

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Stunting merupakan salah satu permasalahan gizi kronis pada balita yang ditandai dengan tinggi badan yang berada di bawah standar usia. Kondisi ini tidak hanya berdampak pada pertumbuhan fisik, tetapi juga memengaruhi perkembangan kognitif dan produktivitas individu di masa depan. Oleh karena itu, deteksi dini stunting menjadi sangat penting sebagai dasar pengambilan keputusan dalam intervensi kesehatan masyarakat. Seiring berkembangnya teknologi, pendekatan machine learning mulai banyak diterapkan dalam bidang kesehatan, khususnya untuk klasifikasi dan prediksi penyakit. Berbagai algoritma klasifikasi seperti *Random Forest*, *Support Vector Machine*, dan *XGBoost* telah digunakan untuk mengolah data kesehatan dengan tujuan meningkatkan akurasi diagnosis. Salah satu algoritma yang terbukti memiliki performa tinggi adalah *XGBoost (Extreme Gradient Boosting)* karena kemampuannya menangani data kompleks, tidak seimbang, dan berukuran besar.

Namun, keberhasilan model machine learning tidak hanya ditentukan oleh algoritma yang digunakan, tetapi juga oleh metode evaluasi model. Salah satu tantangan utama dalam klasifikasi data kesehatan adalah

ketidakseimbangan kelas, di mana jumlah data pada setiap kelas tidak sebanding. Untuk mengatasi permasalahan tersebut, metode *Stratified K-Fold Cross Validation* digunakan agar proporsi kelas pada data latih dan data uji tetap terjaga.

Beberapa penelitian terdahulu telah menunjukkan efektivitas penggunaan Stratified K-Fold Cross Validation dalam evaluasi model machine learning. Penelitian oleh *Widodo et al. (2022)* menerapkan Stratified K-Fold Cross Validation untuk optimasi model machine learning dalam "*prediksi penyakit kanker serviks* dan menunjukkan bahwa metode ini mampu menghasilkan evaluasi model yang lebih stabil dan representatif dibandingkan K-Fold biasa.[1] Selanjutnya, penelitian "*SKCV pada jurnal Frontiers in Nanotechnology (2022)* menggunakan Stratified K-Fold Cross Validation pada beberapa algoritma klasifikasi, termasuk XGBoost, untuk "*prediksi risiko kanker serviks*. Hasil penelitian tersebut menunjukkan bahwa Stratified K-Fold mampu meningkatkan konsistensi performa model, khususnya pada dataset medis yang bersifat tidak seimbang.[2] Penelitian lain oleh *IJSAT (2025)* melalui kajian komprehensif terhadap berbagai teknik cross validation menegaskan bahwa Stratified K-Fold merupakan metode yang paling direkomendasikan untuk kasus klasifikasi dengan distribusi kelas yang tidak merata, karena mampu mengurangi bias evaluasi model.[3] Hal ini diperkuat oleh studi *SSRN (2024)* yang membandingkan berbagai teknik cross validation dan menyimpulkan bahwa Stratified K-Fold memberikan hasil evaluasi yang lebih akurat dan reliabel pada data

klasifikasi dibandingkan metode cross validation konvensional.[4] Selain itu, penelitian *arXiv (2025)* yang menerapkan Stratified K-Fold Cross Validation pada klasifikasi serangan jaringan juga menunjukkan bahwa metode ini efektif dalam menjaga keseimbangan kelas selama proses pelatihan dan pengujian model, sehingga performa model dapat dinilai secara lebih objektif.[5]

Berdasarkan uraian tersebut, dapat disimpulkan bahwa penggunaan *Stratified K-Fold Cross Validation* sangat relevan untuk penelitian klasifikasi data kesehatan, termasuk klasifikasi status stunting balita. Oleh karena itu, penelitian ini mengusulkan penerapan algoritma *XGBoost* dengan pendekatan *Stratified K-Fold Cross Validation* untuk mengklasifikasikan status stunting balita, dengan harapan mampu menghasilkan model yang akurat, stabil, dan dapat dijadikan sebagai alat bantu dalam pengambilan keputusan di bidang kesehatan.

1.2 RUMUSAN MASALAH

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana proses pengolahan data balita untuk klasifikasi status stunting menggunakan algoritma XGBoost?
2. Bagaimana kinerja algoritma XGBoost dalam mengklasifikasikan status stunting balita?

3. Bagaimana pengaruh penggunaan Stratified K-Fold Cross Validation terhadap hasil evaluasi model klasifikasi?
4. Sejauh mana kemampuan model dalam mendeteksi balita yang mengalami stunting berdasarkan hasil evaluasi klasifikasi?

1.3 BATASAN MASALAH

Agar penelitian ini lebih terfokus dan terarah, maka ditetapkan batasan masalah sebagai berikut:

1. Dataset yang digunakan merupakan data balita yang memuat atribut usia, jenis kelamin, dan tinggi badan.
2. Klasifikasi dilakukan dalam bentuk klasifikasi biner, yaitu stunting dan tidak stunting.
3. Pembagian data dilakukan menggunakan pendekatan Stratified K-Fold Cross Validation.
4. Evaluasi model menggunakan metrik akurasi, *classification report*, dan *confusion matrix*.

1.4 TUJUAN PENELITIAN

Tujuan dari penelitian ini adalah sebagai berikut:

1. Membangun model klasifikasi status stunting balita menggunakan algoritma XGBoost.
2. Menganalisis performa model XGBoost dalam mengklasifikasikan balita stunting dan tidak stunting.

3. Mengetahui pengaruh penggunaan Stratified K-Fold Cross Validation terhadap stabilitas dan keandalan model klasifikasi.
4. Mengevaluasi kemampuan model dalam mendeteksi balita yang berisiko mengalami stunting.

1.5 MANFAAT PENELITIAN

Adapun manfaat dari penelitian ini adalah:

1. Manfaat Teoretis : Memberikan kontribusi ilmiah mengenai penerapan algoritma XGBoost dan Stratified K-Fold Cross Validation dalam klasifikasi data kesehatan.
2. Manfaat Praktis : Memberikan referensi bagi pihak terkait sebagai alat bantu dalam deteksi dini stunting balita.
3. Manfaat Akademis : Menjadi bahan acuan bagi penelitian selanjutnya yang berkaitan dengan klasifikasi data tidak seimbang menggunakan machine learning.

1.6 SISTEMATIKA PENULISAN

Sistematika penulisan laporan penelitian ini adalah sebagai berikut:

1. **Bab I Pendahuluan**, membahas latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

2. **Bab II Tinjauan Pustaka**, berisi landasan teori mengenai stunting, *machine learning*, algoritma XGBoost, Stratified K-Fold Cross Validation, serta penelitian terkait.
3. **Bab III Metodologi Penelitian**, menjelaskan tahapan penelitian, dataset, *preprocessing* data, metode klasifikasi, dan teknik evaluasi.
4. **Bab IV Hasil dan Pembahasan**, menyajikan hasil eksperimen, evaluasi model, *confusion matrix*, serta analisis hasil penelitian.
5. **Bab V Kesimpulan dan Saran**, berisi kesimpulan penelitian dan saran untuk pengembangan penelitian selanjutnya.