

KLASIFIKASI TINGKAT OBESITAS MENGGUNAKAN METODE GBM DAN CONFUSION MATRIX

Farhan Reynaldi Valerian, Mohammad Syarief, Doni Abdul Fatah

Sistem Informasi, Universitas Trunojoyo Madura

Jl. Raya Telang, Kecamatan Kamal, Bangkalan, Jawa Timur, Indonesia

farhan.reynaldi.valerian@gmail.com

ABSTRAK

Di zaman yang serba teknologi ini, semua hal dapat diperoleh dengan mudah, mulai dari informasi, jasa serta layanan. Dari segi informasi, semua orang dapat memperoleh berbagai data dan panduan, seperti resep makanan, layanan pengiriman barang, hingga layanan kesehatan. Dalam jurnal ini, topik yang dibahas adalah tentang obesitas, sebuah masalah kesehatan yang semakin meningkat secara global. Metode GBM dipilih karena kemampuannya dalam menangani data non-linear dan memberikan akurasi tinggi dalam masalah klasifikasi. Dengan memanfaatkan dataset yang terdiri dari 2111 entri dan 16 atribut yang mencakup faktor-faktor fisik dan gaya hidup, penelitian ini berhasil mengidentifikasi pola-pola yang signifikan dalam klasifikasi obesitas. Hasil analisis menunjukkan bahwa model GBM tidak hanya mencapai akurasi tinggi sebesar 95%, tetapi juga menunjukkan keseimbangan yang baik antara precision, recall, dan F1-Score, yang masing-masing mencapai 95%, 96%, dan 95%.

Kata kunci : *Obesitas, Machine learning, Klasifikasi, GBM, Confusion matrix*

1. PENDAHULUAN

Perkembangan teknologi informasi yang pesat telah mengubah cara kita hidup secara signifikan. Akses mudah terhadap informasi, layanan, dan berbagai kemudahan lainnya telah menjadi bagian tak terpisahkan dari kehidupan modern. Namun, di balik segala kemudahan tersebut, kita juga dihadapkan pada tantangan baru, salah satunya adalah peningkatan kasus obesitas. Gaya hidup modern yang semakin sedentary, pola makan yang tidak sehat, serta pengaruh lingkungan yang memicu konsumsi makanan olahan berlebih menjadi beberapa faktor utama yang berkontribusi pada masalah makanan global [1].

Kebiasaan makan yang tidak sehat, seperti makan berlebihan, makan terlalu cepat, mengonsumsi makanan yang terlalu banyak mengandung bumbu, atau makanan yang terkontaminasi oleh bakteri. [2]. mengalami gangguan makan dan peningkatan resiko gangguan makan diantaranya dengan memiliki status gizi lebih [3].

Obesitas merupakan kondisi yang ditandai dengan penumpukan lemak berlebih dalam tubuh, yang dapat meningkatkan risiko berbagai penyakit kronis seperti diabetes, hipertensi, dan penyakit jantung. Dengan prevalensi obesitas yang terus meningkat, diperlukan metode yang efektif untuk mendeteksi dan mencegah kondisi ini sejak dini.

Klasifikasi adalah teknik dalam machine learning yang bertujuan untuk memetakan data ke dalam kelas-kelas yang telah ditentukan sebelumnya. Proses ini melibatkan pelatihan model untuk mempelajari pola dalam data pelatihan, sehingga model dapat memprediksi kelas dari data baru yang belum pernah dilihat sebelumnya [4].

Klasifikasi merupakan teknik dalam machine learning yang bertujuan untuk mengelompokkan data ke dalam kategori yang spesifik berdasarkan

karakteristiknya. Klasifikasi juga berfungsi dalam berbagai tahap dalam proses analisis data, mulai dari eksplorasi data hingga proses pembuatan model. Dengan mengidentifikasi pola dalam data, klasifikasi membantu kita memahami data dengan lebih baik dan membuat keputusan yang lebih informatif. Salah satu jenis klasifikasi adalah GBM memiliki kepanjangan Gradient Boosting Machine.

GBM adalah teknik yang menggabungkan banyak pohon keputusan untuk membuat model prediksi yang kuat. Masing-masing pohon keputusan dalam GBM belajar dari kesalahan pohon sebelumnya, sehingga model secara keseluruhan menjadi semakin akurat. Proses pembelajaran ini dipandu oleh fungsi kerugian yang dapat di optimasi secara matematis [5]. GBM dapat dimanfaatkan untuk mengklasifikasikan status obesitas seseorang berdasarkan atribut seperti indeks massa tubuh (IMT), lingkar pinggang, kadar lemak tubuh, dan riwayat penyakit. GBM, sebagai salah satu algoritma ensemble, terbukti efektif dalam menangani data kompleks dan memberikan akurasi yang tinggi dalam tugas klasifikasi.

