e-ISSN 2962-1968

Implementasi Algoritma K-Nearest Neighbor untuk Klasifikasi Obesitas

Adi Tiyas Ahmad Fathoni[1*], Dwi Hartanti[2]

[1,2]*Informatika*
Univesitas Duta Bangsa Surakarta
[1*]*202020357@mhs.udb.ac.id,* [2]*dwihartanti@udb.ac.id*

Abstrak— **Obesitas merupakan masalah kesehatan yang semakin meningkat di banyak negara, termasuk di seluruh dunia. Mengkategorikan obesitas menurut faktor individu seperti indeks massa tubuh (BMI), lingkar pinggang, tinggi badan, usia, dan jenis kelamin dapat mengarah pada pemahaman yang lebih baik tentang pola dan karakteristik obesitas. Pada penelitian ini, kami mengimplementasikan algoritma K-Nearest Neighbor (KNN) untuk mengklasifikasikan obesitas menggunakan data yang mengandung atribut tersebut. Langkah-langkah dalam penelitian ini meliputi pengolahan data, pemilihan parameter KNN yang optimal, dan evaluasi kinerja model. Tujuan utama dari penelitian ini adalah untuk mengklasifikasikan obesitas berdasarkan karakteristik yang relevan dan untuk mengidentifikasi pola dan hubungan antara faktor-faktor tersebut. Dengan penerapan algoritma KNN, kami mencoba memprediksi kelas obesitas dari data individu yang status obesitasnya tidak diketahui. Hasil penelitian ini dapat membantu mengembangkan metode yang lebih akurat untuk mengklasifikasikan obesitas. Selain itu, model klasifikasi yang dihasilkan dapat digunakan sebagai alat yang berguna di bidang kesehatan, termasuk diagnosis dini, perencanaan intervensi, dan pengembangan program penanganan obesitas. Diharapkan penelitian ini dapat membantu masyarakat dan tenaga kesehatan dalam upaya pencegahan dan penanganan obesitas secara efektif.**
Kata kunci— K-Nearest Neighbor, klasifikasi, obesitas.

Abstract— **Obesity is a growing health problem in many countries, including around the world. Classification of obesity according to individual factors such as body mass index (BMI), waist circumference, height, age, and gender provides a better understanding of obesity patterns and characteristics. In this study, we implemented the K Nearest Neighbors (KNN) algorithm to classify obesity from data containing these attributes. The steps in this study include data processing, selection of optimal KNN parameters, and evaluation of model performance. The primary aim of this study is to classify obesity based on relevant features and identify patterns and associations between these factors. By applying the KNN algorithm, we try to predict the obesity class from single data with unknown obesity status. The results of this study may help develop more accurate methods for classifying obesity. Furthermore, the resulting classification model can be used as a useful tool in the healthcare field, including early diagnosis, intervention planning, and development of obesity treatment programs. It is hoped that this research will contribute to effective prevention and treatment of obesity in the community and health care workers.**
Keywords— K-Nearest Neighbor, classification, obesity.

I. PENDAHULUAN

Di zaman modern ini, masalah obesitas telah menjadi tantangan kesehatan utama di seluruh dunia. Obesitas dapat menyebabkan berbagai gangguan kesehatan seperti penyakit jantung, diabetes dan penyakit pernapasan, serta dapat berdampak negatif pada kualitas hidup seseorang. Oleh karena itu, penting untuk mengembangkan metode yang efektif untuk mengklasifikasikan obesitas untuk mengambil tindakan pencegahan dan kuratif yang tepat. Salah satu metode yang digunakan untuk mengklasifikasikan obesitas adalah algoritma K-Nearest Neighbor (K-NN). Algoritma K-NN adalah metode machine learning yang digunakan untuk mengklasifikasikan data berdasarkan hubungannya dengan data training yang ada. Sebagai bagian dari klasifikasi obesitas, algoritma K-NN dapat digunakan untuk mengklasifikasikan orang obesitas atau non-obesitas berdasarkan karakteristik dan karakteristik yang relevan seperti indeks massa tubuh (BMI), dan tinggi badan.

