

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penelitian tentang stunting sangat penting karena stunting memiliki dampak jangka panjang pada pertumbuhan dan perkembangan anak, termasuk penurunan kemampuan kognitif dan fisik yang bisa mempengaruhi produktivitas mereka di masa dewasa. Stunting adalah masalah kekurangan gizi akibat asupan gizi yang tidak mencukupi dalam jangka waktu panjang, yang mengakibatkan gangguan pertumbuhan pada anak dengan tinggi badan di bawah standar usianya (Rufina et., al 2023). Stunting di Indonesia adalah masalah gizi yang mendesak dan membutuhkan penanganan khusus. Meskipun prevalensi stunting di Indonesia telah menurun setiap tahun, angkanya masih di atas batas WHO yaitu di bawah 20%. Berdasarkan Studi Status Gizi Indonesia (SSGI) 2021, angka prevalensi stunting di Indonesia adalah 24,4% (Intan, 2023). Berdasarkan data dari Dinas Kesehatan Kota Samarinda tahun 2021, prevalensi stunting pada balita di Kota Samarinda mencapai 24,7%. Pada tahun 2020, tercatat sebanyak 1.402 balita di Kota Samarinda mengalami stunting, dengan rincian 403 balita dalam kategori sangat pendek dan 999 balita dalam kategori pendek (Rufina et., al 2023). Mengingat konsekuensi serius yang diakibatkan oleh stunting, tentu menjadi kekhawatiran tersendiri untuk masa depan, terutama karena banyaknya individu yang masih teridentifikasi dalam kategori stunting. Oleh karena itu, perlu adanya penggunaan teknik analisis data canggih seperti machine learning untuk membantu pengambilan keputusan dalam mendeteksi stunting di masa depan. Penggunaan teknik ini membuka peluang untuk menerapkan metode data mining dan klasifikasi yang lebih spesifik, yang akan meningkatkan akurasi dalam mengidentifikasi kasus stunting, serta memungkinkan intervensi yang lebih tepat sasaran dan efektif.

Dari penelitian yang telah dilakukan sebelumnya dalam bidang data mining, klasifikasi status stunting telah dijajaki dengan menggunakan beberapa pendekatan, seperti algoritma *Random Forest*, *Logistic Regression*, *Support Vector Classification*, dan lain-lain. Dan dalam beberapa tahun terakhir, penelitian telah menunjukkan tingkat akurasi yang cukup baik, dengan persentase akurasi sekitar 79.21%(RF), 44.71(LR), 64.02%(SVC) dan 63.85(XG Boost). Namun, penelitian terkait masih menggunakan data berdimensi rendah, yang memiliki atribut terbatas untuk analisis. Data dengan lima hingga delapan atribut atau bahkan kurang dari sepuluh atribut dapat dianggap sebagai data berdimensi rendah (Obvious et., al 2023). Data berdimensi rendah memiliki kelemahan dalam hal kemampuan untuk menangani kompleksitas dan variasi data yang besar. Hal ini dapat menyebabkan masalah seperti overfitting, di mana model terlalu spesifik terhadap data latih dan tidak dapat generalisasi dengan baik terhadap data baru. Selain itu, pengurangan dimensi yang terlalu drastis dapat mengakibatkan hilangnya informasi penting yang diperlukan untuk pemodelan yang akurat (Dylan et., al 2023). Berdasarkan Penelitian yang dilakukan Oleh (Estiyak et., al 2023) *Algoritma Random Forest(RF)*, *Naive Bayes(NB)*, dan *Logistic Regression(LR)* dapat mengalami penurunan performa ketika dihadapkan pada data dengan dimensi tinggi, yang dapat mempengaruhi akurasi yang diperoleh 60%, 58%, dan 58%.

