

# Klasifikasi Tingkat Obesitas Menggunakan Algoritma KNN

Sandra Yoseba Sibi<sup>1</sup>, Anastasia Rita Widiarti<sup>2</sup>

Informatika

Universitas Sanata Dharma

Yogyakarta, Indonesia

e-mail: <sup>1</sup>Sandrasibi3@gmail.com, <sup>2</sup>rita\_widiarti@usd.ac.id

## Abstrak

Tingkat obesitas merupakan suatu gangguan yang terjadi akibat penumpukan lemak yang berlebihan pada tubuh, sehingga memiliki resiko dalam masalah kesehatan. Diperlukan upaya pencegahan, misalnya dengan melakukan pengecekan tingkat obesitas secara mandiri. Penelitian ini mencoba mengembangkan sistem sederhana untuk mengetahui status seseorang berkategori obesitas atau tidak dengan memasukkan beberapa informasi yang diperlukan. Sistem dicoba dikembangkan dengan menggunakan metode klasifikasi *k*-nearest neighbor (*k*-nn). Proses klasifikasi dilakukan dalam beberapa tahap yaitu, tahap seleksi atribut, transformasi data, dan normalisasi. Data yang digunakan untuk percobaan diambil dari uci machine learning repository yang berjumlah 2111 data dengan 16 atribut dan 1 label. Percobaan dilakukan dengan mengubah variabel *k* pada KNN yaitu *k*=1,3,5,7,9, menggunakan uji validasi silang 3,5, dan 7-fold, dan mengubah atribut-atribut yang digunakan. Dalam penelitian ini ditemukan model klasifikasi obesitas menggunakan KNN yang optimal yaitu dengan *k*=1 dengan atribut yang paling berpengaruh yaitu atribut berat badan dan usia. Akurasi maksimal yang diperoleh adalah sebesar 79,96%, sehingga dapat disimpulkan bahwa KNN mampu digunakan untuk klasifikasi tingkat obesitas.

**Kata kunci:** klasifikasi, *k*-nearest neighbour, obesitas.

## Abstract

The level of obesity is a disorder that occurs due to the accumulation of excessive fat in the body, so it has a risk of health problems. Prevention efforts are needed, for example by checking the level of obesity independently. This study tries to develop a simple system to find out whether someone is categorized as obese or not by entering some necessary information. The system was developed using the *k*-nearest neighbor (*k*-nn) classification method. The classification process is carried out in several stages, namely, the attribute selection stage, data transformation, and normalization. The data used for the experiment was taken from the UCI machine learning repository, totaling 2111 data with 16 attributes and 1 label. The experiment was carried out by changing the *k* variable on KNN, namely *k*=1,3,5,7,9, using the 3.5, and 7-fold cross-validation test, and changing the attributes used. In this study, it was found that the obesity classification model used the optimal KNN with *k*=1 with the most influential attributes, namely weight and age attributes. The maximum accuracy obtained is 79.96%, so it can be concluded that KNN can be used for classification of obesity levels.

**Keywords:** classification, *k*-nearest neighbour, obesity

## 1. Pendahuluan

Obesitas adalah sebuah status kondisi gangguan kesehatan tubuh yang diakibatkan oleh adanya penumpukan lemak, dan seringkali bisa diukur dari berat badan yang di atas normalnya [1]. Asupan kalori yang berlebihan apabila dibandingkan dengan aktivitas pembakaran kalori mengakibatkan kalori yang berlebihan itu menumpuk dalam bentuk lemak.

Keadaan obesitas dalam jangka panjang akan berdampak pada kualitas hidup yang bersangkutan dalam aktivitas hariannya karena munculnya ketidaknyamanan atau nyeri badan. Selain itu, obesitas juga dapat memicu munculnya penyakit kronis, di antaranya *diabetes melitus*, *jantung koroner*, *stroke iskemik* dan *hemoragik*. Di sisi lain, adanya obesitas juga dapat meningkatkan jumlah biaya pengeluaran [2].

Obesitas menjadi salah satu masalah kesehatan yang tidak biasa lagi, dikarenakan angka kasus tingkat obesitas semakin meningkat, dan dapat diidap oleh setiap orang tanpa melihat usia [3]. Penelitian yang dilakukan dengan melakukan pemantauan berkala untuk kelebihan berat badan dan obesitas pada semua populasi di dunia menunjukkan bahwa penderita obesitas terdapat di banyak negara, seperti di negara Peru sebanyak 19,7 % [4], Kolombia sebanyak 22,3% [5], termasuk di Indonesia sebanyak 6,9% [6].

