

BAB 1

PENDAHULUAN

1.1.Latar Belakang Masalah

Perkembangan teknologi saat ini semakin meningkat dan memberi pengaruh yang besar hampir disetiap sektor kehidupan dan kenegaraan. Proses globalisasi yang terjadi disetiap negara di dunia saat ini juga mendukung perkembangan dan penggunaan teknologi. Kebutuhan informasi akan semakin meningkat bersamaan dengan perkembangan teknologi. Semakin banyak informasi yang dibutuhkan maka data yang dibutuhkan juga semakin banyak dan jumlahnya akan semakin besar. Kebutuhan akan jumlah data yang besar dapat kita temukan dalam dunia kesehatan. Hal ini dikarenakan, dunia kesehatan memiliki banyak cabang ilmu yang masing – masing memiliki banyak data dan mungkin akan terus meningkat.

Jumlah data yang terus meningkat ini memerlukan beberapa metode untuk mengolah dan mengambil kesimpulan dan informasi dari data tersebut. Beberapa metode yang digunakan untuk mengolah data yang sifatnya besar untuk menemukan pola yang terdapat didalamnya diantaranya adalah: teknik klustering, analisis diskriminan, *teorema bayes*, *decision tree*, *artificial neural networks*, *support vector machine*, regresi linear, *support vector regresi*. Setiap metode tersebut memiliki algoritma-algoritma yang digunakan untuk memproses data yang ada. Namun pada kesempatan kali ini penulis mengangkat mengenai penggunaan *Algoritma C4.5* dan *K-Nearest Neighbor (KNN)* yang merupakan algoritma dari metode *decision tree* dan *algoritma nearest neighbor* untuk pengklasifikasian status gizi balita. Algoritma C4.5 merupakan algoritma yang memiliki proses *learning* dan klasifikasi yang sederhana dan cepat. Secara umum, model algoritma C4.5 mempunyai tingkat akurasi yang tinggi (*Han & Kamber, 2006*). Sedangkan *K-Nearest Neighbor (KNN)* merupakan algoritma machine learning sederhana dan mudah diterapkan yang dapat digunakan untuk menyelesaikan masalah klasifikasi dan regresi. Algoritma ini termasuk dalam jenis *supervised learning*. Algoritma KNN bertujuan untuk mengidentifikasi tetangga terdekat dari titik kueri yang diberikan, sehingga kita dapat menetapkan label kelas ke titik tersebut. Alasan

kenapa penulis memilih algoritma C4.5 dan algoritma KNN untuk mengklasifikasi status gizi adalah karena kedua algoritma tersebut memiliki perhitungan yang sederhana dan cepat serta memiliki tingkat akurasi yang tinggi. Sebelumnya sudah pernah ada penelitian tentang perbandingan algoritma C4.5 dan algoritma KNN yaitu penelitian yang dilakukan oleh Yholanda Maldini dkk pada tahun 2021 yang berjudul “ *Perbandingan Algoritma C 4.5 dan Algoritma KNN Untuk Menentukan Pemberian Kredit Bagi Nasabah Koperasi*”. Penelitian tersebut mengangkat tema tentang pemberian simpan pinjam kepada nasabah yang ingin mengajukan pinjaman yang bertujuan untuk memenuhi kebutuhan rumah tangga pada salah satu Bank BPR di Tasikmalaya. Algoritma KNN digunakan untuk mengklasifikasikan data obyek atau nasabah berdasarkan data pembelajaran yang jaraknya paling dekat. dari data training dan data testing. Algoritma C4.5 digunakan untuk memprediksi dengan mengolah variabel pendapatan, pinjaman, lama_kredit, angsuran_perbulan. Percobaan dilakukan dengan data training yang diambil dari dataset yaitu 1-39, karena dari data nomor 40 yaitu dijadikan sebagai data testing. Menghasilkan nilai akurasi paling tinggi bagi kedua algoritma. Algoritma C 4.5 mampu memprediksi tingkat akurasi sebesar 57.5% sedangkan Algoritma KNN menghasilkan nilai akurasi yang lebih besar yaitu sebesar 62.5%.

Setiap orang memiliki status gizi yang berbeda sesuai dengan usianya. Status gizi dapat ditentukan melalui pemeriksaan laboratorium maupun secara antropometri. Antropometri merupakan cara penentuan status gizi yang paling mudah dan murah. Pengukuran antropometri adalah pengukuran yang digunakan untuk menentukan keadaan gizi seseorang. Agar memperoleh hasil yang tepat, diberikan suatu patokan sebagai pedoman. Adapun pedoman antropometri bagi penentuan keadaan gizi merupakan parameter yang dipilih dan dianjurkan, yang meliputi penilaian terhadap usia dan berat badan, panjang badan, atau tinggi badan, dan lingkaran lengan atas (DKK, 2007).