Implementasi algoritma K-NN untuk mengklasifikasikan obesitas melibatkan beberapa langkah. Pertama, data latihan dikumpulkan dalam kaitannya dengan orang berlabel obesitas dan non-obesitas. Kemudian, atribut yang relevan seperti BMI, dan tinggi diekstraksi dari masing-masing orang. Data pelatihan ini kemudian digunakan untuk melatih model K-NN. Pada fase uji, orang yang baru diklasifikasikan dimasukkan kedalam model KNN. Model kemudian menghitung jarak antara pasien baru dan data pelatihan menggunakan metrik jarak seperti jarak Euclidean atau jarak Manhattan. K-kemiripan terdekat dengan orang baru dipilih dan sebagian besar label tetangga terdekat ditentukan sebagai label klasifikasi orang baru. Dengan demikian, seseorang dapat diklasifikasikan sebagai

obesitas atau nonobesitas berdasarkan model K-NN yang telah dilatih sebelumnya.

Adapun maksud dari penelitian ini, yaitu bertujuan untuk mengklasifikasi obesitas menggunakan metode algoritma K-Nearest Neighbor. Dengan menggunakan dataset yang terdiri dari tinggi badan, dan berat badan, kita dapat mengklasifikasikan individu sebagai obesitas atau tidak obesitas berdasarkan atributatribut tersebut. Implementasi algoritma KNN ini diharapkan dapat memberikan akurasi yang baik dalam mengklasifikasikan obesitas.

II. LANDASAN TEORI

A. Obesitas

Obesitas adalah penyakit yang ditandai dengan kelebihan lemak yang tidak normal. Kondisi ini telah menjadi masalah kesehatan global yang berkembang pesat selama beberapa dekade terakhir. Landasan teori obesitas adalah untuk memahami faktor-faktor yang berkontribusi terhadap perkembangan obesitas, termasuk faktor genetik, lingkungan, gaya hidup, dan psikologis. Pada artikel ini kita melihat secara rinci dasar teori obesitas.

1. Definisi dan klasifikasi obesitas

   Obesitas merupakan suatu kondisi dimana seseorang memiliki lemak yang berlebihan di dalam tubuh, yang dapat berdampak negatif bagi kesehatan. Metode yang umum digunakan untuk mengklasifikasikan obesitas adalah indeks massa tubuh (BMI), yang menggambarkan hubungan antara berat dan tinggi badan. Klasifikasi obesitas berdasarkan BMI mencakup kategori seperti obesitas Kelas I, obesitas Kelas II, dan obesitas Kelas III.
2. Faktor genetik

   Studi pada keluarga identik dan kembar telah menunjukkan adanya faktor genetik yang berkontribusi terhadap perkembangan obesitas. Beberapa gen terlibat dalam pengaturan berat badan dan metabolisme, seperti gen FTO, MC4R dan POMC. Variasi gen ini dapat memengaruhi kecenderungan seseorang untuk mengalami obesitas. Namun, faktor genetik hanya berperan kecil dalam perkembangan obesitas, sedangkan faktor lingkungan dan gaya hidup memiliki pengaruh lebih besar.
3. Faktor lingkungan

   Faktor lingkungan seperti diet, aktivitas fisik dan pengaruh sosial memainkan peran penting dalam perkembangan obesitas. Mengonsumsi makanan berkalori tinggi, rendah serat, konsumsi makanan olahan dan minuman manis dapat meningkatkan risiko obesitas. Selain itu, prevalensi obesitas juga didukung oleh lingkungan yang mendukung gaya hidup sedenter, seperti B. kurangnya ruang gerak, lalu lintas yang tidak sesuai untuk pejalan kaki, dan peningkatan penggunaan teknologi.
4. Gaya hidup

   Gaya hidup modern yang umumnya kurang melibatkan aktivitas fisik juga menjadi faktor penting dalam perkembangan obesitas. Kurangnya aktivitas fisik dan duduk lama menyebabkan penumpukan lemak tubuh. gaya hidup dan pola makan yang tidak sehat dan asupan kalori dan pengeluaran energi juga dapat menyebabkan kelebihan berat badan dan obesitas.
5. Faktor psikologi

   Aspek psikologis seperti stres, depresi, dan kecemasan juga dapat berkontribusi pada perkembangan obesitas. Beberapa orang mungkin menggunakan makanan sebagai mekanisme koping untuk mengatasi emosi negatif. Selain itu, gangguan makan seperti binge eating dan gangguan makan lainnya juga dapat menyebabkan kenaikan berat badan dan obesitas.
6. Komplikasi dan Efek Kesehatan

   Obesitas meningkatkan risiko banyak komplikasi kesehatan yang serius, termasuk diabetes tipe 2, penyakit jantung, tekanan darah tinggi, masalah pernapasan, gangguan tidur, osteoartritis, dan beberapa jenis kanker. Obesitas juga dapat mempengaruhi kualitas hidup seseorang dan meningkatkan risiko masalah psikososial seperti rendah diri dan depresi [5][6][7].

