

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Salah satu dari tujuh belas poin pembangunan berkelanjutan atau *Sustainable Development Goals* (SDGs) adalah tanpa kelaparan. Goals pada SDGs nomor dua ini yaitu menghilangkan kelaparan dan segala bentuk malnutrisi dan mencapai ketahanan pangan sesuai yang tertera pada target SDG tanpa kelaparan. Pada poin nomor 2.2 menegaskan visi di tahun 2030 untuk menghilangkan segala bentuk kekurangan gizi untuk anak pendek dan kurus di bawah usia 5 tahun, dan memenuhi kebutuhan gizi remaja perempuan, ibu hamil dan menyusui, serta manula. Karena itu masalah gizi masih menjadi masalah yang krusial saat ini. Salah satu masalah gizi yang akan dibahas dalam penelitian ini adalah *stunting* (Badan Perencanaan Pembangunan Nasional, 2023).

Definisi *stunting* menurut WHO (2015) adalah keadaan malnutrisi atau kurang gizi pada masa pertumbuhan dan perkembangan anak, yang ditandai dengan panjang atau tinggi badannya di bawah standar. Penentuan kondisi *stunting* tidak bisa terlepas dari status gizi, status gizi anak dapat ditentukan dengan menggunakan pengukuran antropometri. Menurut Saptawati Bandarsono (2019), perhitungan antropometri dalam status gizi anak digunakan sebagai indikator utama untuk menilai kecukupan gizi dan pertumbuhan bayi dan balita.

Standar antropometri anak di Indonesia yang tercantum pada Peraturan Menteri Kesehatan Republik Indonesia nomor 2 tahun 2020 menjadi acuan, di mana standar antropometri sendiri adalah kumpulan data tentang ukuran, proporsi, komposisi tubuh sebagai rujukan untuk menilai status gizi dan tren pertumbuhan anak dengan menggunakan indikator seperti berat badan, tinggi badan, usia, serta jenis kelamin. Apabila nilai z-score antropometri anak kurang dari minus 2 standar deviasi (SD) berdasarkan indeks panjang badan menurut umur (PB/U) atau tinggi badan menurut umur (TB/U) maka anak termasuk dalam indikasi *stunting*. Padahal masa kecil sangat membutuhkan lebih banyak nutrisi dari segi kuantitas dan kualitas, karena anak usia dini pada umumnya memiliki tingkat aktivitas fisik yang cukup tinggi karena masih dalam proses perkembangan. Adapun beberapa faktor penyebab

kekurangan gizi pada balita, mulai dari faktor pangan seperti kekurangan protein, karbohidrat, kualitas makanan, hingga faktor sosial-ekonomi, lingkungan, genetik dan lain sebagainya (Zogara & Pantaleon, 2020). Permasalahan gizi juga dapat berasal dari lingkup terdekat yaitu rumah tangga atau keluarga. Masalah pada tingkat rumah tangga seringkali terjadi karena ketidakmampuan dalam memperoleh akses terhadap pangan yang berkualitas untuk semua anggota keluarga, serta bagaimana mengolah, dan menyajikan makanan yang bergizi (Sediaoetama, 2009). Karena itu asupan gizi yang buruk, terutama pada periode pertumbuhan dan perkembangan anak yang kritis (*golden age*) dapat menyebabkan *stunting* atau gagal tumbuh. Berdasarkan studi yang dilakukan oleh (Soliman et al., 2021), dan juga oleh (Dewey & Begum, 2011) *stunting* memiliki efek jangka pendek seperti, keterlambatan perkembangan kognitif, motorik, dan bahasa, risiko kecacatan, penyakit menular, dan mortalitas. Adapun juga efek jangka panjang, risiko penyakit degeneratif, seperti tekanan darah tinggi, diabetes, koroner, penyakit jantung, dan stroke, rentan mengalami gangguan mental dan emosional. Dampak lebih lanjut tentu akan sangat terasa pada masa dewasa misalnya berkurangnya efisiensi kerja atau produktivitas karena keterbatasan fisik dan kognitif.

