E-ISSN 3032-601X & P-ISSN 3032-7105
Vol. 1, No. 4, Tahun 2024

MISTER

Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research

Jurnal Penelitian Multidisiplin dalam Ilmu Pengetahuan, Teknologi dan Pendidikan

UNIVERSITAS SERAMBI MEKKAH
KOTA BANDA ACEH
mister@serambimekkah.ac.id

Journal of Multidisciplinary Inquiry in Science Technology and Educational Research

Journal of MISTER

Vol. 1, No. 4, Tahun 2024

Pages: 1938-1949

Analisis Perbandingan Klasifikasi Obesitas Menggunakan Metode K-Nearest Neighbor dan Ensemble Learning

Dian Maharani, Dwitamara Septyana, Caritta Elizabeth, Salsa Pramudhita Agustiardani, Anggraini Puspita Sari

Prodi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Tulungagung, Indonesia

Article in Journal of MISTER

Available at	:	https://jurnal-serambimekkah.org/index.php/mister/index
DOI	:	https://doi.org/10.32672/mister.v1i4.2161

How to Cite this Article

APA	:	Septyana, D., Maharani, D., Elizabeth, C., Agustiardani, S. P., & Sari, A. P. (2024). Analisis Perbandingan Klasifikasi Obesitas Menggunakan Metode K-Nearest Neighbor dan Ensemble Learning . *MISTER: Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research*, *1*(4), 1938 – 1949. https://doi.org/10.32672/mister.v1i4.2161
Others Visit	:	https://jurnal-serambimekkah.org/index.php/mister/index

ISSN 3032-7105 ISSN 3032-601X

Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research
UNIVERSITAS SERAMBI MEKKAH
https://jurnal-serambimekkah.org/index.php/mister

Vol. 1 No. 4, Tahun 2024
e-ISSN 3032-601X & p-ISSN
Doi: 10.32672/mister.v1i4.2161
Hal. 1938-1949

Analisis Perbandingan Klasifikasi Obesitas Menggunakan Metode K-Nearest Neighbor dan Ensemble Learning

Dian Maharani[1*], Dwitamara Septyana[2] , Caritta Elizabeth[3] , Salsa Pramudhita Agustiardani[4] , Anggraini Puspita Sari[5]
Prodi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Tulungagung, Indonesia[1,2,3,4,5]

[*]Email Korespodensi: 22081010184@student.upnjatim.ac.id

Diterima: 15-08-2024 | Disetujui: 16-08-2024 | Diterbitkan: 17-08-2024

ABSTRACT

Obesity is not a problem that is easily avoided by humans because it is related to health and daily lifestyle. Therefore, it is important to contribute to the development of smarter, data-based diagnostic tools that can be utilized by health professionals in identifying individuals at risk of developing obesity problems and designing more personalized and effective interventions. The aim of this research is to develop an obesity prediction model by implementing the K-Nearest Neighbor (KNN) method and Ensemble Learning from Naive Bayes and Random Forest. This research indicates that the KNN algorithm produces an accuracy value of 89.60%, while the Ensemble Learning method from Naive Bayes and Random Forest produces an accuracy value of 89.36%. Of the two methods used, KNN is more effective in detecting the type of obesity based on the dataset used, compared to the Ensemble Learning methods from Naive Bayes and Random Forest. These results indicate that the selection of an appropriate algorithm is critical in the development of a more accurate and efficient obesity classification system.

Keywords: *Obesity; Lifestyle; K-Nearest Neighbor; Ensemble Learning.*

ABSTRAK

Penyakit obesitas adalah suatu permasalahan yang tidak mudah dihindari oleh manusia karena berkaitan dengan kesehatan serta gaya hidup sehari hari. Oleh karena itu, penting untuk memberikan kontribusi dalam pengembangan alat diagnostik yang lebih cerdas dan berbasis data yang dapat dimanfaatkan oleh tenaga kesehatan profesional dalam mengidentifikasi seorang individu yang berisiko mempunyai permasalahan obesitas dan merancang intervensi yang lebih personal dan efektif. Tujuan dari penelitian ini untuk mengembangkan model prediksi obesitas dengan mengimplementasikan metode K-Nearest Neighbor (KNN) dan Ensemble Learning dari Naive Bayes dan Random Forest. Pada penelitian ini mengindikasikan bahwa algoritma KNN menghasilkan nilai akurasi sebesar 89,60%, sedangkan metode Ensemble Learning dari Naive Bayes dan Random Forest menghasilkan nilai akurasi sebesar 89,36%. Dari kedua metode yang digunakan, KNN lebih efektif dalam mendeteksi jenis obesitas berdasarkan dataset yang digunakan, dibandingkan dengan metode Ensemble Learning dari Naive Bayes dan Random Forest. Hasil ini menunjukkan bahwa pemilihan algoritma yang tepat sangat penting dalam pengembangan sistem klasifikasi obesitas yang lebih akurat dan efisien

