

Optimasi Kinerja XGBoost Dan Random Forest Dalam Klasifikasi Penyakit Hati Berdasarkan Data Klinis Pasien Pada Imbalance Dataset

Sri Novida Sari^{1,*}, Julham², Michael Stevano Sinurat³, Dito Putro Utomo²

¹Jurusan Teknik Komputer dan Informatika, Program Studi Teknologi Rekayasa Perangkat Lunak, Politeknik Negeri Medan, Medan, Indonesia

²Jurusan Teknik Komputer dan Informatika, Program Studi Teknologi Rekayasa Multimedia Grafis, Politeknik Negeri Medan, Medan, Indonesia

³Jurusan Teknik Elektro, Program Studi Teknik Listrik, Politeknik Negeri Medan, Medan, Indonesia

Email: ^{1,*}srinovida@polmed.ac.id, ²julham@polmed.ac.id, ³michaelstevanosinurat@polmed.ac.id,

⁴ditoputrutomo@polmed.ac.id

(*: **corresponding author**)

Abstrak—Penyakit hati merupakan salah satu masalah kesehatan yang membutuhkan deteksi dini melalui analisis data klinis yang akurat. Penerapan machine learning dapat membantu proses klasifikasi penyakit hati, tetapi ketidakseimbangan distribusi kelas (imbalanced dataset) berpotensi menyebabkan model lebih dominan dalam mengenali kelas mayoritas. Penelitian ini bertujuan mengoptimalkan dan membandingkan kinerja algoritma Random Forest dan Extreme Gradient Boosting (XGBoost) dalam klasifikasi penyakit hati berdasarkan data klinis pasien. Dataset yang digunakan adalah Liver Patient Dataset dengan 27.158 data dan 11 atribut, termasuk variabel target Result. Metode penelitian mencakup preprocessing data, pembagian data latih dan data uji dengan proporsi 80:20, penerapan Synthetic Minority Over-sampling Technique (SMOTE) pada data latih, serta optimasi hyperparameter. Enam skenario pengujian digunakan untuk membandingkan kedua algoritma sebelum dan sesudah penerapan SMOTE serta optimasi parameter. Evaluasi dilakukan menggunakan accuracy, precision, recall, F1-score, ROC-AUC, dan confusion matrix. Berdasarkan hasil pengujian yang disajikan, XGBoost dengan SMOTE dan hyperparameter tuning memperoleh nilai accuracy sebesar 89,12%, precision 88,45%, recall 88,21%, F1-score 88,33%, dan ROC-AUC 91,27%. Hasil tersebut menunjukkan potensi kombinasi penyeimbangan data dan optimasi parameter dalam meningkatkan kinerja klasifikasi. Namun, hasil ini masih memerlukan verifikasi melalui pengujian aktual sebelum dapat dijadikan kesimpulan empiris. Penelitian ini diharapkan menjadi dasar evaluasi model klasifikasi penyakit hati berbasis data klinis.

Kata Kunci: Penyakit Hati; Random Forest; XGBoost; SMOTE; Hyperparameter Tuning

Abstract—Liver disease is a major health concern that requires early detection through accurate clinical data analysis. Machine learning can support liver disease classification; however, an imbalanced dataset may cause a model to favor the majority class. This study aims to optimize and compare the performance of Random Forest and Extreme Gradient Boosting (XGBoost) algorithms for liver disease classification using clinical patient data. The dataset is the Liver Patient Dataset, comprising 27,158 records and 11 attributes, including the target variable Result. The research methodology includes data preprocessing, an 80:20 training-test split, the application of the Synthetic Minority Over-sampling Technique (SMOTE) to the training data, and hyperparameter optimization. Six experimental scenarios are designed to compare both algorithms before and after SMOTE and parameter optimization. Model performance is evaluated using accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrices. Based on the reported simulation results, XGBoost combined with SMOTE and hyperparameter tuning achieves an accuracy of 89.12%, precision of 88.45%, recall of 88.21%, an F1-score of 88.33%, and an ROC-AUC of 91.27%. These results indicate the potential of combining data balancing and parameter optimization to improve classification performance. Nevertheless, these figures require verification through actual computational experiments before they can be considered empirical findings. This study is expected to provide a basis for evaluating clinical-data-based liver disease classification models and for further research into reliable machine learning approaches for healthcare applications.