Berdasarkan data dari hasil Survei Status Gizi Indonesia (SSGI) Tahun 2022, kesungguhan dalam menangani permasalahan *stunting* dari pemerintah terlihat dari tren *stunting* yang dari tahun ke tahun menunjukkan penurunan sejak 2013 yaitu 37,6% hingga 2022 yaitu 21,6%. Meskipun terjadi penurunan, namun *stunting* masih menjadi tantangan pemerintah karena target angka prevalensi *stunting* dalam Rencana Pembangunan Jangka Menengah Nasional yaitu sebesar 14% di tahun 2024. Angka prevalensi *stunting* di Jawa Timur adalah 19,2% dimana tahun sebelumnya 23,5%. Sementara itu, menjadi kota terbesar di Indonesia bukan berarti akan mempunyai kesejahteraan gizi balita yang selalu baik karena data mencatat angka prevalensi di Surabaya sendiri adalah 28,9% di tahun 2021. Hebatnya, penurunan yang luar biasa drastis ditunjukkan Surabaya dengan berhasil mencapai angka prevalensi sebesar 4,8% pada tahun 2022. Meskipun tren menunjukkan penurunan, *monitoring* tetap perlu dilakukan untuk tetap menjaga hal baik ini. Dengan demikian, diperlukan sebuah metode pengolahan data yang dapat mengidentifikasi status gizi balita *stunting*.

*Data mining* dengan teknik klasifikasi pada status gizi balita *stunting* dapat dilakukan untuk membantu dalam mengidentifikasi balita yang mengalami *stunting* dan memberikan pengukuran yang objektif tentang status gizi mereka. Selain itu, klasifikasi gizi balita *stunting* juga membantu dalam memonitor dan mengevaluasi efektivitas program-program gizi serta intervensi yang telah dilakukan untuk mengatasi masalah *stunting*. Dan juga manfaat lain dari klasifikasi ini yaitu untuk dapat membantu tenaga kesehatan dalam mendeteksi balita yang mengalami *stunting* di bawah unit kegiatan yang menjadi tanggung jawabnya (Saeful Bachri et al., 2021). Dalam hal ini, metode yang digunakan dalam klasifikasi status gizi balita *stunting* adalah algoritma C4.5 yang termasuk dalam pendekatan pohon keputusan dan *naive bayes* yang menggunakan pendekatan berbasis probabilitas kemunculan kelas.

*Naive bayes* adalah satu algoritma klasifikasi statistik yang mempunyai tujuan untuk memperkirakan peluang yang akan terjadi di masa depan berdasarkan pengalaman masa lalu (*historical data*). *Naive bayes* dinamakan "*naive*" karena metodenya memiliki asumsi sederhana dan idealistik bahwa setiap fitur atau atribut dalam data independen satu sama lain. Karena sifat konstruksinya yang sederhana dan tidak membutuhkan estimasi parameter yang kompleks sehingga menghasilkan komputasi yang cepat (Wu et al., 2008a). *Naive bayes* pada beberapa kasus bisa jadi tidak realistis dan terkesan membatasi karena asumsi independensi pada setiap fitur. Hal ini terutama terlihat ketika variabel-variabel data yang digunakan saling terkait dan saling mempengaruhi satu sama lain (Ian H. Witten et al., 2011). Oleh karena itu, dibutuhkan metode klasifikasi lain untuk menangani kasus data yang memiliki keterikatan antara variabel dalam dataset.

Berbeda dengan model klasifikasi *naive bayes*, C4.5 membangun pohon keputusan dari data pelatihan sementara *Naive Bayes* menggunakan probabilitas untuk menentukan label kelas. C4.5 juga tidak mengasumsikan independensi fitur seperti yang dilakukan oleh *Naive Bayes*. C4.5 termasuk model klasifikasi yang dapat menangani data yang memiliki keterikatan atau hubungan antara variabel. Meskipun C4.5 secara *default* tidak membuat asumsi tentang keterikatan antara variabel dalam dataset, algoritma ini dapat menangkap hubungan antara variabel yang saling terkait. C4.5 melakukan pemilihan fitur dengan menggunakan

pengukuran seperti gain information, yang dapat menangkap hubungan antara fitur dengan kelas target. Dengan karakteristik pendekatan pohon keputusan algoritma C4.5 memiliki beberapa kelebihan, yaitu dapat menangani data kategorik dan kontinu, tidak memerlukan asumsi tertentu tentang hubungan antara variabel, dapat menangani data yang hilang atau tidak lengkap, cocok untuk data yang sedikit maupun kompleks, dapat memangkas cabang yang tidak dibutuhkan. Salah satu kelemahan C4.5 adalah penggunaannya yang membutuhkan waktu dan memori CPU yang cukup besar untuk melakukan pembelajaran dan konstruksi pohon keputusan (Wu et al., 2008a).

