

## Optimasi AdaBoost dan XGBoost untuk Klasifikasi Obesitas Menggunakan SMOTE

Cici Emilia Sukmawati<sup>1</sup>, Adi Rizky Pratama<sup>2</sup>, Hanny Hikmayanti<sup>3</sup>, Ayu Ratna Juwita<sup>4</sup>  
<sup>1,2,3,4</sup>Fakultas Ilmu Komputer, Universitas Buana Perjuangan Karawang, Jl. Ronggo Waluyo, Karawang, 41361, Indonesia

### Info Artikel

#### Riwayat Artikel:

Received 2025-03-03

Revised 2025-07-06

Accepted 2025-05-28

**Abstract** – Obesity is a condition in which a person's weight exceeds the normal limit due to excessive accumulation of fat tissue. Thus, obesity is considered a global public health challenge. This is evidenced by the latest data from the World Health Organization (WHO) in 2022, namely that 2.5 billion adults aged 18 years and over are overweight and 890 million of them are obese. Therefore, it is very important to accurately identify these risk factors in order to implement effective interventions in the prevention and management of obesity. However, in previous studies there has been no application of SMOTE with the AdaBoost and XGBoost algorithms, so this study aims to compare the performance of the AdaBoost and XGBoost algorithms with SMOTE. The stages of this research begin with problem identification, data collection, preprocessing and model evaluation and model comparison. This study also applies the SMOTE technique to balance unbalanced data. Based on the results of the research that has been carried out, it shows that the accuracy and recall values of the XGBoost algorithm with SMOTE are 0.945 and precision 0.947. Meanwhile, the accuracy and recall values on AdaBoost with SMOTE are 0.388. Then, the precision is 0.371. Thus, it is expected that the results of the XGBoost model with SMOTE can be a source for other research and can help in efforts to prevent and manage obesity.

**Keywords:** AdaBoost; Obesity; SMOTE; XGBoost

### Corresponding Author:

Cici Emilia Sukmawati

Email:

cici.emilia@ubpkarawang.ac.id



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

**Abstrak** – Obesitas ini adalah kondisi di mana berat badan seseorang melebihi batas normal akibat penumpukan jaringan lemak yang berlebihan. Sehingga, Obesitas dianggap sebagai tantangan kesehatan masyarakat dunia. Hal tersebut dibuktikan dengan data terbaru dari Organisasi Kesehatan Dunia (WHO) pada tahun 2022, yaitu sebanyak 2,5 miliar orang dewasa berusia 18 tahun ke atas mengalami kelebihan berat badan serta 890 juta di antaranya menderita obesitas. Oleh karena itu, sangat penting untuk secara akurat mengidentifikasi faktor-faktor risiko ini agar dapat menerapkan intervensi yang efektif dalam pencegahan dan pengelolaan obesitas. Namun, pada penelitian sebelumnya telah belum adanya penerapan SMOTE dengan algoritma AdaBoost dan XGBoost, maka penelitian ini bertujuan untuk membandingkan kinerja Algoritma AdaBoost dan XGBoost dengan SMOTE. Tahapan penelitian ini dimulai dengan identifikasi masalah, pengumpulan data, preprocessing dan evaluasi model serta perbandingan model. Pada penelitian ini juga menerapkan Teknik SMOTE untuk menyeimbangkan data yang tidak seimbang. Berdasarkan hasil penelitian yang telah dilakukan menunjukkan bahwa nilai akurasi dan recall pada algoritma XGBoost dengan SMOTE sebesar 0,945 serta presisi 0,947. Sedangkan nilai akurasi dan recall pada AdaBoost dengan SMOTE sebesar 0,388. Kemudian, presisi sebesar 0,371. Dengan demikian, diharapkan hasil dari model XGBoost dengan SMOTE ini dapat menjadi sumber bagi penelitian lain serta dapat membantu dalam upaya pencegahan serta pengelolaan obesitas.

**Kata Kunci:** Obesitas, SMOTE, AdaBoost, XGBoost

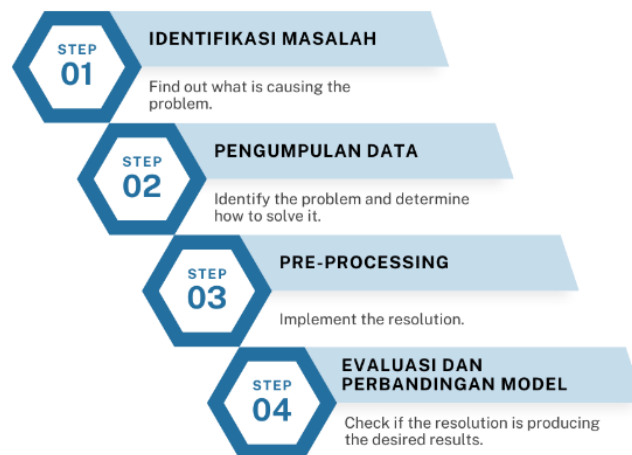
## I. PENDAHULUAN

Obesitas dikatakan sebagai masalah kesehatan global yang terus meningkat prevalensinya [1]. Selain itu, Obesitas ini adalah kondisi di mana berat badan seseorang melebihi batas normal akibat penumpukan jaringan lemak yang berlebihan[2], [3]. Sehingga, Obesitas dianggap sebagai tantangan kesehatan masyarakat dunia[4]. Hal tersebut dibuktikan dengan data terbaru dari Organisasi Kesehatan Dunia (WHO) pada tahun 2022, yaitu sebanyak 2,5 miliar orang dewasa berusia 18 tahun ke atas mengalami kelebihan berat badan serta 890 juta di antaranya menderita obesitas. Hal tersebut disebabkan oleh jumlah energi yang dibakar oleh tubuh yang lebih sedikit dibandingkan dengan jumlah kalori yang dikonsumsi. Seseorang dianggap mengalami obesitas jika dilihat dari hasil pengukuran Indeks Massa Tubuh (IMT) [5]. Adapun Obesitas tersebut ditandai dengan akumulasi lemak tubuh yang berlebihan, termasuk penyakit jantung dan diabetes tipe 2 serta kanker [6] [7]. Kondisi ini umumnya disebabkan oleh konsumsi kalori yang melebihi kebutuhan tubuh. Sehingga kemudian diubah menjadi lemak jenuh [8]. Selain itu, terdapat beberapa faktor yang dapat menyebabkan obesitas yaitu pola makan yang tidak sehat, kurangnya aktivitas fisik, dan pengaruh genetik yang dapat meningkatkan risiko [9], [10]. Selain itu, faktor penyebab lainnya ialah rendahnya konsumsi sayur dan buah [11]. Oleh karena itu, sangat penting untuk secara akurat mengidentifikasi faktor-faktor risiko ini agar dapat menerapkan intervensi yang efektif dalam pencegahan dan pengelolaan obesitas. Adapun salah satu pendekatan yang sering digunakan yaitu teknik data mining [12].