Penelitian ini berfokus pada klasifikasi tingkat obesitas menggunakan metode Gradient Boosting Machine (GBM), yang dikenal karena kemampuannya dalam menangani data non-linear dan memberikan akurasi tinggi dalam masalah klasifikasi. Dataset yang digunakan mencakup berbagai atribut yang relevan, seperti data demografis, kebiasaan makan, aktivitas fisik, dan gaya hidup, dengan tujuan untuk memahami faktor-faktor yang berkontribusi terhadap obesitas. Hasil dari penelitian ini diharapkan dapat memberikan wawasan baru yang berguna untuk pencegahan obesitas dan pengembangan strategi intervensi yang lebih efektif.

Tujuan dari penelitian ini adalah untuk mengetahui hasil klasifikasi data, data yang digunakan

ini memiliki 2111 data dengan 16 atribut serta 1 kolom target atau label. Hasil klasifikasi di hitung menggunakan evaluasi model Confusion Matrix untuk menghasilkan nilai *accuracy*, *recall*, *precision*, dan *f1-score*.

2. TINJAUAN PUSTAKA

2.1. Obesitas

Obesitas adalah kondisi medis yang ditandai oleh penumpukan lemak tubuh yang berlebihan, yang dapat berdampak negatif pada kesehatan seseorang. Dampak dari obesitas adalah begitu luas dan kompleks, mulai dari masalah kesehatan fisik hingga psikologis. Obesitas dapat meningkatkan risiko berbagai penyakit kronis seperti penyakit jantung, stroke, diabetes tipe 2, beberapa jenis kanker, serta gangguan sendi dan pernapasan. Selain itu, obesitas juga dapat menurunkan kualitas hidup seseorang, menyebabkan rendah diri, dan menghambat aktivitas sehari-hari.

Namun, sebelum obesitas menjadi masalah serius, deteksi dini dapat dilakukan dengan mengukur menggunakan Indeks Massa Tubuh (IMT), yang merupakan perbandingan antara berat badan (dalam kilogram) dan tinggi badan kuadrat (dalam meter). Seseorang dikategorikan obesitas jika IMT-nya lebih dari 30 kg/m². Umumnya Obesitas disebabkan oleh ketidakseimbangan antara jumlah energi yang masuk dengan yang dibutuhkan oleh tubuh untuk berbagai fungsi biologis seperti pertumbuhan fisik, perkembangan, aktivitas, pemeliharaan kesehatan[6].

2.2. Machine Learning

Machine learning adalah salah satu teknologi dalam *artificial intelligence* atau bahasa Indonesianya kecerdasan buatan yang dirancang untuk meniru kemampuan berpikir manusia. Teknologi ini digunakan untuk menciptakan sistem yang dapat beroperasi dan mengambil keputusan secara mandiri tanpa memerlukan pemrograman ulang. Machine learning bekerja dengan menganalisis data input untuk mengidentifikasi pola tertentu, terutama dalam jumlah data yang besar atau big data. Dalam prosesnya, terdapat dua konsep utama: data training dan data testing. Data training digunakan untuk melatih algoritma agar dapat mengenali pola, sementara data testing berfungsi menguji kemampuan algoritma dalam memproses data baru yang belum pernah digunakan selama pelatihan[7].

Dalam machine learning, hasil yang akurat dapat dicapai melalui penerapan teknik-teknik cerdas. Beberapa teknik yang digunakan dalam machine learning meliputi supervised learning, unsupervised learning, semi-supervised learning, dan reinforcement learning. Salah satu teknik tersebut adalah unsupervised learning, yang merupakan metode untuk menganalisis dataset tanpa adanya label respons yang terdefinisi, sehingga bisa mendapatkan kesimpulan secara mandiri dari data tersebut[8].

2.3. Data Mining

Data mining adalah proses yang bertujuan untuk mengekstrak informasi dan mendapatkan wawasan dari kumpulan data yang berukuran besar. Proses dalam data mining meliputi beberapa tahapan, seperti pemahaman data (Data Understanding), persiapan data (Data Preparation), pemodelan data (Data Modeling), evaluasi hasil (Evaluation), hingga penerapan model (Deployment).[9].

Dalam data mining, informasi diambil dari basis data yang besar dengan menggabungkan pendekatan dari berbagai bidang ilmu, seperti pembelajaran mesin, pengenalan pola, statistik, basis data, dan visualisasi.

Data mining memanfaatkan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk menggali informasi dari basis data yang besar. Proses ini melibatkan berbagai metode untuk mengidentifikasi pola, hubungan, dan tren tersembunyi dalam data. Berdasarkan tugasnya, data mining dapat dibagi menjadi lima kategori utama: Exploratory Data Analysis (EDA) untuk memahami karakteristik data, Descriptive untuk mendeskripsikan data melalui model statistik, Modelling untuk memprediksi nilai atau kelas variabel tertentu, Predictive Modelling, serta penemuan pola dan aturan yang mendasari data, termasuk pemanggilan konten. [10].

