

# Penggunaan Data Mining untuk Prediksi tingkat Obesitas di Meksiko Menggunakan Metode Random Forest

<sup>1</sup>Ersa Dwi Nur Aini, <sup>2</sup>Reka Ainul Khasanah, <sup>3</sup>Aidina Ristyawan, <sup>4</sup>Erna Diniati

<sup>1-4</sup>Universitas Nusantara PGRI Kediri

**Diterima:**

10 Juni 2024

**Revisi:**

10 Juli 2024

**Terbit:**

1 Agustus 2024

[ersadwinuraini@gmail.com](mailto:ersadwinuraini@gmail.com), [rekaainulkhasanah@gmail.com](mailto:rekaainulkhasanah@gmail.com),  
[aidinaristi@unpkediri.ac.id](mailto:aidinaristi@unpkediri.ac.id), [ernadaniati@unpkediri.ac.id](mailto:ernadaniati@unpkediri.ac.id)

**Abstrak**— Metode KDD (*Knowledge Discovery in Databases*) digunakan untuk pemahaman masalah, pemilihan data yang relevan, dan preprocessing sebelum analisis data. Selanjutnya, algoritma KNN digunakan untuk klasifikasi obesitas berdasarkan faktor risiko. Algoritma ini mengelompokkan data baru berdasarkan tetangga terdekatnya dalam ruang fitur. Hasil evaluasi menunjukkan bahwa model *Random Forest* mencapai akurasi keseluruhan sebesar 93.38%. Model ini terutama unggul dalam memprediksi kelas *Obesity\_Type\_I* dan *Obesity\_Type\_III*, meskipun kinerjanya sedikit menurun pada kelas *Overweight\_Level\_I*. Meskipun demikian, *Random Forest* terbukti sebagai alat yang efisien dalam memproyeksikan tingkat obesitas, menunjukkan potensi dalam upaya pencegahan dan penanganan obesitas di Meksiko.

**Kata Kunci**— obesitas, K-Nearest Neighbors, Random, Forest KDD, Klasifikasi

**Abstract**— *The use of data mining techniques to predict obesity levels in Mexico using the K-Nearest Neighbors (KNN) and Random Forest algorithms is explored in this study. The data utilized is sourced from a dataset containing information about individuals from Mexico, Peru, and Colombia, including obesity risk factors and the levels of obesity they experience. The aim of this research is to develop an accurate predictive model to aid in addressing the obesity issue in Mexico. The KDD (Knowledge Discovery in Databases) method is employed for problem understanding, relevant data selection, and preprocessing prior to data analysis. Subsequently, the KNN algorithm is utilized for obesity classification based on risk factors. This algorithm groups new data based on their nearest neighbors in feature space. Evaluation results indicate that the Random Forest model achieves an overall accuracy of 93.38%. This model excels particularly in predicting Obesity\_Type\_I and Obesity\_Type\_III classes, albeit with a slight decrease in performance for the Overweight\_Level\_I class. Nonetheless, Random Forest proves to be an efficient tool in projecting obesity levels, demonstrating potential in obesity prevention and management efforts in Mexico.*

**Keywords**—obesity, K-Nearest Neighbors, Rndom Forest, KDD, Classification

This is an open access article under the CC BY-SA License.



---

## Penulis Korespondensi:

Erna Daniati  
Sistem Informasi,  
Universitas Nusantara PGRI Kediri  
Email: [ernadaniati@unpkediri.ac.id](mailto:ernadaniati@unpkediri.ac.id),  
ID Orcid: [<https://orcid.org/0009-0008-9471-4421>]  
Handphone: 081335242202

## I. PENDAHULUAN

Obesitas merupakan masalah kesehatan global yang besar dengan konsekuensi yang signifikan[1].Di Meksiko, prevalensi obesitas meningkat pesat dalam beberapa dekade terakhir, menjadikannya masalah kesehatan masyarakat yang utama. Di Meksiko, angka obesitas mencapai 74,5% pada wanita dewasa dan 69,1% pada pria dewasa (Organisasi Kesehatan Dunia, 2023). Hal ini memicu berbagai masalah kesehatan, seperti diabetes tipe 2, penyakit jantung, dan stroke.

