



## Efektivitas algoritma *AdaBoost* dan *XGBoost* pada dataset obesitas populasi dewasa

Cici Emilia Sukmawati, Anis Fitri Nur Masruriyah, Ayu Ratna Juwita,  
Rigger Damaiarta Tejayanda, Trisya Nurmayanti

Universitas Buana Perjuangan Karawang, Indonesia

### Riwayat Artikel:

Diterima 17 September 2024

Direvisi 22 Oktober

Disetujui 29 Oktober

### Kata Kunci:

*AdaBoost*

Obesitas

*XGBoost*

**ABSTRACT.** This research discusses the performance evaluation of *AdaBoost* and *XGBoost* on obesity disease. The *AdaBoost* and *XGBoost* algorithms were chosen because they improve prediction accuracy on complex data. *AdaBoost* iteratively combines simple models to produce more robust models. At the same time, *XGBoost* is known for its speed, efficiency, and ability to handle large datasets with complex features. Combining these two algorithms allows for a more accurate analysis of obesity disease data with varying risks and factors. Testing is done using accuracy, precision, and recall. Based on the results of the research, the *XGBoost* method is proven to be superior to *AdaBoost*. The accuracy, precision, and recall are 92%. Meanwhile, the accuracy and recall for the *AdaBoost* method is 40%, and the precision is 39%. The *AdaBoost* method uses a shallow decision tree and has difficulty handling complex datasets and overfitting. Meanwhile, the *XGBoost* method matches the dataset used in this study.

**ABSTRAK.** Penelitian ini membahas evaluasi kinerja *AdaBoost* dan *XGBoost* pada penyakit obesitas. Algoritma *AdaBoost* dan *XGBoost* dipilih karena keduanya unggul dalam meningkatkan akurasi prediksi pada data kompleks. *AdaBoost* menggabungkan model sederhana secara iteratif untuk menghasilkan model yang lebih kuat, sedangkan *XGBoost* dikenal dengan kecepatan dan efisiensinya, serta kemampuannya menangani dataset besar dengan fitur-fitur rumit. Kombinasi kedua algoritma ini memungkinkan analisis yang lebih akurat pada data penyakit obesitas dengan risiko dan faktor yang bervariasi. Pengujian dilakukan menggunakan *accuracy*, *precision* dan *recall*. Berdasarkan hasil penelitian yang telah dilakukan, bahwa metode *XGBoost* terbukti lebih unggul dibandingkan dengan *AdaBoost*. Adapun *accuracy*, *precision* dan *recall* sebesar 92%. Sedangkan untuk *accuracy* dan *recall* untuk metode *AdaBoost* sebesar 40% serta *precision* 39%. Hal ini disebabkan metode *AdaBoost* cenderung menggunakan pohon keputusan yang dangkal dan cenderung mengalami kesulitan dengan dataset yang sangat kompleks serta menangani overfitting. Sedangkan, metode *XGBoost* tersebut cenderung cocok dengan dataset yang digunakan pada penelitian ini.

This is an open-access article under the [CC-BY-SA](#) license.



### Penulis Korespondensi:

Cici Emilia Sukmawati,

Universitas Buana Perjuangan Karawang

Jl. H.S. Ronggowaluyo, Teluk Jambe, Karawang, 41361, Indonesia

Email: [cici.emilia@ptk.ubpkarawang.ac.id](mailto:cici.emilia@ptk.ubpkarawang.ac.id)

## PENDAHULUAN

Obesitas merupakan masalah kesehatan global yang terus meningkat prevelensinya (Ella dkk., 2022; Safitri & Rahayu, 2020) tidak hanya di negara-negara maju tetapi juga di negara-negara berkembang. Menurut *World Health Organization* (WHO), obesitas adalah kondisi medis yang ditandai dengan

penumpukan lemak, yang berpotensi menyebabkan berbagai komplikasi Kesehatan, termasuk penyakit jantung, diabetes tipe 2 (Hamza & Alsolami, 2023) dan kanker. Kondisi tersebut umumnya diakibatkan oleh asupan kalori yang melebihi kapasitas yang dibutuhkan oleh tubuh. Sehingga, berubah menjadi lemak jenuh (Erwis dkk., 2022).

Menurut studi epidemiologi, terdapat beberapa penyebab obesitas, di antaranya adalah pola makan yang tidak sehat, minimnya aktivitas fisik, serta faktor genetik dapat berkontribusi pada risiko (Amrynia & Prameswari, 2022; Hanani dkk., 2021; Syifa & Djuwita, 2023). Identifikasi yang tepat terhadap faktor-faktor risiko ini sangat penting untuk mengembangkan intervensi yang efektif dalam mencegah dan mengelola obesitas. Salah satu pendekatan yang dapat digunakan untuk mengidentifikasi faktor risiko tersebut adalah teknik data mining.

