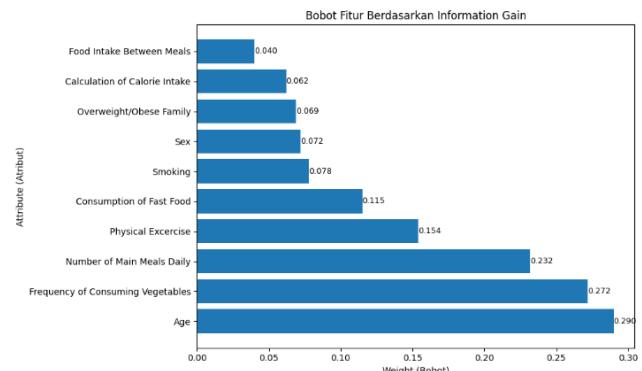
TABEL XI
RINCIAN HASIL EVALUASI GAUSSIAN NAÏVE BAYES (DENGAN SMOTE)

Kelas Target	Precision	Recall	F1-Score
Underweight	0.16	0.87	0.27
Normal	0.69	0.41	0.51
Overweight	0.62	0.47	0.53
Obesity	0.52	0.68	0.59

Kinerja Gaussian Naïve Bayes yang kurang optimal dapat dijelaskan melalui karakteristik dataset. Model ini mengasumsikan bahwa setiap fitur bersifat independen dan mengikuti distribusi Gaussian. Namun, sebagian besar fitur pada dataset obesitas merupakan variabel kategorikal hasil *label encoding*, sehingga tidak memenuhi asumsi distribusi normal. Selain itu, banyak fitur saling bergantung dan memiliki hubungan non-linear misalnya kebiasaan pola makan, aktivitas fisik, dan konsumsi makanan sehat yang melanggar asumsi *conditional independence*.

Analisis *feature importance* melalui *Information Gain* (Gambar 5) menunjukkan bahwa fitur tertentu, seperti *Age* (0.290), memiliki dominasi kuat terhadap hasil prediksi. Kondisi ini menyulitkan model probabilistik yang sensitif terhadap ketergantungan antarfitur. Hal ini diperparah dengan kenyataan bahwa sebagian besar fitur dalam dataset ini, seperti kebiasaan makan dan aktivitas fisik, memiliki korelasi linear dan non-linear yang kuat satu sama lain. Keberadaan fitur dominan seperti *Age* (0.290) menciptakan bias pada

model probabilistik yang mengasumsikan setiap fitur berdiri sendiri, sehingga prediksi menjadi tidak akurat.



Gambar 5. Bobot Fitur Berdasarkan *Information Gain* (Bagian dari Analisis Awal)

Hasil *confusion matrix* juga memperlihatkan bahwa *Gaussian Naïve Bayes* kesulitan dalam membedakan kelas *Normal*, *Overweight*, dan *Obesity* yang memiliki karakteristik perilaku sangat berdekatan. Ketiga kelas tersebut saling berbagi sejumlah pola gaya hidup yang serupa, seperti frekuensi konsumsi makanan cepat saji, aktivitas fisik, serta kebiasaan *ngemil*. Akibatnya, model sering kali menghasilkan kesalahan klasifikasi antar kelas tersebut, terutama pada sampel yang memiliki nilai fitur di area *borderline*. Hal ini menunjukkan bahwa batas keputusan (*decision boundary*) antar tingkat obesitas pada dataset ini sangat tipis, sehingga kesalahan klasifikasi sering terjadi pada kelas-kelas yang berdampingan secara klinis. Interpretasi ini membuktikan bahwa faktor perilaku manusia tidak selalu berkorelasi linier dengan kenaikan kategori berat badan secara kaku. Fenomena salah klasifikasi ini juga menunjukkan bahwa *Naïve Bayes* kurang sesuai untuk *dataset* yang memiliki hubungan antar fitur yang saling bergantung dan bersifat *non-linear*, karena model tersebut mengasumsikan independensi antar fitur serta *distribusi Gaussian*. Ketidaksesuaian antara asumsi model dan struktur data menjadikan *Gaussian Naïve Bayes* tidak mampu menangkap kompleksitas yang diperlukan untuk menghasilkan prediksi obesitas yang lebih akurat.

