

BAB I PENDAHULUAN

1.1 Latar Belakang

Stunting adalah permasalahan gizi kronis yang mencerminkan hambatan pertumbuhan anak dalam jangka panjang. Kondisi ini tidak hanya memengaruhi tinggi badan, tetapi juga berdampak negatif pada perkembangan kecerdasan, kesehatan di masa depan, hingga potensi produktivitas saat anak tumbuh dewasa [1,2]. Dalam pengklasifikasian status gizi, anak dapat digolongkan ke dalam empat kategori, yaitu *tall*, *normal*, *stunting*, dan *severely stunted*, berdasarkan acuan pertumbuhan dari WHO [3]. Khususnya di Provinsi Sumatera Utara, angka kejadian stunting masih tergolong tinggi sebagaimana tercatat dalam Survei Kesehatan Indonesia (SKI) tahun 2023. Stunting tidak hanya berdampak pada individu yang mengalaminya, tetapi juga memberi konsekuensi lebih luas terhadap dunia pendidikan dan ekonomi. Oleh karena itu, penting untuk melakukan deteksi stunting sejak dini guna mendukung penanganan yang lebih tepat sasaran [2].

Saat ini, perkembangan teknologi informasi khususnya dalam bidang kecerdasan buatan seperti *Machine Learning* telah banyak dimanfaatkan dalam dunia kesehatan untuk mendeteksi kondisi atau penyakit tertentu. Salah satu algoritma yang dikenal mampu memberikan hasil klasifikasi yang baik adalah *Random Forest*, yang unggul dalam menangani data kompleks dan menghasilkan prediksi yang akurat [4]. Namun, penerapan algoritma klasifikasi seperti *Random Forest* juga memiliki tantangan. Pertama, data yang digunakan sering kali tidak seimbang antar kelas. Kedua, banyak model hanya dijalankan dengan parameter standar (*default*), yang belum tentu menghasilkan kinerja terbaik. [3,4,5].

Beberapa penelitian sebelumnya menunjukkan upaya mengatasi permasalahan tersebut. Misalnya, algoritma seperti Naive Bayes dan Decision Tree telah digunakan, tetapi hasilnya bervariasi. Salah satu studi menunjukkan bahwa *Random Forest* mampu meningkatkan akurasi dari 0,7730 menjadi 0,9578 setelah dilakukan penyesuaian parameter menggunakan *GridSearchCV* [3]. Sayangnya, metode tersebut membutuhkan waktu komputasi yang cukup lama, terutama pada dataset berukuran besar [5,6].

Salah satu cara untuk meningkatkan performa model adalah melalui *Hyperparameter Tuning*, yaitu proses mencari dan mengatur kombinasi parameter terbaik agar akurasi dan efisiensi prediksi dapat ditingkatkan. Proses ini dikenal juga sebagai *fine-tuning* [8]. Beberapa metode umum dalam *fine-tuning* antara lain *Grid Search*, *Randomized Search*, dan metode

yang lebih efisien seperti *Optuna*, yang semuanya digunakan untuk mengeksplorasi ruang parameter secara sistematis [8,9].

Untuk meningkatkan efisiensi, penelitian ini menggunakan *Optuna*, yaitu suatu teknik *tuning hyperparameter* yang lebih cepat dan fleksibel dibandingkan metode tradisional. *Optuna* telah terbukti mampu meningkatkan performa model klasifikasi seperti KNN, SVM, dan *Naive Bayes* dalam berbagai studi [11,12]. Sebagai contoh, akurasi SVM dalam salah satu penelitian meningkat dari 82,80% menjadi 86,20% setelah dioptimasi dengan *Optuna* [11].

Untuk mengatasi permasalahan ketidakseimbangan data antar kelas, digunakan pendekatan SMOTE (Synthetic Minority Oversampling Technique). Teknik ini secara khusus dirancang untuk menyeimbangkan jumlah data pada kelas minoritas dengan menciptakan data sintetis baru berdasarkan pola yang ada. Dengan cara ini, model dapat belajar dari distribusi data yang lebih adil tanpa risiko overfitting akibat duplikasi data. SMOTE telah terbukti efektif dalam meningkatkan performa model klasifikasi pada kasus serupa [13], [14].