Banyak hal yang dapat dilakukan untuk mencegah munculnya obesitas, misalnya melalui edukasi di masyarakat mengenai gejala, dan akibat kalau mengidap obesitas. Salah satu cara edukasi adalah dengan menyediakan sistem sederhana yang dapat digunakan oleh masyarakat guna melakukan cek secara mandiri mengenai kondisi kesehatannya. Apabila dari hasil pengecekan secara mandiri ditemukan gejala adanya penyakit seperti obesitas, maka masyarakat bisa melanjutkan pemeriksaan ke dokter. Guna mewujudkan sistem tersebut, dapat dimanfaatkan algoritma pembelajaran mesin terawasi, misalnya menggunakan algoritma klasifikasi jika sudah ada data yang digunakan sebagai basis pelatihan.

K-Nearest Neighbour atau biasa disingkat KNN merupakan salah satu algoritma yang dapat digunakan untuk mengklasifikasikan suatu data yang sebelumnya belum diketahui kelasnya, ke kelas terdekat dari sekitar data baru tersebut yang mempunyai kemiripan. Terdapat beberapa penelitian yang telah dilakukan dengan menerapkan algoritma KNN pada persoalan-persoalan terkait gizi maupun lainnya, dan mendapat akurasi di 78% ke atas. Dewi dan Dwidasmara yang melakukan klasifikasi tingkat obesitas dengan menggunakan data dari kagle.com dari 10 atribut yang ada, nilai  $k = 7$ , dan uji silang 10-fold mendapatkan hasil akurasi 78,98% [7]. Penelitian lainnya oleh Kartini untuk klasifikasi status gizi balita berdasarkan indeks antropometri BB/U (Berat Badan menurut Umur) mendapatkan hasil akurasi sebesar 80% [8]. Yunita mengembangkan sistem klasifikasi penyakit diabetes dengan akurasi 96% [9]. KNN untuk klasifikasi akreditasi pada data sekolah menengah atas, mendapatkan hasil akurasi 80,7051% [10]. Paper ini menyajikan hasil penelitian klasifikasi KNN untuk masalah obesitas dengan data diambil dari UCI *machine learning repository* [11], dan menggunakan fungsi-fungsi yang sudah tersedia pada python.

## 2. Metode Penelitian

Bagian ini memaparkan mengenai penjelasan tentang data yang digunakan sebagai bahan input, pengembangan alat uji untuk mengubah data input menjadi data output, dan skenario pengujian untuk mendapatkan berbagai output guna mendapatkan jawaban atas pertanyaan-pertanyaan dalam penelitian.

### 2.1. Data Input

Data yang digunakan pada penelitian ini adalah data umum yang diambil dari UCI *machine learning repository* dengan judul “Estimation of obesity levels based on eating habits and physical condition data set” [11]. Jumlah cacah data 2111 records dengan 16 atribut dan 1 label. Terdapat 7 kelas yang tersimpan dalam data label yaitu berat badan tidak mencukupi, berat badan normal, obesitas tipe 1, obesitas tipe 2, obesitas tipe 3, kelebihan berat badan level 1 atau kelebihan berat badan level 2.

Atribut-atribut yang dipunyai oleh data dapat dikelompokkan ke dalam 3 hal, yaitu atribut yang berhubungan dengan kebiasaan makan, atribut yang berhubungan dengan kondisi fisik dan atribut lainnya. Atribut-atribut yang berhubungan dengan kebiasaan makan yaitu atribut FAVC yang berisikan frekuensi kebiasaan orang mengkonsumsi makanan yang memiliki kalori tinggi, FCVC yaitu frekuensi orang mengkonsumsi sayuran, NCP yaitu jumlah makan utama yang dimakan, CAEC yaitu kebiasaan orang mengkonsumsi makanan diantara waktu makan, CH20 yaitu kebiasaan orang mengkonsumsi air setiap harinya, dan atribut CALC yaitu kebiasaan orang mengkonsumsi minuman beralkohol. Atribut-atribut yang berhubungan dengan kondisi fisik adalah SCC yaitu kebiasaan orang dalam melakukan pemantauan mengenai perihal konsumsi kalori, FAF yaitu frekuensi aktivitas fisik yang dilakukan, TUE yaitu lama waktu menggunakan perangkat teknologi misal tv, games, ponsel dan atribut MTRANS yaitu jenis transportasi yang digunakan. Atribut lain terdiri dari jenis kelamin, umur, tinggi dan berat badan, status Riwayat obesitas, dan kebiasaan merokok.