Dalam data mining ada klasifikasi, clustering, asosiasi, regresi, anomali detection, serta reduksi dimensi. Dalam penelitian ini membahas dalam lingkup klasifikasi.

2.4. Klasifikasi

Klasifikasi adalah metode analisis yang bertujuan untuk mengelompokkan data ke dalam kategori-kategori berdasarkan sifat atau karakteristik yang dimilikinya. [11].

Klasifikasi sering diterapkan untuk membangun model prediksi yang dapat membantu dalam pengambilan keputusan. Proses klasifikasi adalah mengelompokkan data ke dalam kategori tertentu berdasarkan aturan yang telah ditetapkan. Berbeda dengan clustering, yang tidak membutuhkan label pada hasil data, klasifikasi memerlukan data berlabel untuk proses pembelajarannya. Beberapa algoritma yang digunakan dalam klasifikasi antara lain decision tree, naïve bayes, k-nearest neighbor (KNN), dan support vector machine (SVM). Algoritma-algoritma tersebut mempelajari data untuk mengidentifikasi pola yang menghubungkan fitur data dengan kategori yang sesuai[12].

Dalam klasifikasi, terdapat beberapa tahapan seperti pembuatan model, penerapan model, dan evaluasi. Model dibangun dengan menggunakan data latih yang sudah memiliki atribut dan kelas. Kemudian, data baru diterapkan pada model untuk menentukan kelasnya. Pada tahap terakhir, model dievaluasi untuk mengukur akurasi dalam membangun dan menerapkan model pada data baru. Proses klasifikasi terdiri dari dua tahap, yaitu tahap pelatihan

dan tahap pengujian. Pada tahap pelatihan, data digunakan untuk mengembangkan model, sementara tahap pengujian digunakan untuk menguji model dengan data berbeda untuk menilai tingkat akurasi[13].

Salah satu metode yang jarang digunakan dalam proses klasifikasi adalah GBM atau *Gradient Booster Machine*.

2.5. Gradient Booster Machine

Gradient Boosting Machine (GBM) adalah salah satu algoritma ensemble yang digunakan untuk tugas klasifikasi dan regresi. GBM bekerja dengan cara membangun model prediksi secara bertahap, di mana setiap model baru berfokus untuk mengoreksi kesalahan (residual) dari model sebelumnya. Proses ini dilakukan dengan menambahkan model-model sederhana (biasanya decision tree) secara berurutan, sehingga model final merupakan kombinasi dari pohon keputusan yang saling memperbaiki kesalahan masing-masing.

GBM mengoptimalkan fungsi kerugian menggunakan gradient descent, yang mendorong algoritma ini untuk meningkatkan akurasi dengan cepat. Proses pelatihan dilakukan dengan cara memberi bobot lebih pada data yang sulit diprediksi oleh model sebelumnya, dan semakin banyak model yang ditambahkan, semakin akurat prediksi yang dihasilkan. Keunggulan dari GBM adalah kemampuannya untuk menangani data dengan berbagai jenis distribusi dan memiliki kinerja yang sangat baik pada berbagai masalah machine learning.

Gradient Boosting Machine (GBM) adalah metode ensemble yang membangun model secara bertahap, di mana setiap model baru berfokus pada kesalahan yang dibuat oleh model sebelumnya. Proses dimulai dengan membuat model dasar yang sederhana, seperti prediksi rata-rata, lalu model ditingkatkan dengan membangun pohon keputusan untuk meminimalkan kesalahan residual. Setiap model baru ditambahkan secara berturut-turut, dan hasil prediksi diperbarui dengan menambahkan output model baru yang telah diperbaiki dengan learning rate η . Proses ini berlanjut hingga mencapai jumlah iterasi yang ditentukan atau konvergen.

Pada setiap iterasi, GBM menggunakan fungsi kerugian (seperti MSE untuk regresi) untuk menghitung kesalahan dan gradien, yang digunakan untuk memperbarui model. Update dilakukan dengan menambahkan model baru yang dikalibrasi dengan gradien tersebut, sehingga model secara keseluruhan dapat menghasilkan prediksi yang lebih akurat. Formula umum untuk pembaruan model adalah $F_t(x) = F_{t-1}(x) + \eta h_t(x)$, di mana $h_t(x)$ adalah model baru dan η adalah learning rate yang mengontrol pembaruan tersebut.

2.6. Evaluasi Model

Pada tahap evaluasi model, digunakan Confusion Matrix untuk menilai kinerja model klasifikasi.

Confusion Matrix adalah tabel yang membandingkan hasil prediksi model dengan data asli, memberikan informasi tentang jumlah prediksi yang benar (True Positive dan True Negative) serta kesalahan prediksi (False Positive dan False Negative).