Kelemahan yang ditemukan pada *Gaussian Naïve Bayes* ini sekaligus menegaskan efektivitas pendekatan *stacking*. Secara ilmiah, efektivitas ini tercapai melalui prinsip keragaman model (*model diversity*), di mana *Random Forest* bekerja dengan menangkap pola non-linear yang kompleks, sementara *Naïve Bayes* memberikan estimasi probabilistik. *Meta-learner* (Logistic Regression) kemudian berfungsi sebagai pengambil keputusan linier yang mengintegrasikan probabilitas prediksi dari kedua model dasar, sehingga mampu meminimalisir kesalahan klasifikasi pada sampel yang berada di area *borderline* yang sulit dipisahkan oleh satu model tunggal.

Penelitian ini juga mengamati adanya gejala *overfitting*, di mana terdapat selisih akurasi yang cukup mencolok antara data latih (99.84%) dan data uji (86.34%). Hal ini merupakan implikasi dari penggunaan SMOTE pada dataset yang relatif

kecil, Secara teknis, SMOTE menciptakan sampel sintetis berdasarkan tetangga terdekat yang dalam dataset skala terbatas cenderung menyebabkan redundansi informasi. Akibatnya, model kehilangan sebagian kemampuan generalisasinya karena terlalu terbiasa dengan pola sintetis di tahap pelatihan. Temuan ini memberikan interpretasi penting bahwa dalam dataset medis berskala kecil, keseimbangan kelas harus diimbangi dengan teknik regularisasi yang lebih ketat.

Hasil akhir menunjukkan bahwa akurasi *Stacking* mencapai 86.34%, melampaui performa individual *Random Forest* maupun *Gaussian Naïve Bayes*. Selain itu, peningkatan nilai *precision*, *recall*, dan *F1-score* pada sebagian besar kelas mengindikasikan bahwa model gabungan ini lebih mampu menangani variasi karakteristik perilaku dan kondisi fisik pada setiap kategori obesitas. Dengan demikian, *stacking* terbukti lebih stabil dalam proses klasifikasi *multikelas* serta memiliki ketahanan yang lebih baik terhadap data yang bersifat kompleks dan tidak seimbang. Secara praktis, model *Stacking Classifier* yang dihasilkan memiliki potensi besar untuk diintegrasikan ke dalam Sistem Pendukung Keputusan Medis (*Clinical Decision Support System* - CDSS). Dalam skenario operasional layanan kesehatan, integrasi ini dapat diwujudkan melalui antarmuka API (*Application Programming Interface*) yang menghubungkan mesin inferensi (*inference engine*) berbasis Python dengan sistem informasi rumah sakit atau aplikasi kesehatan seluler.

Alur kerjanya dimulai dengan pengumpulan data variabel gaya hidup pasien melalui formulir digital. Data tersebut kemudian dikirim ke model *Stacking* untuk diproses menggunakan nilai probabilitas dari *base learners*. Jika sistem mendeteksi risiko *Obesity* dengan tingkat keyakinan tinggi (merujuk pada nilai *recall* 0,86 pada Tabel VI), CDSS dapat secara otomatis memicu protokol intervensi dini, seperti pengiriman notifikasi jadwal konsultasi gizi atau rekomendasi pemeriksaan metabolisme lebih lanjut kepada tenaga medis secara *real-time*. Dengan demikian, model ini tidak hanya berfungsi sebagai alat prediksi statistik, tetapi sebagai komponen aktif dalam sistem pencegahan penyakit kronis yang mendukung efisiensi skrining kesehatan masyarakat.

Penelitian ini mengidentifikasi beberapa keterbatasan yang berpotensi memengaruhi generalisasi hasil. Pertama, terdapat risiko bias data karena atribut yang digunakan bersifat *self-reported* (dilaporkan mandiri oleh responden). Hal ini membuka peluang terjadinya bias subjektivitas dibandingkan dengan data hasil observasi klinis langsung.