Langkah paling awal untuk mengolah data sebelum diolah ke dalam sistem dengan melakukan proses seleksi atribut menggunakan bantuan weka tools. Metode seleksi yang digunakan yaitu *information gain*, dan diperoleh hasil perangkian atribut seperti terlihat pada Gambar 1.

```

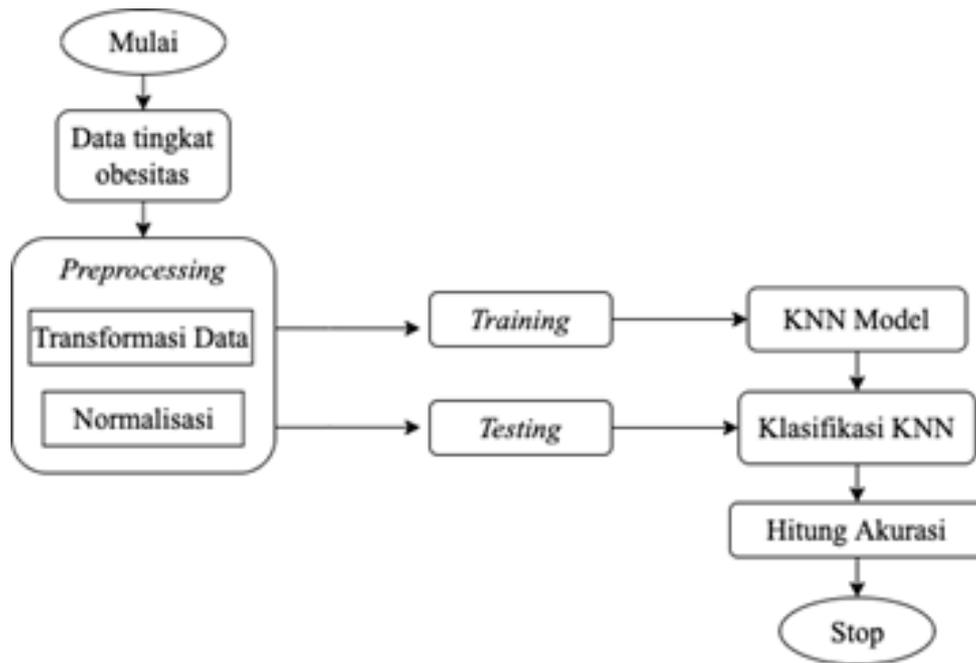
Ranked attributes:
1.7036 4 Weight
0.8038 2 Age
0.566 7 FCVC
0.4472 13 FAF
0.4008 14 TUE
0.3925 11 CH20
0.3693 8 NCP
0.2982 1 Gender
0.2939 3 Height
0.2317 9 CAEC
0.2301 5 family_history_with_overweight
0.1466 15 CALC
0.1111 16 MTRANS
0.0863 6 FAVC
0.0459 12 SCC
0.0108 10 SMOKE
    
```

Gambar 4. Gambaran umum aliran data dan proses yang terjadi pada data.

Informasi mengenai nilai gain setiap atribut seperti tertera pada Gambar 1 akan digunakan sebagai bahan pertimbangan dalam skenario percobaan.

**2.2. Desain Proses**

Proses untuk mengubah data input menjadi data output guna kepentingan analisis data sehingga dapat disimpulkan mengenai pemanfaatan KNN untuk klasifikasi tingkat obesitas mengikuti diagram seperti tergambar pada Gambar 2.



Gambar 2. Gambaran umum aliran data dan proses yang terjadi pada data.

Setelah data awal melalui tahap seleksi atribut di luar sistem, selanjutnya akan dilakukan proses *Preprocessing* yang terdiri dari Transformasi Data untuk mengubah data dari data belum numerik menjadi data numerik dan Normalisasi dengan menggunakan algoritma min-max. Setelah itu data akan dibagi untuk kepentingan uji silang *cross validation*, dengan jumlah k-fold (3,5,7). Dimana data latih akan digunakan sebagai pemodelan KNN, sedangkan pada tahap testing akan dilakukan proses klasifikasi berdasarkan model KNN yang sudah dilakukan pada tahap latih. Lalu di langkah terakhir akan dilakukan tahap perhitungan akurasi.

### 2.3. Skenario Pengujian

Pengujian akan dilakukan dengan mengubah-ubah atribut, jumlah tetangga  $k$  dalam KNN, dan  $k$  dalam  $k$ -fold cross validation. Nilai  $k$  dalam KNN adalah 3,5,7,9. Nilai  $k$ -fold yang dipakai adalah 3,5, dan 7 yang akan di bagi menjadi data training dan data testing. Ketentuan pembagian data pada 3-fold yaitu 2/3 dari data total akan menjadi data latih dan 1/3 dari data total akan menjadi data uji. Lalu untuk pembagian data di 5-fold yaitu 4/5 dari data total akan menjadi data latih dan 1/5 dari data total akan menjadi data uji. Untuk skenario 7-fold pembagiannya akan menjadi 6/7 dari data total adalah data latih dan 1/7 dari data total akan menjadi data uji. Untuk setiap kali percobaan pada  $k$  di klasifikasi KNN, atribut yang digunakan akan diubah jumlahnya dari 1 atribut sampai 16 atribut digunakan semua.