.**Kata kunci**: Obesitas; Gaya Hidup; K-Nearest Neighbor; Ensemble Learning

e-ISSN3032-601X & p-ISSN3032-7105

PENDAHULUAN

Obesitas merupakan masalah kesehatan global yang terus meningkat dalam satu dekade terakhir [1]. Kelebihan berat badan dan obesitas, menurut Organisasi Kesehatan Dunia (WHO), dapat didefinisikan sebagai penumpukan lemak yang berlebihan di berbagai bagian tubuh [2]. Beberapa peneliti telah berusaha keras untuk mengidentifikasi faktor-faktor yang mempengaruhi timbulnya obesitas sejak dini, bahkan menciptakan alat web seperti kalkulator indeks massa tubuh (BMI) [3]. Faktor yang mempengaruhi obesitas adalah gaya hidup, termasuk pola makan dan aktivitas fisik [4]. Beberapa hal yang dapat menyebabkan obesitas pada remaja antara lain yaitu konsumsi makronutrien yang berlebihan, seringnya mengonsumsi makanan cepat saji, kurangnya kegiatan fisik, kebiasaan makan yang tidak teratur, faktor genetik (keturunan), serta tidak rutin sarapan [5]. Pemahaman yang lebih baik mengenai hubungan antara gaya hidup dan tingkat obesitas dapat membantu dalam merancang intervensi yang lebih tepat dan efektif.

Penelitian ini disusun berdasarkan dengan sumber penelitian sebelumnya sebagai rujukan dan perbandingan. Dalam penelitian [6] tentang Implementasi Algoritma KNN, Random Forest, Naive Bayes dan Decision Tree untuk mengklasifikasikan tingkat obesitas ditunjukkan hasil akurasi Naive Bayes sebesar 65,12%. Merujuk pada penelitian tersebut, kami menggunakan metode Ensemble dengan menggabungkan metode Naive Bayes dan Random Forest untuk meningkatkan kinerja Naive Bayes agar menghasilkan akurasi yang lebih optimal.

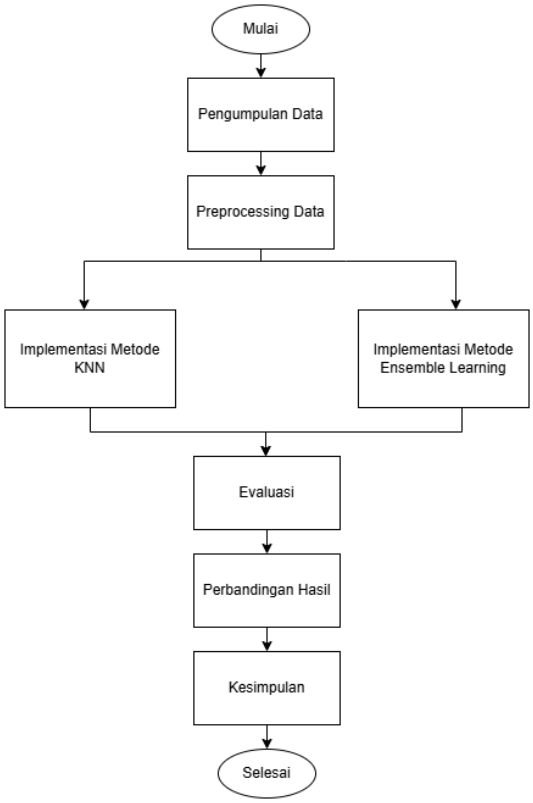
Dalam penelitian ini, kami akan membandingkan kinerja dua metode algoritma, yaitu K-Nearest Neighbor (KNN) dan Ensemble Learning dalam mengklasifikasikan tingkat obesitas berdasarkan gaya hidup. Pemilihan kedua metode ini didasarkan pada karakteristik unik masing-masing. KNN dikenal dengan kesederhanaannya dan efektivitas dalam klasifikasi berdasarkan kedekatan data, sedangkan Ensemble Learning menggabungkan kekuatan model probabilistik dan ansambel yang dapat memberikan performa yang lebih baik dalam mengelola data yang kompleks dan bervariasi. Analisis data performa pengklasifikasian ini merupakan perhitungan akurasi, presisi, recall, dan F1-score [7].

Alasan utama kami melakukan penelitian ini adalah untuk menentukan metode mana yang lebih akurat dan efisien dalam mengklasifikasikan tingkat obesitas berdasarkan gaya hidup. Hasil dari penelitian ini diharapkan bagi para pembaca untuk mengambil peran penting dalam mengembangkan alat diagnostik yang lebih kompatibel dan berbasis data yang dapat dimanfaatkan oleh tenaga kesehatan profesional dalam mengidentifikasi individu yang berisiko obesitas dan merancang intervensi yang lebih personal dan efektif.

