

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG MASALAH

Data mining merupakan sebuah cara dalam memperoleh sebuah informasi yang terkandung dalam *database* dan menarik kesimpulan untuk dijadikan sebagai referensi dalam pengambilan sebuah keputusan, *data mining* juga bisa digunakan dalam melakukan prediksi dan klasifikasi terhadap data yang jumlahnya besar[1]. Dengan begitu, pengertian *data mining* merupakan sebuah teknik yang bisa digunakan untuk memperoleh informasi bermanfaat terhadap sebuah data. *Data mining* secara umum diketahui sebagai *Knowledge Discovery in Database* (KDD)[2]. Dalam penerapannya, KDD merupakan tahapan dimana nantinya akan berguna untuk mencari informasi yang bermanfaat pada data. Proses pencarian informasi yang tersembunyi dari sejumlah besar data menggunakan teknik *data mining* dilakukan melalui algoritma[3]. Algoritma ini nantinya akan melakukan ekstraksi pada pengolahan data menjadi informasi atau pengetahuan. Adapun algoritma dalam *data mining* terdiri dari *Association*, *Clustering*, dan Klasifikasi data[4]. Pada penelitian ini menggunakan algoritma klasifikasi.

Klasifikasi termasuk ke dalam metode *supervised learning*, yaitu pendekatan yang berusaha mengidentifikasi korelasi antara atribut *input* dan atribut target[5]. Metode klasifikasi akan menentukan label pada kelas terhadap sampel yang ada. Dalam konteks klasifikasi, tujuannya yaitu untuk melakukan

pengembangan terhadap suatu model yang mampu mengkategorikan item data ke dalam kelompok-kelompok tertentu berdasarkan informasi yang ada pada data tersebut [6]. Model tersebut merupakan suatu aspek yang krusial dalam menjalankan pengambilan keputusan. Berbagai metode klasifikasi yang umumnya dikenal meliputi *decision tree*, klasifikasi *Bayesian*/klasifikasi *Naïve Bayes*, jaringan saraf, analisis statistik, algoritma genetika, *rough sets*, *k-nearest neighbor*, metode *rule based*, *memory based reasoning*, dan *Support Vector Machines (SVM)*[7]. Namun, umumnya algoritma digunakan yaitu *Decision tree* dan *Naïve Bayes*. Pengklasifikasian menggunakan *decision tree* mendapatkan hasil akurasi yang sama, bahkan bisa lebih baik dari pengklasifikasian menggunakan algoritma lainnya[8]. Hal ini dikarenakan pada daerah pengambilan keputusan menggunakan *decision tree* yang kompleks bisa disederhanakan ke dalam bentuk yang lebih spesifik. Dalam melakukan klasifikasi, *decision tree* merupakan teknik yang dibangun oleh algoritma C4.5[9]. Dimana, *decision tree* akan berbentuk *flowchart* atau akar untuk menghasilkan sebuah keputusan.

Algoritma C4.5 digunakan untuk mengelompokkan data melalui pembangunan *decision tree*. Sebelum *decision tree* dibangun, langkah yang paling krusial yaitu menetapkan atribut sebagai simpul akar[10]. Atribut tersebut nantinya akan memiliki cabang yang saling terhubung satu sama lain. Algoritma C4.5 menjadi bagian dari beberapa metode perhitungan yang dimanfaatkan untuk tujuan klasifikasi atau pembagian dengan sifat prediktif[11]. Salah satu contohnya yaitu pada klasifikasi terhadap status gizi balita. *Decision tree* pada algoritma C4.5 termasuk sebuah metode klasifikasi dan prediksi yang banyak diketahui,

hal ini dikarenakan *decision tree* merupakan algoritma yang mudah dipahami serta dapat diinterpretasikan dengan baik[12]. Oleh karena itu, algoritma C4.5 sesuai untuk memasukkan status gizi balita ke dalam kategori atau golongan.

Status gizi balita bisa didapatkan melalui survei dengan memperkirakan massa tubuh manusia, biasanya diketahui dengan Antropometri. Dimana, Antropometri dikenal sebagai suatu metode yang biasa digunakan dengan tujuan untuk mengevaluasi dimensi, rasio, dan struktur tubuh pada manusia[13]. Antropometri diukur dengan mempertimbangkan berat badan sesuai dengan usia (BB/U), panjang atau tinggi badan relatif terhadap usia (PB/U atau TB/U), berat badan yang proporsional bagi panjang/tinggi badan (BB/PB atau BB/TB), serta ukuran berat badan relatif sesuai dengan umur (IMT/U). Kategori status gizi mencakup gizi buruk, gizi kurang, gizi baik, dan gizi lebih[14]. Klasifikasi untuk status gizi balita di Puskesmas Lambur akan menggunakan kategori, yakni gizi kurang, gizi baik, dan gizi lebih.