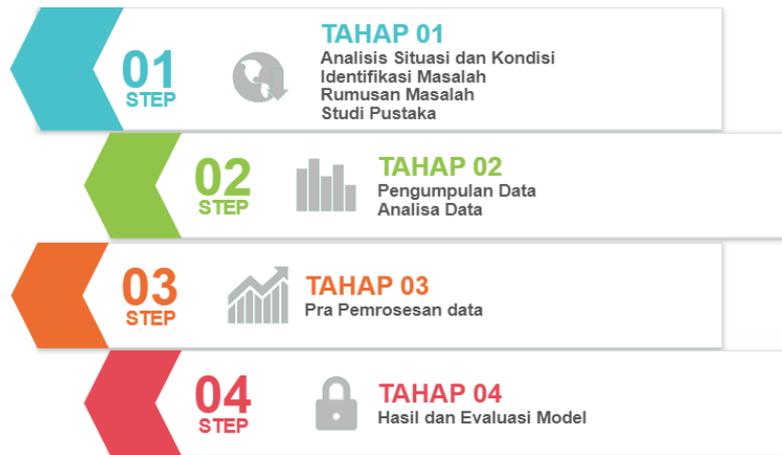
Data mining adalah proses ekstraksi pengetahuan yang tersembunyi dalam data dengan menggunakan berbagai teknik komputasi (Gede Iwan Sudipa dkk., 2023; Masruriyah dkk., 2024). Dalam konteks obesitas, teknik data mining bisa diterapkan untuk menganalisis pola makan, aktivitas fisik, serta faktor-faktor lain yang terkait dengan obesitas pada populasi dewasa. Hal ini dibuktikan oleh penelitian yang dilakukan oleh (Dwi dkk., 2024) untuk mengetahui tingkat obesitas berdasarkan faktor-faktor faktor resiko. pada penelitian tersebut menggunakan algoritma *random forest* dan menghasilkan akurasi sebesar 93.38%. Selain itu, Iqbal dkk. menggunakan berbagai algoritma seperti XGBoost, AdaBoost, Gradient Boosting Classifier, dan Extra Trees Classifier untuk prediksi obesitas, dengan akurasi tertinggi ditemukan pada model XGBoost, yaitu 99.61%. Namun, meskipun perbandingan antara algoritma seperti AdaBoost dan XGBoost sering dilakukan, studi ini berusaha memperkenalkan pendekatan baru yang lebih signifikan daripada sekadar membandingkan algoritma yang ada.

Dalam penelitian ini, dilakukan eksplorasi pengembangan dan modifikasi dari algoritma AdaBoost dan XGBoost untuk meningkatkan akurasi prediksi obesitas. Pendekatan ini diharapkan dapat memberikan nilai kebaruan yang lebih dalam penelitian terkait obesitas. Adapun perbandingan ini penting dikarenakan obesitas merupakan salah satu penyebab utama berbagai penyakit kronis yang membebani sistem kesehatan di seluruh dunia. Dengan meningkatkan akurasi prediksi, profesional kesehatan dapat membuat keputusan yang lebih tepat dalam pencegahan dan penanganan obesitas. Selain itu, teknik ini juga dapat diadaptasi untuk diterapkan dalam skenario kesehatan lainnya, seperti pada studi penyakit ginjal kronis oleh farin, i mana XGBoost terbukti memberikan performa unggul dibandingkan algoritma lain dengan akurasi mencapai 96% (De Guia dkk., 2019). Sedangkan, algoritma penelitian oleh (Nikmah dkk., 2023) menunjukkan bahwa meskipun algoritma *random forest* lebih unggul dalam prediksi harga mobil bekas, studi ini memberikan gambaran mengenai penggunaan algoritma AdaBoost dan XGBoost dalam konteks yang berbeda.

Penelitian ini bertujuan untuk mengevaluasi efektivitas algoritma AdaBoost dan XGBoost dalam menganalisis pola makan, aktivitas fisik, serta faktor-faktor risiko lain yang berkontribusi terhadap obesitas pada populasi dewasa. Diharapkan hasil penelitian ini dapat memberikan wawasan yang signifikan bagi para peneliti, profesional kesehatan, dan pembuat kebijakan dalam upaya pencegahan dan pengelolaan obesitas dengan memanfaatkan pendekatan berbasis data yang lebih akurat dan efisien.

## METODE

Adapun penelitian ini menggunakan beberapa tahapan, yaitu Tahap pertama terdiri dari analisis situasi dan kondisi. kemudian, identifikasi masalah. Selanjutnya, rumusan masalah dan studi pustaka. Sedangkan, tahap kedua terdiri dari pengumpulan data dan analisa data. Selanjutnya, untuk tahap ketiga terdiri dari prapemrosesan data. Sedangkan tahap terakhir yaitu hasil dan evaluasi model. hal tersebut ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Berdasarkan Gambar 1 menunjukkan tahapan yang digunakan pada penelitian ini. Adapun tahapan pertama yang dilakukan yaitu melakukan analisis mendalam terhadap literatur yang relevan untuk memahami konteks masalah dan mengidentifikasi kerangka kerja serta metodologi yang sesuai untuk mengatasi permasalahan yang dihadapi. Adapun hal tersebut dapat membantu dalam merumuskan pertanyaan penelitian yang tepat serta memastikan bahwa pendekatan yang digunakan akan memberikan solusi yang efektif. Selanjutnya tahapan kedua yaitu tahap pengumpulan data dan analisa data yang melibatkan proses pengumpulan data dari *kaggle* dan analisis data untuk mengidentifikasi pola dan hubungan antara atribut yang ada dalam dataset. Analisis data ini merupakan langkah penting dalam memahami karakteristik data serta mempersiapkan data untuk tahap selanjutnya. Langkah berikutnya adalah tahapan ketiga yaitu. pra pemrosesan data. Pada tahap ini dilakukan proses penanganan data, seperti mengatasi missing value, duplikat data, noise dan outlier. Kemudian selanjutnya dilakukan proses transformasi data. Adapun proses tersebut dilakukan untuk meningkatkan kualitas data serta memudahkan analisis. Selanjutnya adalah proses normalisasi data. Proses tersebut dilakukan untuk mengubah nilai data ke dalam rentang 0 hingga 1. Hal tersebut dilakukan agar memperbaiki interpretasi data serta meningkatkan kinerja model. Kemudian, tahap terakhir yaitu tahap implementasi dan evaluasi model.

