IMPLEMENTASI DECISION TREE UNTUK KLASIKASI OBESITAS

Ahlul Khikam, Nanda Martyan Anggadimas, Muhammad Udin
Teknik Informatika, Universitas Merdeka Pasuruan
Jl.Ir.H.Juanda No.68, Tapaan, Kec. Bugul Kidul, Kota Pasuruan, Jawa Timur 67129
ahlulkhikam25@.gmail.com

ABSTRAK

Angka obesitas telah meningkat secara signifikan dalam beberapa dekade terakhir di seluruh dunia. Aspek seperti pola makan buruk, kurangnya olahraga, faktor genetik, lingkungan, dan masalah psikologis bisa mempengaruhi berkembangnya obesitas. Obesitas dapat menyebabkan berbagai masalah kesehatan seperti penyakit jantung, diabetes, dan gangguan pernapasan, serta bisa berdampak buruk pada kualitas hidup seseorang. Tujuan penelitian untuk mengimplementasi metode *Decision Tree* untuk klasifikasi obesitas. *Decision Tree* merupakan metode yang sangat efektif dalam melakukan klasifikasi. Dataset dalam penelitian ini, dilakukan dengan mengambil data *Obesity Levels* dari website *Arcihive.ics.uci.edu.* Dataset yang digunakan terdapat 2111 data dan 17 kolom. Hasil dari prediksi menggunakan metode *Decision Tree* mendapatkan nilai *accuracy* sebesar 94%, *pecision* sebesar 94%, *recall* sebesar 94%, Dan *F1-score* sebesar 94%. Meskipun hasil ini cukup memuaskan, penelitian ini juga menunjukkan adanya peluang untuk pengembangan lebih lanjut, seperti mempertimbangkan aspek – aspek lain seperti menambahkan jumlah dan variasi data, teknik *preprocessing one-hot encoding*, dan menggunakan metode klasifikasi lain untuk membandingkan kinerja model guna mendapatkan hasil yang lebih optimal dalam klasifikasi obesitas.

Kata kunci : *Klasifikasi, Prediksi, Obesitas, Decision Tree.*

1. PENDAHULUAN

Obesitas telah menjadi perhatian dunia karena dampaknya yang signifikan pada kesehatan masyarakat. Menurut *World Health Organization* (WHO), angka obesitas telah meningkat secara signifikan dalam beberapa dekade terakhir di seluruh dunia. Aspek seperti pola makan buruk, kurangnya olahraga, faktor genetik, lingkungan, dan masalah psikologis bisa mempengaruhi berkembangnya obesitas [10]. Obesitas tidak hanya terjadi pada orang dewasa, dan remaja, tetapi juga pada anak-anak [3].

Berdasarkan data terbaru dari WHO pada tahun 2022, sebanyak 2,5 miliar orang dewasa berusia 18 tahun ke atas mengalami kelebihan berat badan dan 890 juta termasuk menderita obesitas. Jika tingkat pertumbuhan ini bertahan, proporsi orang yang masuk dalam kategori obesitas akan meningkat pada tahun 2030. Di Indonesia sendiri, menurut hasil Survei Kesehatan Indonesia 2023, angka obesitas di Indonesia meningkat menjadi 23% [8].

Obesitas terjadi karena asupan kalori yang lebih banyak daripada aktivitas membakar kalori, atau dengan kata lain, kalori yang dikonsumsi tidak seimbang dengan kalori yang digunakan [1]. Tubuh membutuhkan kalori untuk bertahan hidup dan menjalani aktivitas sehari-hari, namun untuk mempertahankan berat badan diperlukan keseimbangan antara energi yang dikonsumsi dan energi yang dibakar [9].

Aktivitas fisik berperan penting dalam kejadian obesitas dan regulasi tubuh karena dengan aktivitas fisik bisa mengakibatkan pengeluaran energi harian, meningkatkan pembakaran lemak, menurunkan kadar leptin, meningkatkan sensitivitas reseptor leptin, serta bisa meningkatkan massa otot dan mengurangi massa lemak.

Kemajuan dalam bidang teknologi dan transportasi juga dapat menyebabkan berkembangnya obesitas, seperti berkurangnya ruang untuk beraktivitas dan berolahraga, serta terbatasnya fasilitas untuk bergerak menyebabkan seseorang memiliki lebih sedikit kesempatan untuk membakar energi. Selain itu, kemajuan dalam dunia kuliner, serta pertumbuhan restoran cepat saji, turut berkontribusi.

Obesitas dapat menyebabkan berbagai masalah kesehatan seperti penyakit jantung, diabetes, dan gangguan pernapasan, serta bisa berdampak buruk pada kualitas hidup seseorang. Menurut data, kematian yang diakibatkan oleh obesitas tercatat sekitar 30 per 100.000 penduduk Indonesia [5].