### 3. Hasil dan Pembahasan

Setelah data mengalami proses transformasi dan normalisasi, proses pengujian akurasi dilakukan di dalam sistem dengan menggunakan fungsi `KNeighborsClassifier`, `K-fold` dan `cross_val_score` yang sudah disediakan pada python. Percobaan sebanyak 16 kali dilakukan dengan menggunakan  $k$ -fold cross validation dengan nilai  $k = 3,5,7$  dan nilai  $k=1,3,5,7,9$  pada klasifikasi KNN, untuk mendapatkan hasil akurasi yang tertinggi.

Hasil percobaan untuk dengan jumlah tetangga 1 atau nilai  $k = 1$  pada KNN atau disebut 1NN dengan menggunakan bantuan google colab dapat dilihat pada Tabel 1.

Tabel 14. Prosentase akurasi klasifikasi pada skenario 1NN.

Jumlah Atribut	3-fold	5-fold	7-fold
1	61,49%	60,16%	62,72%
2	79,16%	79,58%	79,96%
3	78,78%	78,45%	78,16%
4	74,75%	75,6%	75,98%
5	71,01%	72,72%	73,09%
6	69,26%	72,05%	72,24%
7	68,78%	71,44%	71,72%
8	69,16%	71,53%	71,62%
9	71,01%	73,43%	73,28%
10	72,38%	74,14%	73,8%
11	73,05%	74,56%	74,8%
12	73,76%	75,27%	75,7%
13	75,6%	76,55%	77,07%
14	76,84%	77,78%	77,59%
15	76,98%	78,16%	77,97%
16	76,93%	78,16%	78,02%

Proses pada Tabel 1 di atas, dapat disimpulkan bahwa dengan menggunakan klasifikasi 1NN yang memiliki akurasi tertinggi terdapat pada penggunaan 2 atribut saja dan pada uji silang 7-fold, dengan hasil akurasi 79,96%.

Cara yang sama pada pengujian 1NN kemudian diterapkan untuk 3NN, 5NN, 7NN, dan 9NN, sehingga diperoleh data akurasi tertinggi pada setiap nilai  $k$  pada KNN dengan berbagai uji silangnya seperti terlihat pada Tabel 2.

Tabel 2. Rangkuman nilai akurasi tertinggi pada berbagai skenario  $k$  dalam KNN.

Jumlah Tetangga KNN	3-fold	5-fold	7-fold
1	79,16%	79,58%	79,96%
3	78,02%	77,5%	78,96%
5	76,69%	77,78%	78,63%
7	76,03%	76,03%	76,83%
9	75,56%	75,56%	75,74%

Menggunakan berbagai skenario jumlah tetangga dalam proses klasifikasi dengan algoritma KNN 1, 3, 5, 7, 9 ternyata hasil akurasi tertinggi tetap di kondisi seperti pengujian pertama yaitu di jumlah tetangga 1, jumlah atribut 2 dan pada uji validasi 7 fold.

### 3.1. Akurasi Data Uji Tunggal

Dari pengujian pada data latih, di peroleh atribut yang optimal untuk mendapatkan hasil klasifikasi yaitu atribut berat badan dan usia. Maka dua atribut tersebut digunakan pada sistem untuk pengujian data tunggal.

Data yang diujikan pada sistem merupakan data yang diambil dari kuesioner pada google form yang sudah disebarakan pada 10 responden dengan meminta data mengenai berat badan dan usia. Untuk melihat unjuk kerja sistem, kemudian data-data tersebut diolah menggunakan sistem dan hasilnya seperti terlihat rangkumannya di Tabel 3.

Tabel 3. Data hasil uji tunggal.