Studi yang dilaksanakan oleh beberapa peneliti sebelumnya terhadap pengklasifikasian status gizi pada balita sudah pernah dilakukan. Seperti studi yang telah dijalankan oleh Sri Lestari dan Amanda Amalia yaitu (Penerapan Algoritma C4.5 Pada Klasifikasi Status Gizi Balita di Posyandu Desa Sukalilah Cibatu Kabupaten Garut Jawa Barat), dimana akurasi yang didapatkan sebesar 95,50% terhadap status gizi baik dan gizi buruk[13]. Kategori status gizi yang digunakan pada penelitian tersebut hanya gizi baik dan gizi buruk. Selain itu terdapat juga penelitian dengan penggunaan algoritma berbeda, yaitu pada penelitian yang dikerjakan Monica Yoshe Titimeidara dan Wiwien

Hadikurniawati dengan judul (Implementasi Metode *Naïve Bayes Classifier* Untuk Klasifikasi Status Gizi Stunting Pada Balita) yang menghasilkan akurasi yang berbeda dari C4.5, dengan tingkat akurasi mencapai angka 88% dari jumlah 300 data yang telah dilakukan pembagian data dengan ketentuan 275 sebagai *data training* dan 25 sebagai *data testing*[15]. Masalah status gizi balita juga terdapat pada wilayah Puskesmas Lambur.

Puskesmas Lambur sendiri merupakan sebuah puskesmas yang terletak di Jl.Beringin N0.4, Kec.Muara Sabak Timur, Kab.Tanjung Jabung Timur. Dalam aktivitasnya, Puskesmas Lambur aktif melakukan pelayanan terhadap masyarakat mulai dari promosi kesehatan, imunisasi serta kesehatan ibu dan anak, membantu puskesmas lain dalam rangka jejaring sosial, hingga pengukuran status gizi pada balita.

Puskesmas Lambur selalu melakukan pengukuran status gizi terhadap balita tiap bulannya. Hasil dari pengukuran tersebut direpresentasikan melalui dataset status gizi balita Puskesmas Lambur. Dengan begitu, akan terjadi penumpukan data pada histori status gizi di Puskesmas Lambur, dan tidak terdapat pelaksanaan manfaat lebih lanjut dari data-data yang telah dikumpulkan. Apabila data-data tersebut dapat dilakukan proses menggunakan sebuah teknik pengolahan data, maka data-data tersebut dapat menjadi pengetahuan dan informasi yang bermanfaat bagi Puskesmas Lambur, sehingga pihak Puskesmas Lambur dapat memperoleh referensi dalam menangani masalah gizi buruk pada Puskesmas Lambur. Dengan adanya permasalahan tersebut, maka dibutuhkan pengolahan data lebih lanjut untuk mendapatkan informasi dari dataset status gizi balita

Puskesmas Lambur. Pengolahan dataset status gizi balita bisa dijalankan berdasarkan pemanfaatan teknik *data mining* serta algoritma C4.5 untuk pengklasifikasian dataset status gizi balita Puskesmas Lambur. Dalam penelitian ini, total data yang akan digunakan sebanyak 800 data status gizi balita pada bulan Juli hingga September tahun 2023 dengan 12 atribut dan 1 label.

Berdasarkan permasalahan yang telah diketahui, maka ditetapkan sebuah judul penelitian yang melatar belakangi tugas akhir ini yaitu **“Analisa Dan Penerapan Algoritma C4.5 Dalam Melakukan Klasifikasi Terhadap Status Gizi Balita (Studi Kasus : Puskesmas Lambur)”**.

1.2 RUMUSAN MASALAH

Dengan diketahuinya masalah penelitian yang dilakukan, untuk itu ditetapkan beberapa rumusan masalah penelitian. Adapun rumusan masalah yang ditetapkan yaitu :

1. Bagaimana memperoleh dan menganalisa dataset status gizi balita Puskesmas Lambur ?
2. Bagaimana algoritma C4.5 diterapkan dalam melakukan klasifikasi pada data status gizi balita di Puskesmas Lambur ?
3. Seberapa akurat hasil dari algoritma C4.5 dalam pengklasifikasian dataset status gizi balita Puskesmas Lambur ?