Dari berbagai penyebab terjadinya resiko obesitas. Penelitian ini bertujuan untuk mengembangan model untuk prediksi klasifikasi obesitas dengan menganalisis faktor-faktor risiko yang berkontribusi terhadap terjadinya kondisi tersebut. Penelitian ini dapat memberikan wawasan yang lebih mendalam tentang bagaimana pola makan, aktivitas fisik, kebiasaan sehari-hari, dan faktor genetik dapat meningkatkan kemungkinan seseorang mengalami obesitas. Informasi yang diperoleh dari penelitian ini diharapkan dapat digunakan untuk pencegahan yang efektif sehingga dapat membantu mengurangi peningkatan angka obesitas.

Pada penelitian sebelumnya telah dilakukan membahas dan mengklasifikasi dengan beberapa metode dengan menggunakan data obesitas. hasil dari metode *Decision Tree* mendapatkan hasil akurasi yang baik dari pada beberapa penelitian menggunakan metode yang lain. Karena *Decision Tree* merupakan

metode yang sangat efektif dalam melakukan klasifikasi. Sehingga pada peneliltian ini peneliti ingin mengimplementasi metode *Decision Tree* untuk klasifikasi obesitas. Hasil dari penelitian ini dapat mengetahui tingkat akurasi metode *Decision Tree* dalam klasifikasi obesitas.

2. TINJAUAN PUSTAKA

2.1. Obesitas

Obesitas adalah kondisi kelebihan berat badan yang dapat memengaruhi kesehatan dan kualitas hidup seseorang. Hal ini bisa meningkatkan risiko penyakit seperti diabetes tipe 2, masalah jantung, dan beberapa jenis kanker. Selain itu, Obesitas juga bisa memengaruhi kesehatan tulang, sistem reproduksi, dan kemampuan untuk tidur atau bergerak dengan nyaman. Untuk mendiagnosis kelebihan berat badan, biasanya dilakukan pengukuran berat badan dan tinggi badan seseorang, kemudian dihitung indeks massa tubuh (IMT): berat badan (kg)/tinggi badan^2 (m^2).

2.2. Klasifikasi

Klasifikasi merupakan metode pengelohan data di mana dataset yang tersimpan di analisis untuk menghasilkan aturan yang membagi dataset kedala m kelompok kelas tertentu. Dalam klasifikasi, model atau algoritma dilatih menggunakan data yang sudah diberi label (dikenal sebagai data training) sehingga dapat memprediksi kelas dari data baru yang belum diketahui. Algoritma klasifikasi, seperti *Decision Tree*, *Support Vector Machine* (SVM), dan *K-Kearest Neighbors* (KNN), bekerja dengan mempelajari pola dari data pelatihan dan menggunakan pola tersebut untuk mengkategorikan data baru secara akurat. Klasifikasi banyak digunakan dalam berbagai aplikasi, termasuk deteksi spam, diagnosis medis, dan pengenalan wajah.

2.3. *Python*

Python adalah bahasa pemrograman tingkat tinggi yang dirancang untuk memberikan sintaks yang mudah dibaca dan dipahami, membuatnya ideal bagi pemula maupun profesional. Diciptakan oleh Guido van Rossum dan dirilis pertama kali pada tahun 1991, *Python* mendukung berbagai paradigma pemrograman, termasuk pemrograman berorientasi objek, prosedural, dan fungsional. Salah satu kekuatan *Python* terletak pada ekosistem pustakanya yang luas dan komunitas pengembang yang aktif, yang menjadikannya pilihan utama untuk berbagai aplikasi, mulai dari pengembangan *web*, analisis data, hingga kecerdasan buatan. Popularitas *Python* terus meningkat, menjadikannya salah satu bahasa pemrograman paling digunakan di dunia.

2.4. *Decision Tree*

Decision Tree adalah struktur diagram yang menyerupai pohon, digunakan untuk memecah data besar menjadi kelompok yang lebih kecil dengan menerapkan aturan keputusan. Setiap kelompok kemudian dikelompokkan berdasarkan kesamaan yang ada. Dalam pohon keputusan, atribut dibagi menjadi *node* untuk klasifikasi sesuai dengan label yang telah ditentukan. Algoritma pembentukan pohon keputusan terdiri dari dua langkah yaitu pencarian *root* dan pembuatan cabang. Pencarian *root* dilakukan untuk mempertimbangkan setiap atribut dari data latih yang telah diberi label.

Decision Tree merupakan teknik klasifikasi yang membangun model berdasarkan informasi dari *entropy*. Model ini dibuat dengan menghitung nilai *entropy* dan *gain* dari sampel data yang digunakan. *Entropy* adalah indikator statistik yang mengukur tingkat ketidakpastian dalam data, sedangkan *gain* adalah ukuran yang menilai seberapa efektif suatu atribut dalam membagi contoh pelatihan sesuai dengan kelas target [6]. Berikut ini adalah langkah langkah untuk menghitung nilai *entropy*:

$$Entropy\ (s) = \sum_{i=1}^{n} - pi * \log_2 pi \quad (1)$$

Keterangan:
S = Himpunan kasus
A = Fitur
N = Jumlah Partisi S
Pi = Proporsi Dari Si terhadap S

Selanjutnya adalah langkah langkah untuk menghitung nilai *Gain*:

$$Gain\ (SA) = Entropy\ (s) - \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy \quad (2)$$

