

BAB I

PENDAHULUAN

1.1 Latar Belakang

Gagal jantung adalah kondisi ketika jantung tidak dapat memompa darah untuk dialirkan ke seluruh tubuh secara optimal akibat adanya masalah pada otot dari organ. Gagal jantung merupakan kondisi medis dimana kemampuan jantung dalam memompa darah ke seluruh tubuh mengalami penurunan akibat gangguan pada otot jantung [1]. Sebagai bagian dari kelompok penyakit kardiovaskular yang luas, kondisi ini menjadi fokus utama dalam dunia kesehatan mengingat peningkatan kasus yang signifikan secara global [2]. Data Organisasi Kesehatan Dunia (WHO) mencatat lebih dari 17 juta kematian tahunan akibat penyakit kardiovaskular di seluruh dunia. Di Indonesia sendiri tercatat 651.481 kematian per tahun dengan komposisi 331.349 kasus stroke, 245.343 penyakit jantung koroner, dan 50.620 kasus gagal jantung terkait hipertensi [3]. Angka ini menunjukkan urgensi deteksi dini untuk mengurangi dampak kesehatan masyarakat.

Dalam analisis data medis, ketidakseimbangan kelas sering kali menjadi tantangan utama ketika jumlah sampel antar kategori tidak proporsional [4]. Situasi ini menyebabkan bias model prediktif terhadap kelas mayoritas sehingga akurasi prediksi untuk kelompok minoritas seperti kasus gagal jantung cenderung rendah. Penelitian Wang dkk (2021) dalam studi tentang klasifikasi data tidak seimbang menggunakan kombinasi SMOTE-ENN dan XGBoost menunjukkan hasil yang menjanjikan. Meskipun XGBoost unggul dalam kecepatan klasifikasi dan SMOTE-ENN efektif membersihkan noise data, kombinasi ini masih menunjukkan keterbatasan pada parameter evaluasi seperti presisi dan F1-score [5].

Masalah performa model tersebut terkait erat dengan kompleksitas hyperparameter dalam algoritma machine learning [6]. Proses penyetulan hyperparameter melalui metode seperti Optimasi Bayesian telah terbukti meningkatkan akurasi model secara signifikan dibandingkan teknik konvensional seperti Grid Search atau Random Search [7]. Beberapa studi

komparatif oleh Xia dkk (2017) dan Hnin-Jeeanunta (2019) menunjukkan bahwa pendekatan probabilistik pada Optimasi Bayesian mampu menemukan kombinasi parameter optimal dengan efisiensi waktu lebih baik daripada algoritma evolusioner seperti Particle Swarm Optimization.

Kelebihan utama Optimasi Bayesian terletak pada kemampuannya mengadaptasi parameter berdasarkan evaluasi iteratif sebelumnya sehingga mengurangi risiko overfitting [7]. Berbeda dengan metode trial-error konvensional yang memerlukan waktu komputasi panjang tanpa jaminan optimalisasi maksimal, pendekatan ini menggunakan proses pembelajaran aktif untuk mencapai konfigurasi terbaik. Dalam konteks klasifikasi data medis yang kompleks dan sensitif seperti diagnosis gagal jantung, presisi dalam penentuan hyperparameter menjadi faktor krusial untuk memastikan akurasi prediksi sekaligus menjaga keseimbangan performa antar metrik evaluasi.

Upaya integrasi antara teknik penanganan data tidak seimbang dan optimasi parameter tampaknya menjadi solusi potensial untuk meningkatkan kualitas model prediktif di bidang kardiologi. Implementasi kombinasi SMOTE-ENN dengan XGBoost yang diperkuat Optimasi Bayesian mungkin dapat menghasilkan sistem klasifikasi yang lebih robust meskipun diperlukan kajian lebih mendalam mengenai implementasinya pada dataset spesifik pasien Indonesia mengingat karakteristik populasi yang unik [5].

Dengan demikian, Optimasi Bayesian tidak hanya efisien dalam hal waktu, tetapi juga lebih efektif dalam meningkatkan akurasi dan kinerja model secara signifikan. Hal ini menjadikannya pilihan utama dalam proses penyetulan *hyperparameter*, terutama ketika menghadapi masalah kompleks yang memerlukan presisi tinggi [7].

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, berikut adalah tiga rumusan masalah yang dapat diajukan:

1. Bagaimana pengaruh penggunaan teknik SMOTE-ENN terhadap performa model XGBoost dalam mengatasi permasalahan class imbalance?

2. Model manakah yang paling optimal digunakan dalam prediksi data tidak seimbang antara XGBoost tanpa SMOTE, XGBoost SMOTE-ENN Default, dan XGBoost SMOTE-ENN Fine-Tuned?

1.3 Tujuan Penelitian

Berdasarkan permasalahan yang telah diuraikan, maka tujuan yang ingin dicapai pada penelitian ini adalah sebagai berikut:

1. Untuk mengidentifikasi dan menerapkan metode yang efektif dalam mengatasi masalah ketidakseimbangan data (imbalance data) pada kasus gagal jantung, guna meningkatkan akurasi prediksi khususnya pada kelas minoritas.
2. Untuk mengevaluasi performa kombinasi metode SMOTE-ENN dan XGBoost dalam mengklasifikasikan data tidak seimbang terkait gagal jantung, serta menganalisis faktor-faktor yang memengaruhi efektivitas model tersebut.
3. Untuk menguji efektivitas Optimasi Bayesian sebagai solusi dalam penyetelan *hyperparameter* guna meningkatkan performa model klasifikasi gagal jantung, khususnya dalam mengatasi kompleksitas *hyperparameter* pada XGBoost.

1.4 Manfaat Penelitian

1. Penelitian ini mengeksplorasi kombinasi teknik SMOTE-ENN dan algoritma XGBoost yang disempurnakan melalui optimasi Bayesian untuk penyetelan parameter. Temuannya menunjukkan potensi signifikan dalam mengatasi tantangan klasifikasi data tidak seimbang sekaligus meningkatkan akurasi prediksi. Pendekatan ini bisa menjadi kerangka acuan bagi pengembangan model prediktif di berbagai bidang yang menghadapi masalah serupa.
2. Dari perspektif praktis implementasinya di dunia kesehatan model ini mungkin membantu mengurangi risiko kesalahan diagnosis pasien penyakit kardiovaskular. Bukti awal mengindikasikan bahwa peningkatan akurasi klasifikasi berpotensi berdampak pada penurunan angka mortalitas melalui intervensi medis yang lebih tepat waktu. Menariknya studi ini juga

menyoroti perlunya kesadaran akan bias data dalam sistem prediktif berbasis AI dimana ketimpangan distribusi kelas sering kali diabaikan dalam pengembangan teknologi kesehatan. Temuan tersebut menegaskan bahwa penanganan data yang komprehensif tetap menjadi faktor krusial meskipun kemajuan algoritma terus terjadi.

1.5 Batasan Masalah

Agar penelitian ini lebih terarah, beberapa batasan masalah yang ditetapkan adalah:

1. Penelitian ini hanya menggunakan dataset sekunder dari Kaggle dengan jumlah yang sudah ditentukan.
2. Penelitian ini hanya berfokus pada pengujian perbandingan metode, tidak berupa aplikatif
3. Penelitian ini hanya menggunakan optimasi bayesian untuk penyetelan *hyperparameter*, tidak menggunakan optimasi lain.