Kedua, meskipun teknik SMOTE berhasil menyeimbangkan distribusi kelas, ditemukan gejala *overfitting* yang tercermin dari selisih akurasi (gap) sebesar 13,50% antara data latih (99,84%) dan data uji (86,34%) pada Tabel X. Fenomena ini mengindikasikan bahwa pada dataset berukuran relatif kecil ($n=1610$), SMOTE cenderung menciptakan redundansi informasi yang membuat model terlalu mempelajari pola spesifik dari data sintetis daripada pola umum populasi.

Ketiga, keterbatasan skala dataset membatasi validitas eksternal model untuk populasi global yang memiliki variasi genetik dan budaya makan yang berbeda. Oleh karena itu, penelitian selanjutnya disarankan untuk mengintegrasikan teknik regularisasi yang lebih ketat atau penggunaan dataset klinis berskala besar untuk meningkatkan ketahanan model terhadap bias.

IV. KESIMPULAN

Penelitian ini berorientasi pada pengembangan model klasifikasi tingkat obesitas melalui pendekatan *ensemble learning*, yaitu *Stacking Classifier*, yang mengintegrasikan *Random Forest* dan *Gaussian Naïve Bayes* dengan *Logistic Regression* sebagai *meta-learner*. Berdasarkan hasil evaluasi yang diperoleh, model *Stacking* menunjukkan performa superior dengan akurasi mencapai 86.34%. Angka ini tidak hanya melampaui akurasi *Random Forest* yang sudah di-*tuning* (85.71%), tetapi juga jauh meninggalkan kinerja *Gaussian Naïve Bayes* (50%). Hasil tersebut mengindikasikan bahwa kombinasi beberapa algoritma dengan keunggulan yang saling melengkapi mampu menghasilkan model prediksi yang lebih konsisten dan memiliki tingkat akurasi lebih baik.

Rangkaian proses *pra-pemodelan* yang meliputi pembersihan dataset, transformasi fitur kategorikal, pemisahan data, serta penyeimbangan distribusi kelas melalui SMOTE, terbukti memberikan kontribusi signifikan dalam memperbaiki kualitas data dan meningkatkan kinerja model secara menyeluruh. Selain itu, proses *tuning* yang diterapkan pada *Random Forest* memberikan dampak positif signifikan dalam meningkatkan kemampuannya dalam mengekstraksi pola kompleks dari dataset. Meskipun *Gaussian Naïve Bayes* menunjukkan performa yang rendah karena adanya keterbatasan asumsi model terhadap fitur yang saling berkorelasi, hasil ini justru memperkuat justifikasi penggunaan teknik *stacking*.

Secara umum, hasil penelitian menunjukkan bahwa penerapan *Stacking Classifier* menjadi strategi yang sangat efisien dalam melakukan prediksi tingkat obesitas dengan memanfaatkan data terkait kebiasaan hidup dan perilaku individu. Model ini berhasil memitigasi isu utama yang dihadapi, seperti permasalahan ketidakseimbangan kelas, hubungan *non-linear* antar fitur, serta keterbatasan model individual. Keberhasilan ini membuka peluang implementasi lebih lanjut, terutama dalam pengembangan sistem pendukung keputusan di bidang kesehatan preventif.

Pada penelitian mendatang, pengembangan model dapat diarahkan dengan memanfaatkan algoritma boosting yang lebih mutakhir seperti XGBoost maupun LightGBM, atau dengan memperkaya data menggunakan variabel antropometri yang lebih detail. Eksplorasi teknik *feature selection* lebih lanjut juga dapat dilakukan untuk mengidentifikasi subset fitur yang paling berpengaruh. Selain itu, penggunaan teknik *explainable AI* dapat memberikan wawasan tambahan mengenai interpretasi hasil prediksi yang dihasilkan oleh model.