Beberapa penelitian telah memanfaatkan teknik data mining untuk menangani obesitas [9], [13], [14], [15], [16]. Adapun penelitian yang pertama yang dilakukan siti dkk dengan membandingkan empat algoritma. Algoritma tersebut terdiri dari KNN, SVM, *Decision Tree* dan Regresi Logistik. Data yang digunakan diperoleh dari *kaggle*. Pada penelitian tersebut menunjukkan bahwa *Decision Tree* lebih unggul dibandingkan dengan yang lainnya yaitu akurasi 99,3%, *precision* 97-100%, *recall* 98-100% serta *f1-score* 98-100% [15]. Dengan menggunakan data yang sama Dian dkk melakukan perbandingan KNN dan *Ensemble Learning*. Berdasarkan hasil perbandingan yang dilakukan menunjukkan bahwa KNN lebih efektif dengan akurasi 89,60% [17]. Selanjutnya, masih dengan menggunakan data yang sama Eggy dkk melakukan perbandingan antara *Support Vector Machine* (SVM) dan *Naïve Bayes* (NB). Hasil evaluasi menunjukkan bahwa metode SVM mencapai akurasi sebesar 89%, sedangkan metode *Naive Bayes* hanya mencapai akurasi sebesar 65% [18]. Bukan hanya itu, terdapat penelitian lain yang dilakukan oleh Heru dkk untuk membandingkan beberapa algoritma yaitu *Naive Bayes*, *K-Nearest Neighbors* (KNN), *Decision Trees*, *Random Forest*, dan C4.5. Adapun hasil penelitian tersebut menunjukkan bahwa algoritma *random forest* memiliki performa paling baik dibandingkan algoritma lainnya, dengan akurasi 96%, standar deviasi antar kelipatan 0,02, dan nilai AUC 1 atau sempurna dibanding algoritma lainnya [19]. Dikarenakan penelitian sebelumnya belum adanya penerapan SMOTE dengan algoritma *AdaBoost* dan *XGBoost*, maka penelitian ini bertujuan untuk membandingkan kinerja Algoritma *AdaBoost* dan *XGBoost* dengan SMOTE. Adapun penelitian sebelumnya juga menerapkan Teknik SMOTE untuk klasifikasi jantung coroner yang dilakukan oleh Mariana dewi dkk. Pada penelitian tersebut menunjukkan bahwa penggunaan Teknik SMOTE dapat meningkatkan nilai akurasi dari 76.60% menjadi 80,85% [20]. Selain itu, penelitian lain yang dilakukan oleh Debi dkk menerapkan Teknik SMOTE untuk klasifikasi penyakit kanker. Pada penelitian tersebut menunjukkan hasil klasifikasi tergolong klasifikasi terbaik dengan AUC 1.00% [21]. Demikian hal itu, penelitian ini diharapkan dapat memberikan pemahaman yang berharga bagi peneliti, praktisi kesehatan, dan pembuat kebijakan dalam upaya pencegahan serta pengelolaan obesitas. Adapun perbedaan utama pada penelitian ini dengan penelitian sebelumnya yaitu terletak pada penggunaan SMOTE dengan algoritma boosting yaitu *AdaBoost* dan *XGBoost* untuk klasifikasi pada data obesitas yang tidak seimbang.

## II. METODE

Adapun penelitian ini menggunakan beberapa tahapan, yaitu Identifikasi masalah, pengumpulan data, preprocessing, evaluasi dan perbandingan. hal tersebut dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Berdasarkan Gambar 1. menunjukkan beberapa tahapan yang digunakan pada penelitian. Tahap pertama yaitu identifikasi masalah. Adapun identifikasi masalah tersebut digunakan untuk memastikan bahwa metode yang diterapkan akan menghasilkan solusi yang efektif. Kemudian, tahap selanjutnya yaitu pengumpulan data. Adapun tahap tersebut digunakan dalam mengumpulkan, mengukur, dan menganalisis informasi untuk mendapatkan wawasan, membuat keputusan, atau menjawab pertanyaan tertentu. Selain itu, pengumpulan data merupakan langkah penting dalam penelitian. Adapun data yang dikumpulkan diperoleh dari sumber data sekunder yaitu *kaggle* sebanyak 2111 dengan 17 atribut. Isi atribut yaitu tersebut Usia, Jenis Kelamin, Tinggi Badan, Berat Badan, CALC (konsumsi alkohol), FAVC (konsumsi makanan berlemak), FCVC (konsumsi sayur/buah), NCP (frekuensi makan di luar), SCC (kebiasaan merokok), Smoke (status merokok), CH20

(konsumsi air putih), Riwayat Keluarga dengan Kelebihan Berat Badan, FAF (aktivitas fisik), TUE (penggunaan teknologi), CAEC (konsumsi kalori), dan MTRANS (moda transportasi).

Selanjutnya yaitu tahap preprocessing. Pada tahap tersebut data yang diperoleh dari tahap sebelumnya akan dilakukan proses cleaning data. Proses cleaning data tersebut terdiri dari pembersihan data dari missing value, noise, outlier serta duplikat. Selain itu, pada tahap ini juga diterapkan Teknik SMOTE. Teknik tersebut SMOTE pertama kali diperkenalkan oleh Chawla pada tahun 2002 [22]. SMOTE tersebut merupakan suatu metode untuk menjaga proporsi kelas agar lebih seimbang [23]. Selain itu, SMOTE dapat meningkatkan performa klasifikasi [24]. Adapun SMOTE tersebut bekerja secara acak memilih data dari kelas minoritas, kemudian mencari k data terdekat dari kelas yang sama. Dari k data tersebut, satu data akan dipilih secara acak dan dihubungkan dengan data awal yang telah dipilih untuk membentuk segmen garis dalam ruang fitur. Data sintetik dihasilkan melalui kombinasi convex antara kedua data yang dipilih secara acak tersebut [25]. Sehingga, memungkinkan model pembelajaran untuk menghasilkan area pembentukan model klasifikasi yang lebih luas cakupannya [22].