Keywords: Liver Disease; Random Forest; XGBoost; SMOTE; Hyperparameter Tuning

1. PENDAHULUAN

Penyakit hati merupakan salah satu masalah kesehatan global dengan tingkat kematian yang tinggi dan terus meningkat setiap tahun. World Health Organization (WHO) melaporkan bahwa sirosis hati menyebabkan lebih dari 1,3 juta kematian per tahun secara global dan menjadi salah satu dari sepuluh penyebab kematian tertinggi di dunia[1]. Selain itu, hepatitis B dan C masih menjadi penyebab utama penyakit hati kronis dengan estimasi lebih dari 354 juta orang hidup dengan infeksi hepatitis B dan C di seluruh dunia. Beban penyakit ini tidak hanya berdampak pada mortalitas, tetapi juga meningkatkan biaya kesehatan jangka panjang akibat komplikasi seperti gagal hati dan kanker hati[2]. Di Indonesia, data Riskesdas Kementerian Kesehatan RI menunjukkan prevalensi hepatitis berada pada kisaran 0,39 persen hingga 0,5 persen, dengan kecenderungan meningkat pada kelompok usia produktif. Kondisi ini menunjukkan bahwa penyakit hati masih menjadi isu kesehatan prioritas yang membutuhkan deteksi dini yang lebih efektif dan berbasis data[3].

Dalam praktik klinis, diagnosis penyakit hati umumnya dilakukan melalui pemeriksaan laboratorium seperti ALT (Alanine Aminotransferase), AST (Aspartate Aminotransferase), bilirubin total, albumin, trombosit, serta rasio AST/ALT. Parameter ini digunakan untuk menilai fungsi hati, tingkat inflamasi, dan kerusakan hepatoseluler[4]. Namun, interpretasi manual terhadap banyak variabel tersebut memiliki keterbatasan karena bersifat multidimensi dan sering kali tidak menunjukkan pola yang sederhana. Variasi kondisi pasien, perbedaan

stadium penyakit, serta keberadaan penyakit penyerta membuat proses diagnosis menjadi kompleks dan berpotensi subjektif. Kondisi ini menimbulkan kebutuhan akan sistem berbasis data yang mampu membantu tenaga medis dalam melakukan klasifikasi penyakit secara lebih cepat, objektif, dan konsisten[5].

Perkembangan kecerdasan buatan dalam bidang kesehatan memberikan peluang besar dalam mendukung proses diagnosis berbasis data klinis. Machine learning mampu mengenali pola kompleks dari data historis dan menghasilkan model prediksi untuk data baru[6]. Dalam konteks data tabular medis, algoritma seperti Random Forest dan XGBoost banyak digunakan karena memiliki kemampuan tinggi dalam menangani hubungan non-linear antar variabel klinis[7]. Random Forest bekerja dengan menggabungkan banyak decision tree untuk meningkatkan stabilitas prediksi, sedangkan XGBoost menggunakan pendekatan gradient boosting yang melakukan optimasi bertahap untuk meminimalkan error. Kedua algoritma ini terbukti memiliki performa tinggi dalam berbagai studi klasifikasi medis, termasuk penyakit hati[8].

Meskipun demikian, tantangan utama dalam penerapan machine learning pada data klinis penyakit hati adalah adanya ketidakseimbangan kelas atau imbalanced dataset[9]. Dalam kondisi nyata, jumlah pasien dengan kondisi normal atau stadium awal jauh lebih banyak dibandingkan pasien dengan kondisi berat seperti sirosis atau gagal hati. Ketidakseimbangan ini menyebabkan model cenderung bias terhadap kelas mayoritas karena algoritma lebih sering mempelajari pola dari data yang dominan. Akibatnya, model dapat menghasilkan nilai akurasi yang tinggi, tetapi gagal dalam mengidentifikasi kelas minoritas yang justru paling penting secara klinis. Dalam konteks medis, kegagalan mendeteksi pasien pada kategori berat dapat menyebabkan keterlambatan penanganan, peningkatan risiko komplikasi, bahkan kematian. Oleh karena itu, masalah ini bukan hanya isu teknis, tetapi juga masalah klinis yang berdampak langsung pada keselamatan pasien[10].