Data Responden	Berat Badan	Usia	Kelas Target	Kelas Prediksi	Penilaian
1	39	23	Berat badan kurang	Normal	Salah
2	73	23	Kelebihan bb level 1	Obesitas tipe 1	Salah
3	40	12	Berat badan kurang	Normal	Salah
4	80	43	Obesitas tipe 1	Normal	Salah
5	77	41	Obesitas tipe 1	Normal	Salah
6	85	42	Obesitas tipe 1	Normal	Salah
7	55	12	BB normal	Normal	Benar
8	80	60	Kelebihan bb level 2	Normal	Salah
9	80	30	Kelebihan bb level 2	Normal	Salah
10	78	33	Obesitas tipe 1	Obesitas tipe 1	Benar

Dari Tabel 3 di mana informasi mengenai berat badan dan usia berasal dari dari responden secara langsung. Informasi kelas target diperoleh langsung dari responden juga dan bukan dari para medis. Dari hasil keluaran sistem setelah dicocokkan dengan kenyataan, dari data 10 responden pada sistem yang benar hanya berjumlah 2 data, maka akurasi yang didapat dari uji data tunggal adalah sebesar 20%.

### 4. Kesimpulan

Dari berbagai pengujian yang dilakukan, model yang paling optimal untuk klasifikasi tingkat obesitas menggunakan metode KNN dalam penelitian ini hanya memerlukan atribut berat badan dan usia, dan hanya menggunakan 1 data terdekat saja yang berpengaruh. Mempergunakan data dari dataset UCI *machine learning repository*, akurasi data latih tertinggi yang diperoleh pada penelitian ini adalah sebesar 79,96%. Dengan nilai akurasi yang masih di bawah 80%, dapat disimpulkan pula bahwa variabel-variabel yang diterapkan dalam percobaan-percobaan di penelitian ini masih belum memberikan hasil yang optimal. Masih terbuka penelitian lanjutan mengenai klasifikasi tingkat obesitas dengan menggunakan metode penyiapan data lain dan properti lain yang dapat dilakukan dari algoritma KNN.

### Daftar Pustaka

- [1] D. S. Ayu and O. W. K. Handayani, "Diary Teratas (Terapi Anak Obesitas) Dalam Perubahan Perilaku Gizi Siswa Sekolah Dasar," *Journal of Public Health*, pp. 167-175, 2016.
- [2] Rohayati, W. Wiarsih dan A. Y. Nursasi, "'Frustrasi versus Puas": Pengalaman Keluarga dalam Merawat Anggota Keluarga Obesitas," *Jurnal Penelitian Kesehatan SUARA FORIKES (Journal of Health Research Forikes Voice)*, vol. 12, no. 3, pp. 332-336, Juli 2021.
- [3] Nadimin, Ayumar dan Fajarwati, "Obesitas Pada Orang Dewasa Anggota Keluarga Miskin Di Kecamatan Lembang Kabupaten Pinrang Obesity an Adults Poor Family in Sub District Lembang District Pinrang," *Media Kesehatan Masyarakat Indonesia*, vol. 11, no. 3, p. 9–15, 2015.
- [4] NoName, "Ulasan Peru: 58 fakta dan sorotan," <https://versus.com/id/peru>, Peru, NoYear.
- [5] NoName, "Ulasan Kolumbia: 58 fakta dan sorotan," <https://versus.com/id/kolumbia>, NoYear.
- [6] NoName, "Ulasan Indonesia: 58 fakta dan sorotan," <https://versus.com/id/indonesia>, NoYear.
- [7] A. M. S. I. Dewi dan I. B. G. Dwidasmara, "Implementation Of The K-Nearest Neighbor (KNN) Algorithm For Classification Of Obesity Levels," *Jurnal Elektronik Ilmu Komputer Udayana*, vol. 9, no. 2, pp. 277-284, 2020.
- [8] P. D. Kartini, "Klasifikasi Status Gizi Balita Berdasarkan Indeks Anthropometri Bb/U Menggunakan Metode KNN (K Nearest Neighbor),"

---

[http://simki.unpkediri.ac.id/mahasiswa/file\\_artikel/2017/099e77c6013da6ced14ec1e90917ce8.pdf](http://simki.unpkediri.ac.id/mahasiswa/file_artikel/2017/099e77c6013da6ced14ec1e90917ce8.pdf), 2017.

- [9] A. Y. N. C. N. Umikulsum Indah Lestari, "Penerapan Metode K-Nearest Neighbor Untuk Sistem Pendukung Keputusan Identifikasi Penyakit Diabetes Melitus," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 4, pp. 2071-2082, 2021.
- [10] A. D. P. Sinaga, "Klasifikasi Akreditasi Menggunakan Metode K-Nearest Neighbor (KNN) Pada Data Sekolah Menengah Atas. Skripsi.," Universitas Sanata Dharma, Yogyakarta, 2020.
- [11] F. M. & d. l. H. M. A. Palechor, "Estimation of obesity levels based on eating habits and physical condition Data Set," UC Irvine Machine Learning Repository, 2019.