B. K-Nearest Nighbor

KNN adalah salah satu metode klasifikasi algoritma yang paling populer dalam pembelajaran mesin. Algoritma ini didasarkan pada gagasan bahwa objek serupa biasanya berada di lingkungan yang sama dalam ruang fitur. Landasan teori KNN

adalah memahami prinsip operasi algoritma ini, memilih parameter yang tepat dan mengevaluasi kinerjanya [2].

C. *Prinsip kerja KNN*

Prinsip dasar KNN adalah mengklasifikasikan objek berdasarkan ID kelas dari tetangga terdekatnya. Algoritma ini bekerja dalam langkah-langkah berikut:

1. Persiapan Data
   Pertama-tama, data yang akan digunakan harus dipersiapkan. Ini melibatkan pemisahan data menjadi data pelatihan yang digunakan untuk melatih model dan data pengujian (test data) yang digunakan untuk menguji model. Memilih K tetangga terdekat berdasarkan jarak [8].
2. Penghitungan Jarak
   Setelah data dipersiapkan, algoritma KNN menghitung jarak antara contoh yang tidak diketahui (contoh pengujian) dengan semua contoh pelatihan yang ada. Umumnya, jarak Euclidean digunakan sebagai metrik jarak, tetapi metrik jarak lain seperti jarak Manhattan juga dapat digunakan [9].
3. Penentuan Tetangga Terdekat
   Setelah jarak dihitung, langkah selanjutnya adalah menentukan K tetangga terdekat dari contoh pengujian. Nilai K harus ditentukan sebelumnya dan biasanya dipilih melalui validasi silang (cross-validation).
4. Pemilihan Kelas Mayoritas
   Setelah K tetangga terdekat ditemukan, kelas mayoritas dari tetangga-tetangga tersebut ditentukan. Artinya, jika mayoritas tetangga adalah kelas A, contoh pengujian diklasifikasikan sebagai kelas A.
5. Prediksi Kelas
   Setelah kelas mayoritas ditentukan, contoh pengujian diklasifikasikan sebagai kelas tersebut [10][11].

D. *Parameter KNN*

KNN memiliki beberapa parameter penting yang perlu diperhatikan:

1. Nilai K:
   Menentukan jumlah tetangga terdekat yang digunakan untuk pemungutan suara keputusan mayoritas.
2. Pengukur dispersi:
   Mendefinisikan metode untuk mengukur jarak antara objek. Contoh ukuran jarak adalah Euclidean, Manhattan dan Minkowski.
3. Fungsi cetak:
   Opsional digunakan untuk menetapkan bobot yang berbeda ke tetangga terdekat.

E. *Pilihan parameter K*

Memilih nilai K yang tepat sangat penting dalam KNN. Jika nilai K terlalu kecil, algoritme mungkin sensitif terhadap derau data. Sebaliknya, jika nilai K terlalu besar, algoritme dapat kehilangan kemampuannya untuk mengenali pola yang lebih baik. Pemilihan nilai K optimal dapat dilakukan dengan validasi silang atau dengan menggunakan metode pemilihan parameter yang lebih canggih [1].

F. *Kinerja KNN*

Untuk mengukur kinerja KNN, beberapa ukuran evaluasi yang umum digunakan meliputi presisi, akurasi, daya ingat, skor F1, dan kurva ROC. Precision mengukur tingkat akurasi klasifikasi secara keseluruhan, sedangkan precision dan recall memberikan informasi tentang kinerja dalam mengklasifikasikan kelas tertentu. Skor F1 adalah ukuran komposit yang menggabungkan presisi dan daya ingat [2].

III. METODOLOGI PENELITIAN

A. *Pengumpulan data*

Kami mengumpulkan dataset yang terdiri dari informasi nama. Tinggi badan, dan berat badan. Jumlah total dataset adalah sebanyak 10. Kami mendapatkan data tersebut dari dari website kaggle.com.

B. *Pra-pemrosesan Data*

Sebelum menerapkan algoritma KNN, data subjek harus disiapkan terlebih dahulu. Kami memproses data sebelumnya dengan menghapus nilai yang hilang, mengonversi variabel kategori menjadi numerik, dan menormalisasikan skala jika diperlukan.

e-ISSN 2962-1968

C. Implementasi Algoritma K-Nearest Neighbor

Algoritma K-NN sangat sederhana dan bergantung pada jarak terpendek dari instance kueri ke sampel pelatihan untuk menentukan KNN. Sampel pelatihan diproyeksikan ke ruang multidimensi, di mana setiap dimensi mewakili fitur data. Ruang ini dibagi berdasarkan klasifikasi sampel pelatihan. Suatu titik dalam ruang ini disebut kelas c jika kelas c adalah klasifikasi yang paling sering terjadi di k tetangga terdekat [1].