2.7. Google Colaboratory

Google Colaboratory (Colab) adalah sebuah platform yang digunakan untuk menulis dan menjalankan kode Python secara langsung di browser, tanpa perlu menginstal perangkat lunak tambahan di komputer Anda. Colab menyediakan lingkungan pengembangan yang lengkap, termasuk akses ke GPU (Graphics Processing Unit) yang sangat berguna untuk menjalankan tugas komputasi intensif seperti machine learning.

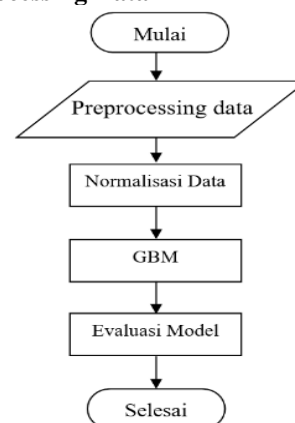
2.8. Penelitian Terdahulu

Havidez Yasaf Al-Afghoni meneliti tentang klasifikasi jenis kacang menggunakan Decision tree C4.5, SMOTE, Confusion Matrix dan dari penelitian itu menghasilkan tingkat akurasi sebesar 94% presisi sebesar 94%, recallnya 94%, dan f1- scorenya adalah 94%[14], pada tahun 2024 rif'at chusnul maafi telah melakukan penelitian tentang processing gambar menggunakan metode cnn, dengan menggunakan evaluasi model Confusion matrix, dan penelitian tersebut menghasilkan nilai tingkat akurasi sebesar 76.92%.[15], tahun 2023 kurniawan melakukan penelitian dengan Tujuan menemukan akurasi tertinggi di antara algoritma klasifikasi prediksi yang diusulkan penerima bantuan raskin memakai tools python machine learning dan diimplementasikan melalui suatu website, menunjukkan algoritma klasifikasi RF memiliki hasil precision, recall, f-measure yaitu nilai 97%, nilai accuracy sebesar 97,26% dan nilai ROC 0,970[16].

3. METODE PENELITIAN

Pada penelitian ini menggunakan pemrosesan data dengan metode GBM, tapi sebelum itu data di normalisasi dan dipisahkan menjadi data training dan evaluasi model menggunakan confusion matrix

3.1. Preprocessing Data



Gambar 1. flowchart Penelitian

Pada tahap ini data yang digunakan di inputkan ke dalam sistem untuk di proses perhitungan, data yang digunakan adalah data tentang apel dan gambarannya sebagai berikut:

Tabel 1. Penjelasan Data

Atribut	keterangan
Gender	Jenis kelamin
Age	Umur (th)
Height	Tinggi badan (cm)
Weight	Berat badan (kg)
family_history_with_overweight	Riwayat keluarga terkait obesitas (ya/tidak).
FAVC	Frekuensi makan cepat saji (fast food).
FCVC	Frekuensi konsumsi sayuran.
NCP	Jumlah makanan per hari.
CAEC	Konsumsi alkohol setiap hari.
SMOKE	Kebiasaan merokok (ya/tidak).
CH2O	Konsumsi air harian. (1)
SCC	Aktifitas Fisik.
FAF	Frekuensi makan buah.
TUE	Waktu penggunaan perangkat elektronik.
CALC	Konsumsi kalsium.
MTRANS	Mode transportasi yang sering digunakan.
NObesyedad	Label kelas (tingkat obesitas).

3.2. Normalisasi Data

Normalisasi data adalah proses mengubah nilai-nilai data menjadi skala tertentu, biasanya dalam rentang tertentu, seperti 0 hingga 1 atau -1 hingga 1. Normalisasi digunakan untuk memastikan bahwa semua fitur atau variabel memiliki skala yang sebanding, sehingga tidak ada fitur yang mendominasi perhitungan dalam model machine learning karena memiliki nilai yang jauh lebih besar daripada yang lain. Normalisasi dilakukan menggunakan MinMaxScaling, Rumusnya sebagai berikut :

$$X = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Dimana X adalah data yang di pilih, Xmin adalah data terkecil, dan Xmax data terbesar.

3.3. GBM (Gradient Boosting Machine)

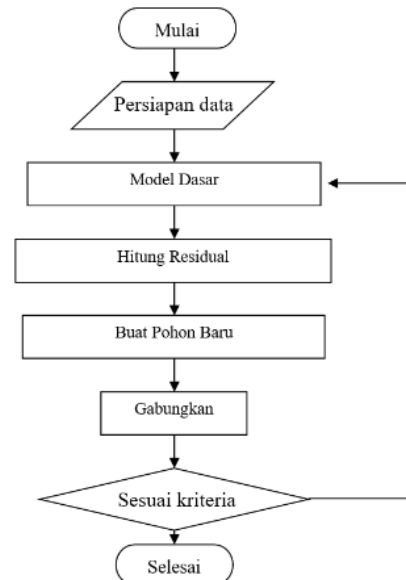
GBM (Gradient Boosting Machine) adalah salah satu algoritma machine learning yang sangat populer, terutama digunakan untuk tugas regresi dan klasifikasi. Intinya, GBM adalah metode yang secara bertahap membangun model prediksi yang kuat dengan menggabungkan banyak model yang lebih sederhana.

a. Persiapan Data

Dari data yang telah di proses dan di persiapan, data dilakukan pemisahan data menjadi data testing dan data training, dan dilakukan perhitungan klasifikasi.

b. Inisialisasi Model Dasar

Model dasar dalam GBM biasanya adalah pohon keputusan (decision tree) yang sederhana. Pohon keputusan ini terus diperbaiki dan digabungkan dengan pohon-pohon lainnya untuk membentuk model akhir yang lebih kompleks.