Keterangan :
S = Himpunan kasus
A = Atribut
N = Jumlah partisi atribut A
|Si| = Jumlah atribut pada partisi ke-i
|S| = Jumlah kasus pada S

2.5. *Confusion Matrix*

Confusion Matrix merupakan metode yang digunakan untuk mengevaluasi sejauh mana model pengelompokan mampu mengenali data dari berbagai tingkatan. Kinerja model klasifikasi dapat dinilai berdasarkan Tingkat *Accurraci*, *Precison*, *Recall*, dan *F1-score* [2]. Rumus untuk *Confusion Matrix* sebagai berikut:

Tabel 1. *Confusion Matrix*

Confusion Matrix	***Predictive Positive***	***Predictive Negative***
Actual Positive	TP	FP
Actual Negative	TN	FN

$$Accurracy = \frac{TP+TN}{TP+FP+TN+FN} \quad \ldots\ldots (3)$$

$$Precision = \frac{TP}{TP+FP} \quad \ldots\ldots (4)$$

$$Recall = \frac{TP}{TP+FN} \quad \ldots\ldots (5)$$

$$F1\text{-}scorel = \frac{TP}{TP+FN} \quad \ldots\ldots (6)$$

Keterangan :

TP = Nilai prediksi positif dengan nilai sebenarnya positif.

TN = Nilai prediksi positif dengan nilai sebenarnya negative.

FP = Nilai prediksi negatif dengan nilai sebenarnya positif.

FN = Nilai prediksi negatif dengan nilai sebenarnya negative.

2.6. Penelitian Terkait

Penelitian lain yang berjudul "Analisis Prediksi Level Obesitas Menggunakan Perbandingan Algoritma *Machine learning* dan *Deep learning*". Tujuan Penelitian untuk membandingkan tingkat akurasi antara algoritma *Machine learning* dengan *Deep learning*. Hasil dari analisa algoritma *Machine learning* menghasilkan akurasi dari *Random Forest* sebesar 96,37%, *Decision Tree classifier* sebesar 88,33%, *Logistic Regression* sebesar 73,66%, *Naïve Bayes* sebesar 56% dan KNN sebesar 88,96%, Sedangkan *Deep learning* didapati akurasi sebesar 86,05% [3].

Penelitian lain yang berjudul "Implementasi Algoritma *Decision Tree* dan *Support Vector Machine* (SVM) untuk Prediksi Risiko *Stunting* pada Keluarga". Tujuan penelitian yaitu untuk mengetahui algoritma mana yang cocok untuk Prediksi Risiko *Stunting* pada Keluarga. Hasil penelitian menunjukkan algoritma *Decision Tree* mendapatkan nilai akurasi sebesar 96,15%, Presisi nilai *recall* Tidak sebesar 92.06%, serta Ya sebesar 97.34%, dan nilai presisi Tidak sebesar 90.99%, serta Ya sebesar 97.68%. sedangkan untuk Algoritma SVM mendapatkan nilai akurasi sebesar 62.48%, nilai *recall* Tidak sebesar 99.12%, serta Ya sebesar 51.80%, dan nilai presisi Tidak sebesar 37.49%, serta Ya sebesar 99.51% [11].

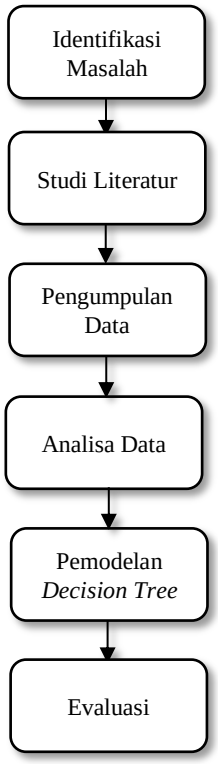
Penelitian selanjutnya yang berjudul "Prediksi Tingkat Obesitas Menggunakan *Neural Network*: Pendekatan Klasifikasi Biner". Tujuan penelitian yaitu untuk mengembangkan model prediksi dengan jaringan saraf tiruan (*Neural Network*) untuk memprediksi tingkat obesitas pada setiap individu berdasarkan kebiasaan makan dan kondisi fisik. Hasil yang telah diuji menghasilkan akurasi sebesar: 0.9684, presisi sebesar: 0.9669, dan *recall* sebesar: 0.9915 [10].

Penelitian lain yang berjudul "Analisis Perbandingan KNN, SVM, *Decision Tree* dan *Regresi Logistik* Untuk Klasifikasi Obesitas Multi Kelas" Tujuan penelitian adalah untuk membandingkan kinerja empat algoritma *Machine learning*, KNN, SVM, *Decision Tree*, dan *Regresi Logistik* dalam mengklasifikasikan dari beberapa tingkatan obesitas. Hasil penelitian menunjukkan *Decision Tree* mendapatkan hasil akurasi sebesar 99.3%, presisi sebesar 0.97-1.00, dan *recall* sebesar 0.98-1.00, Metode KNN mendapatkan hasil akurasi sebesar 99.0%, presisi sebesar 0.98-1.00, dan *recall* sebesar 0.98-1.00, Hasil Algoritma *Regresi Logistik* mendapatkan akurasi sebesar 98%, presisi sebesar 0.95-1.00, Dan *recall* sebesar 0.95-1.00, Dan Metode SVM hasil akurasi sebesar 96.6%, presisi sebesar 0.90-0.99, Dan *recall* sebesar 0.94-1 [4].