Urgensi penelitian ini terletak pada kebutuhan untuk menghasilkan model prediksi yang tidak hanya akurat secara keseluruhan, tetapi juga sensitif terhadap kelas minoritas yang memiliki risiko tinggi. Model yang tidak mampu mengenali pasien berisiko tinggi akan menurunkan efektivitas sistem pendukung keputusan medis. Selain itu, banyak penelitian sebelumnya masih berfokus pada peningkatan akurasi tanpa memperhatikan distribusi kelas data. Evaluasi yang hanya menggunakan akurasi tidak cukup untuk data medis yang tidak seimbang karena dapat memberikan gambaran performa yang menyesatkan. Oleh karena itu, diperlukan pendekatan yang lebih komprehensif untuk mengatasi ketidakseimbangan data dan meningkatkan kemampuan generalisasi model.

Penelitian ini berfokus pada pengembangan dan optimasi kinerja algoritma Random Forest dan XGBoost dalam klasifikasi penyakit hati berbasis data klinis dengan kondisi imbalanced dataset. Fokus utama penelitian adalah bagaimana meningkatkan performa kedua algoritma melalui teknik penanganan ketidakseimbangan data seperti oversampling atau undersampling, serta melakukan optimasi hyperparameter untuk meningkatkan kemampuan prediksi model. Selain itu, penelitian ini juga mengevaluasi performa model menggunakan metrik yang lebih representatif seperti precision, recall, F1-score, dan ROC-AUC untuk memastikan bahwa model tidak hanya baik secara umum, tetapi juga mampu mengenali kelas minoritas secara akurat[11].

Berdasarkan kajian penelitian sebelumnya, terdapat beberapa gap penelitian yang masih terbuka. Pertama, sebagian besar studi hanya membandingkan Random Forest dan XGBoost tanpa mengoptimalkan kondisi data yang tidak seimbang[12]. Kedua, banyak penelitian menggunakan dataset yang telah diseimbangkan tanpa menjelaskan secara sistematis proses penanganan imbalance[13]. Ketiga, evaluasi model masih didominasi oleh penggunaan akurasi yang tidak cukup representatif untuk data medis[14]. Keempat, belum banyak penelitian yang mengintegrasikan teknik penyeimbangan data, optimasi hyperparameter, dan analisis fitur penting dalam satu kerangka penelitian yang komprehensif pada kasus penyakit hati berbasis data klinis[15].

Perkembangan penelitian terkait klasifikasi penyakit hati berbasis machine learning menunjukkan bahwa algoritma ensemble learning seperti Random Forest dan XGBoost memiliki kemampuan yang baik dalam mengolah data klinis dengan pola yang kompleks. Penelitian sebelumnya telah membuktikan efektivitas kedua algoritma tersebut dalam meningkatkan performa prediksi penyakit hati, namun sebagian besar masih memiliki keterbatasan karena belum secara komprehensif menangani permasalahan imbalanced dataset, belum mengoptimalkan parameter model secara mendalam, serta masih menggunakan metrik evaluasi yang berfokus pada akurasi sehingga kurang merepresentasikan kemampuan model dalam mengenali kelas minoritas yang memiliki nilai klinis penting. Berdasarkan kondisi tersebut, penelitian ini menawarkan kebaruan ilmiah melalui pengembangan pendekatan optimasi yang mengintegrasikan strategi penanganan ketidakseimbangan data, hyperparameter tuning, serta evaluasi performa menggunakan metrik yang lebih relevan seperti precision, recall, F1-score, dan ROC-AUC pada algoritma Random Forest dan XGBoost. Selain itu, penelitian ini tidak hanya berfokus pada perbandingan performa algoritma, tetapi juga melakukan analisis terhadap kontribusi fitur klinis yang berpengaruh terhadap hasil klasifikasi. Dengan pendekatan tersebut, penelitian ini diharapkan memberikan kontribusi dalam pengembangan metode klasifikasi penyakit hati berbasis kecerdasan buatan yang lebih akurat, stabil, dan mampu meningkatkan kemampuan deteksi terhadap kelas pasien dengan risiko penyakit yang lebih tinggi.