Gambar 2. flowchart alur GBM

• Pemilihan Akar

Dimulai dengan memilih atribut (fitur) yang paling baik untuk membagi data menjadi dua kelompok. Pemilihan atribut ini biasanya berdasarkan pada metrik ketidakmurnian seperti Gini impurity atau entropy.

• Pembentukan Cabang

Setiap cabang dari akar mewakili nilai yang berbeda dari atribut yang dipilih. Proses ini diulang secara rekursif untuk setiap cabang sampai mencapai kondisi berhenti (misalnya, kedalaman maksimum pohon tercapai atau jumlah sampel pada setiap daun minimal).

• Penentuan Nilai Daun

Setiap daun pada pohon berisi nilai prediksi. Untuk masalah regresi, nilai ini bisa berupa rata-rata target variabel pada data yang masuk ke daun tersebut. Untuk masalah klasifikasi, nilai ini bisa berupa kelas yang paling sering muncul pada data di daun tersebut.

c. Menghitung Nilai Residual

residual adalah selisih antara nilai prediksi yang dihasilkan oleh model pada iterasi sebelumnya dengan nilai aktual. Sederhananya, residual menunjukkan seberapa besar kesalahan yang dibuat oleh model pada iterasi tersebut.

d. Bangun Pohon Baru

Dalam setiap iterasi pada algoritma GBM, model diperbarui dengan menambahkan sebuah pohon keputusan baru. Proses pembaruan ini bertujuan

- untuk meningkatkan kinerja model secara bertahap.
- Gabungkan
Dalam GBM diperlukan menggabungkan model lama dengan model baru, hal ini dilakukan untuk membentuk model yang lebih baik, jadi setiap pengulangan memperbaiki model sehingga hasil lebih akurat.
 - Kriteria
Mengulangi langkah mulai dari inialisasi model apabila belum mencapai kriteria yang di tetapkan seperti mencapai pohon tertentu ataupun hasil tertentu.

3.4. Evaluasi Model

Dalam penelitian ini menggunakan confusion matrix, Confusion Matrix adalah metode evaluasi model klasifikasi yang berbentuk tabel matriks untuk membandingkan hasil prediksi model dengan data aktual. Matriks ini memberikan gambaran detail mengenai prediksi yang benar dan salah untuk setiap kelas, sehingga memudahkan analisis performa model.

Dari Confusion Matrix, dapat menghitung metrik evaluasi seperti:

- True Positive (TP): Data aktual "Good" dan diprediksi sebagai "Good".
- True Negative (TN): Data aktual "Bad" dan diprediksi sebagai "Bad".
- False Positive (FP): Data aktual "Bad" tetapi diprediksi sebagai "Good".
- False Negative (FN): Data aktual "Good" tetapi diprediksi sebagai "Bad".

Akurasi: Persentase prediksi yang benar. (2)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Presisi: Tingkat ketepatan prediksi positif. (3)

$$Precision = \frac{TP}{TP + FP}$$

Recall: Kemampuan model mendeteksi data positif.

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

F1-Score: Harmonik rata-rata antara presisi dan recall.

$$F1 - Score = 2 \times \frac{precision \times recall}{recall + precision} \times 100\% \quad (5)$$

Confusion Matrix memberikan hasil menyeluruh terhadap performa model, baik pada dataset seimbang maupun tidak seimbang, sehingga menjadi metode evaluasi yang umum digunakan.

4. HASIL DAN PEMBAHASAN

4.1. Preprocessing Data

Pada tahap preprocessing data, dilakukan beberapa langkah untuk mempersiapkan dataset agar dapat digunakan dalam proses klasifikasi. Langkah-langkah yang dilakukan adalah sebagai berikut:

- Penanganan Data Kategorikal
Kolom-kolom dengan tipe data kategorikal seperti Gender, family_history_with_overweight, FAVC, dan lainnya diubah menjadi nilai numerik menggunakan metode Label Encoding. Proses ini dilakukan agar data dapat dipahami oleh algoritma machine learning.
- Split data
Dataset dibagi menjadi dua bagian, yaitu data latih (80%) dan data uji (20%), menggunakan metode train_test_split. Data latih digunakan untuk membangun model, sementara data uji digunakan untuk mengevaluasi performa model.