METODE PENELITIAN

Penelitian ini dilakukan untuk menguji dan menganalisis kinerja dua metode dalam pengklasifikasian obesitas, yaitu metode *K-Nearest Neighbors (KNN)* dan metode *Ensemble Learning* yang menggabungkan *Naive Bayes* dan *Random Forest.* Penelitian ini memiliki tujuan untuk melakukan perbandingan klasifikasi antara dua metode dari dataset yang sama untuk menemukan metode klasifikasi yang lebih efektif dari keduanya. Tahapan atau alur yang terlaksana dalam penelitian ini ditunjukkan dalam Gambar 1.

e-ISSN3032-601X & p-ISSN3032-7105



Gambar 1. Tahapan Penelitian

A. Pengumpulan data

Data dalam penelitian ini mengambil bagian dari sumber dataset UCI Machine Learning Repository yang dapat diakses pada tautan https://archive.ics.uci.edu/. Selain itu, dilakukan juga studi pustaka dengan mempelajari buku, jurnal referensi, artikel *online*, dan sumber literatur lainnya untuk memperoleh data pendukung yang dapat memperkuat argumentasi, hasil analisis, dan evaluasi. Data dalam penelitian ini diperoleh dari sumber dataset publik yaitu UCI Machine Learning Repository. UCI Machine Learning merupakan forum yang berisi kumpulan *database* yang digunakan oleh komunitas belajar *machine learning*. Penelitian ini menggunakan dataset yang mencakup data untuk estimasi tingkat obesitas yang dikaji melalui kebiasaan makan, gaya hidup, dan kondisi fisik. Dataset ini disumbangkan pada tahun 2019 yang berisi 17 atribut dan 2111 baris data, dengan variabel kelas *NObesity* atau Tingkat Obesitas yang mengklasifikasikan data ke dalam beberapa kategori yaitu : *Insufficient Weight, Normal Weight, Overweight Level* I, *Overweight Level II, Obesity Type I, Obesity Type II, dan Obesity Type III* [8].

B. Preprocessing data

Data yang diperoleh akan diproses agar data sesuai dengan kebutuhan penelitian. Data akan dilakukan pembersihan untuk memastikan data bahwa tidak terdapat data kosong maupun data terduplikasi.

e-ISSN3032-601X & p-ISSN3032-7105

Setelah data dipastikan telah bersih, data tersebut akan diuraikan melalui perbandingan 80 % data training dan 20 % data *testing*. Terdapat 1688 data yang akan dilatih (data *training*) dan 423 data yang akan menjadi data testing. Selain itu, dalam proses *preprocessing data* juga dilakukan transformasi data, yaitu mengubah data non numerik ke dalam data numerik.

C. Implementasi metode

Pengimplementasian kedua metode tersebut dilakukan menggunakan pemrograman python melalui Google Colab dengan *library* tertentu.

1. Metode KNN

   Penerapan metode knn menggunakan library skicit-learn dengan jarak tetangga k dalam KNN yang digunakan yaitu 3. Dalam menghitung ukuran jarak antara data *testing* dan *training* dapat menggunakan rumus *manhattan distance* dan *euclidean distance*. Pada penelitian ini, penulis menggunakan *Euclidean Distance* untuk proses pengukuran jarak yang menentukan jarak terdekat antara titik data baru dengan titik titik data dalam training set [9] .

$$di = \sqrt{\sum} (x1 - x2\ )\ p\ 2\ i{=}1 \qquad (1)$$

   Keterangan :
   d = jarak antara dua titik
   i = variabel data yang mewakili setiap dimensi
   p = dimensi data
   x1 = sampel data
   x2 = data yang diuji

2. Metode *Naive Bayes*

   Metode Naïve bayes adalah suatu metode algoritma dalam proses pengklasifikasian yang melewati beberapa fase sehingga mendapatkan hasil yang akurat. Naive Bayes mempunyai metode yang sederhana dan dapat menampung jumlah data yang cukup besar membantu untuk menghitung probabilitas melalui cara penjumlahan dari frekuensi dan nilai lain dari kelas kelas yang berasal dari dataset [10].

$$P(Ci|X) = P(X|Ci).\ P(Ci)/P(X) \qquad (3)$$

   Keterangan :
   X = sampel data yang kelasnya tidak diketahui
   Ci = hipotesis (asumsi) bahwa X adalah data kelas tertentu
   P(Ci) = probabilitas dari hipotesis C
   P(Ci) = probabilitas dari data sampel yang diamati
   P(Ci|X) = probabilitas kondisi data X pada hipotesis