Selanjutnya yaitu tahap evaluasi dan perbandingan. Pada tahap tersebut dilakukan proses evaluasi model berdasarkan metode atau algoritma yang digunakan. Adapun algoritma yang diterapkan pada penelitian ini ialah AdaBoost dan XGBoost. Algoritma AdaBoost merupakan representatif dari semua algoritma boosting [26]. Algoritma ini dapat membangun pengklasifikasian yang kuat dari pengklasifikasian yang lemah. AdaBoost memiliki peran sebagai alat yang dikombinasikan dengan algoritma klasifikasi seperti pohon keputusan, regresi logistik, Naive Bayes, dll., dalam meningkatkan akurasi model tunggal. Pada fase pelatihan data dari AdaBoost, bobot pengklasifikasi dasar dan distribusi probabilitas dari sampel pelatihan dihitung berdasarkan tingkat kesalahan pelatihan [27]. Berikut persamaan dari AdaBoost:

$$H(x) = \sum_{t=1}^T \alpha_t h_t(x) \quad (1)$$

Dimana:

$h_t(x)$	: Pengklasifikasi dasar atau lemah
$\alpha_t$	: Tingkat pembelajaran ( <i>learning rate</i> )
$H(x)$	: Hasil, berupa pengklasifikasi kuat atau akhir

Sedangkan, algoritma XGBoost adalah suatu grup *Decision Tree* yang didasarkan pada *Gradient Boosting* dan dirancang untuk menjadi sangat *scalable*. Seperti *gradient boosting*, XGBoost melakukan ekspansi tujuan secara additif dengan menjaga minimisasi fungsi kerugian. Oleh karena itu, XGBoost hanya menangani *Decision Tree* sebagai *Classifier* Dasar, serta variasi fungsi kerugian digunakan untuk mengontrol kompleksitas pohon [23]. Rumus XGBoost memperkenalkan istilah regularisasi dalam fungsi objektif untuk mencegah *overfitting*. Fungsi objektif didefinisikan pada persamaan (1) : [28].

$$0 = \sum_{i=1}^n L(y_i, F(x_i)) + \sum_{k=1}^t R(f_k) + C \quad (2)$$

Pada persamaan (2) terdapat beberapa penjelasan sebagai berikut:

$L(y_i, F(x_i))$	: fungsi kerugian ( <i>loss function</i> ) yang digunakan untuk mengukur seberapa baik model dalam memprediksi data.
$R(f_k) = \alpha H + \frac{1}{2} n \sum \omega_j^2 H_j = 1$	: merupakan istilah regularisasi yang digunakan untuk mencegah <i>overfitting</i> .

Penjelasan lebih lengkap sebagai berikut:

$\alpha$	: mewakili kompleksitas daun.
$H$	: menunjukkan jumlah daun.
$n$	: menunjukkan parameter penalti.
$\omega_j^2$	: menunjukkan hasil output dari setiap simpul daun.
$C$	: merupakan konstanta yang dapat dihilangkan secara efektif.

Selanjutnya, penerapan algoritma XGBoost dan AdaBoost dengan SMOTE. Namun, sebelum penerapan algoritma tersebut dilakukan pembagian dataset. Dataset dibagi menggunakan perbandingan 80:20 dengan data sebelum SMOTE dan sesudah SMOTE. Kemudian, dataset tersebut dilakukan penerapan algoritma dan di evaluasi dengan menggunakan *confusion matrix*. Confusion Matrix merupakan salah satu metode evaluasi dalam klasifikasi yang disajikan dalam bentuk tabel matriks. Matriks ini digunakan untuk membandingkan hasil prediksi model dengan data sebenarnya, serta memberikan informasi rinci tentang jumlah prediksi yang tepat dan keliru pada tiap kelas, sehingga membantu dalam menganalisis kinerja model [29]

### III. HASIL DAN PEMBAHASAN

#### A. Hasil Identifikasi Masalah

Berdasarkan hasil identifikasi masalah yang telah dilakukan menunjukkan bahwa penyakit obesitas ini tidak hanya terus berkembang, namun berpotensi menimbulkan dampak negatif jangka panjang. Hal tersebut ditunjukkan dengan prevalensi obesitas yang terus meningkat berdasarkan data statistik terbaru. Adapun obesitas ini dapat mengakibatkan dampak kesehatan, dampak psikologis dan kesehatan mental.

#### B. Hasil Pengumpulan Data

Adapun dataset yang digunakan diperoleh dari sumber data sekunder yaitu kaggle sebanyak 2111 dengan 17 atribut. Isi atribut yaitu tersebut Usia, Jenis Kelamin, Tinggi Badan, Berat Badan, CALC (konsumsi alkohol), FAVC (konsumsi makanan berlemak), FCVC (konsumsi sayur/buah), NCP (frekuensi makan di luar), SCC (kebiasaan merokok), Smoke (status merokok), CH20 (konsumsi air putih), Riwayat Keluarga dengan Kelebihan Berat Badan, FAF (aktivitas fisik), TUE (penggunaan teknologi), CAEC (konsumsi kalori), dan MTRANS (moda transportasi). Potongan data dapat dilihat pada Tabel 3. Sedangkan atribut target adalah *Nobeyesdad* yang terdiri dari beberapa kategori: Berat Badan Tidak Cukup (berat badan di bawah normal), Berat Badan Normal (ideal), Kelebihan Berat Badan Tingkat 1, Obesitas Tipe 1 (obesitas ringan), Obesitas Tipe 2 (obesitas sedang), dan Obesitas Tipe 3 (obesitas berat).