Karena itu, menjadi penting untuk mengembangkan metode yang efisien dalam memprediksi dan mencegah obesitas. Data mining merupakan teknik yang berguna untuk menganalisis kumpulan data besar dan rumit guna mengenali pola dan anomali. Salah satu alat yang sering digunakan dalam prediksi adalah algoritma *K-Nearest Neighbors* (KNN).[2]

Algoritma Naive Bayes adalah algoritma pembelajaran mesin untuk masalah klasifikasi, yang terutama digunakan untuk klasifikasi teks yang berisi kumpulan data pelatihan berdimensi tinggi.[3] Decision tree adalah metode yang sangat efektif dan terkenal untuk klasifikasi dan prediksi. Teknik ini mengubah data yang sangat kompleks menjadi pohon keputusan yang menggambarkan aturan-aturan.[4]

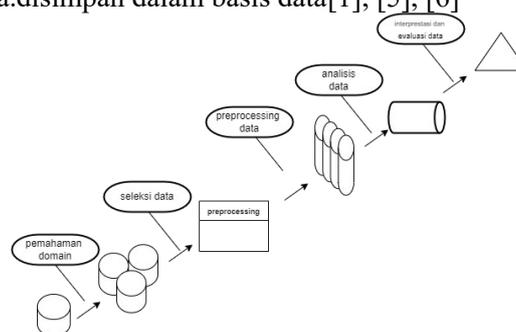
Tujuan penelitian ini adalah untuk memanfaatkan teknik data mining, terutama algoritma KNN dan *Random Forest* , untuk melakukan prediksi tingkat obesitas di Meksiko dengan menggunakan *Phyton*. Data yang digunakan dalam penelitian berasal dari dataset <https://www.kaggle.com/datasets/fatemehmehrpavar/obesity-levels>. Dataset ini mencakup informasi tentang individu dari Meksiko, Peru, dan Kolombia, termasuk faktor risiko obesitas serta tingkat obesitas yang mereka alami.

Penelitian ini bertujuan untuk memberikan kontribusi signifikan dalam penanggulangan masalah obesitas di Meksiko melalui pengembangan model prediksi yang akurat. Dengan model ini, individu yang rentan terhadap obesitas dapat diidentifikasi dan intervensi pencegahan yang tepat dapat diterapkan. Artikel ini menyajikan hasil penelitian tentang penggunaan algoritma *K-Nearest Neighbors* (KNN) untuk klasifikasi obesitas, dengan data yang diambil dari [sumber ini](#) dan memanfaatkan fungsi-fungsi yang tersedia di Python.[5]

## II. METODE

### 2.1. KDD

KDD (*Knowledge Discovery in Databases*) atau penemuan pengetahuan dalam basis data adalah proses mengekstraksi informasi yang tersembunyi atau tidak jelas dari informasi yang disimpan dalam basis data. disimpan dalam basis data[1], [5], [6]



*gambar 1. Pendekatan KDD*

Pendekatan KDD terdiri dari lima tahap utama yaitu :

**1. Pemahaman Domain**

Pada tahap awal Knowledge Discovery in Databases (KDD), terdapat pemahaman mendalam tentang masalah obesitas di Meksiko yang mencakup identifikasi faktor risiko seperti usia, jenis kelamin, indeks massa tubuh (BMI), pola makan, gaya hidup, dan genetika. Selain itu, perlu ditentukan jenis data yang diperlukan untuk penelitian, termasuk data demografi, data medis, dan gaya hidup. Penelitian ini bertujuan untuk mengembangkan model prediktif obesitas, mengidentifikasi kelompok risiko, serta memahami faktor penyebab obesitas.