DAFTAR PUSTAKA

- [1] Prakoso, R. N., Rochim, S. I., Subarnas, A., & Kurniawan, M. E. (2025). Perbandingan Algoritma Naïve Bayes Dan Random Forest Dalam Klasifikasi Obesitas Berdasarkan Faktor Gaya Hidup. *Journal of Information Engineering and Educational Technology*, 9(1), 11–18. <https://doi.org/10.26740/jeet.v9n1.p11-18>
- [2] Dwi, E., Aini, N., Khasanah, R. A., Ristyawan, A., Diniati, E., Nusantara, U., & Kediri, P. (2024). Penggunaan Data Mining untuk Prediksi tingkat Obesitas di Meksiko Menggunakan Metode Random Forest. In *Agustus* (Vol. 8). Online.
- [3] Maryani, I., & Irmayansyah, I. (2023). Penerapan Algoritma Naïve Bayes Untuk Penentuan Diagnosa Obesitas Pada Peserta Sosialisasi Deteksi Dini Penyakit Tidak Menular (PTM). *TeknoIS : Jurnal Ilmiah Saraswati*, S. K., Rahmalingrum, F. D., Pahsya, M. N. Z., Paramitha, N., Wulansari, A., Ristantya, A. R., Sinabutar, B. M., Pakpahan, V. E., & Nandini, N. (2021). Literature Review : Faktor Risiko Penyebab Obesitas. *MEDIA KESEHATAN MASYARAKAT INDONESIA*, 20(1), 70–74. <https://doi.org/10.14710/mkmi.20.1.70-74>
- [4] Emilia Sukmawati, C., Fitri Nur Masruriyah, A., Ratna Juwita, A., Damaiarta Tejayanda, R., Nurmayanti, T., Korespondensi, P., & Buana Perjuangan Karawang Jl Ronggowaluyo, U. H. (2024). Efektivitas algoritma AdaBoost dan XGBoost pada dataset obesitas populasi dewasa. *Jambura Journal of Informatics*, 6(2), 101–111. <https://doi.org/10.37905/jji>
- [5] Novianti, N., Alkadri, S. P. A., & Fakhruzi, I. (2024). Klasifikasi Penyakit Hipertensi Menggunakan Metode Random Forest. *Progresif: Jurnal Ilmiah Komputer*, 20(1), 380. <https://doi.org/10.35889/progresif.v20i1.1663>
- [6] Aulia, Y., Andriyansyah, A., Suharjito, S., & Nensi, S. W. (2024). Analisis Prediksi Stroke dengan Membandingkan Tiga Metode Klasifikasi Decision Tree, Naïve Bayes, and Random Forest. *Jurnal Ilmu Komputer Dan Informatika*, 3(2), 89–98. <https://doi.org/10.54082/jiki.90>
- [7] K. Dhibi, M. Mansouri, K. Bouzrara, H. Nounou and M. Nounou, "An Enhanced Ensemble Learning-Based Fault Detection and Diagnosis for Grid-Connected PV Systems," in IEEE Access, vol. 9, pp. 155622-155633, 2021, doi: 10.1109/ACCESS.2021.3128749.
- [8] Mustaqim, A. Z., Fadil, N. A., & Tyas, D. A. (2023). Artificial Neural Network for Classification Task in Tabular Datasets and Image Processing: A Systematic Literature Review. *Jurnal Online Informatika*, 8(2), 158–168. <https://doi.org/10.15575/join.v8i2.1002>
- [9] Sofiyah, W., Negara, B. S., Irsyad, M., Iskandar, I., & Yanto, F. (2025). Lung Disease Detection Using Gradient-Weighted Class Activation Mapping (Grad-CAM). *Journal of Artificial Intelligence and Software Engineering*, 5(2), 720–730. <https://doi.org/10.30811/aise.v5i2.7041>
- [10] Lutfi, M., Arsanto, A. T., Amrulloh, M. F., & Kulsum, U. (2023). Penanganan Data Tidak Seimbang Menggunakan Hybrid Method Resampling Pada Algoritma Naive Bayes Untuk Software Defect Prediction. *INFORMAL: Informatics Journal*, 8(2), 119. <https://doi.org/10.19184/isi.v8i2.41090>