Metode yang diimplementasikan dalam penelitian ini adalah algoritma AdaBoost dan XGBoost. AdaBoost adalah teknik dalam data mining yang dirancang untuk meningkatkan tingkat akurasi pada metode klasifikasi. (Febianto dkk., 2024). Selain itu, *AdaBoost* dikatakan sebagai ensemble learning yang sering digunakan pada algoritma boosting (Siddik Hasibuan & Harahap, 2024). Adapun langkah pertama algoritma tersebut yaitu menggunakan algoritma dasar. Kemudian, menyesuaikan bobot sampel dengan hasil klasifikasi dasar. Setelah itu, sampel tersebut digunakan untuk melatih base learner berikutnya. setelah iterasi, pembobotan ditambahkan ke base learner untuk membentuk klasifikasi akhir (Sanjaya dkk., 2020). Sedangkan, Algoritma *XGBoost* metode yang dapat menemukan solusi optimal untuk mengatasi permasalahan klasifikasi dan regresi (Alnur dkk., 2023). Selain itu, Metode tersebut ditemukan oleh Friedman yang merupakan pengembangan dari algoritma GBDT (Gradient Boosting Decision Tree)(Givari dkk., 2022). Langkah awal dalam algoritma ini biasanya dimulai dengan penggunaan algoritma klasifikasi sederhana. Setelah itu, bobot sampel disesuaikan berdasarkan hasil klasifikasi awal, sehingga sampel yang salah klasifikasi mendapat perhatian lebih. Kemudian, sampel yang telah disesuaikan tersebut digunakan untuk melatih model dasar berikutnya. Setelah melalui beberapa iterasi, pembobotan tambahan diterapkan pada model dasar untuk menghasilkan klasifikasi akhir (Sanjaya dkk., 2020). Berdasarkan hal tersebut, algoritma AdaBoost dan XGBoost dipilih dalam penelitian ini untuk mengevaluasi efektivitas kedua metode pada dataset obesitas yang digunakan.

Kemudian, setelah metode tersebut diterapkan, maka proses selanjutnya yaitu evaluasi model. Model-model tersebut dievaluasi menggunakan confusion matrix yaitu akurasi, presisi dan recall. Adapun akurasi tersebut merupakan tolak ukur yang digunakan untuk mengetahui seberapa tepat pola memprediksi data pada masa selanjutnya (Musu dkk., 2021). Sementara itu, Recall mengacu pada proporsi dokumen yang berhasil ditemukan kembali oleh sistem dalam proses pencarian informasi. (Prabowo, 2021). Kemudian, *precision* adalah Rasio dari jumlah prediksi yang benar-benar positif terhadap semua kelas positif yang diprediksi dengan benar (Wie & Siddik, 2022). Selanjutnya, dari hasil evaluasi tersebut dilakukan perbandingan kinerja dari model *AdaBoost* dan *XGBoost*. Dengan demikian, hasil perbandingan tersebut dapat menunjukkan sejauh mana efektivitas algoritma *AdaBoost* dan *XGBoost* dalam diterapkan pada dataset obesitas populasi dewasa.

## HASIL DAN DISKUSI

### Hasil Pengumpulan Data

Dataset yang dipakai dalam penelitian ini adalah hasil pengumpulan data dari Kaggle. Data disajikan dalam format .csv dengan jumlah total 2111 data dan terdiri dari 16 atribut serta 1 atribut target. Atribut ini terdiri Gender, Age, Height, Weight, Family\_history\_with\_overweight, FAVC, FCVC, NCP, CAEC, SMOKE, CH20, SCC, FAF, TUE, CALC, MTRANS serta Nobeyesdad. Adapun cuplikan dari dataset tersebut ditunjukkan pada Tabel 1.

Tabel 1. Cuplikan dataset

Gender	Age	Height	Weight	..	TUE	CALC	MTRANS	NObesyesdad
0	21	2	64	..	1	3	4	Normal_Weight
0	21	2	56	..	0	4	4	Normal_Weight
1	23	2	77	..	1	2	4	Normal_Weight
1	27	2	87	..	0	2	5	Overweight_Level_I
...	...	...	...	..	...	...	...	...
0	21	2	132	..	1	4	4	Obesity_Type_III
0	22	2	134	..	1	4	4	Obesity_Type_III
0	23	2	134	..	1	4	4	Obesity_Type_III
0	25	2	134	..	1	4	4	Obesity_Type_III

Berdasarkan Tabel 1 menunjukkan dari cuplikan data yang digunakan pada penelitian. Adapun cuplikan data tersebut memiliki 1 atribut target yaitu Nobeyesdad. Atribut tersebut terdiri dari Atribut tersebut terdiri dari Normal\_Weight, Overweight\_Level\_I, Overweight\_Level\_II, Obesity\_Type\_I, Insufficient\_Weight, Obesity\_Type\_II, Obesity\_Type\_III. Selanjutnya, penjelasan dari setiap atribut ditunjukkan pada Tabel 2.