**2. Seleksi Data**

Pemilihan data yang relevan untuk penelitian. Dalam penelitian ini, data yang digunakan berasal dari dataset <https://www.kaggle.com/datasets/fatemehmehrpavar/obesity-levels>. Dataset ini berisi informasi tentang individu dari Meksiko, Peru, dan Kolombia, termasuk faktor risiko obesitas dan tingkat obesitas mereka.

**3. Preprocessing Data**

Pembersihan dan persiapan data untuk analisis. Dalam penelitian ini, dilakukan dengan penamaan ulang kolom. Data juga dinormalisasi untuk memastikan bahwa semua fitur memiliki skala yang sama.[7], [8]

**4. Analisis Data**

Penelitian ini menggunakan teknik data mining untuk mengidentifikasi pola dan wawasan dalam data, dengan algoritma K-Nearest Neighbors (KNN) dan Random Forest untuk memprediksi tingkat obesitas individu berdasarkan faktor risiko mereka. KNN mengelompokkan data baru berdasarkan jumlah dan jarak tetangga terdekatnya, dimana pemilihan nilai k mempengaruhi akurasi prediksi. Algoritma ini menghitung jarak antara data uji ( $x'$ ,  $y'$ ) dan data latih ( $x'$ ,  $x$ ), mengurutkan hasilnya dari yang terkecil hingga terbesar, lalu memilih k data terdekat untuk menentukan kelas data uji berdasarkan kelas tetangga terdekatnya.

**5. Interpretasi dan Evaluasi**

Interpretasi hasil analisis data dan evaluasi kinerja model prediksi. Dalam penelitian ini, akurasi model prediksi dievaluasi menggunakan metrik seperti akurasi, presisi, dan recall.

2.2.Data Mining

Data mining adalah rangkaian proses untuk mengekstraksi nilai tambah dari sebuah database dengan cara menemukan pola-pola yang sebelumnya tidak diketahui. Tujuannya adalah mengubah data tersebut menjadi informasi yang lebih berharga dengan mengidentifikasi pola informasi penting atau menarik yang terdapat dalam database.[9], [10], [11]

2.3. Klasifikasi

Klasifikasi adalah salah satu model penambangan data. Model klasifikasi adalah suatu teknik prediksi data yang membuat prediksi tentang nilai data yang ditemukan hasilnya berasal dari berbagai kumpulan data. Tujuan dari model ini adalah untuk memprediksi nilai suatu variabel yang tidak diketahui berdasarkan variabel lain yang telah ditentukan.[12], [13]

2.4. Prediksi

Prediksi adalah proses memperkirakan kejadian yang paling mungkin terjadi di masa depan berdasarkan informasi yang tersedia saat ini, dengan tujuan mengurangi kesalahan (perbedaan antara kejadian aktual dan hasil prediksi). Dalam konteks ilmu data, prediksi sering kali dibuat menggunakan algoritme yang dilatih berdasarkan data historis atau data yang dikumpulkan sebelumnya.[14], [15]

### 2.5. Random Forest

Random Forest adalah algoritma pembelajaran ensemble yang menggabungkan beberapa pohon keputusan untuk meningkatkan akurasi prediksi. [16], [17] Dalam konteks ini, model Random Forest diterapkan untuk memprediksi tingkat obesitas individu berdasarkan berbagai fitur dalam dataset.

## III. HASIL DAN PEMBAHASAN

### 3.1. Input Data

Penelitian ini menggunakan data umum dari situs Kaggle [www.kaggle.com/datasets/obesity-levels](http://www.kaggle.com/datasets/obesity-levels) yang terdiri dari 2.111 rekaman dengan 16 atribut dan 1 label. Data label 'Nobeyesded' mencakup 7 kelas: berat badan tidak mencukupi, berat badan normal, obesitas tipe 1, obesitas tipe 2, obesitas tipe 3, kelebihan berat badan level 1, dan kelebihan berat badan level 2. Atribut-atribut yang dimiliki oleh data dapat dikelompokkan ke dalam 3 hal, yaitu:

1. Atribut yang berhubungan dengan kebiasaan makan . table 1

*Table 1. Atribut kebiasaan makan*

Atribut	Keterangan
FAVC	kebiasaan orang mengkonsumsi makanan ber kalori tinggi..
FCVC	frekuensi orang mengkonsumsi sayuran.
NCP	jumlah makan utama yang dimakan.
CAEC	kebiasaan orang mengkonsumsi makanan diantara waktu makan.
CH20	kebiasaan orang mengkonsumsi air setiap harinya.
CALC	Kebiasaan orang mengkonsumsi minuman beralkohol.

2. Atribut-atribut yang berhubungan dengan kondisi

*Table 2. Atribut Kondisi*

Atribut	Keterangan
SCC	kebiasaan orang dalam melakukan pemantauan mengenai perihal konsumsi
FAF	frekuensi aktivitas fisik yang dilakukan
TUE	lama waktu menggunakan perangkat teknologi
MTRANS	jenis transportasi yang digunakan

3. Atribut lain terdiri dari. Table 3

*Table 3. Atribut lain*

Atribut	Keterangan
Gender	Jenis Kelamin
Age	Umur
Height	Tinggi Badan
Weight	Berat Badan
Family History With Overweight	Riwayat Obesitas
Smoke	Kebiasaaan Merokok

### 3.2. Preprocessing

#### a. Penamaan ulang kolom

Parameter 'inplace=True' berarti perubahan akan dilakukan langsung pada DataFrame asli, 'df', tanpa membuat DataFrame baru dengan kolom yang diganti namanya. Setelah skrip dijalankan, nama kolom pada 'df' akan diperbarui sesuai dengan nama baru yang ditentukan dalam kamus seperti gambar 1.

```
In [27]: # Rename the columns
df.columns[4:]
df.rename(columns={'CALC':'Alcohol', 'FAVC': 'High caloric food', 'FCVC':'vegetables meals', 'NCP':'No of meals', 'SCC':'Monitor
'family_history_with_overweight':'family_history', 'FAF':'Physical activity', 'TUE':'tech usage', 'CAEC':'Snack', 'MTRANS
'NObeysdad':'Obesity_level'}, inplace = True])

In [28]: df.head(4)
Out[28]:
```

	Age	Gender	Height	Weight	Alcohol	High caloric food	vegetables meals	No of meals	Monitor calories	Smoke	Water intake frequency	family_history	Physical activity	Tech usage	Snack	Transpor
0	21.0	Female	1.62	64.0	no	no	2.0	3.0	no	no	2.0	yes	0.0	1.0	Sometimes	Public_Transpc
1	21.0	Female	1.52	50.0	Sometimes	no	3.0	3.0	yes	yes	3.0	yes	3.0	0.0	Sometimes	Public_Transpc
2	23.0	Male	1.80	77.0	Frequently	no	2.0	3.0	no	no	2.0	yes	2.0	1.0	Sometimes	Public_Transpc
3	27.0	Male	1.80	87.0	Frequently	no	3.0	3.0	no	no	2.0	no	2.0	0.0	Sometimes	V

Gambar 1. Rename atribut

#### b. Distribusi presentase Tingkat obesitas

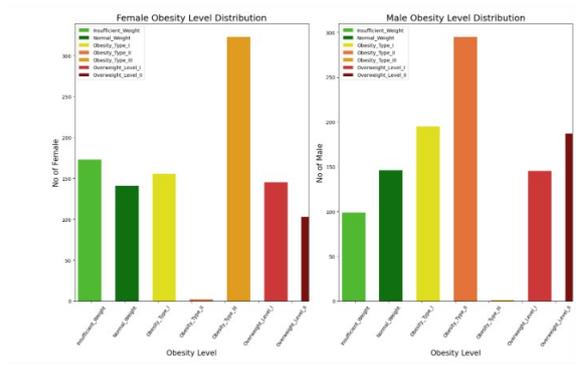
Diagram ini menggambarkan distribusi persentase untuk setiap tingkat obesitas dalam dataset. Pilihan warna digunakan untuk membedakan setiap kategori, dan persentasenya ditampilkan pada masing-masing segmen. seperti gambar 2.