Penelitian lain yang berjudul "Implementasi Metode *Decision Tree* pada Prediksi Penyakit *Diabetes Melitus* Tipe 2". Tujuan Penelitian adalah untuk mengimplementasikan dan membangun sebuah model prediksi penyakit diabetes. Hasil dari implementasi metode *Decision Tree* mendapatkan nilai akurasi sebesar sebesar 92% , presisi sebesar 0,92, Dan *recall* sebesar 0,915 [7].

3. METODE PENELITIAN

Dalam metode penelitian ini terdapat beberapa alur yang dilakukan yaitu identifikasi masalah, studi literatur, pengumpulan data, analisa data, pemodelan menggunakan *Decision Tree*, dan evaluasi model. Gambar 1 dibawah ini merupakan diagram alur metode penelitian:



Gambar 1. Diagram Alur Metode Penelitian

3.1. Identifikasi Masalah

Identifikasi masalah pada penelitian ini adalah obesitas yang merupakan kondisi kelebihan berat badan yang dapat memengaruhi kesehatan dan kualitas hidup seseorang serta dapat menyebabkan beberapa penyakit hingga kematian. Tingkat penderita obesitas di dunia terus meningkat. Oleh karena itu, penelitian ini bertujuan untuk mengidentifikasi dan mengklasifikasi tingkat obesitas dengan menganalisis faktor-faktor risiko yang berkontribusi terhadap terjadinya kondisi tersebut sehingga dapat membantu dalam pencegahan untuk mengurangi peningkatan angka obesitas.

3.2. Studi Literatur

Dalam tahap berikutnya pada penilatian ini, dilakukan eksplorasi literatur dengan melakukan pencarian refrensi dari beberapa jurnal maupun situs internet lainnya yang relavan dengan topik dan judul yang telah ditetapkan yaitu " Implementasi medote *Decision Tree* untuk klasifikasi Obesitas".

3.3. Pengumpulan Data

Pada tahap pengumpulan data dalam penilitian ini, dilakukan dengan cara mengambil data *Obesity Levels* dari website *Arcihive.ics.uci.edu*. Dataset yang digunakan adalah data tingkat obesitas pada individu dari negara Meksiko, Peru, dan Kolombia berdasarkan pola makan dan keadaan fisiknya dengan 77% data dihasilkan secara sintesis menggunakan alat Weka dan Filter SMOTE, 23% data dikumpulkan dari pengguna platform web. Dataset tersebut diambil karena memiliki fitur yang berupa faktor yang menyebabkan obesitas dan memiliki label dengan tingkatan obesitas. Dataset yang diperoleh terdapat 2111 data dan 17 kolom,. Berikut ini adalah beberapa tabel dari contoh perolehan data:

Tabel 2. Data Dengan Kolom Fitur *Gender, Age, Height,* dan *Weight*

No	Gender	Age	Height	Weight
1.	Female	21	1,62	64
2.	Female	21	1,52	56
..	…	…	…	…
2111.	Female	23,664709	1,738836	133,472641

Tabel 3. Data dengan fitur Kolom Family_history _ with_overweigth dan FAVC

No	Family_history_with_overweigth	FAVC
1.	Yes	No
2.	Yes	No
..	…	…
2111.	Yes	Yes

Tabel 4. Data dengan fitur Kolom *FCVC, CAEC, SMOKE,* dan *NCP*

No	FCVC	CAEC	SMOKE	NCP
1.	2	Sometimes	No	3
2.	3	Sometimes	Yes	3
..	…	…	…	…
2111.	3	Sometimes	No	3

Tabel 5. Data dengan fitur Kolom *FAF* dan *CALC*

No	FAF	CALC
1.	0	No
2.	3	Sometimes
..	…	…
2111.	1,026452	Sometimes

Tabel 6. Data dengan fitur Kolom *MITRANS* dan *NObeyesdad*

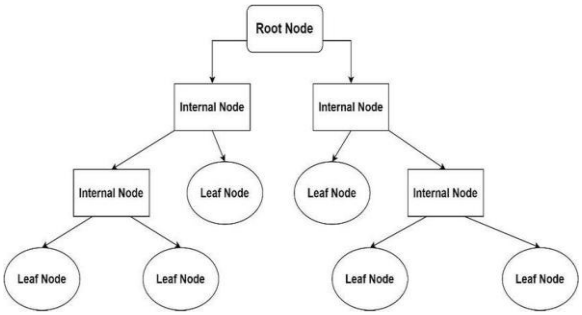
	MITRANS	NObeyesdad
1.	Public_Transportation	Normal_Weight
2.	Public_Transportation	Normal_Weight
..	…	…
2111.	Public_Transportation	Obesity_Ttpe_III

3.4. Analisa Data

Pada tahap analisa data pada penelitian ini dilakukan dengan memeriksa data untuk mendapatkan informasi data yang akan digunakan dan mengubah data untuk memastikan kualitas data yang digunakan serta memahami pola yang ada di dalam data. Langkah-langkah dalam analisa data yaitu melihat informasi dataset seperti tipe data untuk setiap kolom, pengecekan *missing values*, dan melihat kolerasi data menggunakan grafik.