### Hasil Pra-Pemrosesan Data

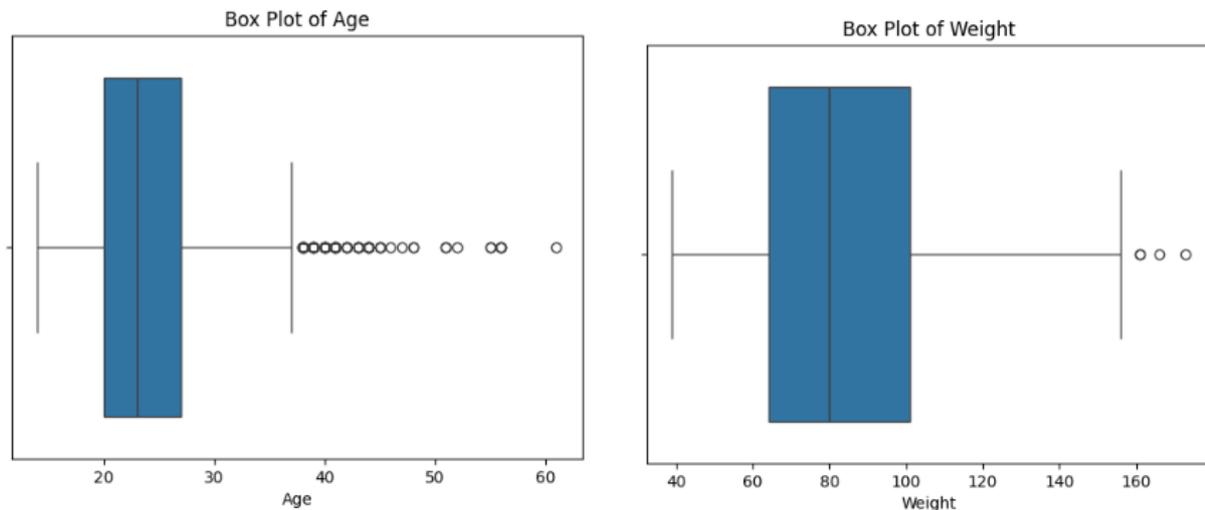
Pada tahap ini dilakukan pra pemrosesan data yang dimulai dengan memeriksa missing value dan duplikat data. Adapun fungsi yang digunakan ialah `isnull()`. Fungsi tersebut bekerja dengan cara memeriksa setiap elemen dalam dataset dan mengevaluasi apakah elemen tersebut merupakan nilai yang hilang atau tidak. Nilai yang hilang biasanya direpresentasikan sebagai NaN (Not a Number) atau None. Kemudian fungsi tersebut akan mengembalikan nilai boolean True jika elemen adalah NaN/None dan False jika tidak. Berdasarkan hasil pemeriksaan menggunakan fungsi tersebut, dataset yang digunakan tidak memiliki missing value tau nilai yang hilang.

Selanjutnya yaitu pemeriksaan data duplikat dengan menggunakan fungsi `is duplicated()`. Adapun fungsi tersebut digunakan untuk mendeteksi nilai yang atau elemen yang sama. Oleh karena itu, data yang sama atau data dupliat dapat mempengaruhi analisis atau hasil model. Berdasarkan pengecekan tersebut menunjukkan adanya data yang sama sebanyak 558 data. Sehingga data yang sama tersebut

dilakukan penghapusan data. Kemudian, data berkurang menjadi 1553 dari 2111 data. Berikut visualisasi cuplikan dataset yang digunakan ditunjukkan pada Gambar 2.

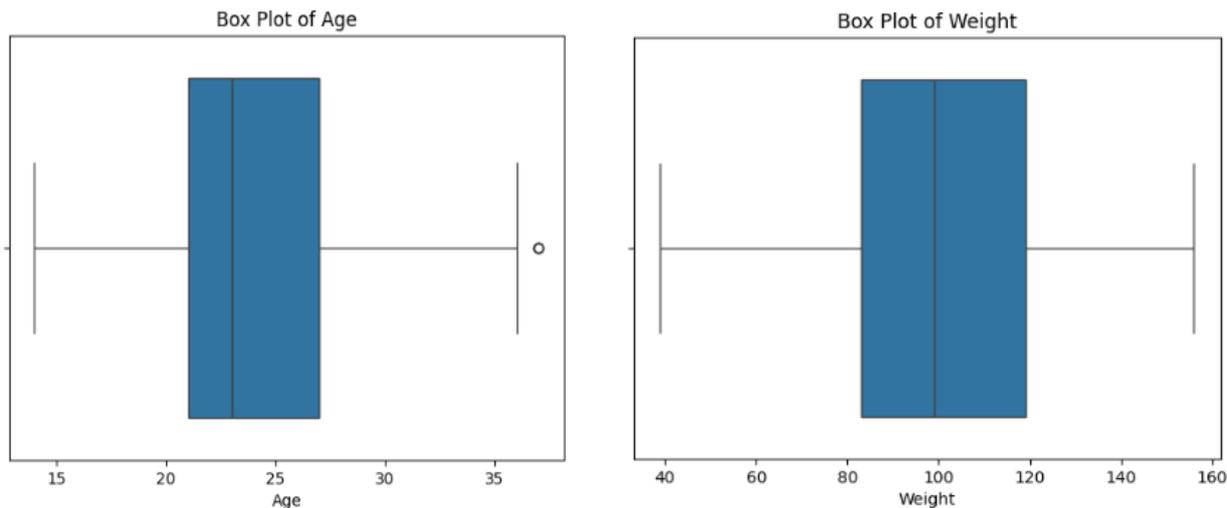
Tabel 2. Deskripsi atribut

Atribut	Deskripsi Atribut
Gender	Jenis Kelamin
Age	Umur
Height	Tinggi Badan
Weight	Berat Badan
family_history_with_overweight	Riwayat keluarga obesitas
FAVC	Frekuensi konsumsi makanan berkalori tinggi
FCVC	Frekuensi konsumsi sayur
NCP	Jumlah makanan utama
CAEC	Konsumsi makanan diantara waktu makan
SMOKE	Merokok
CH2O	Konsumsi aair setiap hari
SCC	Pemanrauan konsumsi kalori
FAF	Frekuensi aktivitas fisik
TUE	Waktu penggunaan teknologi
CALC	Konsumsi alkohol
MTRANS	Transfortasi yang digunakan
NObesydad	Kategori Obesitas



Gambar 2. Visualisasi dataset

Berdasarkan Gambar 2. visualisasi data menunjukkan bahwa dataset yang digunakan memiliki noise dan outlier. Oleh karena itu, selanjutnya yaitu memeriksa data noise dan outlier. Adapun pemeriksaan noise tersebut menggunakan pendekatan metode Z-Score. Adapun fungsi tersebut bekerja dengan cara mengukur seberapa jauh suatu nilai dari mean atau satuan standar deviasi. Berdasarkan hal tersebut, dataset yang digunakan memiliki noise sebanyak 180 data dengan menggunakan rentang nilai  $<-3$  dan  $>3$ . Selanjutnya, data noise tersebut dihapus untuk meningkatkan kuliatas data serta analisis yang akurat. Kemudian, tahap selanjutnya yaitu pemeriksaan outlier dengan menggunakan fungsi IQR. Fungsi tersebut digunakan untuk mengatasi data anomali, sehingga data lebih akurat dan relevan. Hasil dari penerapan tersebut menunjukkan sebanyak 782 data dianggap sebagai outlier. Oleh karena itu, data tersebut dihapus. Sehingga, data menjadi 591 data. Adapun dataset tersebut akan digunakan pada tahap selanjutnya. Berikut visualisasi data dari cuplikan dataset setelah pembersihan noise dan outlier (Gambar 3).