#### C. Hasil Preprocessing Data

Pada tahap ini, dilakukan pra-pemrosesan data yang dimulai dengan memeriksa missing value dan duplikat data. Hasil dari pemeriksaan tersebut, tidak terdapat missing value pada dataset yang digunakan. Namun, terdapat data duplikat sebanyak 558 data. Oleh karena itu, data tersebut dihapus dengan menggunakan fungsi *drop\_duplicates()*. Sehingga jumlah data berkurang menjadi 1553 dari total 2111 data. Selanjutnya dilakukan pengecekan noise dan *outlier*. Berdasarkan hasil pengecekan tersebut terdapat adanya noise sebanyak 180 data dan 782 *outlier*. Kemudian didapatkan hasil akhir 591 data.

Adapun tahap selanjutnya adalah transformasi atau mengubah atribut (*Nobeyesdad*) dari kategori menjadi numerik dengan label encoding. Berikut cuplikan data sebelum transformasi data ditunjukkan pada Gambar 2. Sedangkan, hasil dari transformasi data ditunjukkan pada Gambar 3.

FAF	TUE	CALC	MTRANS	Nobeyesdad
1	0	4	3	Normal_Weight
1	1	2	4	Normal_Weight
1	1	3	4	Normal_Weight
1	0	4	1	Normal_Weight
0	1	4	4	Normal_Weight
...	...	...	...	...
2	1	4	4	Obesity_Type_III

Gambar 2. Cuplikan data sebelum transformasi

FAF	TUE	CALC	MTRANS	Nobeyesdad
1	0	4	3	1
1	1	2	4	1
1	1	3	4	1
1	0	4	1	1
0	1	4	4	1
...	...	...	...	...
2	1	4	4	4

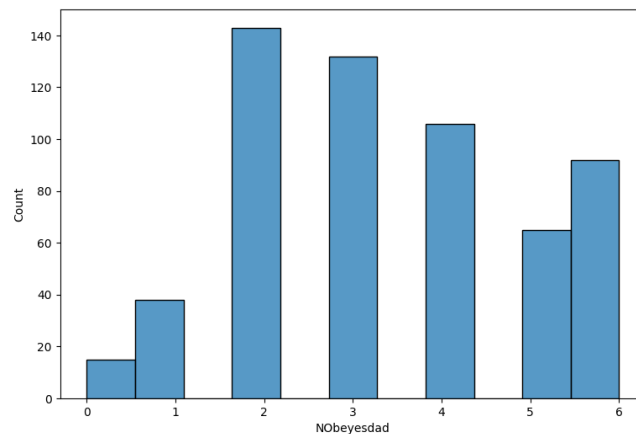
Gambar 3. Cuplikan data setelah transformasi

Adapun tahap selanjutnya yaitu normalisasi data. Proses normalisasi ini diterapkan pada semua atribut, kecuali atribut *Nobeyesdad*. Berikut hasil dari normalisasi data yang ditunjukkan pada Gambar 4.

FAF	TUE	CALC	MTRANS
0.333333	0.0	1.0	0.50
0.333333	0.5	0.0	0.75
0.333333	0.5	0.5	0.75
0.333333	0.0	1.0	0.00
0.000000	0.5	1.0	0.75

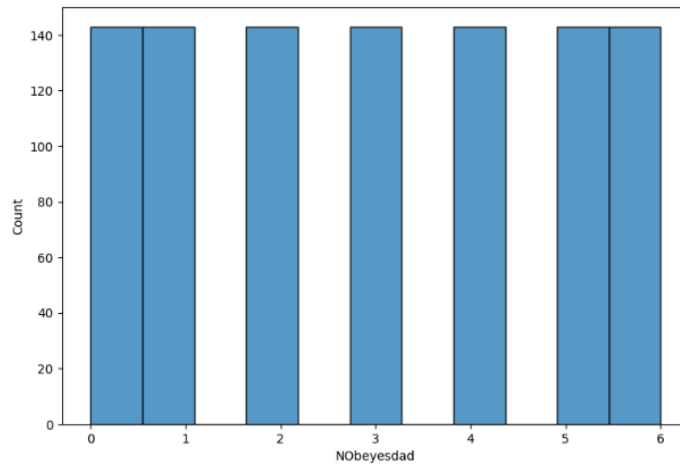
Gambar 4. Cuplikan data setelah normalisasi

Berdasarkan Gambar 4, menunjukkan bahwa atribut-atribut tersebut dilakukan normalisasi dengan rentang nilai 0-1. Demikian, antar atribut memiliki nilai rentang yang sama. Selanjutnya visualisasi nilai dari atribut *Nobeyesdad* ditunjukkan pada Gambar 5.



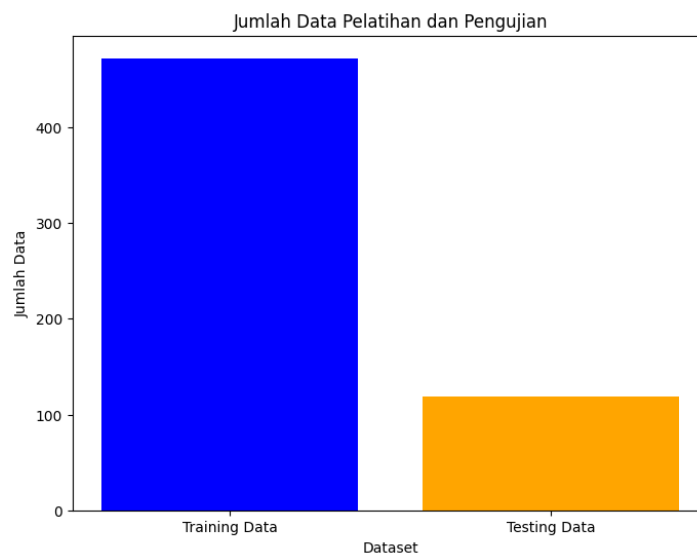
Gambar 5. Data sebelum di SMOTE

Berdasarkan Gambar 5 menunjukkan visualisasi dari nilai atribut *Nobeyesdad*. Terdapat adanya ketidakseimbangan data. Oleh karena itu, dalam penelitian ini perlu adanya penerapan metode teknik SMOTE untuk menyeimbangkan data tersebut. Sehingga, data seimbang dapat menghasilkan model yang lebih baik. Adapun hasil dari penerapan teknik SMOTE tersebut ditunjukkan pada Gambar 6.



