

PROPOSAL
MODEL MACHINE LEARNING



PENERAPAN METODE K-NEAREST NEIGHBOR (KNN)
DALAM MEMPREDIKSI LEVEL OBESITAS SESEORANG

Disusun Oleh :

Laurentius Joandanu

A11.2022.14179

FAKULTAS ILMU KOMPUTER
UNIVERSITAS DIAN NUSWANTORO

SEMARANG

2024

RINGKASAN

Obesitas adalah kondisi berat badan melebihi batas normal karena adanya penumpukan lemak berlebih di dalam tubuh.⁽¹⁾ Obesitas terjadi karena asupan kalori lebih banyak dibandingkan jumlah kalori yang dibakar, sehingga tubuh akan menyimpan kalori yang tidak digunakan dalam bentuk lemak.⁽²⁾ Untuk mengetahui apakah seseorang mengalami obesitas atau tidak maka perlu melakukan penelitian dengan model prediksi yang akurat dan tepat menggunakan metode K-Nearest Neighbors (KNN). K-Nearest Neighbors (KNN) adalah metode pembelajaran mesin terawasi yang digunakan untuk mengatasi masalah klasifikasi dan regresi.⁽³⁾ Penelitian ini bertujuan untuk memprediksi kondisi obesitas seseorang dengan menerapkan metode KNN pada dataset Obesity Levels dari Kaggle.

Kata Kunci : Obesitas; Prediksi; KNN; Dataset

DAFTAR ISI

| | |
|---------------------------------------|------------|
| HALAMAN JUDUL | i |
| RINGKASAN | ii |
| DAFTAR ISI..... | iii |
| BAB I PENDAHULUAN..... | 1 |
| 1.1 Latar Belakang..... | 1 |
| 1.2 Rumusan Masalah | 1 |
| 1.3 State Of The Art | 1 |
| BAB II DATASET | 3 |
| BAB III METODE PENELITIAN..... | 6 |
| BAB IV HASIL..... | 7 |

BAB I

PENDAHULUAN

1.1 Latar Belakang

Obesitas adalah kondisi gangguan kesehatan yang ditunjukkan dengan berat badan yang berlebih, kegemukan, dan mengandung banyak lemak pada tubuh.⁽⁴⁾ 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.⁽¹⁾ Obesitas dapat meningkatkan risiko beberapa penyakit seperti jantung, diabetes, tekanan darah tinggi, dan kanker. Penyebab obesitas karena kadar kalori masuk lebih banyak dari yang dibutuhkan oleh tubuh. Selain itu, obesitas juga dapat disebabkan oleh pengaruh genetik, perilaku, dan hormonal.⁽⁴⁾ Obesitas dipengaruhi oleh jenis kelamin, tingkat pengetahuan, tingkat pendidikan, jenis pekerjaan, status pernikahan, Riwayat keturunan, aktivitas fisik, status merokok dan intake kalori.⁽⁵⁾

Prediksi tingkat obesitas seseorang dapat dilakukan dengan menggunakan metode K-Nearest Neighbors (KNN). KNN merupakan algoritma berdasarkan pada konsep nearest neighbors, yang memprediksi kelas data berdasarkan mayoritas kelas tetangga terdekatnya. Dalam konteks prediksi obesitas, KNN membantu mengidentifikasi faktor risiko dan memprediksi tingkatan berat badan seseorang yang termasuk dalam tingkatan obesitas. Penelitian ini memberikan kontribusi baru terhadap penggunaan K-Nearest Neighbors (KNN) dalam prediksi tingkat obesitas dengan memperdalam pemahaman kita tentang metode K-Nearest Neighbors (KNN) dan menerapkan algoritma ini pada kumpulan data yang relevan.⁽¹⁾

1.2 Rumusan Masalah

Bagaimana menerapkan metode K-Nearest Neighbors (KNN) untuk memprediksi seseorang mengalami obesitas ?

1.3 State Of The Art

Metode KNN telah dilakukan oleh beberapa peneliti untuk memprediksi obesitas. Salah satunya yang berjudul *“Klasifikasi Tingkat Obesitas Menggunakan Algoritma KNN”* dan disusun oleh Anastasia Rita Widiarti. Penelitian ini mencoba

mengembangkan sistem sederhana untuk mengetahui apakah seseorang sudah 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.⁽¹⁾

BAB II

DATASET

Dataset yang saya pilih berasal dari platform Kaggle yaitu dataset Obesity Levels yang disusun oleh Fatemeh Mehrparvar. Dataset ini tentang data yang memiliki atribut seperti jenis kelamin, umur, tinggi badan, berat badan, riwayat diabetes pada keluarga, FAVC, FCVC, NCP, CAEC, merokok, CH2O, SCC, FAF, TUE, CALC, MTRANS, NObeyesdad. Dataset ini berisi 2111 data mentah sesuai dengan atribut yang disebutkan sebelumnya. Contoh isi dari dataset tersebut sesuai pada Gambar 1 dan 2.