Gambar 2. Diagram distribusi presentase

#### c. Distribusi Tingkat Obesitas Berdasarkan Gender

Segmen ini menampilkan distribusi tingkat obesitas berdasarkan jenis kelamin (wanita dan pria) yang ditunjukkan pada Gambar 3. Data dibagi menjadi dua bagian: Diagram Batang untuk Female yang menampilkan jumlah wanita pada setiap tingkat obesitas, dan Diagram Batang untuk Male yang menampilkan jumlah pria pada setiap tingkat obesitas.

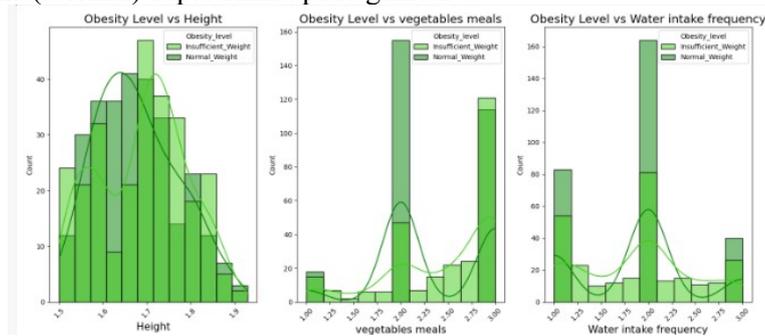


*gambar 3. Diagram Distribusi Tingkat Obesitas Berdasarkan Gender*

d. Analisis Tingkat Obesitas Berdasarkan Tinggi Badan, Konsumsi Sayuran, dan Frekuensi Minum Air

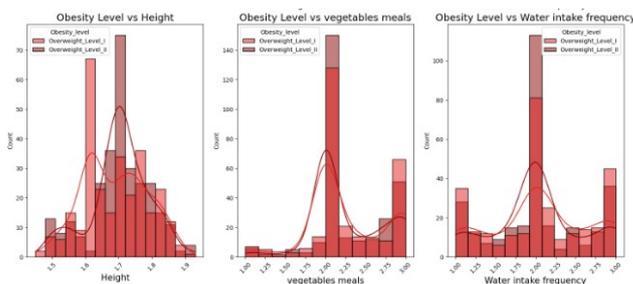
Grafik ini terdiri dari sembilan histogram yang menggambarkan distribusi tingkat obesitas (Normal, Overweight, dan Obesitas) berdasarkan tiga variabel utama: tinggi badan, konsumsi sayuran, dan frekuensi minum air. Setiap baris dalam subdiagram merepresentasikan satu kelompok obesitas.

Baris Pertama (Normal) dapat dilihat pada gambar 6



*gambar 6. Diagram Tingkat Obesitas Berdasarkan Tinggi Badan, Konsumsi Sayuran, dan Frekuensi Minum Air*

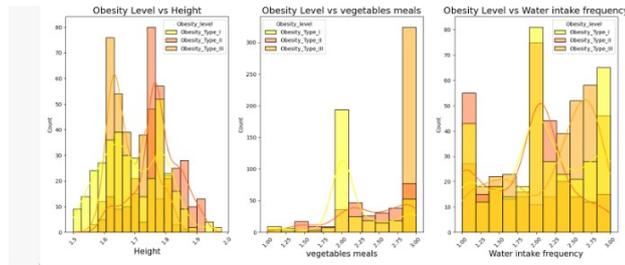
1. Height (Tinggi Badan) : Distribusi tinggi badan dalam kelompok obesitas normal.
  2. Vegetables Meals (Konsumsi Sayuran) : Frekuensi konsumsi sayuran pada kelompok obesitas normal.
  3. Water Intake Frequency (Frekuensi Minum Air) : Frekuensi minum air pada kelompok obesitas normal.
- Baris Kedua (Overweight) dapat dilihat pada gambar 7



*gambar 7. Diagram Overweight*

1. Height (Tinggi Badan) : Distribusi tinggi badan pada kelompok overweight.
2. Vegetables Meals (Konsumsi Sayuran) : Frekuensi konsumsi sayuran pada kelompok overweight.
3. Water Intake Frequency (Frekuensi Minum Air) : Frekuensi minum air pada kelompok overweight.

