

BAB 1. PENDAHULUAN

1.1 Latar Belakang

Kegemukan dan obesitas adalah penumpukan lemak berlebihan yang dapat menyebabkan resiko bagi kesehatan. Pemicu utama obesitas ialah tidak seimbangnya kalori yang masuk dan kalori yang dikeluarkan. Kegemukan dan obesitas menjadi faktor risiko utama yang berpotensi menyebabkan sejumlah penyakit kronis. Diantaranya adalah penyakit jantung dan stroke, yang menjadi penyebab utama kematian. Selain itu, kondisi berat badan berlebih juga dapat memicu diabetes dan masalah terkait, seperti gangguan penglihatan, amputasi, serta kebutuhan untuk menjalani cuci darah (WHO-Obesity, 2021).

Berdasarkan laporan yang disampaikan UNICEF Indonesia tahun 2022, menyatakan bahwa di Indonesia angka *stunting* dan *wasting* menurun, namun angkanya masih cukup tinggi, yaitu sebesar 30,8% dan 10,2%. Sementara itu, angka obesitas ditahun yang sama mengalami peningkatan pada beberapa kelompok usia, yaitu dari 9,2% menjadi 20% pada anak usia 5-12 tahun, 1,9% menjadi 14,8% pada remaja usia 13-18 tahun, dan peningkatan tertinggi terjadi pada dewasa usia 18 tahun ke atas, dari 21,7% menjadi 35,4% (UNICEF Indonesia, 2022). Oleh sebab itu, penting untuk mengklasifikasikan tingkat obesitas menggunakan teknik yang tepat agar dapat mengidentifikasi risiko kesehatan, serta memberikan pencegahan ataupun perawatan yang tepat.

Untuk mengetahui tingkat obesitas yang terjadi pada setiap individu, salah satunya dapat memanfaatkan metode pengolahan data melalui data mining, yaitu metode klasifikasi. Metode klasifikasi digunakan untuk membangun model yang dapat mengklasifikasikan objek atau data baru ke dalam kelas tertentu berdasarkan karakteristik yang dimilikinya. Tahapan penting membangun model klasifikasi adalah melatih model menggunakan algoritma klasifikasi yang tepat. Metode klasifikasi yang digunakan, yaitu *naive bayes* dan *k-nearest neighbor* (Tiyas, dkk. 2023; Wie dan Siddik, 2022)

Berdasarkan cara pelatihannya, algoritma klasifikasi dibagi menjadi dua jenis yaitu *eager learner* dan *lazy learner*. Algoritma *eager learner* adalah algoritma

yang melakukan pelatihan pada data latih terlebih dahulu, sebelum melakukan klasifikasi pada objek baru. Dalam *eager learner*, model pembelajaran dihasilkan pada saat pelatihan dan disimpan dalam memori, sehingga bila ada objek baru, model yang sudah terlatih untuk mengklasifikasikan. Contoh algoritma *eager learner* adalah *naive bayes*, *decision trees*, dan *artificial neural networks* (ANN). Sedangkan, algoritma *lazy learner* adalah algoritma yang tidak melakukan pelatihan pada data latih sebelum objek baru diberikan. Dalam *lazy learner*, model pembelajaran tidak dihasilkan pada saat pelatihan, tetapi model dihasilkan saat objek baru diberikan untuk diklasifikasikan. Contoh algoritma *lazy learner* adalah *k-nearest neighbors* (KNN) dan *case-based reasoning* (CBR).

Beberapa penelitian telah dilakukan untuk mengklasifikasikan obesitas atau penyakit lain menggunakan algoritma klasifikasi. Salah satunya, penelitian yang bertujuan untuk membangun sebuah sistem informasi yang dapat mengklasifikasikan status gizi obesitas pada anak disabilitas di SLB Sri Mujinab Pekanbaru, menggunakan algoritma terbaik antara *naive bayes* dan *k-nearest neighbor*. Hasilnya, algoritma *k-nearest neighbor* memiliki akurasi lebih tinggi yaitu 68%, dibandingkan *naive bayes* yang hanya 48%. Oleh karena itu, *k-nearest neighbor* diterapkan pada sistem informasi yang dibangun (Triwira Lestari, dkk. 2022). Penelitian lain, dilakukan oleh Putry dan Nurina Sari yang tidak hanya memberikan informasi mengenai penyakit diabetes melitus, tetapi juga menyajikan perbandingan dan pengembangan ilmu terkait algoritma *naive bayes* dan *k-nearest neighbor*. Hasil penelitiannya menunjukkan bahwa algoritma *naive bayes* menjadi algoritma terbaik untuk kasus ini (Marito Putry & Nurina Sari, 2022).

Dari paparan diatas, peneliti ingin membandingkan kinerja algoritma klasifikasi antara *eager learner* dengan *lazy learner* menggunakan data yang sama. Data yang digunakan adalah data klasifikasi tingkat obesitas berdasarkan kebiasaan makan dan kondisi fisik pada setiap individu. Algoritma *naive bayes* sebagai perwakilan *eager learner* dan algoritma *k-nearest neighbor* sebagai perwakilan *lazy learner*.