3. Metode *Random Forest*

   Penerapan metode Random Forest menggunakan algoritma Machine Learning yang berguna dalam membantu penyelesaian masalah yang melibatkan klasifikasi dan regresi pada suatu studi kasus [11]. Metode ini menggabungkan sejumlah pohon keputusan (*decision tree*), di mana setiap pohon

e-ISSN3032-601X & p-ISSN3032-7105

dibentuk dari subset acak dari data pelatihan. Setiap pohon dalam hutan beroperasi secara independen dan menggunakan distribusi yang sama, dan hasil akhir prediksi ditentukan berdasarkan suara mayoritas dari semua pohon dalam hutan tersebut. Dengan hasil kinerja yang baik serta memiliki struktur yang sederhana membuat *Random Forest* sering dijadikan metode dalam kasus klasifikasi dan regresi [12]. Dalam algoritma *Random Forest* yang acak ini diperlukan adanya penyesuaian dengan cara menghitung *Mean Squared Error* atau yang sering disebut MSE. MSE akan menunjukkan seberapa dekat garis regresi dengan kumpulan titik data yang sebenarnya [13].

$$MSE = 1/N \sum N\, i = 1\ (fi - yi)2 \quad (2)$$

Keterangan :
N = jumlah data
fi = nilai yang diprediksi oleh model
yi = nilai aktual (nilai sebenarnya) untuk data i

4. *Ensemble VotingClassifer*
Penerapan metode ensemble merupakan teknik yang menggabungkan dua model prediksi untuk mencapai akurasi yang lebih optimal dibandingkan hanya menggunakan satu metode saja. Penelitian pada studi kasus yang mempunyai permasalahan kelas kompleks seperti ini diperlukan adanya model ensemble karena dapat membantu dalam peningkatan kinerja algoritma secara efektif. Proses penggabungan metode *Naive Bayes* dan *Random Forest*, dilakukan dengan menggunakan algoritma VotingClassifier. Implementasi metode Naive Bayes menggunakan model distribusi multinomial dengan memasukkan algoritma MultinomialNB dalam library scikit-learn, sedangkan metode Random Forest menggunakan algoritma RandomForestClassifier.

D. Evaluasi Model
Selanjutnya, pengujian kinerja kedua metode tersebut, baik metode KNN dan metode *ensemble* akan dievaluasi menggunakan indikator evaluasi. Indikator evaluasi tersebut, yaitu akurasi, presisi, *recall*, dan *f1-score*. Rumus perhitungan indikator evaluasi, sebagai berikut [14]:

- Akurasi
Akurasi merupakan evaluasi untuk menghitung banyaknya prediksi yang bernilai benar dari keseluruhan jumlah prediksi.
$$Akurasi = \frac{Jumlah\ prediksi\ benar}{Total\ prediksi} \quad (4)$$
- Presisi
Presisi merupakan nilai sensitivitas dari evaluasi untuk menghitung rasio data yang diprediksi positif untuk meminimalkan jumlah *false positive* [15].
$$Presisi = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (5)$$
- *Recall*
Recall merupakan tingkat keberhasilan dari evaluasi untuk menghitung rasio data yang diprediksi benar untuk meminimalkan *false negative* [15].
$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (6)$$

e-ISSN3032-601X & p-ISSN3032-7105

- *F1-score*
  F1-score adalah nilai rata-rata untuk keseimbangan antara presisi dan *recall*.
  $$F1 - score = \frac{Presisi\ x\ Recall}{Presisi + Recall} \quad (7)$$

E. Perbandingan Hasil
Hasil dari pengujian akan dianalisis dan dibandingkan untuk menentukan metode yang lebih efektif.

F. Kesimpulan
Pada tahap ini dilakukan setelah dilakukannya pengujian dan analisis. Laporan yang telah dibuat sesuai dengan aturan yang telah ditentukan akan diambil kesimpulan dan hasil dari penelitian ini yaitu menentukan metode yang lebih efektif untuk pengklasifikasian obesitas berdasarkan gaya hidup.