Gambar 3. Visualisasi dataset setelah pembersihan noise dan outlier

Berdasarkan Gambar 3 menunjukkan adanya perbedaan visualisasi sebelum dan setelah pembersihan noise dan outlier. Meskipun demikian, dari visualisasi tersebut masih adanya titik diluar rentang tersebut, maka perlu adanya pendekatan lain yang menunjukkan data tersebut dianggap data anomali atau bukan. Namun, pada penelitian ini, data tersebut digunakan pada tahap selanjutnya.

Berdasarkan Tabel 1, nilai atribut NObeyesdad terdiri dari data kategorikal. Sehingga, tahap selanjutnya yaitu transformasi data. Proses transformasi tersebut menggunakan fungsi Labelencoder(). Berikut hasil dari transormasi data yang telah dilakukan dapat dilihat pada Gambar 4.

family_history_with_overweight	FAVC	FCVC	NCP	CAEC	SMOKE	CH2D	SCC	FAF	TUE	CALC	MTRANS	NObeyesdad
1	0	2	3	2	0	2	0	0	1	3	3	1
1	0	2	3	2	0	2	0	2	1	1	3	1
0	0	3	3	2	0	2	0	2	0	1	4	5
0	0	2	1	2	0	2	0	0	0	2	3	6
0	1	2	3	2	0	2	0	0	0	2	0	1
...	...	...	...	...	...	...	...	...	...	...	...	...
1	1	3	3	2	0	1	0	1	0	2	3	4
1	1	3	3	2	0	2	0	1	0	2	3	4
1	1	3	3	2	0	2	0	1	0	2	3	4
1	1	3	3	2	0	2	0	1	0	2	3	4
1	1	3	3	2	0	2	0	1	0	2	3	4

Gambar 4. Hasil transformasi data

Berdasarkan Gambar 4 menunjukkan hasil dari transformasi data yang dilakukan. Adapun atribut yang dilakukan transformasi data tersebut yaitu atribut Nobeyesdad. Transformasi tersebut dilakukan dengan mengubah data kategori menjadi data numerik. Sehingga, data tersebut lebih mudah dilakukan proses modeling. Selain itu, tranformasi tersebut dilakukan dikarenakan metode AdaBoost dan XgBoost hanya menerima data numerik. Setelah, atribut Nobeyesdad nya diubah maka, tahap selanjutnya yaitu normalisasi data. Adapun normalisasi data yang dilakukan tersebut diterapkan pada semua atribut kecuali atribut Nobeyesdad. Pada proses tersebut menggunakan fungsi MinMaxScaler(). Fungsi tersebut bekerja dengan cara mengubah atau rescaling data fitur agar berada

dalam rentang antara 0 dan 1. Prinsip kerja MinMaxScaler adalah menghitung setiap nilai fitur relatif terhadap nilai minimum dan maksimum dari fitur tersebut, lalu menormalisasinya ke dalam rentang yang ditentukan. Berikut hasil dari normalisasi data yang dilakukan ditunjukkan pada Gambar 5.

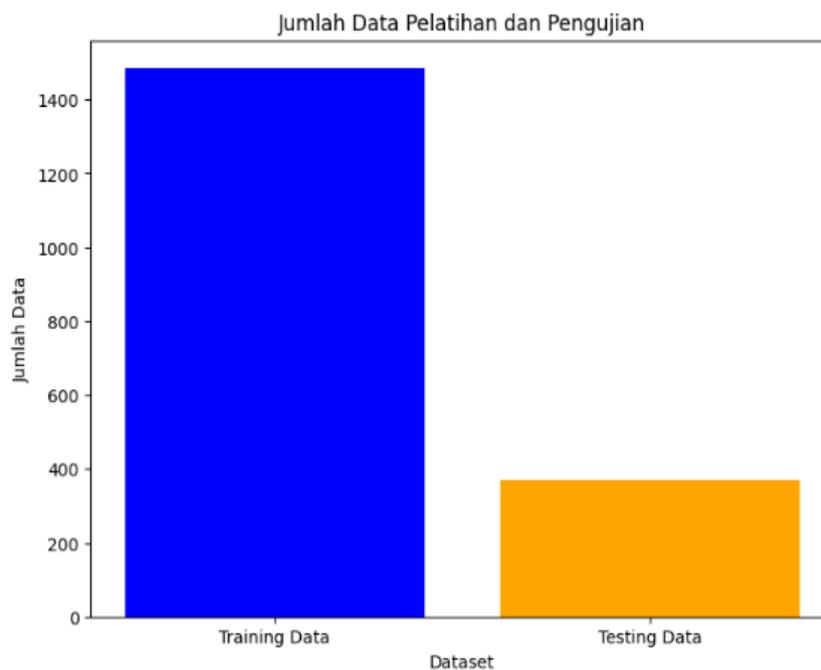
	Gender	Age	Height	Weight	family_history_with_overweight	FAVC	\
0	0.0	0.318182	0.320755	0.206612		1.0	0.0
2	1.0	0.409091	0.660377	0.314050		1.0	0.0
3	1.0	0.590909	0.660377	0.396694		0.0	0.0
4	1.0	0.363636	0.622642	0.413223		0.0	0.0
5	1.0	0.681818	0.320755	0.115702		0.0	1.0