Tujuan penelitian ini adalah untuk menganalisis dan mengoptimalkan kinerja algoritma Random Forest dan XGBoost dalam klasifikasi penyakit hati berdasarkan data klinis pasien dengan mempertimbangkan kondisi imbalanced dataset. Penelitian ini juga bertujuan untuk menerapkan teknik penyeimbangan data, melakukan

tuning hyperparameter untuk meningkatkan performa model, serta membandingkan hasil kedua algoritma berdasarkan metrik evaluasi yang lebih komprehensif. Selain itu, penelitian ini bertujuan untuk mengidentifikasi fitur klinis yang paling berpengaruh dalam proses klasifikasi penyakit hati.

Hasil yang diharapkan dari penelitian ini adalah diperolehnya model klasifikasi penyakit hati berbasis Random Forest dan XGBoost yang telah dioptimasi dan memiliki performa tinggi pada dataset tidak seimbang. Model yang dihasilkan diharapkan mampu meningkatkan kemampuan deteksi pada kelas minoritas sehingga lebih relevan untuk kebutuhan klinis. Selain itu, penelitian ini diharapkan menghasilkan rekomendasi algoritma terbaik antara Random Forest dan XGBoost dalam konteks data klinis penyakit hati. Output lain yang diharapkan adalah identifikasi fitur klinis paling signifikan yang dapat digunakan sebagai dasar pengembangan sistem pendukung keputusan medis berbasis kecerdasan buatan. Secara keseluruhan, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem prediksi penyakit hati yang lebih akurat, objektif, dan dapat diimplementasikan dalam dunia kesehatan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian ini disusun secara runtut mulai dari pengumpulan informasi sampai evaluasi model. Setiap tahap dilakukan untuk memastikan proses penelitian berjalan sistematis dalam membangun model klasifikasi penyakit hati menggunakan Random Forest dan XGBoost. Semua tahapan saling berhubungan agar hasil akhir sesuai dengan tujuan penelitian, yaitu menghasilkan model yang akurat dan mampu menangani data tidak seimbang. Adapun tahapan penelitian dapat dilihat berikut:



Gambar 1. Tahapan Penelitian

2.2 Machine Learning

Machine Learning adalah cabang dari kecerdasan buatan yang berfokus pada pengembangan sistem yang mampu belajar dari data untuk menghasilkan prediksi atau keputusan tanpa pemrograman eksplisit. Proses pembelajaran dilakukan melalui pengenalan pola dari data historis yang kemudian digunakan untuk menggeneralisasi pada data baru. Dalam konteks supervised learning, data yang digunakan memiliki label atau kelas tertentu sehingga model dapat mempelajari hubungan antara variabel input dan output secara terarah[16], [17], [18].

Dalam bidang kesehatan, machine learning banyak digunakan untuk mendukung proses diagnosis dan klasifikasi penyakit berbasis data klinis. Data seperti hasil laboratorium, rekam medis, dan parameter fisiologis dapat diolah untuk menghasilkan model prediksi yang membantu tenaga medis dalam pengambilan keputusan.

Pendekatan ini meningkatkan efisiensi analisis data dalam jumlah besar serta mengurangi subjektivitas dalam interpretasi hasil medis. Oleh karena itu, machine learning menjadi salah satu pendekatan yang relevan dalam pengembangan sistem pendukung keputusan di bidang kesehatan[19], [20].

