

BAB I

PENDAHULUAN

Pada bab ini memaparkan mengenai permasalahan yang terjadi sehingga dapat menjadi landasan alasan penelitian ini dilakukan. Bab ini penting untuk memberikan wawasan dan pemahaman awal kepada pembaca mengenai permasalahan yang diangkat dalam penelitian ini.

1.1 Latar Belakang

Status gizi merupakan indikator penting untuk menilai pertumbuhan dan perkembangan balita. Masalah gizi seperti gizi buruk, wasting, stunting, hingga obesitas masih menjadi tantangan kesehatan di banyak negara termasuk Indonesia [1]. Kelompok yang paling sering mengalami masalah gizi salah satunya adalah balita [2]. Hasil Survei Status Gizi Indonesia (SSGI) 2022 menunjukkan prevalensi stunting 21,6%, wasting 7,1%, dan gizi buruk 3,4% (Kemenkes RI, 2022). Kondisi gizi yang tidak optimal pada balita berpotensi menimbulkan gangguan perkembangan kognitif, penyakit kronis, hingga meningkatkan risiko kematian. Kesalahan diagnosis status gizi – terutama pada kategori ekstrem seperti gizi buruk dan obesitas – dapat berakibat fatal, dan kategori “gizi buruk” maupun “obesitas” sering terabaikan dalam pendekatan manual [3]. Oleh karena itu, pemantauan dan klasifikasi status gizi balita yang akurat menjadi sangat penting sebagai dasar perencanaan intervensi gizi dan peningkatan kualitas sumber daya manusia di masa depan.

Salah satu indikator yang banyak digunakan untuk menilai status gizi balita adalah indeks berat badan terhadap tinggi badan (BB/TB). Indeks ini bersifat independen terhadap umur dan sensitif terhadap perubahan berat badan dalam jangka pendek, sehingga efektif untuk mendeteksi gangguan pertumbuhan akut serta mengidentifikasi kondisi seperti gizi kurang, gizi buruk, maupun gizi lebih [4]. Sensitivitas tersebut menjadikan BB/TB cocok untuk memantau efek intervensi gizi jangka pendek [4]. Namun, proses klasifikasi status gizi menggunakan data antropometri (tinggi badan, berat badan, dan Z-score) memerlukan analisis data yang kompleks. Pada konteks ini, teknik data mining dan machine learning dapat

dimanfaatkan untuk membantu proses klasifikasi secara lebih objektif dan konsisten.

Dalam perkembangan machine learning, pendekatan ensemble dan deep learning semakin mendominasi karena mampu meningkatkan akurasi prediksi pada berbagai domain [5]. Salah satu metode ensemble yang banyak digunakan adalah gradient boosting, yang dalam beberapa tahun terakhir menjadi sangat populer untuk tugas klasifikasi karena mampu meningkatkan akurasi dan kecepatan model [6]. XGBoost (Extreme Gradient Boosting) merupakan varian gradient boosting berbasis ensemble yang dirancang agar lebih efisien dan berperforma tinggi dalam tugas klasifikasi maupun regresi [7]. XGBoost juga memiliki keunggulan dalam menangani dataset berdimensi tinggi, toleran terhadap missing value, dan menyediakan mekanisme regularisasi untuk mengurangi overfitting [7]. Namun, keefektifan XGBoost sangat bergantung pada pemilihan *hyperparameter*, seperti *learning_rate*, *max_depth*, *n_estimators*, dan *subsample*, sehingga diperlukan proses tuning yang tepat.

Optimasi *hyperparameter* dapat dilakukan dengan beberapa pendekatan, di antaranya Grid Search dan Random Search. Grid Search merupakan cara sistematis dengan membentuk kisi (grid) kombinasi nilai *hyperparameter* yang telah ditentukan, kemudian mengevaluasi setiap kombinasi untuk mencari konfigurasi terbaik [8]. Metode ini cocok ketika ruang parameter relatif sempit dan terstruktur, sehingga eksplorasi menyeluruh masih memungkinkan, dan sering digunakan untuk eksplorasi efek *hyperparameter* pada suatu model [8]. Pada model XGBoost, Grid Search menawarkan pendekatan yang sistematis dan menyeluruh untuk mengeksplorasi kombinasi *hyperparameter* utama, sehingga berpotensi meningkatkan kinerja model secara signifikan [9]. Di sisi lain, Random Search memilih kombinasi *hyperparameter* secara acak dari rentang nilai yang sudah ditentukan [10]. Tidak semua kombinasi dievaluasi, namun secara praktik Random Search lebih efisien pada ruang parameter yang besar, karena waktu komputasi lebih rendah dan tetap mampu menemukan area parameter yang baik meskipun eksplorasi tidak dilakukan secara menyeluruh [10]. Metode ini juga dapat

dikombinasikan dengan pendekatan lain untuk memperdalam eksplorasi di sekitar kombinasi yang menjanjikan [10].

Penelitian sebelumnya menunjukkan potensi XGBoost yang kuat untuk klasifikasi status gizi dan permasalahan kesehatan lain. Salah satu penelitian [11] membandingkan AdaBoost, Gradient Boosting, dan XGBoost untuk klasifikasi status gizi balita, dan melaporkan bahwa XGBoost memberikan performa terbaik dengan pembagian data 80:20, yaitu *precision* 0,9849; *recall* 0,9848; *accuracy* 0,9848; F1-score 0,9848; dan ROC-AUC 0,9994 [11]. Keunggulan ini dikaitkan dengan teknik regularisasi yang efektif, kemampuan menangani missing values, dan algoritma boosting yang efisien dengan teknik paralelisasi [11]. Namun, kombinasi *hyperparameter* yang digunakan masih bersifat default sehingga potensi peningkatan performa melalui tuning belum dimanfaatkan secara optimal. Penelitian lain [7] menerapkan optimasi XGBoost menggunakan Grid Search dan Random Search pada klasifikasi penyakit diabetes. Dataset yang digunakan berasal dari UCI Machine Learning Repository dengan 8 variabel input, 1 variabel target, dan 768 record. Hasilnya menunjukkan bahwa tanpa tuning *hyperparameter*, XGBoost menghasilkan log loss negatif sekitar 25% (akurasi $\pm 75\%$), sedangkan setelah dilakukan tuning menggunakan Grid Search dan Random Search, log loss menurun hingga sekitar 5% (akurasi $\pm 95\%$) [7]. Hal ini menunjukkan bahwa kedua metode tuning tersebut mampu meningkatkan kinerja model secara signifikan.