Data berdimensi tinggi terdiri dari berbagai atribut atau fitur yang digunakan untuk analisis. Sebagai contoh, data dengan puluhan atau ratusan atribut dapat diklasifikasikan sebagai data berdimensi tinggi. Data berdimensi tinggi menghadapi beberapa tantangan

utama, antara lain peningkatan kompleksitas model, risiko overfitting, dan kesulitan dalam visualisasi data (Jorg et., al 2023). Algoritma yang mengalami penurunan performa saat menghadapi data berdimensi tinggi, yang berpengaruh pada akurasi klasifikasi. Pemahaman tentang data berdimensi tinggi dan penerapan strategi yang tepat dalam analisisnya menjadi kunci dalam mengoptimalkan kinerja model dan meningkatkan akurasi prediksi stunting di Kota Samarinda(Yoga 2023).

Algoritma klasifikasi dalam penelitian ini akan menggunakan *Random Forest (RF)*, berdasarkan penelitian yang dilakukan oleh (Obvious et., al 2022; Yoga dan Wawan 2023; Khan et., al 2023; Estiyak et al., 2023; Addisalem., al 2023) *Random Forest* memiliki performa terbaik jika dibandingkan dengan algoritma *Support Vector Machine, Logistic Regression, Neural Network, Naïve Bayes, k-nearest neighbors (k-NN), eXtreme Gradient Boosting (Xg boost), Logistic Regression, C5.0, dan CART* terkait klasifikasi data stunting. Penggunaan seleksi fitur *Recursive Feature Elimination (RFE)* juga akan dilakukan untuk mengidentifikasi atribut-atribut yang relevan dalam klasifikasi. Berdasarkan Penelitian Leykun et., al(2024) metode *Recursive Feature Elimination (RFE)* terbukti dapat meningkatkan akurasi dari metode klasifikasi *Random Forest* dimana sebelum menggunakan *RFE* mendapatkan akurasi sebesar 72.41% kemudian, setelah dilakukan penerapan metode *RFE*, terjadi peningkatan signifikan dalam akurasi, di mana akurasi *RF* meningkat menjadi 80.1%. Penelitian lainnya dengan metode klasifikasi yang berbeda (Chandan and Priti 2021), *SVM* dan *LR* sebelum menggunakan Seleksi fitur *RFE* 94% dan 93%, Setelah menggunakannya mendapatkan akurasi yang lumayan tinggi 97% dan 98% yang mana lebih tinggi dibandingkan dengan seleksi fitur yang lain *SVM + Chi-Square* 96.50% - 96.60%, *K-NN + Relif-F* 91.90% - 92.20%, dan *K-NN + Backward Elimination* 95.17 - 90.23%(Yoga et., al 2021; Kemal et., al 2023; Syahrani dan Dwi, 2022).

Namun, setelah menggunakan seleksi fitur *Recursive Feature Elimination (RFE)* dan mencapai akurasi 80,01%(Leykun et al., 2024), akurasi yang didapat masih belum memuaskan. Oleh karena itu, diperlukan optimasi lebih lanjut menggunakan Metode *Genetic Algorithm (GA)* untuk meningkatkan akurasi tersebut. untuk mencari solusi optimal atau mendekati solusi optimal dalam masalah optimasi dengan meniru proses seleksi alamiah evolusi genetik(Parmonagan et al., 2020), Seperti pada penelitian (Parmonagan et al.,2020) dimana sebelum menggunakannya mendapatkan Akurasi 82.5% setelah diterapkan mendapatkan 85.83% begitu juga dengan penelitian (Veeramani et al., 2023) Sebelum menggunakan mendapat 77.58% setelah diterapkan menjadi 79.31%, yang mana juga lebih baik dibandingkan dengan Optimasi yang lain dengan akurasi *PSO + RNN* 96.08%, *PSO + K-NN* 98.9%, *EGA + PSO* 98.97%, *PSO + RF* 99.76% dan *GA + RF* 99.99%(Monire et., al 2021).

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijabarkan, rumusan masalah pada penelitian ini adalah:

- a) Apa saja fitur yang memiliki dampak signifikan terhadap kinerja algoritma *Random Forest* pada dataset Stunting di Kota Samarinda, dengan penerapan seleksi fitur *RFE* ?
- b) Seberapa besar peningkatan tingkat keakuratan yang dicapai oleh algoritma *Random Forest* dalam mengklasifikasikan data stunting di Kota Samarinda setelah implementasi seleksi fitur *RFE* dan optimasi dengan *GA*?