3.5. Pemodelan *Decision Tree*

Setelah analisa data dilakukan pemodelan *Decision Tree* yang digunakan untuk klasifikasi obesitas. *Decision Tree* terdiri dari simpul (node) dan rusuk (edge). Simpul pada sebuah pohon dibedakan menjadi tiga, yaitu simpul akar *root node*, simpul percabangan atau *internal node* dan simpul daun atau *leaf node.*



Gambar 2. Simpul *Decision Tree*

Root Node merupakan node yang paling atas dimana pada node ini tidak ada input dan mempunyai ouput lebih dari satu. internal Node merupakan node percabangan dimana pada node ini hanya terdapat satu input. Leaf Node merupakan node akhir dimana pada node ini hanya terdapat satu input dan tidak mempunyai output**.** Pohon keputusan akan dibangun berdasarkan pemilihan atribut yang paling relavan untuk setiap node, dengan tujuan mengelompokkan data secara efektif ke tingkatan obesitas yang telah ditentukan.

3.6. Evaluasi

Setelah proses pemodelan menggunakan *Decision Tree* dilakukan evaluasi model. Dalam evaluasi model pada penelitian ini metode yang digunakan adalah *confusion matrix*. *Confusion matrix* merupakan proses perhitungan *Accuracy, Precison, Recall,* dan *F1-score.*

4. HASIL DAN PEMBAHASAN

4.1. *Import Dataset*

Import Library adalah proses memasukkan library ke dalam program yang akan dibuat. Sedangkan *import dataset* adalah proses untuk mengumpulkan dan memberikan informasi secara

sistematis. berikut ini adalah langkah-langkah dalam proses melakukan *import library* dan *Input* dataset:

```
# Mengimpor library yang diperlukan
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import classification_report, confusion_matrix

# Input dataset
url = '/content/sample_data/ObesityDataSet_raw_and_data_sinthetic.csv'
data = pd.read_csv(url)
```

Gambar 3. Proses *Import Library* dan Input Dataset

Langkah pertama adalah *import library* yang dibutuhkan, seperti *Pandas, Matplotib*, dan *Seaborn*. *Libary Pandas* (*import pandas as pd*) digunakan untuk memanipulasi dan menganlisa data. *Libary Matpolib* (*import pandas as pd*) dan *Libary Seaborn* (*import pandas as pd*) digunakan untuk visualisasi data. *Labelencording* (*From sklearn.prepocessing import LabelEncording*) adalah fungsi untuk mengubah data kategori ke numerik. *train_test_split* (*From sklearn.prepocessing import train_test_split*) adalah fungsi untuk membagi data latih dan data uji. *DecisionTreeClassiifer* (*From sklearn.prepocessing import DecisionTreeClassiifer*) adalah fungsi untuk pemodelan klasifikasi menggunkan metode *Decision Tree*. *classification_report dan confusion matrix* (*From sklearn.prepocessing import classification_report, confusion matrix*) adalah fungsi untuk menampilkan hasil klasifikasi dan *confusion matrix*. setelah mengimport *libary* dan beberapa fungsi yang digunakan untuk pemodelan *Decision Tree* adalah melakukan *input* dataset dalam bentuk *csv* yaitu *"/content/sample_data/ObesityDataSet_raw_and_data_sinthetic.csv"*

	Gender	Age	Height	Weight	family_history_with_overweight	FAVC	FCVC	NCP	CAEC	SMOKE	CH2O	SCC	FAF	TUE	CALC	MTRANS	NObeyesdad
0	Female	21.0	1.62	64.0	yes	no	2.0	3.0	Sometimes	no	2.0	no	0.0	1.0	no	Public_Transportation	Normal_Weight
1	Female	21.0	1.52	56.0	yes	no	3.0	3.0	Sometimes	yes	3.0	yes	3.0	0.0	Sometimes	Public_Transportation	Normal_Weight
2	Male	23.0	1.80	77.0	yes	no	2.0	3.0	Sometimes	no	2.0	no	2.0	1.0	Frequently	Public_Transportation	Normal_Weight
3	Male	27.0	1.80	87.0	no	no	3.0	3.0	Sometimes	no	2.0	no	2.0	0.0	Frequently	Walking	Overweight_Level_I
4	Male	22.0	1.78	89.8	no	no	2.0	1.0	Sometimes	no	2.0	no	0.0	0.0	Sometimes	Public_Transportation	Overweight_Level_II