4.2. Normalisasi Data

Semua kolom numerik pada dataset dinormalisasi menggunakan metode Min-Max Scaling untuk merentangkan nilai dalam rentang [0, 1]. Normalisasi ini bertujuan untuk mengurangi perbedaan skala antar fitur sehingga algoritma dapat bekerja lebih optimal. Menggunakan rumus:

$$X = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (6)$$

Sebagai contoh menggunakan data kolom CAEC yang telah di wakikan menjadi Always menjadi 1, Frequently menjadi 0.5, dan sometimes menjadi 0. Sehingga dapat dihitung:

$$X = \frac{0.5 - 0}{1 - 0} = 0,5$$

Hasil dari normalisasi data dengan data yang dipilih adalah 0.5.

4.3. Penerapan GBM

Gradient Boosting Machine (GBM) adalah metode pembelajaran mesin yang berbasis pada prinsip boosting. Berikut adalah alur perhitungan dalam penerapan GBM:

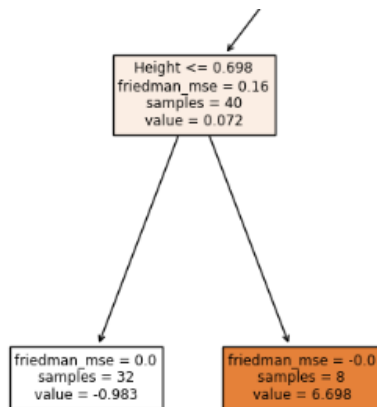
- Persiapan Data
Data yang telah di normalisasi dan di pisah menjadi data testing dan data training dimasukan dalam perhitungan GBM.
- Inisialisasi Model
Model dimulai dengan prediksi yang sederhana dari data, probabilitas kelas dikarenakan untuk perhitungan klasifikasi. Ini berfungsi sebagai titik awal.
- Perhitungan Residual (Error)
Untuk setiap data, hitung error atau residual, yang merupakan selisih antara nilai aktual dan prediksi model saat ini. Menggunakan rumus

$$Residual = y_i - \hat{y}_i \quad (7)$$

y_i = nilai asli

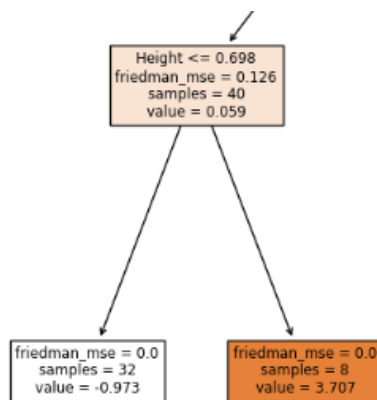
\hat{y}_i = prediksi dari model sebelumnya.

d. Bangun Pohon baru



Gambar 3. Sampel pengulangan ke 1

Pada gambar 3. node dan leaf pada perulangan ke 1 pada height.



Gambar 4. Sampel pengulangan ke 2

Berikutnya gambar 4 adalah perhitungan pada perulangan ke 2 memang berbeda sedikit pada nilai value, namun semakin sedikit nilainya maka hasilnya semakin akurat.

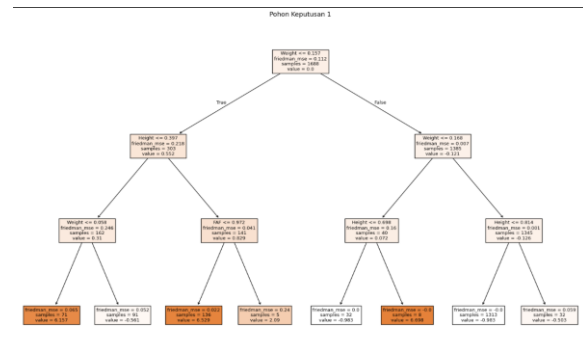
e. Gabungkan

Pada poin ini menggabungkan prediksi model yang ada diperbarui dengan cara menggabungkan prediksi dari kekurangan model sebelumnya ke dalam prediksi yang sudah ada. Pembaruan ini dilakukan dengan menambahkan hasil prediksi dari model baru yang telah dilatih ke prediksi yang sebelumnya dihasilkan oleh model.

f. Kriteria

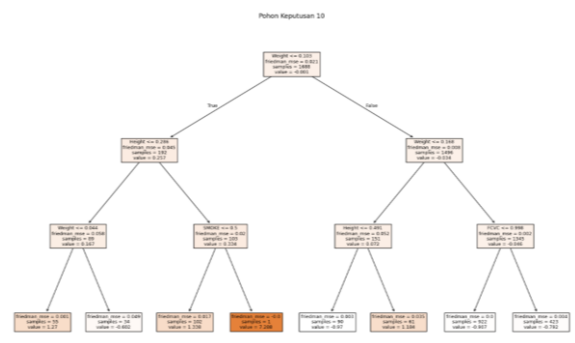
kriteria atau aturan yang digunakan untuk menghentikan proses perulangan pembelajaran dalam Gradient Boosting Machine (GBM). Penelitian ini menggunakan maksimal 10 kali perulangan sehingga menghasilkan 10 pohon keputusan.