Gambar 6. Data sesudah di SMOTE

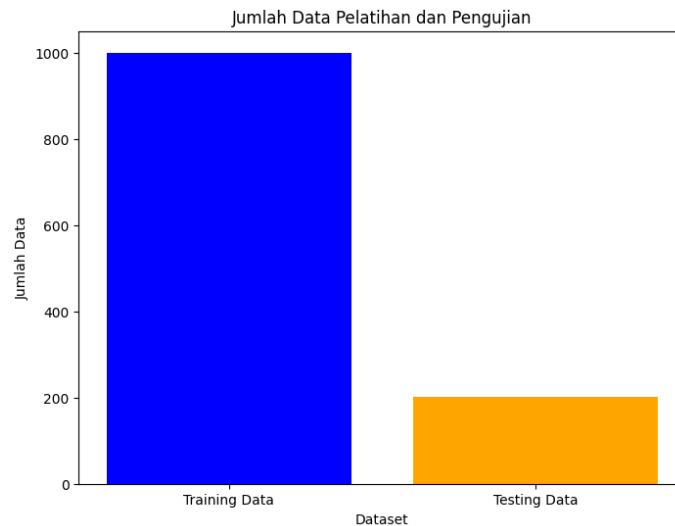
Berdasarkan Gambar 6 menunjukkan hasil dari penerapan teknik SMOTE. Demikian hal itu, nilai dari atribut Nobesyedad menjadi seimbang. Adapun tahap selanjutnya yaitu tahap pembagian data. Adapun pembagian data ini dilakukan dengan perbandingan 80:20 dengan data sebelum SMOTE dan sesudah SMOTE. Berikut adalah visualisasi hasil pembagian dataset sebelum menggunakan smote yang ditampilkan pada Gambar 7.



Jumlah data pelatihan: 472  
Jumlah data pengujian: 119

Gambar 7. Data latih dan uji Sebelum SMOTE

Berdasarkan Gambar 7 menunjukkan visualisasi hasil pembagian dataset yang sebelum menggunakan SMOTE, di mana terdapat 472 data untuk pelatihan dan 119 data untuk pengujian. Data pelatihan akan digunakan untuk melatih model, sedangkan data pengujian akan digunakan untuk mengevaluasi kinerja model tersebut. Sedangkan visualisasi hasil pembagian dataset setelah menggunakan SMOTE yang ditampilkan pada Gambar 8.



Jumlah data pelatihan: 800  
 Jumlah data pengujian: 201

Gambar 8. Data Uji dan Latih setelah SMOTE

Berdasarkan Gambar 8. menunjukkan visualisasi hasil dari pembagian dataset setelah SMOTE. Terdapat 800 data untuk pelatihan dan 201 data untuk pengujian. Selanjutnya, dataset ini sudah siap untuk digunakan dalam tahap berikutnya.

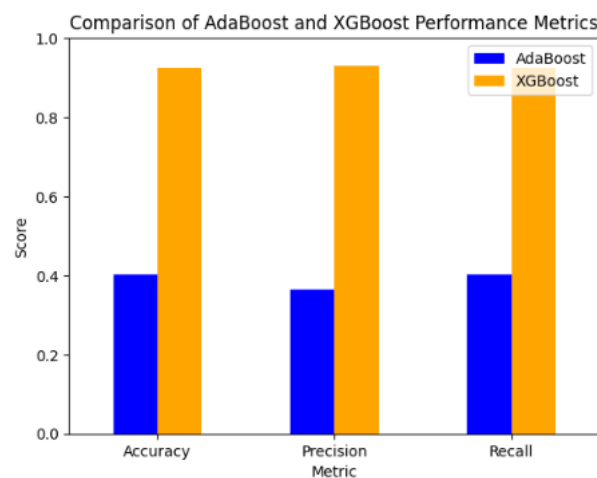
#### D. Evaluasi Model dan Perbandingan

Sebelum dilakukan evaluasi model, *AdaBoost* dan *XGBoost* tersebut diterapkan pada dataset sebelum dan SMOTE. Adapun model model tersebut dievaluasi dengan menggunakan *confusion matrix*. Berikut hasil evaluasi model *AdaBoost* dan *XGBoost* dengan data sebelum SMOTE ditunjukkan pada Tabel 1.

TABEL 1  
 EVALUASI MODEL TANPA SMOTE

Metode	Accuracy	Precision	Recall
<i>AdaBoost</i>	0.403	0.363	0.403
<i>XGBoost</i>	0.924	0.928	0.924

Berdasarkan Tabel 1 menunjukkan hasil dari evaluasi model yang telah dilakukan. Adapun nilai tertinggi terdapat pada model *XGBoost* yaitu akurasi sebesar 0,924, Presisi 0,928 dan recall 0,924. Terdapat perbedaan yang signifikan dalam hasil evaluasi yang dilakukan. Hal tersebut dapat dilihat dari visualisasi hasil evaluasi tersebut pada Gambar 9.



Gambar 9. Hasil Evaluasi

Berdasarkan Gambar 9 menunjukkan bahwa hasil evaluasi model *XGBoost* cenderung lebih tinggi dibandingkan dengan *AdaBoost*. Terdapat perselisihan akurasi dan recall sebesar 0,521 serta presisi sebesar 0,565. Sementara hasil evaluasi dari penerapan data SMOTE ditunjukkan pada Tabel 2.