Gambar 4. Baris Pertama Dataset

Gambar 4 tersebut merupakan tampilan 5 baris pertama datset yang telah import. Dataset tersebut memiliki 16 kolom fitur yaitu *Gender, Age, Height, Weight, Family_history_with_overweight, FAVC, FCVC, NCP, CAEC, SMOKE, CH2O ,SCC, FAF, TUE, CALC, MITRANS* dan 1 label yaitu *NObeyesdad.*

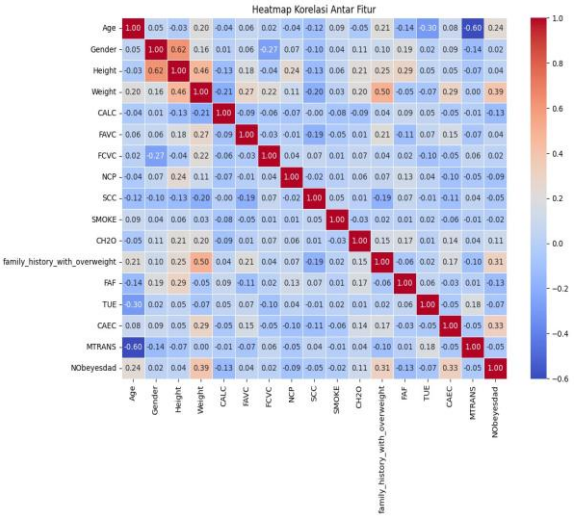
4.2. Analisa Data

Analisa data adalah proses memahami dataset dengan melihat informasi dataset seperti tipe data untuk setiap kolom dan distribusi data, memeriksa *missing values*, melihat kolerasi data menggunakan grafik. Berikut ini adalah langkah-langkah dalam eksplorasi data:

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

Langkah pertama dalam eksplorasi data adalah menampilkan informasi tentang dataset yang menujukkan terdapat 2111 baris tampa *missing value* dengan 16 fitur dan kolom label diberi nama *NObeyesdad*.



Gambar 6. Kolerasi Antar Fitur

Pada kolerasi antar fitur di gambar 6 tersebut menunjukkan kolom fitur *Weight, Family_history_with_overweight,* dan *CAEC* memiliki kolerasi paling tinggi dengan kolom label *NObeyesdad* yang mendapatkan nilai 0,39, 0,31, dan 0,33. Selanjutnya ada beberapa fitur yang memiliki kolerasi paling tinggi seperti fitur Gender dan Height dengan nilai 0,62 , fitur Height dan Weight dengan nilai 0,46 , serta fitur Weight dan *Family_history_with_overweight* dengan nilai 0,50.

4.3. *Preprocessing Data*

Pada *preprocessing data* adalah proses pengubahan data mentah menjadi data yang siap digunakan untuk pemodelan. Reprocesing data pada penelitian ini hanya dilakukan label *encoding* data atau

pengubahan data kategori ke numerik, karena data yang telah diperoleh tidak ada *missing values.* Berikut ini adalah gambar 6 hasil perubahan data kategori ke numerik:

	Age	Gender	Height	Weight	CALC	FAVC	FCVC	NCP	SCC	SMOKE	CH2O
0	21.0	0	1.62	64.0	3	0	2.0	3.0	0	0	2.0
1	21.0	0	1.52	56.0	2	0	3.0	3.0	1	1	3.0
2	23.0	1	1.80	77.0	1	0	2.0	3.0	0	0	2.0
3	27.0	1	1.80	87.0	1	0	3.0	3.0	0	0	2.0
4	22.0	1	1.78	89.8	2	0	2.0	1.0	0	0	2.0

	family_history_with_overweight	FAF	TUE	CAEC	MTRANS	NObeyesdad
0	1	0.0	1.0	2	3	1
1	1	3.0	0.0	2	3	1
2	1	2.0	1.0	2	3	1
3	0	2.0	0.0	2	4	5
4	0	0.0	0.0	2	3	6

Gambar 7. Hasil Perubahan Data Ketegori ke Numerik

Pada gambar 7 terdapat beberapa data yang dilakukan pengubahan data kategoei ke numerik seperti kolom fitur *Gender, CALC, FAVC, SCC, SMOKE, Family_history_with_overweight, TUE, CAEC, MITRANS,* dan label *NObeyesdad.* Katregori *Female* dan *Male* pada kolom fitur *Gender* diubah ke angka 1 dan 2. Katregori *Yes* dan *No* pada kolom fitur *FAVC, SCC, SMOKE, Family_history_with_overweight,* dan *TUE* diubah ke angka 1 dan 2. Katregori *Always, Frequently, Sometimes* dan *No* pada kolom fitur *CALC* dan *CAEC* diubah ke angka 1, 2, 3 dan 4. Katregori Yes dan No pada kolom fitur *FAVC, SCC, SMOKE, Family_history_with_overweight,* dan *TUE* diubah ke angka 1 dan 2. Kategori *Automobile, Motorbike, Publik_transportaion* dan *Walking* pada kolom fitur *CALC* dan *CAEC* diubah ke angka 1, 2, 3 dan 4.

