

BAB I

PENDAHULUAN

1.1 Latar Belakang

Stunting pada balita merupakan isu kesehatan masyarakat yang serius di banyak negara, termasuk Indonesia. Stunting yang ditandai dengan pertumbuhan tinggi badan yang terhambat, dapat berdampak buruk pada perkembangan fisik dan kognitif anak. Penentuan kondisi stunting tidak terlepas dari status gizi anak yang dapat ditentukan dengan menggunakan pengukuran antropometri. Perhitungan antropometri dalam status gizi anak digunakan sebagai indikator utama untuk menilai kecukupan gizi dan pertumbuhan bayi dan balita.

Tingkat penyebaran stunting di Indonesia pada tahun 2023 berdasarkan data yang didapatkan dari hasil penelitian Survei Kesehatan Indonesia (SKI) menunjukkan tingkat yang cukup tinggi berkisaran pada tingkat sebesar 21,4% atau sebanyak 5,33 juta anak [1]. Pengukuran stunting dilandaskan pada indeks PB / U atau TB / U yang mengukur status gizi balita dengan menggunakan standar antropometri. Bila hasil pengukuran tersebut berada di bawah <-2 standar deviasi (SD), maka hal tersebut menunjukkan bahwa anak tersebut bertubuh pendek atau stunting [2].

Dalam upaya menangani masalah stunting, teknologi dapat berperan penting, terutama dalam analisis data yang dapat membantu mengidentifikasi faktor-faktor penyebab stunting. Salah satu teknologi yang dapat digunakan adalah *machine learning*. *Machine learning* memiliki kemampuan untuk menganalisis data dalam jumlah besar dan menemukan pola-pola yang sulit terdeteksi oleh manusia [3]. Dengan memanfaatkan *machine learning*, dapat dilakukan klasifikasi terhadap balita yang berisiko stunting berdasarkan faktor-faktor seperti usia, pertumbuhan fisik, serta status gizi [4].

Machine learning adalah metode komputasi yang menggunakan pengalaman untuk meningkatkan kinerja dan memiliki metode, teknik serta algoritma yang membangunnya [3]. Secara umum, algoritma *machine learning* dibagi menjadi tiga kategori utama, yaitu *supervised learning*, *unsupervised learning* dan *reinforcement learning*. Di antara ketiga kategori tersebut,

supervised learning menjadi salah satu yang paling banyak digunakan dalam proses klasifikasi, karena kemampuannya untuk memanfaatkan berlabel dalam proses pelatihan [5]. Terdapat banyak algoritma *supervised learning* yang populer dan biasa digunakan untuk klasifikasi karena tingkat akurasi yang tinggi diantaranya adalah *Support Vector Machine* (SVM), *Random Forest* dan *Gradient Boosting*.

Dalam penelitian ini, dilakukan perbandingan performa tiga algoritma *machine learning*, yaitu *Support Vector Machine* (SVM), *Random Forest*, dan *Gradient Boosting* untuk mengklasifikasikan status stunting pada balita. *Random Forest* dan *Gradient Boosting* dipilih karena keduanya berada dalam satu rumpun (*family*) *ensemble learning* berbasis *decision tree*, di mana *Random Forest* menggunakan teknik *bagging* dan *Gradient Boosting* menggunakan teknik pengoptimalan bertahap (*boosting*). Sebagai pembanding dari rumpun yang berbeda, SVM digunakan karena memiliki paradigma kerja berbasis pemisahan *hyperplane* pada ruang vektor. Perbandingan antar-rumpun ini bertujuan untuk menguji model mana yang paling optimal dalam memetakan data antropometri anak.

Ketiga algoritma tersebut memiliki karakteristik yang berbeda dalam menangani data. SVM dikenal efektif untuk pemisahan margin optimal pada data non-linier, meskipun cenderung lebih sulit diinterpretasikan [6]. *Random Forest* lebih mudah dipahami, tangguh terhadap *overfitting*, serta mampu menangani variasi data maupun nilai yang hilang (*missing values*) [7]. Sementara itu, *Gradient Boosting* mampu memberikan akurasi tinggi karena secara bertahap belajar dari kesalahan prediksi model sebelumnya [8].