TABEL 2  
 EVALUASI MODEL DENGAN SMOTE

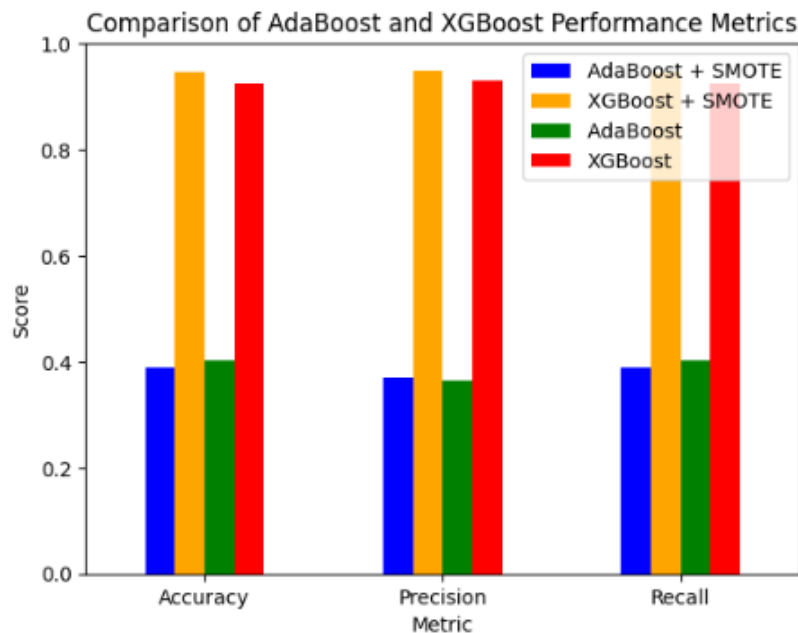
Metode	Accuracy	Precision	Recall
<i>AdaBoost</i>	0.388	0.371	0.388
<i>XGBoost</i>	0.945	0.947	0.945

Berdasarkan Tabel 2 menunjukkan hasil dari evaluasi model yang telah dilakukan. Adapun nilai tertinggi terdapat pada model *XGBoost* yaitu akurasi sebesar 0,945 Presisi 0,947 dan recall 0,945. Terdapat perbedaan yang signifikan dalam hasil evaluasi yang dilakukan. Terdapat perselisihan akurasi dan recall sebesar 0,557. Sedangkan presisi sebesar 0,576. Berikut perbandingan hasil evaluasi dengan data SMOTE atau sebelum SMOTE ditunjukkan pada Tabel 3.

TABEL 3  
 PERBANDINGAN HASIL EVALUASI

Metode	Accuracy	Precision	Recall
<i>AdaBoost</i>	0.403	<b>0.363</b>	0.403
<i>XGBoost</i>	0.924	0.928	0.924
<i>AdaBoost</i> + SMOTE	<b>0.388</b>	0.371	<b>0.388</b>
<i>XGBoost</i> + SMOTE	<b>0.945</b>	<b>0.947</b>	<b>0.945</b>

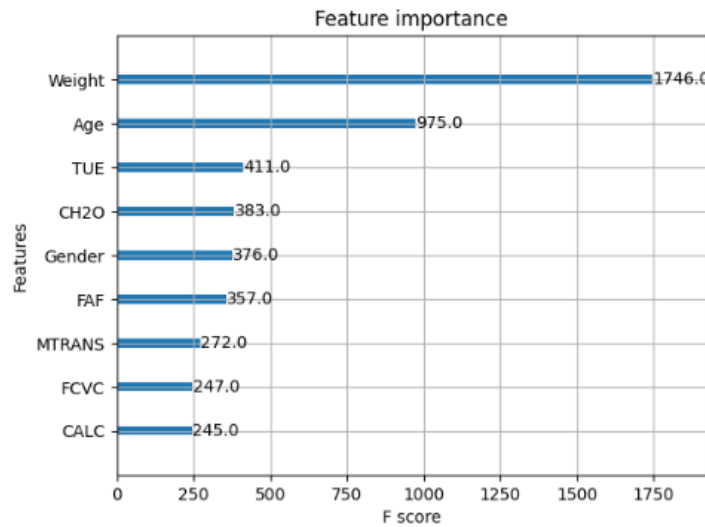
Berdasarkan Tabel 3 menunjukkan hasil dari perbandingan evaluasi model dengan data SMOTE dan tanpa SMOTE. Dari hal tersebut *XGBoost* dengan data SMOTE cenderung lebih tinggi dibandingkan dengan hasil evaluasi model lainnya. Adapun nilai akurasi dan *recall* sebesar 0,945 serta presisi 0,947. Sedangkan nilai akurasi dan recall terendah terdapat pada *AdaBoost* dengan SMOTE sebesar 0,388. Selain itu, nilai presisi terendah terdapat pada *AdaBoost* tanpa SMOTE yaitu sebesar 0,403. Hal tersebut dapat dilihat pada Gambar 10.



Gambar 10. Visualisasi hasil perbandingan model

Berdasarkan visualisasi pada Gambar 10. Hasil evaluasi model dengan algoritma *XGBoost* menunjukkan performa yang lebih baik dibandingkan dengan *AdaBoost*. Adapun hal tersebut disebabkan beberapa factor yaitu Pertama, *XGBoost* memiliki kemampuan yang lebih baik dalam menangani kompleksitas data, termasuk berbagai jenis fitur numerik dan kategorikal seperti pola makan, aktivitas fisik, serta kebiasaan konsumsi. Kedua,

XGBoost menerapkan teknik regularisasi selama proses pelatihan, sehingga mampu mencegah overfitting meskipun data memiliki banyak fitur. Ketiga, kemampuannya dalam mengelola klasifikasi multikelas sangat cocok dengan atribut target NObeyesdad yang terdiri dari enam kategori obesitas. Demikian hal itu, hasil dari penerapan XGBoost dan SMOTE ditunjukkan pada Gambar 11.



Gambar 11. Visualisasi penerapan XGBoost

Gambar tersebut menunjukkan tingkat *importance* (kepentingan) dari masing-masing fitur dalam sebuah model *machine learning*, yang tampaknya menggunakan teknik berbasis pohon, seperti XGBoost atau Random Forest. Importance disini menunjukkan seberapa besar pengaruh setiap fitur terhadap hasil prediksi model. Dari grafik tersebut, fitur *Weight* (berat badan) memiliki pengaruh paling besar dengan nilai F score sebesar 1684.0, menunjukkan bahwa model sangat mengandalkan fitur ini untuk membuat prediksi. Di posisi kedua, fitur *Age* (usia) juga cukup signifikan dengan F score 877.0, menunjukkan bahwa usia juga memainkan peran penting dalam model. Fitur lain seperti *Gender* (jenis kelamin) dengan F score 329.0, serta *CALC* dan *FAF* dengan nilai masing-masing 276.0 dan 263.0, berkontribusi lebih sedikit, tetapi tetap memberikan informasi yang berguna bagi model. Di sisi lain, fitur *MTRANS*, yang memiliki F score 131.0, adalah fitur dengan pengaruh paling kecil terhadap model, menandakan bahwa kontribusinya terhadap prediksi sangat terbatas dibandingkan fitur lainnya.