1.3 Tujuan Penelitian

Berdasarkan latar belakang yang telah dijabarkan, rumusan masalah pada penelitian ini adalah:

- a) Menentukan atribut-atribut yang Mempengaruhi Keakuratan Model *Random Forest* dalam Menganalisis Data Stunting di Kota Samarinda.
- b) Menerapkan Algoritma *Random Forest* untuk Klasifikasi Kasus Stunting di Kota Samarinda, dengan menggunakan *RFE* untuk seleksi fitur dan mengoptimalkan model menggunakan *GA* guna mencapai akurasi prediksi yang lebih tinggi.
- c) Mengevaluasi kinerja Algoritma *Random Forest* dengan seleksi fitur *RFE* dan optimasi *GA* menggunakan *Confusion Matrix*.

1.4 Manfaat Penelitian

Penelitian ini diharapkan memberikan beberapa manfaat signifikan dalam bidang machine learning dan data mining. Berikut adalah beberapa manfaat dari penelitian ini:

- a) Manfaat Akademis

Penelitian ini memberikan kontribusi pada pengembangan dan pemahaman lebih lanjut tentang penggunaan algoritma *Random Forest* dalam klasifikasi data berdimensi tinggi, khususnya dalam konteks analisis data kesehatan, menunjukkan cara-cara efektif untuk meningkatkan akurasi dan kinerja model machine learning melalui penerapan teknik *Recursive Feature Elimination (RFE)* dan optimasi menggunakan *Genetic Algorithm (GA)*, serta memperkaya literatur ilmiah di bidang machine learning dengan menyediakan referensi berharga bagi akademisi dan praktisi yang tertarik dalam meningkatkan performa algoritma klasifikasi pada dataset kompleks dan berdimensi tinggi.

- b) Manfaat Praktis

Penelitian ini menunjukkan penerapan praktis dari teknik machine learning canggih, seperti *RFE* dan *GA*, dalam meningkatkan akurasi model, memberikan panduan praktis bagi data scientist dan praktisi machine learning untuk menerapkan teknik-teknik tersebut pada berbagai jenis data. Menunjukkan bagaimana teknik seleksi fitur dapat digunakan untuk mengidentifikasi atribut yang paling relevan, membantu mengurangi kompleksitas model dan meningkatkan efisiensi pengolahan data tanpa mengorbankan akurasi. Menyediakan teknik dan metodologi yang dapat diterapkan pada berbagai domain lain yang memerlukan analisis data berdimensi tinggi, sehingga memberikan manfaat luas dalam berbagai aplikasi machine learning di industri maupun akademisi. Serta menyediakan benchmark yang berguna bagi pengujian dan pengembangan algoritma klasifikasi lainnya, memungkinkan evaluasi komparatif yang lebih baik di masa depan.

1.5 Batasan Masalah

Agar masalah yang dibahas tidak menjadi lebih luas lagi, penulis membatasi masalah penelitian sebagai berikut:

- a) Data yang digunakan dalam penelitian ini adalah data stunting di Kota Samarinda pada tahun 2023.

- b) Penelitian ini hanya akan menguji efektivitas algoritma Random Forest dalam klasifikasi stunting, dengan fokus pada pemilihan fitur melalui metode Recursive Feature Elimination (RFE) untuk menentukan atribut yang paling relevan dalam prediksi stunting.
- c) Fokus penelitian ini adalah pada peningkatan akurasi prediksi stunting, bukan pada intervensi atau implementasi kebijakan yang dihasilkan dari analisis tersebut.
- d) Atribut yang digunakan pada penelitian ini antara lain adalah Nama, JK, Berat, LiLA, BB/U, ZS BB/U, TB/U, ZS TB/U, BB/TB, ZS BB/TB, Naik Berat B, JmL Vit A, dan Tanggal Pengukuran setelah dilakukan selection data.
- e) Menilai efektivitas model klasifikasi Stunting melalui evaluasi kinerja algoritma Random Forest dengan k-Fold Cross Validation, menggunakan confusion matrix untuk menghitung akurasi yang akurat.