	FCVC	NCP	CAEC	SMOKE	CH2O	SCC	FAF	TUE	CALC	MTRAMS	\
0	0.5	0.666667	0.666667	0.0	0.5	0.0	0.0	0.5	1.000000	0.75	
2	0.5	0.666667	0.666667	0.0	0.5	0.0	1.0	0.5	0.333333	0.75	
3	1.0	0.666667	0.666667	0.0	0.5	0.0	1.0	0.0	0.333333	1.00	
4	0.5	0.000000	0.666667	0.0	0.5	0.0	0.0	0.0	0.666667	0.75	
5	0.5	0.666667	0.666667	0.0	0.5	0.0	0.0	0.0	0.666667	0.00	

Gambar 5. Hasil normalisasi data

Berdasarkan Gambar 5 menunjukkan hasil dari normalisasi data yang dilakukan. Normalisasi data tersebut dilakukan untuk menghindari perbedaan rentang nilai antar atribut. Adapun rentang nilai tersebut dari 0 hingga 1. Selanjutnya, dengan dataset yang telah dinormalisasi tersebut dilakukan proses pembagian dataset dengan perbandingan 80:20. Berikut visualisasi data hasil pembagian dataset tersebut ditunjukkan pada Gambar 6.

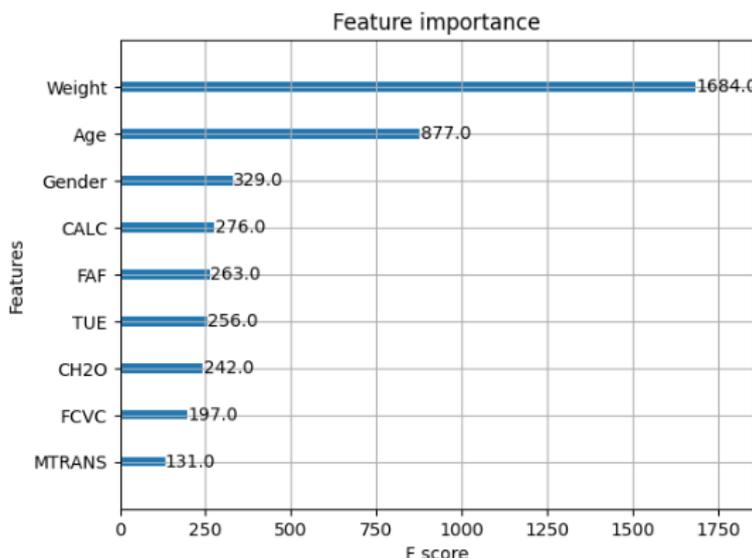


Gambar 6. Visualisasi pembagian dataset

Berdasarkan Gambar 6 menunjukkan visualisasi dari pembagian dataset. Adapun data training sebanyak 472 dan data testing sebanyak 119. Kemudian, data training tersebut akan digunakan untuk melatih model dan data testing digunakan untuk menguji model yang digunakan. Selanjutnya, dataset tersebut siap digunakan pada tahap selanjutnya.

### Hasil Evaluasi Model

Adapun tahap selanjutnya yaitu proses modeling dengan menggunakan metode *AdaBoost* dan *XGBoost*. Berikut hasil dari penerapan *XGBoost* ditunjukkan pada Gambar 7.



Gambar 7. Hasil penerapan *XGBoost*

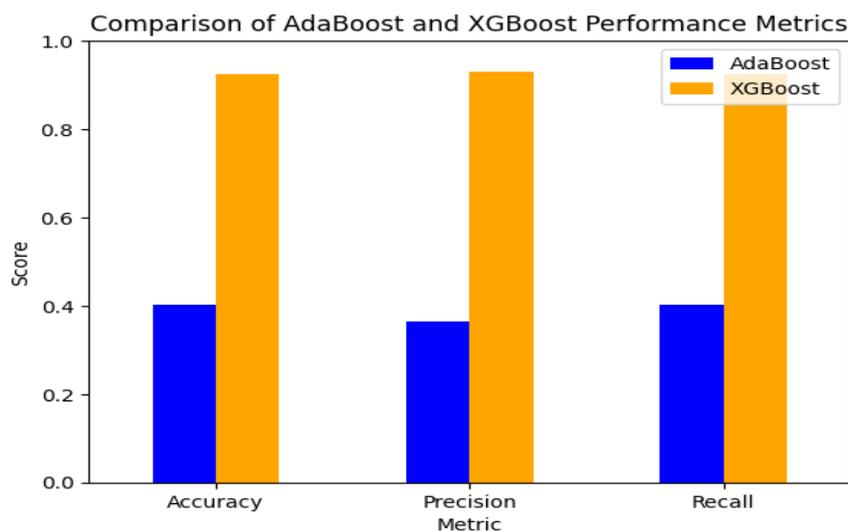
Berdasarkan Gambar 7 menunjukkan hasil dari penerapan metode *XGBoost* yang menghasilkan feature importance. Terdapat beberapa fitur yang memiliki pengaruh besar terhadap model yaitu Weight, Age dan Gender. Sedangkan, fitur yang lain memiliki pengaruh yang lebih rendah dibanding tiga fitur tersebut. Weight adalah indikator langsung dari status obesitas, mencerminkan akumulasi lemak tubuh yang berlebihan. Individu dengan berat badan yang lebih tinggi cenderung mengalami masalah kesehatan terkait obesitas, seperti diabetes tipe 2, hipertensi, dan penyakit jantung. Oleh karena itu, fitur ini memiliki pengaruh yang sangat besar dalam model yang dikembangkan. Selanjutnya, Age memainkan peran penting dalam konteks obesitas. Seiring bertambahnya usia, metabolisme tubuh cenderung melambat, yang dapat menyebabkan penumpukan lemak. Selain itu, perubahan hormonal dan pola aktivitas fisik yang berbeda di berbagai fase kehidupan juga dapat memengaruhi risiko obesitas. Gender juga merupakan faktor penting yang perlu diperhatikan. Jenis kelamin berpengaruh pada pola distribusi lemak tubuh dan respons individu terhadap faktor risiko obesitas. Secara umum, pria dan wanita memiliki pola penyimpanan lemak yang berbeda. Wanita cenderung menyimpan lebih banyak lemak subkutan, sedangkan pria lebih cenderung memiliki lemak visceral. Faktor psikologis dan sosial yang berbeda antara pria dan wanita dapat memengaruhi kebiasaan makan dan aktivitas fisik, berkontribusi pada perbedaan risiko obesitas di antara keduanya.