Secara keseluruhan, narasi dari grafik tersebut adalah bahwa model lebih dipengaruhi oleh fitur-fitur seperti *Weight* dan *Age*, sedangkan fitur lainnya berkontribusi dalam level yang lebih rendah. Informasi ini penting untuk memahami apa yang paling menentukan prediksi model dan bisa menjadi dasar untuk pengambilan keputusan lebih lanjut dalam analisis data. Dengan demikian, hasil penelitian dengan menggunakan algoritma XGBoost dengan Teknik SMOTE memberikan performa klasifikasi obesitas yang lebih akurat dibandingkan dengan AdaBoost. Oleh karena itu, hasil dari penelitian ini diharapkan dapat diterapkan dalam dunia medis. Sehingga, prediksi terhadap individu yang berisiko obesitas dapat dilakukan lebih awal, sehingga intervensi seperti penyuluhan gizi, pengaturan pola makan, atau promosi aktivitas fisik dapat diberikan secara lebih tepat sasaran.

#### IV. SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan menunjukkan bahwa penggunaan metode XGBoost dengan SMOTE cenderung lebih unggul dibandingkan dengan metode lainnya. Adapun nilai akurasi dan *recall* dari metode tersebut sebesar 94,5%. Sedangkan nilai *precision* sebesar 94,7%. Hal tersebut dikarenakan metode XGBoost tersebut cenderung cocok dengan dataset yang digunakan pada penelitian ini. Selain itu, penggunaan SMOTE menunjukkan dapat memperbaiki hasil evaluasi model. Namun, pada metode AdaBoost penggunaan SMOTE nilai akurasi dan *recall* cenderung menurun sebesar 1,5%. Sedangkan nilai *precision* cenderung naik sebesar 0,8% dengan penggunaan SMOTE tersebut. Sehingga, hal tersebut menunjukkan bahwa efektivitas SMOTE tidak seragam untuk semua algoritma, tergantung pada karakteristik model dan data yang digunakan. Dengan demikian, model XGBoost yang dikombinasikan dengan SMOTE direkomendasikan untuk digunakan dalam riset selanjutnya, khususnya pada kasus klasifikasi dengan data yang tidak seimbang. Model ini diharapkan dapat menjadi acuan dalam pengembangan sistem deteksi dan pengelolaan obesitas secara lebih efektif.