HASIL DAN PEMBAHASAN

A. Preprocessing Data

Pada penelitian ini, preprocessing data dilakukan untuk mempersiapkan dataset yang akan digunakan dalam analisis lebih lanjut [16]. Preprocessing data dimulai dengan memastikan bahwa tidak terdapat data yang bernilai kosong seperti pada Gambar 2. Hal ini dilakukan agar data yang diolah dapat menghasilkan nilai yang lebih akurat. Setelah data dipastikan bersih dan tidak memiliki nilai kosong, maka data akan dipisahkan ke dalam data training dan data testing.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2111 entries, 0 to 2110
Data columns (total 17 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Gender                          2111 non-null   object
 1   Age                             2111 non-null   float64
 2   Height                          2111 non-null   float64
 3   Weight                          2111 non-null   float64
 4   family_history_with_overweight  2111 non-null   object
 5   FAVC                            2111 non-null   object
 6   FCVC                            2111 non-null   float64
 7   NCP                             2111 non-null   float64
 8   CAEC                            2111 non-null   object
 9   SMOKE                           2111 non-null   object
 10  CH2O                            2111 non-null   float64
 11  SCC                             2111 non-null   object
 12  FAF                             2111 non-null   float64
 13  TUE                             2111 non-null   float64
 14  CALC                            2111 non-null   object
 15  MTRANS                          2111 non-null   object
 16  NObeyesdad                      2111 non-null   object
dtypes: float64(8), object(9)
memory usage: 280.5+ KB
```

Gambar 2. Informasi bahwa dataset tidak terdapat nilai kosong

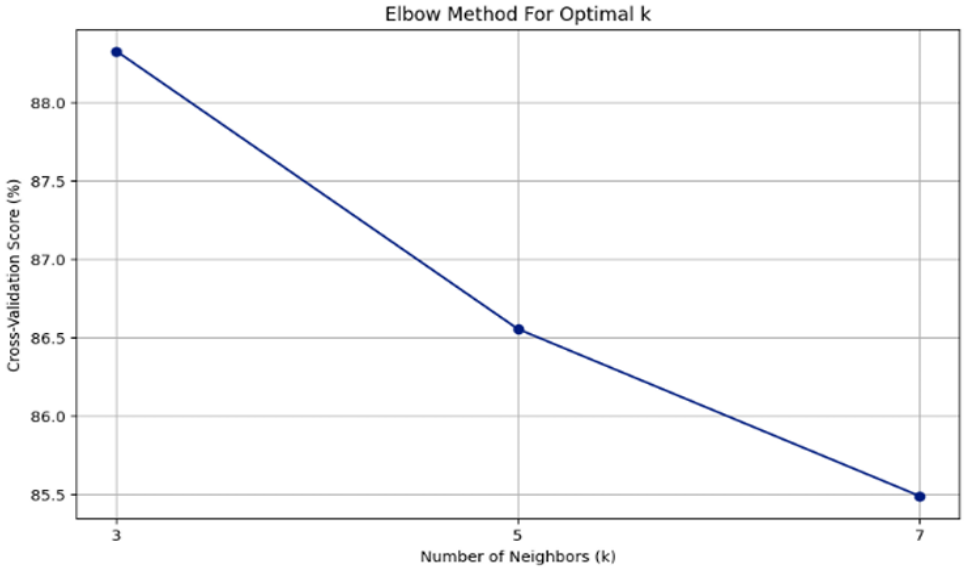
B. Implementasi Metode

Pada tahap ini, diterapkan algoritma metode pada data yang telah dilakukan preprocessing data. Pada setiap metode akan dilakukan cross validation untuk melihat performa setiap metode.

e-ISSN3032-601X & p-ISSN3032-7105

1) Metode KNN

Dalam penelitian ini, data akan dilatih dengan nilai k tetangga terdekatnya, yaitu 3, 5, dan 7. Nilai k optimal adalah nilai yang memberikan performa terbaik pada model tanpa menyebabkan overfitting atau underfitting. Nilai-nilai ini dipilih untuk mengetahui bagaimana perubahan k mempengaruhi performa model.



Gambar 3. Grafik Elbow Metode KNN dengan nilai k = 3, 5, 7

Berdasarkan pada Gambar 3, ditunjukkan hasil dari nilai k = 3, 5, dan 7 dengan masing-masing persentase variasi sebesar 89,60%, 88,42%, dan 87,23%. Kurva elbow menunjukkan bahwa terjadi penurunan signifikan dalam persentase dari nilai k =3 ke nilai k=5 dan 7. Perhitungan ini mengindikasikan bahwa nilai k parameter 3 merupakan nilai yang paling optimal dengan persentase 89,60%. Setiap data yang dilatih akan dilakukan pengujian dengan memberikan penilaian benar dan salah pada klasifikasi obesitas sebenarnya dengan klasifikasi hasil prediksi menggunakan metode KNN.

Positions of Correct Predictions:

	Actual	Predicted	Correct
0	Insufficient_Weight	Insufficient_Weight	True
1	Normal_Weight	Normal_Weight	True
2	Overweight_Level_II	Overweight_Level_II	True
3	Obesity_Type_III	Obesity_Type_III	True
4	Obesity_Type_II	Obesity_Type_II	True
..	...	...	...
417	Normal_Weight	Normal_Weight	True
418	Obesity_Type_II	Obesity_Type_II	True
419	Overweight_Level_II	Overweight_Level_II	True
420	Obesity_Type_II	Obesity_Type_II	True
422	Obesity_Type_I	Obesity_Type_I	True

[379 rows x 3 columns]

Number of Correct Predictions: 379

Positions of False Predictions:

	Actual	Predicted	Correct