Algoritma K-Nearest Neighbor (KNN) merupakan salah satu metode untuk mengklasifikasikan sebuah data yang paling dekat dengan objek tersebut. Algoritma ini merupakan algoritma pembelajaran yang mengklasifikasikan hasil soal yang baru berdasarkan kelas terbanyak dari algoritma tersebut. Kelas yang paling sering muncul kemudian akan menjadi kelas yang dihasilkan dari klasifikasi tersebut [2].

Kedekatan dapat ditentukan dengan menggunakan metrik jarak seperti jarak Euclidean. Jarak Euclidean [3] dapat dicari dengan menggunakan persamaan di bawah ini :

$$Dxy = \sqrt{\sum_{i=1}^{n} (x_i - y_{i)}{}^2}$$

Keterangan :
D = Jarak kedekatan
x = data latih
y = data uji
n = untuk jumlah atribut individu antara 1 s.d. n
f = fungsi similitary atribut i antara kasus X dan kasus Y
i = atribut individu dari 1 hingga n

Berikut adalah langkah langkah untuk menghitung algoritma K-NN [4]:

1. Tentukan jumlah parameter K (jumlah tetangga paling dekat).
2. Hitung kuadrat jarak Euclidean ke objek data sampel apa pun menggunakan Persamaan 1.
3. Urutkan objek-objek ke dalam kelompok yang memliki jarak Euclid terkecil.
4. Kumpulkan kategori Y (klasifikasi tetangga terdekat).
5. Kategori mayoritas tetangga terdekat dapat digunakan untuk memprediksi nilai instans kueri yang dihitung.

IV. HASIL DAN PEMBAHASAN

A. Persiapan data

Kita perlu menyiapkan data pelatihan terlebih dahulu. Data latih merupakan data dari waktu sebelumnya yang sudah diketahui labelnya. Misalnya Kita ingin menggunakan klasifikasi untuk memprediksi apakah siswa A akan lulus atau tidak, Kita perlu mencari beberapa informasi tentang data siswa yang lulus dan tidak lulus dari tahun-tahun sebelumnya.

Dibawah ini merupakan data yang saya gunakan untuk diklasifikasikan menggunakan algoritma K-Nearest Nighbor.

Tabel 1. Dataset

No	Nama	Tinggi badan	Berat Badan	Keterangan
1	Adam	149	50	1
2	Fadilah	163	65	1
3	Yukis	178	80	3
4	Afdillah	130	60	2
5	Medy	157	66	3
6	Patma	172	59	1
7	Asman	133	40	1
8	Belva	169	70	1
9	Anis	165	80	3
10	Bela	140	70	2

Data uji saya memiliki dua atribut data tipe numerik. Dan atribut keterangannya saya menggunakan angka 1 untuk kategori normal, 2 untuk kategori obesitas dan angka 3 untuk kategori kelebihan berat badan.

B. Menghitung jarak euclidean

Tahap selanjutnya pada penelitian ini yaitu menggunakan perhitungan seperti rumus dibawah ini :

e-ISSN 2962-1968

$$d_{Euclidean}(x,y) = \sqrt{\sum_i (x_i - y_i)^2}$$

Keterangan :
Xi = Atribut data yang dinormalisasi
Yi = Data pengujian baru, tidak termasuk data pelatihan.

Dengan menggunakan rumus di atas, dapat dilihat bahwa kita membutuhkan data uji baru untuk mengklasifikasi jarak Euclidean. Data tersebut dapat diambil Dari data dengan kelas atau hasil pelabelan yang tidak diketahui. Data uji yang digunakan adalah sebagi berikut :

Tabel 2. Data Uji

No	Nama	Tinggi Badan	Berat Badan	Keterangan
1	Aditya	161	45	?

Dari data diatas dapat diketahui bahwa data tersebut belum memiliki atribut keterangan. Adapun atribut keterangannya yaitu normal, obesitas, atau kelebihan berat badan. Untuk perhitungan menggunakan excelnya adalah sebagai berikut :

=SQRT((baris satu atribut tinggi badan - nilai data uji atribut tinggi badan)^2 + rumus yang sama untuk atribut yang lainnya.