Namun, banyaknya jumlah balita yang ada membuat data – data gizi balita menjadi sulit untuk diklasifikasikan mana yang status gizinya normal, gemuk, obesitas, kurus, atau sangat kurus. Serta belum adanya analisa hasil dan akurasi dari algoritma C4.5 dengan algoritma KNN untuk klasifikasi status gizi balita.

Berdasarkan masalah yang telah diuraikan, maka penulis mengangkat judul “*KOMPARASI ALGORITMA C4.5 DENGAN K-NEAREST NEIGHBOR (KNN) UNTUK KLASIFIKASI STATUS GIZI BALITA*” agar memudahkan dalam pengklasifikasian gizi balita.

1.2.Rumusan Masalah

Banyaknya jumlah balita yang ada membuat data-data gizi balita menjadi sulit untuk diklasifikasikan mana yang status gizinya normal, gemuk, obesitas, kurus, atau sangat kurus. Serta belum adanya analisa hasil dan akurasi dari algoritma C4.5 dengan algoritma KNN untuk klasifikasi status gizi balita.

1.3.Ruang Lingkup

1. *Output* dari program ini berupa informasi klasifikasi status gizi balita apakah status gizinya normal, gemuk, obesitas, kurus, atau sangat kurus, serta hasil akurasi dari algoritma C4.5 dengan algoritma KNN.
2. Algoritma yang digunakan yaitu algoritma C4.5 yang merupakan algoritma dari metode *Decision Tree* dan algoritma KNN yang merupakan algoritma dari *Nearest Neighbor*.
3. Untuk menguji perbandingan kedua algoritma tersebut menggunakan aplikasi Rapid Miner.

1.4.Tujuan Penelitian

Tujuan yang ingin dicapai adalah mengetahui perbandingan performa algoritma C4.5 dan algoritma *K-Nearest Neighbor (KNN)* dalam pengklasifikasian status gizi balita yang meliputi tingkat akurasi, presisi dan recall kelas.

1.5. Manfaat Penelitian

Manfaat penelitian ini adalah mendapatkan pengetahuan secara mendalam mengenai kinerja relatif dari algoritma C4.5 dan *K-Nearest Neighbor (KNN)* dalam rangka klasifikasi status gizi balita yang didapat adalah lebih mudah dalam klasifikasi status gizi balita sehingga hasilnya bisa digunakan peneliti, praktisi dan pengembang sistem untuk memilih algoritma mana yang paling sesuai untuk klasifikasi status gizi balita ini.

Dengan memahami kelebihan dan kelemahan dari masing-masing algoritma yang diteliti ini dapat dijadikan dasar untuk mengoptimalkan kinerja model klasifikasi. Hal ini dapat mencakup penyesuaian parameter algoritma, pemanfaatan fitur yang lebih efektif, atau strategi lain untuk meningkatkan akurasi prediksi.

1.6. Sistematika Penulisan

Tugas akhir ini ditulis sesuai sistematika penulisan sebagai berikut :

1. BAB I : PENDAHULUAN

Bab ini berisi tentang latar belakang, rumusan masalah, batasan masalah, tujuan manfaat, metodologi penelitian dan sistematika penulisan.

2. BAB II : TINJAUAN PUSTAKA DAN DASAR TEORI

Bab ini menampilkan ringkasan-ringkasan dari jurnal-jurnal yang menjadi acuan penelitian dan menjelaskan tentang teori-teori yang terkait dengan permasalahan yang diambil.

3. BAB III : METODE PENELITIAN

Bab ini menyajikan secara lengkap setiap langkah eksperimen yang dilakukan dalam penelitian.

4. BAB IV : IMPLEMENTASI DAN PEMBAHASAN

Bab ini menguraikan tentang implementasi sistem yang dianggap penting dan hasil pengujian yang dikaitkan dengan penelitian.

5. BAB V : SIMPULAN DAN SARAN

Bab ini berisi hasil dari penelitian yang menjawab masalah dan tujuan serta keunggulan dan kelemahan sistem mengacu pada produk lain (bila sudah ada).

DAFTAR PUSTAKA

Bab ini berisi tentang daftar pustaka yang diambil

LAMPIRAN

Bab ini berisi tentang lampiran-lampiran.