Adapun fitur yang dianggap memiliki pengaruh lebih rendah tersebut antara lain yaitu CALC (kalori), FAF (frekuensi aktivitas fisik), TUE (jumlah tidur), CH2O (konsumsi karbohidrat), FCVC (frekuensi konsumsi sayuran), dan MTRANS (moda transportasi) menunjukkan bahwa berbagai faktor lain juga mempengaruhi risiko obesitas. Kemudian, hasil dari evaluasi *XGBoost* dan *AdaBoost* ditunjukkan pada Tabel 3.

Tabel 3. Hasil evaluasi model

Metode	Accuracy	Precision	Recall
<i>AdaBoost</i>	40%	36%	40%
<i>XGBoost</i>	92%	92%	92%

Berdasarkan Tabel 3 menunjukkan hasil evaluasi model *AdaBoost* dan *XGBoost*. Adapun *AdaBoost* menghasilkan akurasi sebesar 40%, presisi 36% dan *recall* 40%. Kemudian, untuk *XGBoost* menghasilkan akurasi, presisi dan *recall* sebesar 92%. Oleh karena itu, berdasarkan hasil perbandingan tersebut menunjukkan bahwa model dengan metode *XGBoost* cenderung lebih unggul dibandingkan dengan *AdaBoost*. Terdapat selisih yang signifikan antara hasil evaluasi model *AdaBoost* dan *XGBoost*. Adapun selisih dari akurasi dan *recall* sebesar 52% dan selisih dari presisi sebesar 56%. Berikut visualisasi dari perbandingan antar metode *AdaBoost* dan *XGBoost* ditunjukkan pada Gambar 8.



Gambar 8. Visualisasi dari perbandingan antar metode

Berdasarkan Gambar 8 menunjukkan visualisasi dari perbandingan antar metode *AdaBoost* dan *XGBoost*. Adapun warna Biru menunjukkan hasil evaluasi dari *AdaBoost*. Sedangkan, warna kuning menunjukkan hasil evaluasi dari metode *XGBoost*. Dari hal tersebut menunjukkan bahwa hasil evaluasi *XGBoost* cenderung lebih unggul dibandingkan dengan *AdaBoost*. Hal tersebut dikarenakan *AdaBoost* adalah teknik penguatan yang berfokus pada peningkatan model-model lemah secara bertahap. Meskipun metode ini efektif untuk beberapa jenis dataset, ia dapat menghadapi kesulitan dalam menangkap kompleksitas data, terutama jika terdapat interaksi yang rumit antar fitur. Sementara itu, *XGBoost* merupakan metode gradient boosting yang lebih canggih, dirancang untuk menangani interaksi kompleks dan menyediakan lebih banyak pengaturan hyperparameter.

## KESIMPULAN

Berdasarkan hasil penelitian menunjukkan bahwa penggunaan metode *XGBoost* cenderung lebih unggul dibandingkan dengan *AdaBoost*. Adapun nilai akurasi, precision dan *recall* dari *XGBoost* sebesar 92%. Sedangkan nilai akurasi dan *recall* dari *AdaBoost* sebesar 40% serta precision sebesar 39%. Hal tersebut dikarenakan metode *AdaBoost* cenderung menggunakan pohon keputusan yang dangkal dan cenderung mengalami kesulitan dengan dataset yang sangat kompleks serta menangani overfitting. Sedangkan, metode *Xgboost* tersebut cenderung cocok dengan dataset yang digunakan pada penelitian ini. Sehingga, model dapat memanfaatkan berbagai interaksi antara fitur-fitur yang berbeda. Berikut adalah rangkuman dari pernyataan tersebut: Meskipun *XGBoost* menunjukkan keunggulan. Namun, seberapa baik model ini dapat digeneralisasi ke dataset atau populasi lain itu adalah aspek krusial dalam model prediktif, karena kinerja yang baik pada satu dataset tidak menjamin hasil yang sama pada dataset lainnya. Oleh karena itu, pengujian lebih lanjut diperlukan untuk memastikan efektivitas model *XGBoost* di berbagai konteks dan populasi. Selain

itu, penelitian ini memiliki keterbatasan, terutama ukuran dataset yang relatif kecil setelah pra-pemrosesan, yang dapat memengaruhi keandalan hasil evaluasi model. Dataset yang kecil berpotensi membuat model kurang robust dan lebih rentan terhadap overfitting.