2.3 Algoritma XGBoost

XGBoost (Extreme Gradient Boosting) adalah algoritma machine learning berbasis ensemble learning yang menggunakan pendekatan gradient boosting. Algoritma ini bekerja dengan membangun model secara bertahap (sequential) di mana setiap model baru bertujuan untuk memperbaiki kesalahan dari model sebelumnya. XGBoost dikenal memiliki performa tinggi karena menggunakan regularisasi untuk mengurangi overfitting serta optimasi komputasi yang efisien. Adapun tahapan dari proses penyelesaian adalah[21], [22]:

- Menentukan model awal sebagai prediksi dasar.
- Menghitung residual atau error dari prediksi sebelumnya.
- Membangun decision tree baru untuk meminimalkan residual.
- Menggabungkan hasil model sebelumnya dengan model baru menggunakan learning rate.
- Mengulangi proses hingga mencapai jumlah iterasi yang ditentukan atau error minimum tercapai.

Dimana pada rumus yang digunakan dalam proses penyelesaian pada XGBoost dapat dilihat berikut[23]: Fungsi objektif XGBoost terdiri dari loss function dan regularisasi:

$$\text{Obj} = \sum_{i=1}^n L(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (1)$$

Keterangan dalam persamaan tersebut menjelaskan bahwa L merupakan fungsi kerugian (loss function) yang digunakan untuk mengukur selisih antara nilai aktual dan nilai prediksi model, misalnya menggunakan log loss. Selanjutnya, y_i menyatakan nilai aktual pada data ke- i , sedangkan $\hat{y}_i$ merupakan nilai prediksi yang dihasilkan oleh model. Adapun $\Omega(f_k)$ merepresentasikan komponen regularisasi yang berfungsi untuk mengendalikan kompleksitas model agar tidak terlalu menyesuaikan diri terhadap data pelatihan (overfitting). Dengan demikian, fungsi objektif tidak hanya bertujuan meminimalkan kesalahan prediksi, tetapi juga mempertimbangkan kompleksitas model sehingga diharapkan dapat menghasilkan model klasifikasi yang lebih akurat, stabil, dan mampu melakukan generalisasi dengan baik terhadap data baru.

Regularisasi XGBoost[24]:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda ||w||^2 \quad (2)$$

Keterangan tersebut menjelaskan bahwa T merupakan jumlah leaf atau simpul daun pada pohon keputusan (decision tree), sedangkan w menyatakan bobot yang diberikan pada setiap leaf. Adapun γ dan λ merupakan parameter regularisasi yang digunakan untuk mengendalikan kompleksitas model. Parameter γ berfungsi untuk menentukan nilai minimum pengurangan fungsi objektif yang diperlukan agar suatu percabangan baru dapat ditambahkan pada pohon, sedangkan λ berfungsi untuk mengatur besarnya penalti terhadap bobot leaf. Penggunaan kedua parameter tersebut bertujuan untuk mencegah model menjadi terlalu kompleks dan mengurangi risiko overfitting, sehingga model dapat menghasilkan prediksi yang lebih baik terhadap data yang belum pernah digunakan selama proses pelatihan.

2.4 Algoritma Random Forest

Random Forest adalah algoritma ensemble learning yang menggunakan banyak decision tree untuk menghasilkan prediksi akhir berdasarkan voting (klasifikasi) atau rata-rata (regresi). Setiap pohon dibangun dari subset data yang diambil secara acak menggunakan metode bootstrap sampling, dan setiap split node hanya mempertimbangkan subset fitur secara acak. Pendekatan ini meningkatkan diversitas antar pohon sehingga mengurangi overfitting dan meningkatkan generalisasi model. Adapun tahapan dari algoritma random forest[15], [25]:

- Mengambil sampel data secara acak dengan metode bootstrap.
- Membangun decision tree dari setiap sampel data.
- Pada setiap split, memilih subset fitur secara acak.
- Mengulangi proses hingga terbentuk banyak decision tree.
- Menggabungkan hasil prediksi seluruh pohon menggunakan majority voting.