Berdasarkan uraian tersebut, penelitian ini melakukan komparasi Grid Search dan Random Search dalam optimasi *hyperparameter* XGBoost untuk klasifikasi status gizi balita berbasis antropometri BB/TB. Penelitian tidak hanya mengevaluasi peningkatan performa klasifikasi secara keseluruhan, termasuk pada kelas minoritas seperti gizi buruk dan obesitas, tetapi juga menganalisis efisiensi komputasi masing-masing metode sehingga dapat diperoleh rekomendasi praktis dalam pemilihan strategi *hyperparameter* tuning. Tujuannya adalah agar optimasi XGBoost menggunakan Grid Search dan Random Search dapat membantu menemukan kombinasi parameter dengan akurasi terbaik, menurunkan log loss, dan meningkatkan kemampuan model dalam membedakan seluruh kategori status gizi balita. Dengan demikian, diharapkan dapat dihasilkan sistem klasifikasi status gizi

balita yang lebih akurat dan efisien untuk mendukung pengambilan keputusan di bidang kesehatan masyarakat.

1.2. Rumusan Masalah

Untuk meningkatkan akurasi, metode optimasi *hyperparameter* seperti Grid Search dan Random Search dapat digunakan. Grid Search memungkinkan pencarian kombinasi *hyperparameter* terbaik secara sistematis melalui evaluasi berbagai konfigurasi dan Random Search melakukan pencarian kombinasi *hyperparameter* terbaik secara acak. Berdasarkan hal ini, muncul pertanyaan utama

1. Bagaimana pengaruh proses tuning *hyperparameter* menggunakan metode Grid Search dan Random Search terhadap kinerja model XGBoost?
2. Apakah model hasil optimasi dengan Grid Search dan Random Search lebih efisien dalam meningkatkan kinerja klasifikasi dibandingkan model tanpa optimasi?

1.3. Batasan Masalah

Dalam penelitian ini, batasan masalah yang ditetapkan untuk menjaga fokus tujuan utama penelitian adalah sebagai berikut.

1. Dataset yang digunakan dataset status gizi balita yang mencakup indikator antropometri balita usia 24 - 60 bulan.
2. Model yang diteliti difokuskan pada algoritma XGBoost sebagai algoritma utama.
3. Teknik optimasi *hyperparameter* yang digunakan hanya *hyperparameter* Grid Search dan Random Search
4. Penilaian performa model hanya didasarkan metrik evaluasi umum berupa *Confusion matrix*, log loss, dan komputasi waktunya.
5. Penelitian ini hanya berfokus pada peningkatan performa model tanpa terlibat dalam implementasi secara langsung ke dalam sistem kesehatan masyarakat atau aplikasi berbasis teknologi.

1.4. Tujuan

Berdasarkan rumusan masalah sebelumnya yang telah dijelaskan oleh peneliti, berikut adalah tujuan dari penelitian ini :

1. Mengembangkan model klasifikasi yang lebih optimal untuk klasifikasi status gizi balita berbasis BB/TB dengan mengoptimalkan Algoritma XGBoost (*Extreme Gradient Boosting*) menggunakan *hyperparameter* Grid Search dan Random Search.
2. Menerapkan optimasi *hyperparameter* dengan metode Grid Search dan Random Search untuk menemukan kombinasi parameter terbaik
3. Menganalisis pengaruh tuning *hyperparameter* terhadap kinerja model XGBoost dengan membandingkan hasil evaluasi antara parameter default dengan tuning Grid Search dan Random Search.
4. Memberikan konfigurasi *hyperparameter* terbaik untuk meningkatkan performa model.

1.5. Manfaat

Penelitian ini diharapkan dapat memberikan kontribusi dalam bidang kesehatan, akademisi dan peneliti, dan masyarakat dengan beberapa manfaat :

1. Bagi Akademisi dan Peneliti
 - a. Mendorong inovasi dalam pengembangan teknologi untuk penyelesaian masalah kesehatan secara efisien.
 - b. Menjadi referensi dalam pengembangan model klasifikasi berbasis pembelajaran mesin untuk permasalahan kesehatan, khususnya dalam hal pemilihan model, tuning *hyperparameter*, dan penanganan ketidakseimbangan data.
 - c. Menambah wawasan dalam bidang data science dan machine learning, khususnya penerapan XGBoost dalam data klasifikasi multikelas dan pentingnya peran tuning *hyperparameter* dalam meningkatkan performa model.
2. Bagi Bidang Kesehatan
 - a. Mengembangkan sistem klasifikasi status gizi balita yang lebih akurat dan efisien.
 - b. Membantu profesional kesehatan dalam membuat keputusan cepat dan tepat terkait penanganan status gizi balita.
3. Bagi Masyarakat

- a. Membantu menurunkan prevalensi gangguan kesehatan yang disebabkan oleh status gizi buruk pada balita.
- b. Meningkatkan kesadaran akan pentingnya pemantauan status gizi balita sebagai bagian dari kesehatan keluarga.