11	Normal_Weight	Insufficient_Weight	False
14	Normal_Weight	Insufficient_Weight	False
22	Obesity_Type_II	Overweight_Level_II	False
35	Normal_Weight	Overweight_Level_I	False
49	Normal_Weight	Overweight_Level_II	False
..	...	...	...
409	Obesity_Type_I	Obesity_Type_II	False
410	Obesity_Type_I	Overweight_Level_II	False
413	Normal_Weight	Overweight_Level_I	False
414	Insufficient_Weight	Normal_Weight	False
421	Normal_Weight	Insufficient_Weight	False

[44 rows x 3 columns]

Number of False Predictions: 44

Gambar 4. Hasil Prediksi pada Metode KNN

e-ISSN3032-601X & p-ISSN3032-7105

Pada Gambar 4, menunjukkan hasil prediksi yang bernilai salah dan letak indeks data tersebut. Jumlah hasil prediksi yang bernilai salah yaitu 44 data dan hasil prediksi yang bernilai benar yaitu 379 data dari 423 data testing. Jumlah inilah yang nanti akan dihitung hasil akurasinya.

2) Metode Ensemble Learning (Naive Bayes dan Random Forest)

Pada penelitian ini, setiap data yang dilatih akan dilakukan pengujian dengan memberikan penilaian benar dan salah pada klasifikasi obesitas sebenarnya dengan klasifikasi hasil prediksi.

Positions of True Predictions (Sample):

	Actual	Predicted	Correct
572	Insufficient_Weight	Insufficient_Weight	True
370	Normal_Weight	Normal_Weight	True
1002	Overweight_Level_II	Overweight_Level_II	True
1837	Obesity_Type_III	Obesity_Type_III	True
1724	Obesity_Type_II	Obesity_Type_II	True
0	...	...	...
1548	Obesity_Type_II	Obesity_Type_II	True
1161	Overweight_Level_II	Overweight_Level_II	True
1537	Obesity_Type_II	Obesity_Type_II	True
497	Normal_Weight	Normal_Weight	True
1396	Obesity_Type_I	Obesity_Type_I	True

Number of True Predictions: 378

Positions of False Predictions (Sample):

	Actual	Predicted	Correct
1305	Obesity_Type_I	Overweight_Level_II	False
77	Overweight_Level_II	Obesity_Type_I	False
344	Obesity_Type_III	Obesity_Type_II	False
28	Normal_Weight	Overweight_Level_I	False
127	Normal_Weight	Overweight_Level_I	False
0	...	...	...
1376	Obesity_Type_I	Obesity_Type_II	False
210	Obesity_Type_II	Obesity_Type_I	False
231	Normal_Weight	Overweight_Level_I	False
332	Normal_Weight	Overweight_Level_II	False
140	Insufficient_Weight	Normal_Weight	False

Number of False Predictions: 45

Gambar 5. Hasil Prediksi pada Metode Ensemble

Pada Gambar 5 merupakan hasil prediksi obesitas menggunakan metode Ensemble Learning yang bernilai salah beserta dengan letak indeks data tersebut. Hasil prediksi yang bernilai salah terdapat 45 data dan yang benar terdapat 378 data dari 423 data testing.

C. Evaluasi Model

Setiap model yang sudah dihitung prediksinya akan dilakukan perhitungan akurasinya untuk mengukur performa model dan tingkat kesesuaian model dengan data.

1) Metode KNN

Berdasarkan Gambar 6, dapat dilihat hasil akurasi pada metode KNN ini memiliki tingkat akurasi yang tinggi yaitu 89,60%. Hasil presisi pada beberapa kelas, seperti Obesity Type I, Obesity Type II, dan Obesity Type III memiliki presisi yang tinggi yaitu > 0,90. Sedangkan pada Insufficient Weight, Normal Weight, Overweight Level I, dan Overweight Level II memiliki presisi yang lebih rendah < 0,90 yang menunjukkan bahwa model memiliki false positive yang lebih banyak. Hasil recall memiliki tingkat yang cukup tinggi yaitu dengan rata-rata pada seluruh kelasnya sebesar 89,36% sehingga model ini cukup baik dalam meminimalkan false negative. Untuk hasil f1-score memiliki tingkat yang cukup tinggi dengan rata-rata 88,93%.

e-ISSN3032-601X & p-ISSN3032-7105

Classification Report:

	precision	recall	f1-score	support
Insufficient_Weight	0.8667	0.9630	0.9123	54
Normal_Weight	0.8974	0.6034	0.7216	58
Obesity_Type_I	0.9559	0.9286	0.9420	70
Obesity_Type_II	0.9344	0.9500	0.9421	60