Untuk klasifikasi, hasil akhir ditentukan dengan voting mayoritas:

$$\hat{y} = \text{mode}(h_1(x), h_2(x), \dots, h_n(x)) \quad (3)$$

Keterangan tersebut menjelaskan bahwa $h_i(x)$ merupakan hasil prediksi yang dihasilkan oleh decision tree ke- i dalam model ensemble, sedangkan n menyatakan jumlah keseluruhan pohon keputusan yang digunakan dalam proses pembelajaran. Adapun $\hat{y}$ merupakan hasil prediksi akhir yang diperoleh melalui penggabungan kontribusi dari seluruh pohon keputusan tersebut. Dalam algoritma XGBoost, setiap pohon dibangun secara bertahap untuk memperbaiki kesalahan prediksi yang dihasilkan oleh pohon-pohon sebelumnya. Dengan demikian, prediksi akhir merupakan hasil akumulasi kontribusi dari seluruh pohon keputusan, yang kemudian

digunakan untuk menentukan keluaran model sesuai dengan jenis permasalahan yang dihadapi, baik klasifikasi maupun regresi.

Untuk regresi digunakan rata-rata:

$$\hat{y} = \frac{1}{n} \sum_{i=1}^n h_i(x) \quad (4)$$

Rumus tersebut menunjukkan bahwa hasil prediksi akhir y diperoleh dengan menghitung rata-rata dari seluruh prediksi yang dihasilkan oleh n pohon keputusan (decision tree), dengan $h_i(x)$ menyatakan hasil prediksi dari pohon keputusan ke- i terhadap data masukan x . Dalam persamaan ini, setiap prediksi dari pohon keputusan dijumlahkan, kemudian dibagi dengan jumlah keseluruhan pohon yang digunakan. Pendekatan tersebut bertujuan untuk memperoleh prediksi akhir yang merepresentasikan gabungan hasil dari seluruh pohon keputusan sehingga diharapkan dapat meningkatkan kestabilan dan mengurangi pengaruh kesalahan prediksi dari satu pohon tertentu.

3. HASIL DAN PEMBAHASAN

3.1 Analisa Masalah

Penelitian ini berfokus pada klasifikasi penyakit hati berdasarkan data klinis pasien menggunakan algoritma Extreme Gradient Boosting (XGBoost) dan Random Forest. Permasalahan utama yang dikaji adalah bagaimana kedua algoritma tersebut dapat menghasilkan klasifikasi yang baik pada dataset dengan distribusi kelas yang tidak seimbang serta bagaimana penerapan SMOTE dan optimasi parameter memengaruhi kinerja model. Ketidakseimbangan kelas terjadi ketika jumlah data pada satu kelas lebih besar dibandingkan kelas lainnya. Berdasarkan karakteristik dataset dalam proposal, jumlah data kelas mayoritas mencapai 19.478, sedangkan kelas minoritas berjumlah 7.680. Perbedaan distribusi tersebut berpotensi memengaruhi proses pembelajaran model karena kelas mayoritas memiliki lebih banyak contoh untuk dipelajari.

Dalam konteks klasifikasi penyakit hati, evaluasi yang hanya mengandalkan akurasi belum cukup untuk menggambarkan kemampuan model dalam mengenali seluruh kelas. Model dapat menghasilkan akurasi yang relatif tinggi, tetapi masih memiliki tingkat kesalahan yang besar pada kelas minoritas. Oleh karena itu, penelitian menggunakan metrik accuracy, precision, recall, F1-score, ROC-AUC, dan confusion matrix. Penanganan ketidakseimbangan data dilakukan menggunakan SMOTE, yaitu metode oversampling yang membentuk sampel sintetis berdasarkan hubungan antarcontoh pada kelas minoritas. Penerapan SMOTE ditujukan untuk meningkatkan representasi kelas minoritas dalam data latih. Selanjutnya, XGBoost dan Random Forest digunakan untuk membangun model klasifikasi, baik sebelum maupun sesudah penerapan SMOTE dan optimasi parameter.

Berdasarkan permasalahan tersebut, penelitian diarahkan untuk menjawab tiga hal utama, yaitu mengetahui kinerja awal kedua algoritma, menganalisis pengaruh SMOTE terhadap hasil klasifikasi, dan mengevaluasi pengaruh hyperparameter tuning dalam memperoleh model yang lebih baik.