Berdasarkan berbagai temuan tersebut, penelitian ini bertujuan untuk meningkatkan akurasi dan keandalan model klasifikasi *stunting* dengan mengoptimalkan algoritma *Random Forest* menggunakan *Optuna*. Dengan pendekatan ini, diharapkan model mampu menghasilkan prediksi yang lebih akurat dan efisien. Oleh karena itu, penelitian ini berjudul: “OPTIMASI ALGORITMA RANDOM FOREST MENGIDENTIFIKASI STUNTING DENGAN OPTUNA BERBASIS MOBILE.”

1.2 Rumusan Masalah

Dalam proses pengidentifikasian status *stunting* menggunakan algoritma *Random Forest*, ditemukan dua permasalahan utama yang memengaruhi performa klasifikasi, yaitu:

1. Distribusi jumlah dataset *stunting* pada masing-masing kelas target tidak merata, sehingga model cenderung bias terhadap kelas - kelas tertentu
2. Model *Random Forest* yang dijalankan menggunakan parameter *default* dalam mengidentifikasi *stunting* belum menghasilkan performa terbaik dalam segi akurasi, *f1-score*. Oleh karena itu, diperlukan teknik optimasi *hyperparameter* agar model dapat bekerja lebih efektif dan akurat dalam klasifikasi.

1.3 Tujuan

Penelitian ini bertujuan untuk meningkatkan performa algoritma *Random Forest* dalam mengidentifikasi *stunting*, melalui penerapan penanganan ketidakseimbangan data menggunakan SMOTE dan optimasi *hyperparameter* menggunakan metode *Optuna*.

1.4 Manfaat

Adapun manfaat yang diharapkan dari penelitian ini adalah:

1. Menyediakan model klasifikasi *stunting* berbasis *Random Forest* yang telah ditingkatkan performanya melalui penyeimbangan data (SMOTE) dan optimasi *hyperparameter* (*Optuna*), sehingga mampu memberikan hasil klasifikasi yang lebih akurat dan adil antar kelas dalam mengidentifikasi *stunting*.
2. Memberikan kontribusi dalam pengembangan penerapan *machine learning* di bidang kesehatan, khususnya dalam menangani isu ketidakseimbangan data dan *tuning parameter* model untuk deteksi *stunting*.

1.5 Ruang Lingkup

Ruang lingkup tugas akhir ini meliputi:

1. Dataset yang digunakan pada tugas akhir ini terdiri dari 2 sumber yaitu:
 - a. Dataset pertama dari situs Kaggle berjudul “*Stunting Wasting Dataset*” berjumlah 100.000 data, terdiri atas 5 atribut, yaitu umur (bulan), jenis kelamin, tinggi badan (cm), berat badan (kg), serta status *stunting* sebagai variabel target.
 - b. Dataset kedua untuk uji coba model yang ideal dari situs Kaggle dengan judul faktor *stunting* berjumlah 10.000 data terdiri atas 8 atribut, yakni *Gender, Age, Birth Weight, Birth Length, Body Weight, Body Length, Breastfeeding*, dan *Stunting* sebagai variabel target.
2. Model *Random Forest* akan dioptimasi menggunakan teknik *Fine Tuning* dengan metode *OPTUNA*.
3. Pembuatan model dilakukan menggunakan bahasa pemrograman *Python* dengan bantuan platform *Google Colab* sebagai lingkungan pengembangannya.
4. Model *Random Forest* disimpan dalam format *pickle* (.pkl) untuk memudahkan integrasi dengan aplikasi *mobile*.
5. Aplikasi *mobile* dibuat menggunakan *Flutter* untuk membangun aplikasi yang responsif.

6. Desain antarmuka aplikasi dibuat menggunakan *Balsamiq* untuk memastikan tampilan yang sederhana dan intuitif.
7. Model akan dievaluasi menggunakan *Confusion matrix*, *AUC (Area Under Curve)*, *ROC (Receiver Operating Characteristic)* untuk melihat seberapa bagus Model dalam klasifikasi.