- Baris Ketiga (Obesitas) dapat dilihat pada gambar 8



gambar 8. Diagram Obesitas

1. Height (Tinggi Badan) : Distribusi tinggi badan pada kelompok obesitas.
2. Vegetables Meals (Konsumsi Sayuran) : Frekuensi konsumsi sayuran pada kelompok obesitas.
3. Water Intake Frequency (Frekuensi Minum Air) : Frekuensi minum air pada kelompok obesitas.

e. Evaluasi Kinerja Model Klasifikasi

Ada tiga metrik umum untuk mengevaluasi kinerja model klasifikasi: laporan klasifikasi, akurasi, dan matriks kebingungan. Laporan klasifikasi memberikan detail metrik seperti presisi, recall, dan f1-score untuk setiap kelas yang diprediksi. Akurasi mengukur persentase prediksi benar dibandingkan total sampel. Matriks kebingungan menunjukkan jumlah prediksi benar dan salah untuk setiap kelas. Analisis ini membantu menilai kemampuan model dalam mengklasifikasikan kelas target dan mengidentifikasi kelas dengan performa kurang baik. Dapat dilihat di gambar 11

	precision	recall	f1-score	support
0	0.87	0.98	0.92	92
1	0.82	0.73	0.77	77
2	0.88	0.84	0.86	114
3	0.89	0.96	0.93	85
4	0.99	0.99	0.99	92
5	0.81	0.71	0.75	89
6	0.73	0.79	0.76	85
accuracy			0.86	634
macro avg	0.86	0.86	0.85	634
weighted avg	0.86	0.86	0.86	634

```

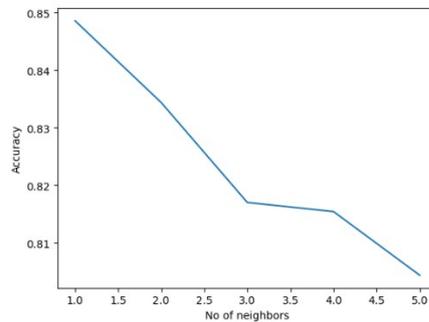
0.8596214511041009
[[90 2 0 0 0 0 0]
 [13 56 0 0 0 0 2]
 [ 0 0 96 10 1 0 7]
 [ 0 0 3 82 0 0 0]
 [ 0 0 1 0 91 0 0]
 [ 0 10 0 0 0 63 16]
 [ 0 0 9 0 0 9 67]]
    
```

gambar 11. Evaluasi Kinerja Model Klasifikasi

f. Pengaruh Jumlah Tetangga Terdekat terhadap Akurasi

Grafik ini menggambarkan hubungan antara jumlah tetangga terdekat yang digunakan dalam algoritma K-Nearest Neighbors (KNN) dengan tingkat akurasi model. Sumbu x menunjukkan jumlah tetangga, sementara sumbu y mewakili tingkat akurasi. Analisis ini penting untuk menentukan jumlah tetangga optimal yang dapat

menghasilkan prediksi dengan akurasi tertinggi menggunakan algoritma KNN. Grafik dapat dilihat di gambar 12