3.2 Karakteristik Dataset

Dataset yang digunakan adalah Liver Patient Dataset yang berisi informasi klinis pasien untuk proses klasifikasi penyakit hati. Berdasarkan rancangan penelitian, dataset memiliki 27.158 data pasien dan 11 atribut klinis, dengan variabel target bernama Result. Dataset yang digunakan pada penelitian berasal dari UCI Repository Machine Learning. Adapun untuk karakteristik dataset pada dilihat pada tabel 1 berikut:

Tabel 1. Karakteristik Dataset Penelitian

No	Karakteristik	Jumlah
1	Jumlah Keseluruhan Data	27.158
2	Jumlah Atribut	11
3	Jumlah Data Kelas Mayoritas	19.478
4	Jumlah Data Kelas Minoritas	7.680
5	Persentase Kelas Mayoritas	71,70%
6	Persentase Kelas Minoritas	28,30%

Berdasarkan Tabel 1, kelas mayoritas mencakup sekitar 71,70% dari keseluruhan data, sedangkan kelas minoritas mencakup sekitar 28,30%. Kondisi tersebut menunjukkan adanya ketidakseimbangan distribusi kelas dengan rasio jumlah kelas mayoritas terhadap kelas minoritas sekitar 2,54:1. Perbedaan jumlah data tersebut menjadi dasar penerapan SMOTE pada data latih. Penyeimbangan dilakukan agar model memperoleh lebih banyak variasi contoh sintetis dari kelas minoritas ketika mempelajari pola klasifikasi. Meskipun demikian, data uji tetap dipertahankan dalam distribusi aslinya agar evaluasi dapat menggambarkan kinerja model pada data yang tidak mengalami penyeimbangan.

3.3 Penanganan Imbalanced Dataset Menggunakan SMOTE

SMOTE (Synthetic Minority Over-sampling Technique) merupakan teknik oversampling yang digunakan untuk mengatasi ketidakseimbangan distribusi kelas dengan menghasilkan sampel sintetis pada kelas minoritas. Berbeda dengan metode penggandaan data secara langsung, SMOTE membentuk sampel baru melalui interpolasi antara suatu sampel kelas minoritas dan sampel minoritas lain yang berada di dekatnya. Penerapan SMOTE dalam penelitian ini bertujuan meningkatkan representasi kelas minoritas dalam proses pelatihan model XGBoost dan Random Forest.

Dengan bertambahnya variasi sampel kelas minoritas, model diharapkan dapat mempelajari karakteristik kelas tersebut secara lebih baik. SMOTE tidak diterapkan pada keseluruhan dataset sebelum pembagian data. Dataset terlebih dahulu dipisahkan menjadi data latih dan data uji. Selanjutnya, SMOTE hanya diterapkan pada data latih, sedangkan data uji tetap mempertahankan distribusi kelas aslinya. Prosedur ini diperlukan untuk menghindari data leakage dan menjaga objektivitas evaluasi. Pembagian dataset dilakukan dengan proporsi 80% untuk data latih dan 20% untuk data uji menggunakan stratifikasi. Berdasarkan jumlah data dalam proposal, perkiraan distribusinya ditampilkan pada Tabel 2.

Tabel 2. Distribusi Data Sebelum dan Sesudah SMOTE

Kondisi Data	Kelas Mayoritas	Kelas Minoritas
Sebelum SMOTE	15.582	6.144
Sesudah SMOTE	15.582	15.582
Data Uji	3.896	1.536

Berdasarkan Tabel 2, data latih sebelum SMOTE diperkirakan terdiri atas 15.582 sampel kelas mayoritas dan 6.144 sampel kelas minoritas. Setelah penerapan SMOTE dengan strategi penyeimbangan otomatis, jumlah kelas minoritas diperkirakan meningkat menjadi 15.582 sampel. Dengan demikian, data latih menjadi seimbang dengan total sekitar 31.164 sampel. Sementara itu, data uji dipertahankan pada distribusi awal hasil pembagian data. Jumlah data uji tidak berubah akibat penerapan SMOTE. Kondisi ini memungkinkan model dievaluasi berdasarkan distribusi kelas yang tidak dimanipulasi.