## REFERENSI

- Alnur, B., Mulyono, Fitri Amillia, & Sutoyo, S. (2023). Rancang bangun sistem informasi absensi mahasiswa politeknik kota malang program studi teknik telekomunikasi berbasis web menggunakan fingerprint. *Journal of Informatics and Telecommunication Engineering*, 7(1), 102–111. [doi:10.31289/jite.v3i1.2494](https://doi.org/10.31289/jite.v3i1.2494)
- Amrynia, S. U., & Prameswari, G. N. (2022). Hubungan pola makan, sedentary lifestyle, dan durasi tidur dengan kejadian gizi lebih pada remaja (Studi Kasus di SMA Negeri 1 Demak). *Indonesian Journal of Public Health and Nutrition*, 2(1), 112–121. [doi: 10.15294/ijphn.v2i1.52044](https://doi.org/10.15294/ijphn.v2i1.52044)
- De Guia, J. D., Concepcion, R. S., Bandala, A. A., & Dadios, E. P. (2019). Performance comparison of classification algorithms for diagnosing chronic kidney disease. *2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management, HNICEM 2019*, 1(August), 68–74. [doi:10.1109/HNICEM48295.2019.9073568](https://doi.org/10.1109/HNICEM48295.2019.9073568)
- Dwi, E., Aini, N., Khasanah, R. A., Ristyawan, A., & Neighbors, K. (2024). Penggunaan data mining untuk prediksi tingkat obesitas di Meksiko menggunakan metode Random. 8, 1256–1265. *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, Kediri: Agustus 2024
- Ella, Y., Noor, I., Sugiarto, E., Fatimah, A. S., Kesehatan, F., Universitas, M., & Surabaya, A. (2022). The description of obesity among housewives in the world. *Jurnal Gizi dan Kesehatan*. 14(1). [doi:10.35473/jgk.v14i1.243](https://doi.org/10.35473/jgk.v14i1.243)
- Erwis, F., Suherdi, D., Pranata, A., & Nasyuha, A. H. (2022). Penerapan metode hybrid case base pada sistem pakar diagnosa penyakit obesitas. *Jurnal Media Informatika Budidarma*, 6(1), 378. [doi:10.30865/mib.v6i1.3491](https://doi.org/10.30865/mib.v6i1.3491)
- Febianto, R. T., Suranti, D., & Alinse, R. T. (2024). Penerapan algoritma adaboost dalam mengetahui pola pengguna kb di puskesmas tanjung harapan. *Journal of Science and Social Research*, 4307(1), 145–155. [doi: 10.54314/jssr.v7i1.1727](https://doi.org/10.54314/jssr.v7i1.1727)
- Gede Iwan Sudipa, I., Andi Putra, T. E., Jurnaidi Wahidin, A., Alfa Syukrilla, W., Khrisna Wardhani, A., Heryana, N., Indriyani, T., Willyanto Santoso Tutuk Indriyani, L., & Willyanto Santoso, L. (2023). Data mining. *Institutional Repository UIN Syarif Hidayatullah Jakarta*.
- Givari, M. R., Sulaeman, M. R., & Umaidah, Y. (2022). Perbandingan algoritma svm, random forest dan xgboost untuk penentuan persetujuan pengajuan kredit. *Nuansa Informatika*, 16(1), 141–149. [doi:10.25134/nuansa.v16i1.5406](https://doi.org/10.25134/nuansa.v16i1.5406)
- Hamza, R. Z., & Alsolami, K. (2023). 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*, 9(8). [doi:10.1016/j.heliyon.2023.e18724](https://doi.org/10.1016/j.heliyon.2023.e18724)
- Hanani, R., Badrah, S., & Noviasy, R. (2021). Pola makan, aktivitas fisik dan genetik mempengaruhi kejadian obesitas pada remaja di SMK wilayah kerja puskesmas Segiri. *Jurnal Kesehatan Metro Sai Wawai*, 14(2), 120–129. [doi: 10.26630/jkm.v14i2.2665](https://doi.org/10.26630/jkm.v14i2.2665)
- Masruriyah, A. F. N., Sukmawati, C. E., & Dermawan, B. A. (2024). Memahami data mining dengan python: implementasi praktis. *Eureka Media Aksara*.
- Musu, W., Ibrahim, A., & Heriadi. (2021). Pengaruh komposisi data training dan testing terhadap akurasi algoritma c4.5. *Prosiding Seminar Ilmiah Sistem Informasi dan Teknologi Informasi*, X(1), 186–195.
- Nikmah, T. L., Syafei, R. M., Muzayanah, R., & Salsabila, A. (2023). Prediction of used car prices using k-nearest neighbour, random forest and adaptive boosting algorithm. *Indonesian Community on Optimization and Computer Application*, 1(1), 17–22.
- Prabowo, T. T. (2021). Efektivitas sistem temu kembali informasi perpustakaan digital institut seni indonesia (isi) yogyakarta dalam tinjauan recall dan precision. *Media Pustakawan*, 28(1), 37–48. [doi: 10.37014/medpus.v28i1.1087](https://doi.org/10.37014/medpus.v28i1.1087)
- Safitri, D. E., & Rahayu, S. (2020). Determinan status gizi obesitas pada orang dewasa di perkotaan: tinjauan sistematis. *ARKEMAS (Arsip Kesehatan Masyarakat)*.
- Sanjaya, J., Renata, E., Budiman, V. E., Anderson, F., & Ayub, M. (2020). Prediksi kelalaian pinjaman bank

- menggunakan Random Forest dan Adaptive Boosting. *Jurnal Teknik Informatika dan Sistem Informasi*, 6(1), 50–60. [doi: 10.28932/jutisi.v6i1.2313](https://doi.org/10.28932/jutisi.v6i1.2313)
- Siddik Hasibuan, M., & Harahap, H. (2024). Penerapan metode haar-like feature dan algoritma adaboost dalam penentuan klasifikasi hama tanaman kopi. *Journal of Science and Social Research*, 4307(1), 87–93. [doi: 10.54314/jssr.v7i1.1718](https://doi.org/10.54314/jssr.v7i1.1718)
- Syifa, E. D. A., & Djuwita, R. (2023). Factors associated with overweight/obesity in adolescent high school students in Pekanbaru city. *Jurnal Kesehatan Komunitas*, 9(2), 368–378. [doi:10.25311/keskom.vol9.iss2.1579](https://doi.org/10.25311/keskom.vol9.iss2.1579)
- Wie, J. V., & Siddik, M. (2022). Penerapan metode naïve bayes dalam mengklasifikasi tingkat obesitas pada pria. *JOISIE Journal of Information System And Informatics Engineering*, 6(Desember), 69–77. [doi:10.35145/joisie.v6i2.2467](https://doi.org/10.35145/joisie.v6i2.2467)