Obesity_Type_III	0.9848	1.0000	0.9924	65
Overweight_Level_I	0.7826	0.9310	0.8504	58
Overweight_Level_II	0.8500	0.8793	0.8644	58
accuracy			0.8960	423
macro avg	0.8960	0.8936	0.8893	423
weighted avg	0.8996	0.8960	0.8926	423

Gambar 6. Classification Report Metode KNN

2) Metode Ensemble Learning

Pada Gambar 7, dapat dilihat hasil akurasi pada metode Ensemble yaitu 89,36%. Dalam model ini diperoleh hasil presisi pada beberapa kelas, seperti Insufficient Weight, Obesity Type II, dan Obesity Type III memiliki presisi yang tinggi yaitu > 0,90. Sedangkan pada Obesity Type I, Normal Weight, Overweight Level I, dan Overweight Level II memiliki presisi yang lebih rendah < 0,90 dan menunjukkan bahwa model pada kelas tersebut memiliki false positive yang lebih banyak. Hasil recall memiliki tingkat cukup tinggi yaitu dengan rata-rata sebesar 89,26% sehingga model ini cukup baik dalam meminimalkan false negative. Untuk hasil f1-score memiliki rata-rata 89,25%

Classification Report:

	precision	recall	f1-score	support
Insufficient_Weight	0.9630	0.9630	0.9630	54
Normal_Weight	0.7705	0.8103	0.7899	58
Obesity_Type_I	0.8986	0.8857	0.8921	70
Obesity_Type_II	0.9355	0.9667	0.9508	60
Obesity_Type_III	0.9846	0.9846	0.9846	65
Overweight_Level_I	0.8136	0.8276	0.8205	58
Overweight_Level_II	0.8868	0.8103	0.8468	58
accuracy			0.8936	423
macro avg	0.8932	0.8926	0.8925	423
weighted avg	0.8944	0.8936	0.8937	423

e-ISSN3032-601X & p-ISSN3032-7105

Gambar 7. Classification Report Metode Ensemble

D. Perbandingan Hasil

Pada tabel 1 menunjukkan perbandingan hasil akurasi pada metode KNN dan Metode Ensemble Learning (Naive Bayes dan Random Forest). Pada tabel tersebut ditunjukkan bahwa pada metode KNN memiliki akurasi lebih besar dibandingkan metode Ensemble Learning dari beberapa aspek, seperti hasil rata-rata precision, recall, maupun hasil akurasi pada data testing-nya. Metode KNN memiliki akurasi 89,60% dengan nilai precision 89,60%, recall 89,36%, dan f1-score 88,93%. Sehingga pada penelitian ini dapat ditentukan metode yang lebih efektif untuk klasifikasi obesitas berdasarkan gaya hidup yaitu metode KNN.

TABEL 1
PERBANDINGAN AKURASI DARI IMPLEMENTASI KEDUA METODE

Metode	*Precision*	*Recall*	*f1-Score*	Akurasi
K-Nearest Neighbor (KNN)	89,60%	89,36%	88,93%	89,60%
Ensemble Learning (*Naive Bayes* dan *Random Forest*)	89,32%	89,26%	89,25%	89,36%

KESIMPULAN

Hasil penelitian ini membandingkan dua metode dalam mengklasifikasikan obesitas berdasarkan gaya hidup, yaitu metode KNN (K-Nearest Neighbor) dan metode Ensemble Learning yang menggabungkan Naive Bayes dan Random Forest. Dari hasil penelitian, diperoleh bahwa metode KNN memiliki akurasi yang lebih tinggi dibandingkan metode Ensemble Learning, yaitu 89,60%, dengan nilai precision sebesar 89,60%, recall 89,36%, dan F1-score 88,93%. Berdasarkan hasil tersebut, dapat disimpulkan bahwa metode KNN lebih cocok dengan dataset yang diuji, sehingga memiliki performa yang lebih efektif untuk pengklasifikasian obesitas berdasarkan gaya hidup dibandingkan dengan metode Ensemble Learning. Metode KNN juga lebih sederhana dibandingkan dengan metode Ensemble. Dengan demikian, KNN dapat diandalkan sebagai metode yang lebih efektif dan efisien untuk mengklasifikasikan tingkat obesitas berdasarkan gaya hidup. Temuan ini dapat membantu dalam pengembangan alat diagnostik yang lebih akurat dan berbasis data, yang bisa digunakan oleh profesional kesehatan untuk mengidentifikasi individu yang berisiko obesitas dan merancang intervensi yang lebih personal dan efektif. Untuk dapat mengoptimalkan metode yang diterapkan ke depannya, perlu dilakukan analisis lebih mendalam mengenai karakteristik dataset.