*gambar 12. Grafik Pengaruh Jumlah Tetangga Terdekat terhadap Akurasi*

g. Evaluasi Model Random Forest untuk Prediksi Tingkat Obesitas

Model Random Forest diterapkan untuk memprediksi tingkat obesitas individu berdasarkan berbagai fitur dalam dataset. Model ini dilatih dengan menggunakan data pelatihan dan dievaluasi dengan data pengujian. Dengan hasil seperti gambar 13:

```

precision    recall  f1-score   support

 0         0.98     0.96     0.97         92
 1         0.77     0.92     0.84         77
 2         0.98     0.93     0.95        114
 3         0.99     1.00     0.99         85
 4         1.00     1.00     1.00         92
 5         0.93     0.80     0.86         89
 6         0.88     0.93     0.90         85

 accuracy          0.93         634
 macro avg         0.93         0.93         0.93         634
 weighted avg      0.94         0.93         0.93         634

0.9337539432176656
[[ 88  4  0  0  0  0  0]
 [ 2 71  0  0  0  3  1]
 [ 0  1 106  1  0  0  6]
 [ 0  0  0 85  0  0  0]
 [ 0  0  0  0 92  0  0]
 [ 0 13  1  0  0 71  4]
 [ 0  3  1  0  0  2 79]]
    
```

*Gambar 13. Hasil evaluasi random forest*

Hasil evaluasi mengindikasikan bahwa model Random Forest mencapai akurasi keseluruhan sebesar 93.38% dalam memprediksi tingkat obesitas. Model ini menunjukkan kinerja prediksi yang sangat baik untuk kelas Obesity\_Type\_I (kelas 0) dan Obesity\_Type\_III (kelas 4), dengan f1-score masing-masing mencapai 0.97 dan 1.00. Namun, untuk kelas Overweight\_Level\_I (kelas 1), kinerja prediksi sedikit lebih rendah dengan f1-score sebesar 0.84.

h. Perbandingan Hasil Accuracy

*Table 4. Perbandingan hasil dari 4 algoritma*

	KNN	DT	NBC	RF
<b>accuracy</b>	85 %	85,79%	65.12%	93.38%
<b>recall</b>	79%	85.32%	65.06%	93%
<b>precision</b>	73%	86.38%	64.80%	88%

Dari perbandingan Machine Learning pada table 4, perbandingan tingkat obesitas di negara Meksiko terdahulu dengan algoritma Decision Tree dan Naïve Bayes dengan pengujian sekarang menggunakan algoritma K Nearest Neighbors dan Random forest, metode Random forest memiliki tingkat akurasi yang lebih tinggi daripada algoritma yang lain.

#### IV. KESIMPULAN

Pemanfaatan metode Random Forest untuk memproyeksikan tingkat obesitas individu, berdasarkan fitur-fitur dalam dataset. Hasil evaluasi mengindikasikan tingkat akurasi keseluruhan sebesar 93.38%, mencerminkan kinerja prediksi yang sangat baik. Model ini terutama unggul dalam memprediksi kelas Obesity\_Type\_I (kelas 0) dan Obesity\_Type\_III (kelas 4), dengan f1-score masing-masing mencapai 0.97 dan 1.00. Namun, kinerja prediksi sedikit menurun pada kelas Overweight\_Level\_I (kelas 1), dengan f1-score sekitar 0.84. Meskipun demikian, model Random Forest terbukti sebagai alat yang efisien dalam memproyeksikan tingkat obesitas, meskipun ada potensi untuk meningkatkan kinerja pada kelas-kelas dengan hasil prediksi yang kurang optimal.