Hasil dari proses GBM. Penelitian ini dilakukan dengan batasan 10 kali pengulangan pohon keputusan.



Gambar 5. Hasil pohon pada pengulangan ke 1

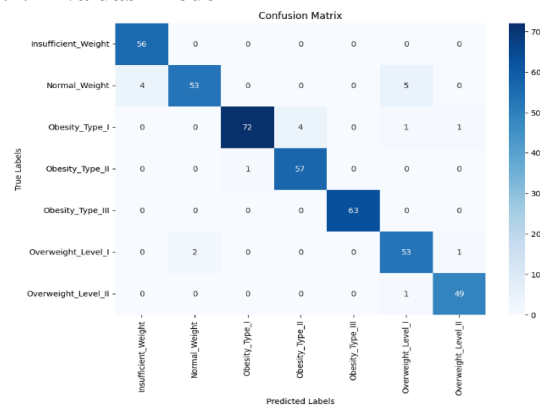
Gambar 5 adalah hasil prediksi dengan data awal dan menghasilkan model dasar.



Gambar 6. Hasil pohon pada pengulangan ke 10

Gambar 6 adalah hasil akhir setelah melakukan 10 pengulangan setiap pengulangan memperbaiki data sebelumnya sehingga mendapat hasil paling optimal.

4.4. Evaluasi Model



Gambar 7. Hasil akhir dari confusion Matrix

Pada gambar 7 dapat disimpulkan nilai dari TP, FP, FN, TN pada tiap label.

Table 2 hasil nilai tiap label.

Label	TP	FP	FN	TN
Insufficient_Weight	56	4	0	306
Normal_Weight	53	2	4	307
Obesity_Type_I	72	0	4	290
Obesity_Type_II	57	4	1	304
Obesity_Type_III	63	0	0	303
Overweight_Level_I	53	1	2	310
Overweight_Level_II	49	1	1	315

Dari tabel 2 sebagai sampel data kelas "Insufficient_Weight" dapat dihitung Akurasi, Presisi, Recall, dan F1-Score.

a. Akurasi (Accuracy)

Akurasi mengukur seberapa sering model memprediksi dengan benar, baik positif maupun negatif, dari seluruh data.

$$Accuracy = \frac{56 + 306}{56 + 306 + 4 + 0} = 0,989$$

Hasil perhitungan akurasinya adalah 0.989 atau 98.9%

b. Presisi (Precision)

Presisi mengukur seberapa akurat prediksi positif, atau semua yang diprediksi sebagai 'Insufficient_Weight'.

$$Precision = \frac{56}{56 + 4} = 0,933$$

Dengan kata lain nilai hasilnya adalah 0.933 atau 93.3%

c. Recall (Sensitivity/True Positive Rate)

Recall mengukur kemampuan model untuk menemukan semua instance positif yang sebenarnya.

$$Recall = \frac{56}{56 + 0} = 1.00$$

Hasilnya recallnya sempurna yaitu 1.00 atau 100%

d. F1-Score

F1-Score adalah rata-rata harmonik dari presisi dan recall. Ini berguna ketika ada ketidakseimbangan kelas. F1-Score juga mencoba menyeimbangkan presisi dan recall.

$$F1 - Score = 2 \times \frac{0.933 \times 1.00}{0.933 + 1.00} = \frac{1,866}{1,933} = 0,965$$

Hasil dari perhitungan f1-scorenya adalah 0.965 atau 96.5%

Kesimpulan untuk Kelas "Insufficient_Weight": Akurasi: 98.9%, Presisi: 93.3, Recall: 100%, F1-Score: 96.5%.

Perlu diingat bahwa perhitungan ini hanya untuk kelas "Insufficient_Weight". proses ini diulangi untuk setiap kelas lainnya dalam confusion matrix untuk mendapatkan gambaran lengkap tentang kinerja model sehingga dapat menghitung nilai rata-rata tiap kelas atau label sehingga hasil rata-ratanya adalah akurasi sebesar 95%, precision 95%, recall 96%, dan F1-Score 95%

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengklasifikasikan tingkat obesitas berdasarkan 16 atribut fisik dan gaya hidup, dengan 1 label kategori obesitas, menggunakan model Gradient Boosting Machine (GBM). Model menunjukkan performa yang sangat baik dengan akurasi sebesar 95%, macro average precision 95%, recall 96%, dan F1-Score 95%. Berdasarkan confusion matrix, model dapat memprediksi kategori obesitas secara andal, meskipun terdapat beberapa kesalahan seperti sejumlah data yang salah diklasifikasikan antar kategori obesitas. Hasil ini membuktikan bahwa GBM

adalah metode yang efektif untuk klasifikasi tingkat obesitas.