**DAFTAR PUSTAKA**

- [1] D. E. Safitri and N. S. Rahayu, "Determinan Status Gizi Obesitas pada Orang Dewasa di Perkotaan: Tinjauan Sistematis," *ARKESMAS (Arsip Kesehatan Masyarakat)*, vol. 5, no. 1, pp. 1–15, 2020, doi: 10.22236/arkesmas.v5i1.4853.
- [2] S. A. Utirahman and A. M. M. Pratama, "Analisis Perbandingan KNN, SVM, Decision Tree dan Regresi Logistik Untuk Klasifikasi Obesitas Multi Kelas," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 6, pp. 3137–3146, 2024, doi: 10.30865/klik.v4i6.1871.
- [3] N. S. Medellu, A. Makmun, N. Laddo, I. D. K. Kamaluddin, Ida Royani, and Zulfahmidah, "Hubungan Antara Obesitas Dengan Kadar Interleukin 6 (IL-6) Pada Populasi Anak Laki-Laki Di Kota Makassar.," *Fakumi Med. J. J. Mhs. Kedokt.*, vol. 3, no. 7, pp. 452–460, 2023, doi: 10.33096/fmj.v3i7.249.
- [4] Z. Helforouh and H. Sayyad, "Prediction and classification of obesity risk based on a hybrid metaheuristic machine learning approach," *Front. Big Data*, vol. 7, 2024, doi: 10.3389/fdata.2024.1469981.
- [5] S. A. Rifqi, R. Masrikhiyah, and Y. D. Rahmawati, "Hubungan aktivitas fisik , asupan makan , konsumsi fast food dengan status gizi pada siswa MTS Assalafiyah," vol. 4, no. 2, pp. 320–327, 2024.
- [6] R. Z. Hamza and K. Alsolami, "Ameliorative effects of Orlistat and metformin either alone or in combination on liver functions, structure, immunoreactivity and antioxidant enzymes in experimentally induced obesity in male rats," *Heliyon*, vol. 9, no. 8, Aug. 2023, doi: 10.1016/j.heliyon.2023.e18724.
- [7] G. K. Kamalam, N. S. Baby, P. Kishore, S. Elango, and C. Hariharan, "Predicting Obesity Risk Using Machine Learning and Deep Learning Techniques," no. M1.
- [8] F. Erwis, D. Suherdi, A. Pranata, and A. H. Nasyuha, "Penerapan Metode Hybrid Case Base Pada Sistem Pakar Diagnosa Penyakit Obesitas," *J. Media Inform. Budidarma*, vol. 6, no. 1, p. 378, 2022, doi: 10.30865/mib.v6i1.3491.
- [9] S. U. Amrnia and G. N. Prameswari, "Hubungan Pola Makan, Sedentary Lifestyle, dan Durasi Tidur dengan Kejadian Gizi Lebih Pada Remaja (Studi Kasus di SMA Negeri 1 Demak)," *Indones. J. Public Heal. Nutr.*, vol. 2, no. 1, pp. 112–121, Mar. 2022, doi: 10.15294/ijphn.v2i1.52044.
- [10] E. D. A. Syifa and R. Djuwita, "Factors Associated with Overweight/Obesity in Adolescent High School Students in Pekanbaru City," *J. Kesehatan Komunitas*, vol. 9, no. 2, pp. 368–378, Aug. 2023, doi: 10.25311/keskom.vol9.iss2.1579.
- [11] M. Iqbal, W. S. Dharmawan, and R. Septian, "Journal of Computer Networks , Architecture and High Performance Computing Prediction of Obesity Categories Based on Physical Activity Using Machine Learning Algorithms Journal of Computer Networks , Architecture and High Performance Computing," vol. 6, no. 3, pp. 1025–1034, 2024.
- [12] A. Nuzulia, "Memahami Data Mining dengan Python: Implementasi Praktis, Eureka Media Aksara," *Memahami Data Min. dengan Python Implementasi Prakt. Eureka Media Aksara*, pp. 5–24, 1967.
- [13] T. L. Nikmah, R. M. Syafei, R. Muzayanah, and A. Salsabila, "Prediction Of Used Car Prices Using K-Nearest Neighbour, Random Forest And Adaptive Boosting Algorithm," *Indones. Community Optim. Comput. Appl.*, vol. 1, no. 1, pp. 17–22, 2023, [Online]. Available: <https://e-journal.ptti.info/index.php/icoca/article/view/68>
- [14] R. J. Alfrdausy and S. Bahri, "Implementasi Algoritma K-Nearest Neighbourfile:///D:/E-Learning 02/Documents/DATA NUNUNG/Bu cici/Jurnal/1.pdf file:///D:/E-Learning 02/Documents/DATA NUNUNG/Bu cici/Jurnal/2.pdf file:///D:/E-Learning 02/Documents/DATA NUNUNG/Bu cici/Jurnal/3.pdf file:///D:/E," *Techno.Com*, vol. 22, no. 3, pp. 635–642, 2023, doi: 10.33633/tc.v22i3.8393.
- [15] R. T. Febianto, D. Suranti, and R. T. Alinse, "Penerapan Algoritma Adaboost Dalam Mengetahui Pola Pengguna Kb Di Puskesmas Tanjung Harapan," *J. Sci. Soc. Res.*, vol. 4307, no. 1, pp. 145–155, 2024, [Online]. Available: <http://jurnal.goretanpena.com/index.php/JSSR>
- [16] M. Siddik Hasibuan and H. Harahap, "Penerapan Metode Haar-Like Feature Dan Algoritma Adaboost Dalam Penentuan Klasifikasi Hama Tanaman Kopi," *J. Sci. Soc. Res.*, vol. 4307, no. 1, pp. 87–93, 2024, [Online]. Available: <http://jurnal.goretanpena.com/index.php/JSSR>
- [17] J. Sanjaya, E. Renata, V. E. Budiman, F. Anderson, and M. Ayub, "Prediksi Kelalaian Pinjaman Bank Menggunakan Random Forest dan Adaptive Boosting," *J. Tek. Inform. dan Sist. Inf.*, vol. 6, no. 1, pp. 50–60, 2020, doi: 10.28932/jutisi.v6i1.2313.
- [18] E. C. Ardianta, P. A. Wibowo, N. Maulana, A. Ristyawan, and E. Daniati, "Optimalisasi Prediksi Tingkat Obesitas di Negara Mexico Menggunakan Perbandingan Support Vector Machine dan Naïve Bayes," vol. 8, pp. 1551–1559, 2024.
- [19] H. T. Santoso, F. A. Felmidi, A. Nur, A. Ristyawan, and E. Daniati, "Analisis Kinerja Algoritma Data Mining pada Klasifikasi Tingkat Obesitas dengan K-Fold Cross Validation dan AUC," vol. 8, pp. 113–122, 2024.
- [20] M. Dewi, T. H. Saragih, and R. Herteno, "Penerapan SMOTE-NCL untuk Mengatasi Ketidakseimbangan Kelas pada Klasifikasi Penyakit Jantung Koroner," *J. Inform. Polinema*, vol. 10, no. 1, pp. 27–34, 2023, doi: 10.33795/jip.v10i1.1394.
- [21] D. Anggitasyah and M. A. P. Siregar, "Penerapan Metode Smote Extreme Gradient Boosting Untuk Klasifikasi Penyakit Kanker Serviks Di Kota Medan," *Jutisi J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 12, no. 2, p. 526, 2023, doi: 10.35889/jutisi.v12i2.1479.
- [22] W. Musu, A. Ibrahim, and Heriadi, "Pengaruh Komposisi Data Training dan Testing terhadap Akurasi Algoritma C4.5 Musu, W., Ibrahim, A., & Heriadi. (2021). Pengaruh Komposisi Data Training dan Testing terhadap Akurasi Algoritma C4.5. Prosiding Seminar Ilmiah Sistem Informasi Dan Teknologi Inf," *Pros. Semin. Ilm. Sist. Inf. Dan Teknol. Inf.*, vol. X, no. 1, pp. 186–195, 2021.
- [23] R. Harahap, M. Irpan, M. A. Dinata, L. Efrizoni, and Rahmaddeni, "Perbandingan Algoritma Random Forest Dan Xgboost Untuk Klasifikasi Penyakit Paru-Paru Berdasarkan Data Demografi Pasien," *J. Ilm. Betrik*, vol. 15, no. 02, pp. 130–141, 2024.
- [24] R. A. Nurdian, Mujib Ridwan, and Ahmad Yusuf, "Komparasi Metode SMOTE dan ADASYN dalam Meningkatkan Performa Klasifikasi Herregistrasi Mahasiswa Baru," *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 1, pp. 24–32, 2022, doi: 10.28932/jutisi.v8i1.4004.
- [25] W. I. Sabilla and C. Bella Vista, "Implementasi SMOTE dan Under Sampling pada Imbalanced Dataset untuk Prediksi Kebangkrutan Perusahaan," *J. Komput. Terap.*, vol. 7, no. 2, pp. 329–339, 2021, doi: 10.35143/jkt.v7i2.5027.
- [26] 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, 2023, doi: 10.3389/fendo.2023.1292167.
- [27] M. Anjas Aprihartha, J. Prasetya, D. Sefri, and I. Fallo, "Implementasi CART-Real Adaboost dalam Memprediksi Minat Pelanggan Membeli Sepatu A B S T R A K INFORMASI ARTIKEL A B S T R A C T," vol. 12, no. 1, pp. 35–46, 2024.
- [28] K. Aditya, A. Wisnu, and A. M. A. Rahim, "Analisis Perbandingan Algoritma XGBoost Dan Algoritma Random Forest Untuk Klasifikasi Data Kesehatan Mental," vol. 2, no. 5, pp. 808–818, 2024.
- [29] F. R. Valerian *et al.*, "Klasifikasi tingkat obesitas menggunakan metode gbm dan confusion matrix," vol. 9, no. 2, pp. 2242–2249, 2025.