DAFTAR PUSTAKA

Sudargo, T., et al. (2024, May 12). *Pola Makan dan Obesitas*. Universitas Gadjah Mada Press. https://ugmpress.ugm.ac.id/userfiles/product/daftar_isi/Pola_Makan_dan_Obesitas.pdf

World Health Organization. (2021). *Obesity and overweight*. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

e-ISSN3032-601X & p-ISSN3032-7105

World Health Organization. (2021). *Body mass index calculator*. https://www.emro.who.int/nutrition/information-resources/bmi-calculator.html

Hanafi, R., Badrah, S., & Noviasty, R. (2021). Pola makan, aktivitas fisik, dan genetik mempengaruhi kejadian obesitas pada remaja. *Jurnal Kesehatan Metro Sai Wawai*, *14*(2), 120-129. https://doi.org/10.26630/jkm.v14i2.2665

Telisa, I., Hartati, Y., & Haripamilu, A. D. (2020). Faktor risiko terjadinya obesitas pada remaja SMA. *Faletheah Health Journal, 7*(3), 124-129. https://media.neliti.com/media/publications/338096-risk-factors-of-obesity-among-adolescent-6bcf1882.pdf

Rifaldi, A., Anang, L., & Satrio, I. D. (2022). Implementasi algoritma k-nearest neighbor (KNN), random forest, Naive Bayes dan decision tree untuk mengklasifikasikan tingkat obesitas. *RG Journal of Applied Science*, 2(2). https://doi.org/10.3140/RG.2.2.29928.03841

Ul Hassan, C. A., Khan, M. S., & Shah, M. A. (2018, September 1). Comparison of machine learning algorithms in data classification. *IEEE Xplore*. https://ieeexplore.ieee.org/abstract/document/8748995/

Palechor, F. M., & Manotas, A. D. (2019). Estimation of obesity levels based on eating habits and physical condition. *UCI Machine Learning Repository*. https://archive.ics.uci.edu/dataset/544/estimation+of+obesity+levels+based+on+eating+habits+and+physical+condition

Cholil, S. R., Handayani, T., Prathivi, R., & Adrianta, T. (2021). Implementasi algoritma klasifikasi k-nearest neighbor (KNN) untuk klasifikasi seleksi penerima beasiswa. *IJCIT : Indonesian Journal of Computer and Information Technology*, *6*(2), 118-127. https://ejournal.bsi.ac.id/ejurnal/index.php/ijcit/article/view/5418

Putro, H. F., Vulandari, R. T., Saptomo, W. L., & Saptomo, W. L. (2020). Penerapan metode naive Bayes untuk klasifikasi pelanggan. *TIKomSin: Jurnal Teknologi Informasi dan Komunikasi Sinar Nusantara*, *8*(2). https://p3m.sinus.ac.id/jurnal/index.php/TIKomSin/article/view/50415

Setiawan, M. J., Nugroho, B., & Sari, A. P. (2023). Klasifikasi penyakit daun tanaman menggunakan algoritma CNN dan random forest: Classification leaf diseases. *Teknologi: Jurnal Ilmiah Sistem Informasi*, 25(2), 197-227. https://journal.unipdu.ac.id/index.php/teknologi/article/view/3739

Biau, G., & Scornet, E. (2016). A random forest guided tour. *Statistics Surveys, 10*, 197-227. https://doi.org/10.1007/s11749-016-0481-7

Ardianasyah, M. (2016). Model ensemble algoritma random forest dalam klasifikasi penyakit paru-paru untuk meningkatkan akurasi. *Jurnal Sains dan Teknologi, 2*(2), 32-38. https://doi.org/10.37476/smartteknol.v2i2.4407

Yuliansyah, H., Imaniat, P. A. P., & Yuliansyah, H. (2022). Predicting students graduate on time using C4.5 algorithm. *Jurnal Informatika Systems Engineering and Business Intelligence*, 8(1), 1-10. https://doi.org/10.20473/jisebi.v7i.67-73

Azhari, M., Situmorang, Z., & Rosnelly, R. (2021). Prediksi klasifikasi pada algoritma C4.5, random forest, SVM dan naive Bayes. *Jurnal Media Informatika Budidarma, 5*(2), 640-647. https://doi.org/10.30865/mib.v5i2.2937

e-ISSN3032-601X & p-ISSN3032-7105

Nurrahman, H., & Susanto, M. (2020). Optimasi penentuan jumlah klaster menggunakan algoritma K-means. *JURIKOM: Jurnal Riset Komputer*, 10(2), 99-107. https://doi.org/10.30865/jurikom.v10i2.5087