Penggunaan ketiga algoritma ini juga didukung oleh beberapa penelitian terdahulu. Nugraha (2021) menemukan bahwa SVM menghasilkan akurasi sebesar 80%, lebih tinggi dibandingkan *Gradient Boosting* (78%) dan LightGBM (76%) [9]. Sebaliknya, penelitian Similien dkk. (2022) menunjukkan bahwa *Gradient Boosting* mampu mencapai akurasi 79,33%, mengungguli *Naive Bayes* dan *Extreme Gradient Boosting* [10]. Perbedaan hasil pada studi-studi terdahulu

ini makin memperkuat perlunya pengujian kembali pada dataset antropometri yang digunakan dalam penelitian ini.

Untuk mengevaluasi kinerja ketiga algoritma secara obyektif, penelitian ini menerapkan metode *10-Fold Cross Validation*. Data dibagi menjadi 10 bagian (*folds*), di mana proses pelatihan dan pengujian diulang sebanyak 10 kali secara bergantian [11]. Metode ini digunakan untuk menjamin pembagian data yang adil, menghindari bias, serta mencegah *overfitting* [13].

Melalui evaluasi yang terstruktur ini, perbandingan ketiga algoritma diharapkan dapat memberikan gambaran yang jelas mengenai keunggulan dan keterbatasan masing-masing model. Hasil akhir penelitian ini tidak hanya menentukan algoritma terbaik untuk deteksi dini stunting, tetapi juga dapat menjadi acuan teknis bagi peneliti maupun praktisi dalam memilih metode analisis data kesehatan anak.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, maka perumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana penerapan algoritma *Random Forest*, *Support Vector Machine* (SVM), dan *Gradient Boosting* dalam mengklasifikasikan status stunting pada balita?
2. Bagaimana perbandingan performa ketiga algoritma tersebut berdasarkan metrik evaluasi *Accuracy*, *Precision*, *Recall*, dan *F1-Score*?
3. Algoritma *Machine Learning* manakah yang memberikan kinerja klasifikasi paling optimal dan presisi dalam mendeteksi status stunting pada balita?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang dijelaskan maka tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Mengimplementasikan algoritma *Random Forest*, *Support Vector Machine* (SVM), dan *Gradient Boosting* untuk mengklasifikasikan status stunting pada balita.
2. Menganalisis dan membandingkan performa ketiga algoritma berdasarkan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score*.
3. Menentukan algoritma terbaik yang memiliki tingkat akurasi dan keandalan paling optimal dalam identifikasi status stunting.

1.4 Batasan Masalah

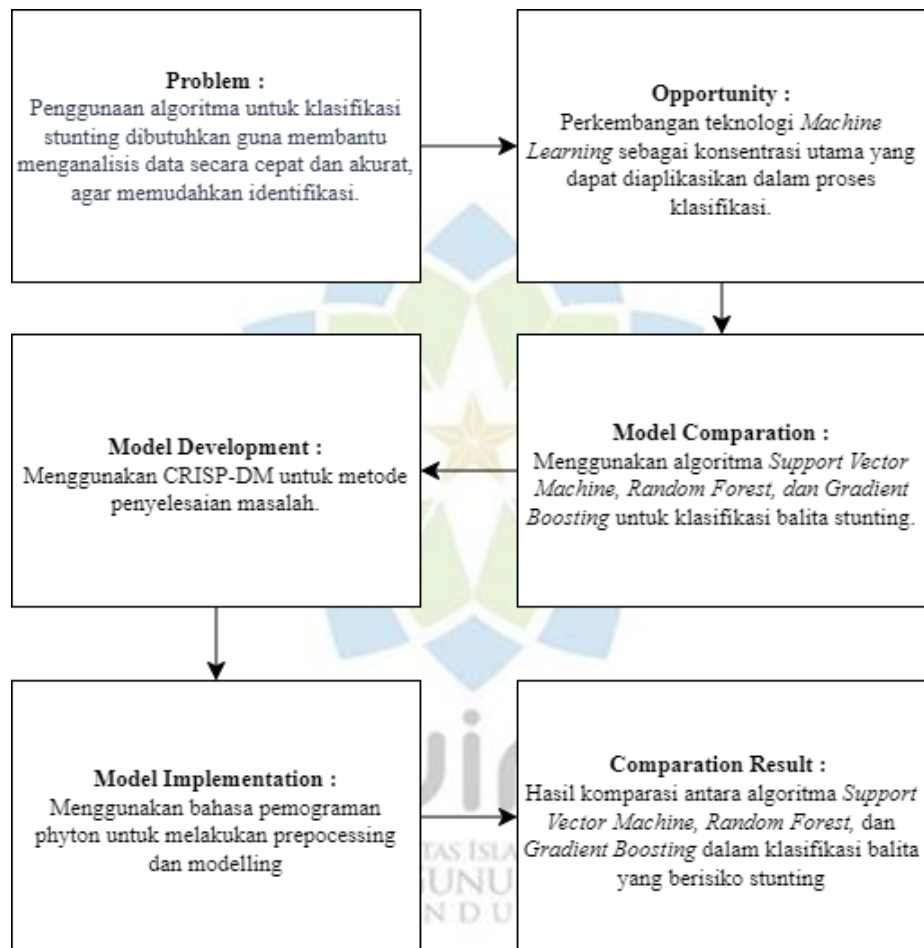
Agar masalah yang dibahas tidak meluas, maka batasan masalah yang akan dibahas pada penelitian ini hanya difokuskan pada:

1. Data yang digunakan merupakan data antropometri balita yang mencakup variabel umur, jenis kelamin, dan tinggi badan.
2. Algoritma *Machine Learning* yang dibandingkan terbatas pada *Random Forest*, *Support Vector Machine* (SVM), dan *Gradient Boosting*.
3. Evaluasi performa model berfokus pada empat metrik utama klasifikasi, yaitu *Accuracy*, *Precision*, *Recall*, dan *F1-Score*.
4. Pengujian keandalan model dilakukan menggunakan metode *Cross-Validation* dan matriks konfusi (*Confusion Matrix*).

UNIVERSITAS ISLAM NEGERI
SUNAN GUNUNG DJATI
BANDUNG

1.5 Kerangka Pemikiran Penelitian

Kerangka pemikiran pada penelitian ini disusun sebagai gambaran alur penelitian yang menunjukkan tahapan penyelesaian permasalahan hingga diperolehnya hasil penelitian.



Gambar 1.1 Kerangka Pemikiran

Gambar 1.1 menunjukkan bahwa penelitian diawali dengan identifikasi permasalahan mengenai kebutuhan penerapan algoritma untuk klasifikasi balita stunting. Selanjutnya, penelitian memanfaatkan perkembangan teknologi *machine learning* dengan menggunakan metode CRISP-DM sebagai kerangka pengembangan model. Proses penelitian dilanjutkan dengan implementasi algoritma *Support Vector Machine* (SVM), *Random Forest*, dan *Gradient Boosting* menggunakan bahasa pemrograman *Python*, kemudian dilakukan

perbandingan kinerja ketiga algoritma untuk memperoleh algoritma dengan performa terbaik dalam klasifikasi balita stunting.

1.6 Manfaat Penelitian

Adapun manfaat yang diharapkan dapat diperoleh dari penelitian ini adalah sebagai berikut:

1. Bagi Akademik

- Memberikan kontribusi ilmiah serta referensi akademis dalam penerapan teknik supervised learning untuk klasifikasi status gizi dan stunting pada balita berbasis data antropometri.
- Menyediakan hasil evaluasi dan perbandingan kinerja secara teruji dan objektif antara algoritma Support Vector Machine (SVM), Random Forest, dan Gradient Boosting dengan menggunakan metode 10-Fold Cross Validation.
- Menjadi acuan bagi penelitian selanjutnya yang mengkaji permasalahan serupa, baik dalam pengembangan pemodelan klasifikasi stunting maupun penerapan komparasi algoritma pada domain kesehatan masyarakat.

2. Bagi Pengguna

- Memberikan gambaran dan acuan teknis bagi instansi/praktisi kesehatan dalam menentukan pilihan algoritma machine learning yang paling optimal untuk klasifikasi stunting.
- Membantu mempermudah proses identifikasi dini risiko stunting pada balita berdasarkan variabel antropometri secara lebih akurat dan efisien.

3. Bagi Penulis

- Menambah wawasan, pemahaman, dan pengalaman praktis dalam mengimplementasikan serta membandingkan performa algoritma Support Vector Machine (SVM), Random Forest, dan Gradient Boosting.

- Melatih kemampuan analisis ilmiah dalam mengaplikasikan metode evaluasi 10-Fold Cross Validation menggunakan bahasa pemrograman Python untuk menyelesaikan permasalahan nyata di bidang kesehatan.