- [11] Maharana, K., Mondal, S., & Nemade, B. (2022). A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*, 3(1), 91–99. <https://doi.org/10.1016/j.gltp.2022.04.020>
- [12] Joseph, V. R. (2022). *Optimal Ratio for Data Splitting*. <https://doi.org/10.1002/sam.11583>
- [13] Diukarev, V., & Starukhin, Y. (2024). Proposed Methods for Preventing Overfitting in Machine Learning and Deep Learning. *Asian Journal of Research in Computer Science*, 17(10), 85–94. <https://doi.org/10.9734/ajrcos/2024/v17i10511>
- [14] Nofianti, A., Yawan, M. Y., & Nazar, M. A. (2023). Implementasi Data Mining dalam Pengolahan Data Transaksi Toko Sembako Menggunakan Algoritma Apriori (Studi Kasus : Toko Devan Mart). *G-Tech: Jurnal Teknologi Terapan*, 7(1), 165–173. <https://doi.org/10.33379/gtech.v7i1.1962>
- [15] Algoritma, A., Pada, K., Rapidminer, S., & Ainurrohmah, W. (2021). Akurasi Algoritma Klasifikasi pada Software Rapidminer dan Weka. *Prosiding Seminar Nasional Matematika*, 4, 493–499. <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [16] Ekin Adhi Guna, M. Davin Diza Ghifary, Esra Fransiska Sihombing, & Age Pius Datubara. (2023). Implementasi Algoritma Decision Tree untuk Klasifikasi Data Evaluation Car Menggunakan Python. *Jurnal Sistem Informasi Dan Ilmu Komputer*, 1(4), 167–177. <https://doi.org/10.59581/jusiik-widyakarya.v1i4.1830>
- [17] Mahmuda, S. (2024). Implementasi Metode Random Forest pada Kategori Konten Kanal Youtube. *JURNAL JENDELA MATEMATIKA*, 2(01), 21–31. <https://doi.org/10.57008/jjm.v2i01.633>
- [18] Alvina Felicia Watratan, Arwini Puspita, B., & Dikwan Moeis. (2020). Implementasi Algoritma Naïve Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia. *Journal of Applied Computer Science and Technology*, 1(1), 7–14. <https://doi.org/10.52158/jacost.v1i1.9>
- [19] Saputro, M. B., & Alamsyah, A. (2024). Comparison of Naive Bayes Classifier and K-Nearest Neighbor Algorithms with Information Gain and Adaptive Boosting for Sentiment Analysis of Spotify App Reviews. *Recursive Journal of Informatics*, 2(1), 37–44. <https://doi.org/10.15294/rji.v2i1.6851>
- [20] Sari, P. W. S., Firmansyah, F., & Kadafi, A. R. K. (2025). Perbandingan Algoritma Random Forest Dan Naïve Bayes Dalam Menganalisis Sentimen Ulasan Pada Produk Skincare Lokal Di Media Sosial Tiktok. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(3S1). <https://doi.org/10.23960/jitet.v13i3S1.8150>
- [21] Author, D. F. A. R., & Author, U. C. (2025). Perancangan Sistem Monitoring Dan Manajemen Proyek Pegawai Berbasis Website Dengan Framework Laravel. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(3S1). <https://doi.org/10.23960/jitet.v13i3S1.7770>
- [22] Sofiyah, W., Negara, B. S., Irsyad, M., Iskandar, I., & Yanto, F. (2025). Lung Disease Detection Using Gradient-Weighted Class Activation Mapping (Grad-CAM). *Journal of Artificial Intelligence and Software Engineering*, 5(2), 720–730. <https://doi.org/10.30811/aise.v5i2.7041>
- [23] Köklü, N., & Sulak, S.A. (2024). *Obesity Dataset*. Kaggle. Available at: kaggle.com/datasets/suleymansulak/obesity-dataset