1.3 BATASAN MASALAH

Agar menunjang keberhasilan dengan merujuk pada hasil studi ini, maka pengertian batasan masalah menjadi diperlukan dengan tujuan penelitian berlangsung dengan lancar serta terfokus pada pembahasan. Adapun batasan masalah yang perlu diketahui yaitu :

1. Penelitian ini hanya memfokuskan pada dataset mengenai status gizi balita di Puskesmas Lambur, yang merupakan objek penelitian ini.
2. Penggunaan data dalam penelitian ini merupakan data primer berdasarkan wawancara kepada pihak Puskesmas Lambur mengenai dataset status gizi balita berumur 0-4 tahun pada bulan Juli-September tahun 2023 sebanyak 800 data.
3. Penelitian ini hanya dilakukan pada wilayah Puskesmas Lambur.
4. Pada penelitian ini menggunakan algoritma C4.5 sebagai pengklasifikasian terhadap dataset status gizi Puskesmas Lambur.
5. Pengukuran kinerja pada algoritma C4.5 ini menggunakan nilai dari *accuracy*, *recall*, dan *precision*.
6. Alat bantu yang digunakan yaitu *Rapidminer*.

1.4 TUJUAN PENELITIAN

Setelah mengidentifikasi permasalahan yang terkandung dalam rumusan masalah, tujuan penelitian yang diinginkan dapat diketahui dengan jelas. Tujuan penelitian ini mencakup :

1. Untuk memperoleh dan menganalisa dataset status gizi balita di wilayah Puskesmas Lambur.
2. Untuk menerapkan algoritma C4.5 dalam melakukan klasifikasi terhadap dataset status gizi balita Puskesmas Lambur.
3. Untuk mengetahui tingkat akurasi algoritma C4.5 ketika melakukan pengklasifikasian dataset status gizi balita Puskesmas Lambur.

1.5 MANFAAT PENELITIAN

Penelitian terhadap klasifikasi status gizi balita Puskesmas Lambur memiliki beberapa manfaat yang berguna bagi pihak tertentu, khususnya pihak Puskesmas Lambur. Dapat diketahui manfaat yang akan diperoleh berdasarkan hasil penelitian ini yaitu :

1. Penelitian ini bisa dijadikan sebagai acuan dalam pengembangan algoritma untuk melakukan klasifikasi terhadap status gizi balita.
2. Dari hasil penelitian ini mampu menyediakan saran kepada Puskesmas Lambur untuk melakukan peningkatan pada keakuratan dalam mengklasifikasikan status gizi balita.
3. Penelitian ini mampu memperlihatkan seberapa efektif tingkat keakuratan yang dihasilkan oleh algoritma C4.5 terhadap dataset mengenai status gizi balita di Puskesmas Lambur.

1.6 SISTEMATIKA PENULISAN

Agar lebih mudah dalam melakukan pemahaman pada Laporan Tugas Akhir, maka terbentuklah sistematika penulisan yang akan diterapkan meliputi :

- BAB I : PENDAHULUAN

Pada bagian pendahuluan terdapat proses mengupas topik mengenai latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, termasuk sistematika penulisan yang menjadi dasar dilaksanakannya penelitian.

-BAB II : LANDASAN TEORI

Bab landasan teori ini berisikan teori-teori yang digunakan juga beberapa sumber yang berkaitan langsung dengan isu yang akan diatasi. Sementara itu, teori yang akan dibahas meliputi *data mining*, klasifikasi, algoritma C4.5, Status Gizi, *Rapidminer*, *Confusion Matrix*, *k-fold cross-validation* serta penelitian sejenis.

-BAB III : METODOLOGI PENELITIAN

Dengan tujuan menjelaskan isi dari penelitian, maka metodologi penelitian memiliki fungsi membahas proses tersebut. Pada bab ini akan dibahas mulai dari kerangka kerja penelitian, tahapan *data mining*, serta alat dengan pemanfaatannya selama penelitian berlangsung.

-BAB IV : ANALISIS DAN HASIL

Dalam hal ini akan dilakukan analisis pada dataset status gizi balita Puskesmas Lambur dan pengukuran nilai dari *recall*, *precision*, serta hasil *accuracy* dataset status gizi balita di Puskesmas Lambur dengan algoritma C4.5 menggunakan *tools Rapidminer*.

-BAB V : PENUTUP

Pada bab ini, bagian akhir dari penelitian akan dibahas. Bab penutup merupakan penjelasan mengenai kesimpulan pada penelitian yang dilaksanakan dan juga terdapat saran yang akan memiliki manfaat atau keuntungan bagi pihak-pihak yang berpartisipasi dan ikut serta dalam proses penelitian ini.