Menurut Gunawan, dkk. (2021), algoritma *naive bayes* membutuhkan sedikit data latih untuk mengestimasi parameter dalam pengklasifikasian. Meskipun

demikian, algoritma *naive bayes* memiliki akurasi dan kecepatan klasifikasi yang baik saat diterapkan pada database besar (Han et al. 2022). Di sisi lain, *k-nearest neighbor* tidak memerlukan proses pembelajaran yang rumit dan tahan terhadap data *noise* (Ariyani, dkk. 2022).

Oleh karena itu, peneliti memilih algoritma *naive bayes* karena algoritma ini mencoba untuk mempelajari data latih, sedangkan untuk algoritma *k-nearest neighbor* menunda pembelajaran hingga saat pengklasifikasian. Selain kedua algoritma tersebut memiliki perbedaan pada cara pelatihan, algoritma *naive bayes* dipilih karena kecepatan dan membutuhkan sedikit data pelatihan untuk melakukan klasifikasi. Sedangkan, algoritma *k-nearest neighbor* cocok untuk dataset berukuran besar, tahan terhadap data pelatihan mengandung *noise*, dan dapat memberikan kinerja yang baik.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, dapat dirumuskan permasalahan sebagai berikut:

- a. Bagaimana cara mengimplementasikan algoritma *naive bayes* dan *k-nearest neighbor* dalam mengklasifikasikan tingkat obesitas berdasarkan kebiasaan makan dan kondisi fisik?
- b. Bagaimana hasil kinerja model klasifikasi dari algoritma *naive bayes* dan *k-nearest neighbor* dalam mengklasifikasikan tingkat obesitas berdasarkan kebiasaan makan dan kondisi fisik?

1.3 Batasan Masalah

Untuk menjaga agar masalah yang dibahas tidak keluar dari pembahasan, maka diperlukan beberapa batasan masalah, yaitu:

- a. Data yang digunakan dalam penelitian ini adalah data klasifikasi tingkat obesitas berdasarkan kebiasaan makan dan kondisi fisik pada setiap individu. Data tersebut memiliki atribut-atribut seperti jenis kelamin, usia, tinggi badan, berat badan, histori keluarga dengan kelebihan berat badan, frekuensi makanan berkalori tinggi, frekuensi konsumsi sayuran, frekuensi makanan utama, konsumsi makanan di antara waktu makan, konsumsi rokok, konsumsi air setiap hari, pemantauan konsumsi kalori, frekuensi aktivitas fisik, waktu

menggunakan perangkat teknologi, konsumsi alkohol, transportasi yang biasa digunakan, dan tingkat obesitas sebagai label. Kelas yang terdapat dalam atribut label meliputi kekurangan berat badan, berat badan normal, kelebihan berat badan, obesitas I, dan obesitas II.

- b. Algoritma yang dibandingkan adalah algoritma *naive bayes* dengan *k-nearest neighbor*.
- c. Sistem yang dibangun telah memilih rasio jumlah data latih dan data uji yang digunakan untuk membandingkan kinerja algoritma *naive bayes* dengan *k-nearest neighbor*. Proses pembagian (*splitting*) data untuk data latih dan data uji telah dilakukan dengan memilih perbandingan frekuensi jumlah yang terbaik antara data latih dan data uji.
- d. Sistem yang dibangun juga telah memilih nilai K untuk algoritma *k-nearest neighbor*. Pemilihan nilai K dipilih berdasarkan nilai K terbaik.
- e. Evaluasi kinerja model klasifikasi dilakukan menggunakan tabel *confusion matrix* dengan menghitung akurasi, presisi, recal, skor F1, serta waktu pelatihan dan pengujian model.

1.4 Tujuan

Tujuan dari penelitian ini adalah sebagai berikut:

- a. Mengetahui parameter mana yang memberikan kinerja model klasifikasi optimal, serta efisiensi dalam pelatihan dan pengujian.
- b. Membangun sebuah sistem yang dapat menampilkan perbandingan hasil kinerja model klasifikasi dari algoritma *naive bayes* dan *k-nearest neighbor* dalam mengklasifikasikan tingkat obesitas berdasarkan kebiasaan makan dan kondisi fisik.

1.5 Manfaat

Berikut beberapa manfaat yang didapat dalam penelitian ini:

- a. Meningkatkan pemahaman tentang algoritma *naive bayes* dan *k-nearest neighbor*, serta perbandingan kinerja keduanya dalam mengklasifikasikan tingkat obesitas berdasarkan kebiasaan makan dan kondisi fisik.

- b. Memperluas pemahaman dan referensi bagi penelitian selanjutnya, khususnya dalam bidang perbandingan algoritma *naive bayes* dan *k-nearest neighbor* menggunakan data yang sama.