

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pendapatan per kapita merupakan indikator penting untuk mengukur tingkat kesejahteraan ekonomi suatu wilayah atau negara. Indonesia telah mengalami kemajuan dalam pertumbuhan ekonomi, namun masih terdapat banyak daerah dengan tingkat pendapatan yang rendah, terutama di wilayah terpencil. Oleh karena itu, rendahnya pendapatan per kapita masih menjadi salah satu permasalahan utama terjadinya ketimpangan ekonomi di Indonesia [1].

Permasalahan ini berdampak pada keterbatasan terhadap kebutuhan pangan yang bergizi, sehingga keluarga dengan pendapatan rendah mengalami kesulitan dalam memenuhi kebutuhan nutrisi yang dibutuhkan untuk pertumbuhan optimal anak-anak mereka. Akibatnya, anak-anak dari keluarga kurang mampu lebih rentan mengalami kekurangan gizi, yang menyebabkan salah satu faktor risiko utama terjadinya *stunting*.

Berdasarkan data yang diperoleh dari statistik *World Health Organization* (WHO) memperkirakan prevalensi balita *stunting* di seluruh dunia sebesar 22,3% atau sebanyak 148,1 juta jiwa pada 2022 [2]. Sementara itu, berdasarkan hasil Survei Kesehatan Indonesia (SKI) 2023 rata-rata nasional mencatat prevalensi *stunting* sebesar 21,5% dan telah terjadi penurunan selama 10 tahun terakhir (2013-2023). Akan tetapi, penurunan tersebut belum dapat memenuhi target Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2020-2024 yang menargetkan prevalensi *stunting* sebesar 14% pada tahun 2024 [3].

Tingginya prevalensi *stunting* memerlukan perhatian lebih terhadap peningkatan efektivitas strategi dalam deteksi *stunting*. Seiring dengan perkembangan teknologi, khususnya dalam bidang pembelajaran *Machine Learning* (ML), berbagai teknologi telah dirancang untuk memperkuat upaya dalam mendeteksi kasus *stunting*. *Machine learning* telah digunakan dalam beberapa studi untuk mendiagnosis, memprediksi, dan merencanakan nutrisi untuk balita yang

mengalami *stunting* [4].

Namun, berdasarkan penelitian sebelumnya, hasil deteksi *stunting* masih menghadapi beberapa permasalahan yang belum terselesaikan, salah satunya adalah ketidakseimbangan data. Hal ini terjadi karena model cenderung memprediksi kelas mayoritas dengan akurasi yang tinggi, namun kurang efektif dalam mengidentifikasi kasus pada kelas minoritas (anak yang tidak mengalami *stunting*). Selain itu, penelitian sebelumnya belum mengimplementasikan optimasi melalui *hyperparameter tuning* dalam proses klasifikasi deteksi *stunting*, sehingga kinerja model belum mencapai hasil yang optimal [5].

Ketidakseimbangan data dapat menyebabkan penurunan kinerja model *machine learning*, sehingga performa dalam klasifikasi menjadi kurang efektif. Selain itu, pemilihan parameter yang tidak tepat dapat mengakibatkan *overfitting* atau mengurangi akurasi kinerja *machine learning*. Beberapa penelitian sebelumnya [6], telah menerapkan metode *Synthetic Minority Over-sampling Technique* (SMOTE) untuk mengatasi ketidakseimbangan data, serta optimasi menggunakan *grid search* sebagai metode penyetelan *hyperparameter* yang dapat meningkat kinerja dalam menemukan konfigurasi parameter terbaik [7].

Penelitian ini menerapkan beberapa algoritma *machine learning* yang sering digunakan dalam berbagai studi, seperti *K-Nearest Neighbor* (KNN), *Random Forest* dan *Decision Tree*. Algoritma *K-Nearest Neighbor* (KNN) dikenal karena mudah dan efektif dalam mengklasifikasikan data, meskipun terkadang rentan terhadap data berdimensi tinggi sehingga menyebabkan kesulitan dalam penyetelan parameter. Sementara itu, *Random Forest* unggul dalam menghasilkan akurasi terbaik dan mampu menangani data dengan banyak fitur, meskipun membutuhkan sumber daya komputasi yang besar. Di sisi lain, *Decision Tree* memudahkan dalam visualisasi hasil, namun terkadang tidak stabil jika terdapat perubahan dalam data.