| # Age | Gender | # Height | # Weight | Δ CALC | ✓ FAVC | # FCVC | # NCP | ✓ SCC | ✓ SMOKE | # CH2O |
|-------|--------|----------|----------|------------|--------|--------|-------|-------|---------|--------|
| 21 | Female | 1.62 | 64 | no | no | 2 | 3 | no | no | 2 |
| 21 | Female | 1.52 | 56 | Sometimes | no | 3 | 3 | yes | yes | 3 |
| 23 | Male | 1.8 | 77 | Frequently | no | 2 | 3 | no | no | 2 |
| 27 | Male | 1.8 | 87 | Frequently | no | 3 | 3 | no | no | 2 |
| 22 | Male | 1.78 | 89.8 | Sometimes | no | 2 | 1 | no | no | 2 |
| 29 | Male | 1.62 | 53 | Sometimes | yes | 2 | 3 | no | no | 2 |
| 23 | Female | 1.5 | 55 | Sometimes | yes | 3 | 3 | no | no | 2 |
| 22 | Male | 1.64 | 53 | Sometimes | no | 2 | 3 | no | no | 2 |
| 24 | Male | 1.78 | 64 | Frequently | yes | 3 | 3 | no | no | 2 |
| 22 | Male | 1.72 | 68 | no | yes | 2 | 3 | no | no | 2 |
| 26 | Male | 1.85 | 185 | Sometimes | yes | 3 | 3 | no | no | 3 |

Gambar 1. Sample dataset Obesity Levels

| ✓ family_hist... | # FAF | # TUE | △ CAEC | △ MTRANS | △ NObeyesd... |
|------------------|-------|-------|------------|---------------------------|-------------------------|
| yes | 0 | 1 | Sometimes | Public_Transpor tation | Normal_Weight |
| yes | 3 | 0 | Sometimes | Public_Transpor tation | Normal_Weight |
| yes | 2 | 1 | Sometimes | Public_Transpor tation | Normal_Weight |
| no | 2 | 0 | Sometimes | Walking | Overweight_Leve l_I |
| no | 0 | 0 | Sometimes | Public_Transpor tation | Overweight_Leve l_II |
| no | 0 | 0 | Sometimes | Automobile | Normal_Weight |
| yes | 1 | 0 | Sometimes | Motorbike | Normal_Weight |
| no | 3 | 0 | Sometimes | Public_Transpor tation | Normal_Weight |
| yes | 1 | 1 | Sometimes | Public_Transpor tation | Normal_Weight |
| yes | 1 | 1 | Sometimes | Public_Transpor tation | Normal_Weight |
| yes | 2 | 2 | Frequently | Public_Transpor tation | Obesity_Type_I |

Gambar 2. Sample dataset Obesity Levels

Dataset tersebut memiliki beberapa atribut sebagai berikut :

a. Gender

Atribut ini berisi tentang data jenis kelamin.

b. Age

Atribut ini berisi tentang data usia.

c. Height

Atribut ini berisi tentang data tinggi badan.

d. Weight

Atribut ini berisi tentang data berat badan.

e. Family_history_with_overweight

Atribut ini berisi tentang data apakah memiliki keturunan yang mengalami obesitas.

f. CALC

Atribut ini berisi tentang data seberapa sering minum minuman berakohol.

- g. FAVC
Atribut ini berisi tentang data apakah sering makan makanan berkalori tinggi.
- h. FCVC
Atribut ini berisi tentang data apakah biasanya makan sayuran dalam makanan.
- i. NCP
Atribut ini berisi tentang data berapa banyak makan dalam sehari.
- j. SCC
Atribut ini berisi tentang data apakah memantau kalori makanan yang dimakan setiap hari.
- k. SMOKE
Atribut ini berisi tentang data apakah merokok.
- l. CH2O
Atribut ini berisi tentang data berapa air yang diminum setiap harinya.
- m. FAF
Atribut ini berisi tentang data seberapa sering melakukan aktivitas fisik.
- n. SEL
Atribut ini berisi tentang data berapa lama menggunakan perangkat seperti ponsel, videogame, komputer, dan lain-lain.
- o. CAEC
Atribut ini berisi tentang data apakah makan makanan ringan sebelum makan makanan utama.
- p. MTRANS
Atribut ini berisi tentang data transportasi yang sering digunakan sehari – hari.
- q. NObeyesdad
Atribut ini berisi tentang data level obesitas.

BAB III

METODE PENELITIAN

Dalam memprediksi level obesitas dapat menggunakan K-Nearest Neighbors. Berikut contoh gambar bagan proses penelitian sebagai berikut.



BAB IV

HASIL

Dalam konteks prediksi level obesitas menggunakan algoritma K-Nearest Neighbor (KNN), hasil yang diharapkan adalah prediksi level obesitas berdasarkan data baru yang diberikan. Prediksi ini didasarkan pada jarak antara data baru dengan data pelatihan yang telah diketahui. Hasil prediksi dapat berupa kategori "obesitas" atau "tidak obesitas", bergantung pada tingkat akurasi model KNN yang digunakan. Dalam beberapa penelitian, KNN telah menunjukkan akurasi yang relatif tinggi dalam memprediksi level obesitas, seperti 88,96% dalam analisis prediksi level obesitas menggunakan perbandingan algoritma machine learning dan deep learning.

DAFTAR PUSTAKA

1. Sibi SY, Widiarti AR. Klasifikasi Tingkat Obesitas Mempergunakan Algoritma KNN. *Semin Nas Corisindo*. 2022;7(2):370–5.
2. Hospitals TMS. *siloamhospitals.com*. 2023 [cited 2024 May 6]. Obesitas - Penyebab, Gejala, dan Cara Mengatasinya. Available from: <https://www.siloamhospitals.com/informasi-siloam/artikel/apa-itu-obesitas>
3. GeeksforGeeks. *geeksforgeeks.org*. 2024 [cited 2024 May 6]. K-Nearest Neighbor(KNN) Algorithm. Available from: <https://www.geeksforgeeks.org/k-nearest-neighbours/>
4. *halodoc.com* [Internet]. 2022 [cited 2024 May 6]. Obesitas. Available from: <https://www.geeksforgeeks.org/k-nearest-neighbours/>
5. Kejadian F, Sentral O, Usia P. *HIGEIA JOURNAL OF PUBLIC HEALTH*. 2018;2(2):249–59.