Jadi hasilnya adalah
=SQRT((149-161)^2+(50-45)^2)
Hasilnya adalah 13

Lakukan perhitungan di atas untuk seluruh data di setiap baris. Dibawah ini adalah hasil perhitungan menggunakan rumus di atas :

Tabel 2. Hasil Distance

No	Nama	Tinggi badan	Berat Badan	Ket	Distance
1	Adam	149	50	1	13
2	Fadilah	163	65	1	20
3	Yukis	178	80	3	39
4	Afdillah	130	60	2	34
5	Medy	157	66	3	21
6	Patma	172	59	1	18
7	Asman	133	40	1	28
8	Belva	169	70	1	26
9	Anis	165	80	3	35
10	Bela	140	70	2	33

C. Menghitung nilai k (nilai tetangga) terdekat.

Pada langkah ini, nilai jarak terkecil ditentukan atau dipilih setelah beberapa nilai k. Misalnya, jika nilai k adalah 2, Anda perlu mencari nilai jarak terkecil antara kedua nilai tersebut.
Contoh:
Nilai k = 1, maka kita ambil nilai jarak terkecil yaitu 13, dan nilai ini memiliki keterangan atribut nomor 1, sehingga modusnya Normal (Tabel 3).

Untuk menentukan atribut kelas, perlu memvoting semua nilai k yang sudah ditentukan. perhitungan menggunakan excelnya adalah sebagai berikut :
=IF(Baris satu nilai Distance <= SMALL (Blok semua kolom disctance, Nilai k), Baris satu atribut Keterangan, "")

Contoh dengan nilai k = 1
=IF(13<=SMALL(blok baris satu sampai terakhir, 1), kolom atribut keterangan baris 1, "")

Jika hasilnya bukan nilai terkecil, kolom kosong. Sebaliknya, jika nilainya adalah nilai terkecil, maka atribut kelas akan ditampilkan.

Nama	Tinggi Badan	Berat Badan	Keterangan	Distance	K-1	K-3	K-5
Adam	149	50	1	13	1	1	1
Fadilah	163	65	1	20		1	1
Yukis	178	80	3	39			
Afdillah	130	60	2	34			
Medy	157	66	3	21			3
Patma	172	59	1	18		1	1
Asman	133	40	1	28			
Belva	169	70	1	26			1
Anis	165	80	3	35			
Bela	140	70	2	33			
Aditva	161	45	?				

Gambar 1. Penentuan Nilai K-1, K-3 dan K-5

Berdasarkan hasil perhitungan diatas data Aditya menunjukan bahwa dia 1 yang berarti **normal**.

V. KESIMPULAN

Dalam penelitian ini, kami berhasil mengimplementasikan algoritma K-Nearest Neighbor (KNN) untuk mengklasifikasikan obesitas berdasarkan karakteristik yang relevan secara klinis.

Hasil percobaan menunjukkan bahwa model KNN dapat mengklasifikasikan obesitas dengan akurasi yang baik. Hasil yang kami dapatkan adalah bahwa data uji yang bernama Aditya menunjukan bahwa dia 1 yang berarti dia normal. Hasil ini bisa sangat membantu dalam mengembangkan metode yang lebih efektif untuk mengklasifikasikan obesitas di masa mendatang

REFERENSI

[1] T. Cover and P. Hart, "Nearest neighbor pattern classification," IEEE Trans. Inf. Theory, vol. 13, no. 1, pp. 21–27, Jan. 1967.
[2] Avelita, B. (2013). A._Klasifikasi_K-Nearest_Neighbor. Dipetik 06 2016, 22, dari www.academia.edu : https://www.academia.edu/91319959/A._Klasifikasi_KNearest_Neighbor
[3] Jiawei Han, M. K. (2012). Data Mining : Concepts and Techniques. United States of America: Elsevier.
[4] Ndaumanu, R. I. (2014). Analisis Prediksi Tingkat Pengunduran Diri Mahasiswa dengan Metode K-Nearest Neighbor. Jatisi Vol 1, 3.
[5] Healthline. Diakses pada 2022. Obesity.
[6] Mayo Clinic. Diakses pada 2022. Obesity.
[7] National Health Services. Diakses pada 2022. Obesity.
[8] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer.
[9] Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques (3rd ed.). Elsevier
[10] Mitchell, T. M. (1997). Machine Learning. McGraw-Hill.
[11] Zhang, T. (2010). Pattern Recognition and Machine Learning. Springer.