Namun, pengembangan *machine learning* masih terbatas dalam mengeksplorasi data untuk menghasilkan akurasi optimal dalam mendeteksi *stunting*. Penelitian ini bertujuan untuk mengimplementasikan teknik SMOTE dalam menangani masalah ketidakseimbangan data. Selain itu, penelitian ini

berfokus pada pengembangan model *machine learning* untuk memprediksi *stunting* pada balita, dengan menguji model melalui penerapan *grid search* sebagai metode penyetelan *hyperparameter*. Dengan memanfaatkan teknik ini, diharapkan dapat dikembangkan model prediksi yang mampu mengidentifikasi balita yang berisiko tinggi mengalami *stunting* secara akurat.

1.2 Rumusan Masalah

1. Bagaimana penerapan metode *Synthetic Minority Over-sampling Technique* (SMOTE) dapat meningkatkan performa algoritma *machine learning* dalam mendeteksi penyakit *stunting* pada balita?
2. Bagaimana penerapan metode *grid search* sebagai penyetelan *hyperparameter tuning* dapat mengoptimalkan akurasi *machine learning* dalam mendeteksi penyakit *stunting* pada balita?

1.3 Batasan Masalah

1. Penelitian ini hanya menggunakan dataset mengenai penyakit *stunting* dari platform Kaggle yang menyediakan data secara online.
2. Fitur dataset yang digunakan dalam penelitian terbatas pada cakupan indikator fisik seperti usia, berat badan saat lahir, berat badan saat ini, panjang badan saat lahir, panjang badan saat ini dan pemberian ASI secara eksklusif.
3. Penelitian ini menerapkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) untuk mengatasi masalah ketidakseimbangan data.
4. Penelitian ini hanya menguji beberapa algoritma *machine learning*, yaitu *K-Nearest Neighbor* (KNN), *Random Forest* dan *Decision Tree* tanpa membandingkan secara mendalam dengan algoritma lainnya.

1.4 Tujuan Penelitian

1. Untuk mengevaluasi efektivitas teknik SMOTE dalam meningkatkan performa algoritma dalam prediksi penyakit *stunting*.
2. Untuk menganalisis pengaruh penyetelan parameter secara otomatis menggunakan *hyperparameter tuning* dengan teknik *grid search* terhadap

peningkatan kinerja algoritma dalam memprediksi penyakit *stunting*.

3. Mengimplementasikan dan membandingkan algoritma *K-Nearest Neighbor* (KNN), *Random Forest* dan *Decision Tree* untuk menentukan algoritma yang paling akurat dan efisien untuk memprediksi *stunting* pada balita.

1.5 Manfaat Penelitian

1. Untuk mengklasifikasikan kasus *stunting* dengan fokus pada pencegahan dan penanganan, sehingga sumber daya kesehatan dapat dialokasikan dengan lebih tepat dan efisien.
2. Untuk mengembangkan sistem diagnosa berbasis algoritma *machine learning* yang canggih dan efektif dalam memprediksi status *stunting*.
3. Untuk mendukung pengambilan kebijakan dalam upaya pencegahan dan pengendalian *stunting*. Dengan mengidentifikasi area prioritas untuk intervensi dan alokasi sumber daya.

1.6 Sistematika Penulisan

Dalam penyusunan skripsi ini terdiri dari lima bab, dengan uraian masing-masing sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan skripsi.

BAB II TINJAUAN PUSTAKA

Bab ini menyajikan studi literatur dan dasar teori yang relevan dengan penelitian ini, mencakup hasil penelitian terdahulu dan konsep-konsep dasar yang digunakan, mencakup *stunting*, data mining, *hyperparameter tuning*, *grid search*, ketidakseimbangan data, metode *resampling* menggunakan SMOTE, klasifikasi, *K-Nearest Neighbor* (KNN), *Random Forest* dan *Decision Tree*.

BAB III METODE PENELITIAN

Bab ini menjelaskan metodologi yang diterapkan dalam penelitian, mulai dari objek penelitian, alur penelitian yang meliputi tahap-tahapan yang dilalui, teknik

pengumpulan data, *preprocessing data* yang mencakup penanganan ketidakseimbangan data, hingga pemodelan algoritma dan evaluasi menggunakan metrik seperti *accuracy*, *precision*, *recall* dan *f1-score*.

BAB IV HASIL DAN PEMBAHASAN

Bab ini memaparkan hasil penelitian yang diperoleh berdasarkan analisis dan pengujian model, termasuk deskripsi dataset, proses *preprocessing data*, pemodelan algoritma dan hasil evaluasi model.

BAB V PENUTUP

Bab ini berisi kesimpulan yang merangkum temuan utama berdasarkan hasil analisis dari penelitian. Selain itu terdapat saran untuk penelitian selanjutnya, yang dapat mengeksplorasi lebih lanjut tentang algoritma atau dataset yang beragam.