#### DAFTAR PUSTAKA

- [1] U. Lisyia Wijaya, B. Widjanarko, R. I. Bagian, P. Kesehatan, I. Perilaku, and K. Masyarakat, "FAKTOR-FAKTOR YANG BERHUBUNGAN DENGAN PERILAKU MAKAN BERISIKO GIZI LEBIH PADA REMAJA SMA DI KOTA SEMARANG," vol. 8, no. 3, 2020, [Online]. Available: <http://ejournal3.undip.ac.id/index.php/jkm>
- [2] 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.
- [3] J. S. Komputer, K. Buatan, and A. Ridwan, "Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus," 2020.
- [4] A. Muzakir and R. A. Wulandari, "Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree," *Scientific Journal of Informatics*, vol. 3, no. 1, 2016, [Online]. Available: <http://journal.unnes.ac.id/nju/index.php/sji>
- [5] S. Y. Sibi and A. R. Widiarti, "Klasifikasi Tingkat Obesitas Mempgunakan Algoritma KNN."
- [6] "252-Article Text-515-1-10-20240330".
- [7] C. Responden and K. Kunci, "Maryanah Safitri 1) , Ardian Dwi Praba 2) 1 Informatika, Universitas Nusa Mandiri," 2024.
- [8] D. Arisandi, T. Sutrisno, and I. Kurniawan, "KLASIFIKASI OPINI MASYARAKAT DI TWITTER TENTANG KEBOCORAN DATA YANG TERJADI DI INDONESIA MENGGUNAKAN ALGORITMA SVM," *Jurnal Teknika*, vol. 15, no. 2, pp. 75–80, Sep. 2023, doi: 10.30736/jt.v15i2.993.
- [9] R. W. Abdullah, D. Hartanti, H. Permatasari, A. W. Septyanto, and Y. A. Bagaskara, "Penerapan Data Mining untuk Memprediksi Jumlah Produk Terlaris Menggunakan Algoritma Naive Bayes Studi Kasus (Toko Prapti)," *Jurnal Ilmiah Informatika Global*, vol. 13, no. 1, Mar. 2022, doi: 10.36982/jiig.v13i1.2060.
- [10] I. A. G. Saputra, N. Soewarno, and Isnalita, "Faktor-faktor yang memengaruhi keputusan pembelian Generasi Z pada kegiatan bisnis berbasis E-commerce," *Journal of research and application: accounting and management*, vol. 4, no. 1, Sep. 2019, doi: 10.18382/jraam.v4i1.003.
- [11] I. Bagus, K. Surya, A. Itb, S. Bali, and J. Raya Puputan, "ANALISIS SENTIMEN PADA MEDIA SOSIAL TERHADAP VAKSIN NUSANTARA MENGGUNAKAN TEXT MINING," *Jurnal informasi dan Komputer*, vol. 11, no. 1, p. 2023.
- [12] P. Simanjuntak, C. E. Suharyanto, S. Sitohang, and K. Handoko, "DATA MINING UNTUK KLASIFIKASI STATUS PANDEMI COVID 19," *Jurnal Teknik Informasi dan Komputer (Tekinkom)*, vol. 5, no. 2, p. 327, Dec. 2022, doi: 10.37600/tekinkom.v5i2.620.

- [13] I. Budiman and R. Ramadina, "Penerapan Fungsi Data Mining Klasifikasi untuk Prediksi Masa Studi Mahasiswa Tepat Waktu pada Sistem Informasi Akademik Perguruan Tinggi," *IJCCS*, vol. x, No.x, no. 1, pp. 1–5.
- [14] E. Retnaningsih, "MODEL PREDIKSI PREVALENSI OBESITAS PADA PENDUDUK UMUR DIATAS 15 TAHUN DI INDONESIA," 2010. [Online]. Available: [www.balitbangdasumsel.net](http://www.balitbangdasumsel.net)
- [15] D. Nur Fitriani *et al.*, "Prediction of Obesity Levels Using Neural Network: Binary Classification Approach," 2024, [Online]. Available: <https://ojs3.unpatti.ac.id/index.php/parameter>
- [16] Y. Sulisty Nugroho and dan Nova Emiliyawati, "Sistem Klasifikasi Variabel Tingkat Penerimaan Konsumen Terhadap Mobil Menggunakan Metode Random Forest." [Online]. Available: <http://archive.ics.uci.edu/ml/>
- [17] S. Devella and F. Novia Rahmawati, "Implementasi Random Forest Untuk Klasifikasi Motif Songket Palembang Berdasarkan SIFT," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 7, no. 2, 2020, [Online]. Available: <http://jurnal.mdp.ac.id>