4.4. Pemodelan *Decision Tree*

Dataset yang telah diproses kemudian dilakukan pembagian data training dan data uji dengan rasio 80% data latih 20% data uji. Setelah dilakukan pembagian data model klasifikasi *Decision Tree*. Pembagian data dapat dilihat pada gambar 8.

```
Jumlah data latih: 1688
Jumlah data uji: 423
```

Gambar 8. Pembagian Data Latih dan Data Uji

Setelah pembagian data dilakukan pemodelan klasifikasi menggunakan metode *Decision Tree*. Hasil dari klasifikasi menunjukkan nilai *accuracy* sebesar 94%, *pecision* sebesar 94%, *recall* sebesar 94%, Dan *F1-score* sebesar 94%. Hasil prediksi lanjutan dari setiap tingkatan seperti *Insffucient_Weight* mendapatkan nilai *pecision* sebesar 92%, *recall* sebesar 96%, Dan *F1-score* sebesar 94%. Hasil prediksi untuk tingkatan *Normal_Weight* dengan prediksi 62 data mendapatkan nilai *pecision* sebesar 88%, *recall* sebesar 85%, Dan *F1-score* sebesar 87%. Hasil prediksi untuk tingkatan *Overweight_Level_I* mendapatkan nilai *pecision* sebesar 96%, *recall* sebesar 94%, Dan *F1-score* sebesar 95%. Hasil prediksi untuk tingkatan *Overweight_Level_II* mendapatkan nilai *pecision* sebesar 95%, *recall* sebesar 95%, Dan *F1-score* sebesar 95%. Hasil prediksi untuk tingkatan *Obesity_Type_II* mendapatkan nilai *pecision* sebesar 100%, *recall* sebesar 100%, Dan *F1-score* sebesar 100%. Hasil prediksi untuk tingkatan *Obesity_Type_II* mendapatkan nilai *pecision* sebesar 89%, *recall* sebesar 91%, Dan *F1-score* sebesar 90%. Hasil prediksi untuk tingkatan *Obesity_Type_II* mendapatkan nilai *pecision* sebesar 96%, *recall* sebesar 96%, dan *F1-score* sebesar 96%. Hasil klasifikasi menggunakan metode *Decision Tree* terlihat pada gambar 9.

	precision	recall	f1-score	support
0	0.92	0.96	0.94	56
1	0.87	0.85	0.86	62
2	0.96	0.94	0.95	78
3	0.95	0.95	0.95	58
4	1.00	1.00	1.00	63
5	0.89	0.91	0.90	56
6	0.98	0.96	0.97	50
accuracy			0.94	423
macro avg	0.94	0.94	0.94	423
weighted avg	0.94	0.94	0.94	423

Gambar 9. Hasil klasifikasi *Decision Tree*

4.6. Evaluasi

Setelah pemodelan menggunakan *Decision Tree* dilakukan evaluasi menggunakan *confusion matrix*. Evaluasi menggunakan *confusion matrix* ditampilkan pada Gambar 10.

```
[[54  2  0  0  0  0  0]
 [ 5 53  0  0  0  4  0]
 [ 0  1 73  3  0  0  1]
 [ 0  0  3 55  0  0  0]
 [ 0  0  0  0 63  0  0]
 [ 0  5  0  0  0 51  0]
 [ 0  0  0  0  0  2 48]]
```

Gambar 10. Evaluasi *Cofusion matrix*

Evaluasi *Cofusion matrix* menunjukkan hasil prediksi dari setiap tingkatan obesitas dan *Confusion matrix.* Prediksi dari label *Insffucient_Weight* mendapatkan 54 data *positif* dan 5 data *negative*. Prediksi dari label *Normal_Weight* mendapatkan 53 data *positif* dan 6 data *negative*. Prediksi dari label *Overweight_Level_I* mendapatkan 73 data *positif* dan 3 data *negative*. Prediksi dari label *Overweight_Level_II* mendapatkan 55 data *positif* dan 3 data *negative.* Prediksi dari label *Obesity_Type_I* mendapatkan 63 data *positif* dan tidak ada data *negative.* Prediksi dari label *Obesity_Type_II* I mendapatkan 51 data *positif* dan 6 data data *negative*. Prediksi dari label *Obesity_Type_III* mendapatkan 48 data *positif* dan 1 data data *negative.*

Hasil dari prediksi menggunakan metode *Decision Tree* mendapatkan nilai *accuracy* sebesar 94%. Hasil keseluruhan dari prediksi menggunakan meode *Decision Tree* mendapatkan.