Untuk meningkatkan performa, penelitian lanjutan dapat menerapkan optimasi hyperparameter, meningkatkan jumlah data pelatihan, atau mengeksplorasi algoritma lain seperti Random Forest atau Neural Networks. Selain itu, penambahan atribut baru atau penerapan teknik pra-pemrosesan data yang lebih canggih dapat membantu meningkatkan akurasi dan keandalan model secara keseluruhan.

DAFTAR PUSTAKA

- [1] Rizka, I. Fatmawati, T. Maryusman, and N. Nasrullah, "The Relationship of Body Image, Anxiety Level and Balanced Nutritional Knowledge with Eating Behavior among Students at 45 Vocational High School in Jakarta," *J. Gizi dan Kesehat.*, vol. 15, no. 1, pp. 45–54, 2023.
- [2] I. Khoirunnisa and M. Saparwati, "Pola Makan pada Penderita Gastritis di Wilayah Kerja Puskesmas Gunung Pati Kota Semarang," *J. Gizi Dan Kesehat.*, vol. 12, no. 1, pp. 19–25, 2020, doi: 10.35473/jgk.v12i1.67.
- [3] H. Cynthia and S. F. Simanungkalit, "The Relationship Between Body Image, Eating Disorders and Stress Levels With Nutritional Status of Nutrition Students UPN Veteran Jakarta," *J. Gizi dan Kesehat.*, vol. 15, no. 1, pp. 33–44, 2023.
- [4] F. M. Hana, "Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 32–39, 2020, doi: 10.47970/siskom-kb.v4i1.173.
- [5] V. Atlantic, E. Sulistianingsih, and H. Perdana, "Gradient Boosting Machine Pada Klasifikasi Kelulusan Mahasiswa," *Bul. Ilm. Math. Stat. dan Ter.*, vol. 13, no. 2, pp. 165–174, 2024.
- [6] R. A. D. Sartika, "Faktor Risiko Obesitas Pada Anak 5-15 Tahun," *Makara Kesehat.*, vol. 1, no. 1, pp. 37–43, 2011, [Online]. Available: https://www.academia.edu/17631598/FAKTOR_RISIKO_OBESITAS_PADA_ANAK_5-15_TAHUN
- [7] E. Retnoningsih and R. Pramudita, "Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python," *Bina Insa. Ict J.*, vol. 7, no. 2, p. 156, 2020, doi: 10.51211/biict.v7i2.1422.
- [8] P. Studi *et al.*, "339072-Implementasi-Decision-Tree-C45-Untuk-Men-Ce21E9Fb," vol. 7, no. November, pp. 5–10, 2018.
- [9] R. Ordila, R. Wahyuni, Y. Irawan, and M. Yulia Sari, "PENERAPAN DATA MINING UNTUK PENGELOMPOKAN DATA REKAM MEDIS PASIEN BERDASARKAN JENIS PENYAKIT DENGAN ALGORITMA CLUSTERING (Studi Kasus : Poli Klinik PT.Inecda)," *J. Ilmu Komput.*, vol. 9, no. 2, pp. 148–153, 2020, doi:

- 10.33060/jik/2020/vol9.iss2.181.
- [10] P. B. N. Setio, D. R. S. Saputro, and Bowo Winarno, "Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5," *Prism. Pros. Semin. Nas. Mat.*, vol. 3, pp. 64–71, 2020.
- [11] Marlina Haiza, Elmayati, Zulus Antoni, and Wijaya Harma Oktafia Lingga, "Penerapan Algoritma Random Forest Dalam Klasifikasi Penjurusan Di SMA Negeri Tugumulyo," *Penerapan Kecerdasan Buatan*, vol. 4, no. 2, pp. 138–143, 2023.
- [12] D. A. Nasution, H. H. Khotimah, and N. Chamidah, "Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN," *Comput. Eng. Sci. Syst. J.*, vol. 4, no. 1, p. 78, 2019, doi: 10.24114/cess.v4i1.11458.
- [13] I. Monika Parapat and M. Tanzil Furqon, "Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 10, pp. 3163–3169, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [14] J. M. H. Y. Al-afghoni *et al.*, "KLASIFIKASI JENIS BENIH KACANG MENGGUNAKAN SMOTE DAN DECISION TREE C4.5," vol. 9, no. 1, pp. 462–469, 2025.
- [15] W. Agustiono *et al.*, "MODEL PENERJEMAH BAHASA ISYARAT HURUF HIJAIYAH," vol. 8, no. 6, pp. 12335–12342, 2024.
- [16] I. Kurniawan, D. C. P. Buani, A. Abdussomad, W. Apriliah, and R. A. Saputra, "Implementasi Algoritma Random Forest Untuk Menentukan Penerima Bantuan Raskin," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 2, pp. 421–428, 2023, doi: 10.25126/jtiik.20231026225.