Terdapat beberapa penelitian terkait penggunaan C4.5 ataupun *naive bayes* untuk klasifikasi. Pada penelitian yang dilakukan oleh (Herliansyah et al., 2021) diperoleh hasil akurasi dari metode algoritma *naive bayes* sebesar 64.02% pada pengujian partisi data 90:10 dan pada pengujian *K-Fold Cross Validation* didapatkan akurasi sebesar 64.36% dalam klasifikasi balita stunting. Dengan menggunakan atribut jenis kelamin, tinggi badan, dan umur balita. Menyambung dari penelitian tersebut, penulis ingin menambahkan metode *discretization* pada tahap *preprocessing* dengan mengelompokkan data menjadi interval kategori yang diharapkan dapat mempengaruhi kinerja model. Selain itu, menambahkan atribut yang digunakan dan hanya mengimplementasikan *naive bayes* saja untuk dibandingkan nantinya dengan *decision tree* C4.5. Algoritma C4.5 juga pernah dilakukan penelitian untuk mengidentifikasi status gizi balita oleh (Firdausia Ismi Nurhayati et al., 2021) yang ditentukan berdasarkan 11 atribut, yaitu jenis kelamin, usia, berat badan, tinggi badan, intensitas asi, penghasilan keluarga, status kerja ibu, frekuensi sakit yang diderita balita dalam 1 bulan, jumlah anak dalam keluarga, pendidikan ibu, serta akses kesehatan yang dimiliki dengan didapatkan hasil akurasi sistem sebesar 85% dengan perbandingan data latih dan data uji sebesar 70 : 30. Komparasi kedua algoritma tersebut juga pernah dilakukan dalam penelitian sebelumnya dalam penentuan status kelayakan donor darah oleh (Handayani et al., 2021) menggunakan atribut jenis kelamin, umur, HB, tensi, dan berat badan dengan hasil akurasi 93,22% untuk C4.5 dan 93,26% untuk *Naive Bayes*.

Dalam penelitian ini akan dibandingkan hasil klasifikasi dari dua metode algoritma, yakni C4.5 dengan pendekatan berbasis pohon keputusan dan *naive*

*bayes* yang menggunakan pendekatan berbasis probabilitas kemunculan kelas. Terdapat asumsi bahwa teknik *discretization* pada tahap proses *preprocessing* dan juga perbedaan pendekatan masing-masing metode akan berpengaruh pada hasil klasifikasi. Analisis metode terbaik untuk mengklasifikasikan status gizi balita *stunting* dapat dilakukan dengan metode *confusion matrix* untuk mengukur efektivitas suatu metode klasifikasi. Dari dasar tersebut, lebih lanjutnya dapat dianalisis menggunakan *matrix* akurasi, presisi, *recall*, *F1-score*, dan *Area Under the Curve* (AUC) pada kurva *Receiver Operating Characteristic* (ROC) untuk membandingkan performa klasifikasi antara dua metode tersebut.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan di atas, maka berikut ini rumusan masalah yang akan diselesaikan pada penelitian ini:

1. Bagaimana implementasi metode *decision tree* C4.5 dan *naive bayes* untuk mengklasifikasikan status gizi *stunting* balita?
2. Bagaimana perbandingan performa algoritma *decision tree* C4.5 dan *naive bayes* untuk mengklasifikasikan status gizi *stunting* balita?

## 1.3 Tujuan Penelitian

Berdasarkan latar belakang dan permasalahan yang telah dijelaskan, maka tujuan dari tugas akhir ini adalah:

1. Mengetahui implementasi *decision tree* C4.5 dan *naive bayes* untuk mengklasifikasikan status gizi balita *stunting*.
2. Membandingkan hasil klasifikasi dan mengukur performa serta tingkat validasi keputusan yang dihasilkan dari model metode C4.5 dan *naive bayes* dalam klasifikasi status gizi balita *stunting*.

## 1.4 Manfaat Penelitian

Harapannya hasil dari penelitian ini akan memberikan manfaat diantaranya:

1. Menambah pengetahuan akan implementasi *data mining* menggunakan metode *decision tree* C4.5 dan *naive bayes* pada studi kasus klasifikasi status gizi balita *stunting*.
2. Mengetahui algoritma manakah diantara C4.5 dan *naive bayes classifier* yang menghasilkan performa terbaik dalam mengklasifikasikan status *stunting*.

3. Memberikan dukungan penelitian kepada tenaga ahli dan masyarakat untuk menjadi pertimbangan preventif dalam mencegah risiko kejadian balita *stunting* menggunakan *data mining*.

### **1.5 Batasan Masalah**

Penulis merumuskan batasan permasalahan agar ketidakpastian yang timbul dapat dikurangi, yakni sebagai berikut:

1. Data yang digunakan berupa data sekunder yang diperoleh dari puskesmas Jagir Surabaya, yaitu data gizi balita tahun 2021.
2. Atribut data yang digunakan untuk proses klasifikasi, yaitu jenis kelamin, umur, berat badan, tinggi badan, berat badan waktu lahir, lingkaran lengan atas.