5. KESIMPULAN DAN SARAN

Hasil penelitian ini menunjukkan bahwa penerapan metode *Decision Tree* memberikan kontribusi signifikan dalam klasifikasi obesitas, dengan data yang diambil dari website *Archive.ics.uci.edu*. Dataset yang digunakan terdapat 2111 data dan 17 kolom. Hasil dari prediksi menggunakan metode *Decision Tree* mendapatkan nilai *accuracy* sebesar 94%, *pecision* sebesar 94%, *recall* sebesar 94%, Dan *F1-score* sebesar 94%. Meskipun hasil ini cukup memuaskan, penelitian ini juga menunjukkan adanya peluang untuk pengembangan lebih lanjut yaitu dengan menambahkan variasi dan jumlah data yang digunakan pada model *decision tree*, serta penggunaan teknik *preprocessing* yang lebih beragam guna meningkatkan akurasi model. Mengingat pada penelitian ini menggunakan label *encoding*, disarankan pada penelitian selanjutnya untuk mempertimbangkan penggunaan *One-hot encoding* terutama jika terdapat kolom fitur kategorikal dengan banyak kategori. Dan juga mempertimbangkan penggunaan metode lain untuk membandingkan kinerja model dalam mengklasifikasi obesitas. Diharapkan dengan mempertimbangkan aspek – aspek tersebut penelitian selanjutnya dapat menghasilkan model yang lebih optimal dalam klasifikasi obesitas.

DAFTAR PUSTAKA

[1] J. V. Wie and M. Siddik, "PENERAPAN METODE NAÏVE BAYES DALAM MENGKLASIFIKASI TINGKAT OBESITAS PADA PRIA," JOISIE Journal Of Information System And Informatics Engineering, vol. 6, no. Desember, pp. 69–77, 2022, [Online]. Available: https://www.kaggle.com/.

[2] D. Fatmawati, W. Trisnawati, Y. Jumaryadi, and G. Triyono, "KLIK: Kajian Ilmiah Informatika dan Komputer Klasifikasi Tingkat Kepuasan Penggunaan Layanan Teknologi Informasi Menggunakan Decision Tree," Media Online), vol. 3, no. 6, pp. 1056–1062, 2023, doi: 10.30865/klik.v3i6.803.

[3] L. Setiyani, A. Nur Indahsari, and R. Roestam, "Analisis Prediksi Level Obesitas Menggunakan Perbandingan Algoritma Machine Learning dan Deep Learning," Jurnal Teknologi Rekayasa), vol. 8, no. 1, pp. 139–146, 2023, doi: 10.31544/jtera.v8.i1.2023.139-146.

[4] S. A. Utiarahman, A. Mulawati, and M. Pratama, "KLIK: Kajian Ilmiah Informatika dan Komputer Analisis Perbandingan KNN, SVM, Decision Tree dan Regresi Logistik Untuk Klasifikasi Obesitas Multi Kelas," Media Online), vol. 4, no. 6, pp. 3137–3146, 2024, doi: 10.30865/klik.v4i6.1871.

[5] T. Hidayatulloh and L. Yusuf, "KLASIFIKASI TIPE BERAT TUBUH MENGGUNAKAN METODE SUPPORT VECTOR MACHINE," INTI Nusa Mandiri, vol. 18, no. 1, pp. 71–77, Aug. 2023, doi: 10.33480/inti.v18i1.4254.

[6] M. Iqbal, S. Miskiyah, S. L. Sham, S. Anwar, and M. H. Fuad, "PERBANDINGAN METODE DECISION TREE DAN NAIVE BAYES PADA TINGKAT PENJUALAN MINUMAN KOPI DI KOPI PAWON NUSANTARA," Jurnal INSAN (Journal of Information Systems Management Innovation, vol. 4, no. 1, 2024.

[7] M. F. Aditya, A. Pramuntadi, D. P. Wijaya, and Y. Wicaksono, "Implementasi Metode Decision Tree pada Prediksi Penyakit Diabetes Melitus Tipe 2," MALCOM: Indonesian Journal of Machine Learning and Computer Science, vol. 4, no. 3, pp. 1104–1110, Jul. 2024, doi: 10.57152/malcom.v4i3.1284.

[8] S. A. Utiarahman, A. Mulawati, and M. Pratama, "KLIK: Kajian Ilmiah Informatika dan Komputer Analisis Perbandingan KNN, SVM, Decision Tree dan Regresi Logistik Untuk Klasifikasi Obesitas Multi Kelas," Media Online), vol. 4, no. 6, pp. 3137–3146, 2024, doi: 10.30865/klik.v4i6.1871.

[9] V. Fresinsya and A. Qoiriah, "Meal-Planning Berbasis Status Gizi dengan Metode Klasifikasi K-nearest neighbors (KNN) untuk Pasien Obesitas," Journal of Informatics and Computer Science, vol. 06, 2024.

[10] D. N. FITRIANI, "Prediksi PREDIKSI TINGKAT OBESITAS MENGGUNAKAN NEURAL NETWORK: PENDEKATAN KLASIFIKASI BINER," PARAMETER: Jurnal Matematika, Statistika dan Terapannya, vol. 3, no. 01, pp. 85–92, Apr. 2024, doi: 10.30598/parameterv3i01pp85-92.

[11] I. Putri, Y. Syarif, P. Jayadi, F. Arrazak, and F. N. Salisah, "Implementasi Algoritma Decision Tree dan Support Vector Machine (SVM) untuk Prediksi Risiko Stunting pada Keluarga," MALCOM: Indonesian Journal of Machine Learning and Computer Science, vol. 3, no. 2, pp. 349–357, Feb. 2024, doi: 10.57152/malcom.v3i2.1228.