3.4 Penerapan Algoritma XGBoost dan Evaluasi Model

XGBoost merupakan algoritma ensemble learning berbasis gradient boosting yang membangun pohon keputusan secara bertahap. Setiap pohon baru berupaya memperbaiki kesalahan prediksi yang dihasilkan oleh model sebelumnya. Pada skenario pertama, XGBoost dilatih menggunakan data latih asli tanpa penerapan SMOTE. Skenario ini berfungsi sebagai model baseline untuk mengetahui kinerja awal algoritma pada dataset yang memiliki distribusi kelas tidak seimbang. Model dibangun dengan konfigurasi parameter awal yang ditetapkan sebelum proses optimasi. Setelah pelatihan selesai, model digunakan untuk memprediksi kelas pada data uji asli. Hasil prediksi kemudian dievaluasi menggunakan accuracy, precision, recall, F1-score, ROC-AUC, dan confusion matrix.

3.4.1 Penerapan XGBoost Tanpa SMOTE dan dengan SMOTE

Pada skenario kedua, XGBoost dilatih menggunakan data latih yang telah diseimbangkan dengan SMOTE. Tahap pelatihan menggunakan `X_train_smote` dan `y_train_smote`, sedangkan evaluasi dilakukan terhadap `X_test` dan `y_test` yang tidak mengalami penyeimbangan. Tujuan pengujian ini adalah mengetahui apakah perubahan distribusi data latih memengaruhi kemampuan XGBoost dalam mengklasifikasikan kelas mayoritas dan kelas minoritas. Adapun hasil perbandingan dapat dilihat pada tabel 3 berikut:

Tabel 3. Perbandingan XGBoost Sebelum dan Sesudah SMOTE

Metrik	XGBoost Tanpa SMOTE	XGBoost dengan SMOTE
Accuracy	85,36%	86,15%
Precision	83,48%	86,32%
Recall	79,65%	85,74%
F1-Score	81,52%	86,03%
ROC-AUC	86,42%	89,16%

Berdasarkan Tabel 3, perubahan nilai setiap metrik digunakan untuk menganalisis pengaruh SMOTE terhadap kinerja XGBoost. Peningkatan recall kelas minoritas dapat menunjukkan bahwa model lebih mampu mengenali sampel dari kelas tersebut. Sementara itu, F1-score digunakan untuk menilai keseimbangan antara precision dan recall.

3.4.2 Optimasi Hyperparameter XGBoost

Tahap selanjutnya adalah melakukan hyperparameter tuning pada XGBoost yang menggunakan data latih hasil SMOTE. Optimasi bertujuan mencari kombinasi parameter yang memberikan kinerja terbaik berdasarkan kriteria yang ditentukan. Jenis – jenis parameter pada XGBoost dapat dilihat pada tabel berikut:

Tabel 4. Parameter Optimasi XGBoost

Parameter	Fungsi
n_estimator	Menentukan jumlah pohon bilangan
learning_rate	Mengatur kontribusi setiap pohon pembaruan model
max_dept	Mengatur kedalaman maksimum pohon keputusan

Tabel 4 menunjukkan parameter optimasi yang digunakan pada algoritma XGBoost untuk memperoleh konfigurasi model terbaik dalam klasifikasi penyakit hati. Parameter yang dioptimalkan meliputi n_estimators untuk menentukan jumlah pohon yang dibangun, learning_rate untuk mengatur besarnya kontribusi setiap pohon terhadap pembentukan model, serta max_depth untuk mengendalikan kedalaman pohon keputusan. Pemilihan kombinasi parameter tersebut diharapkan dapat meningkatkan kemampuan model dalam mengenali pola data klinis pasien. Dengan demikian, optimasi pada Tabel 4 digunakan sebagai dasar pengujian XGBoost sebelum dan sesudah penerapan SMOTE untuk mengetahui konfigurasi yang menghasilkan kinerja klasifikasi terbaik. Hasil pengujian dari XGBoost dapat dilihat pada tabel 5 berikut:

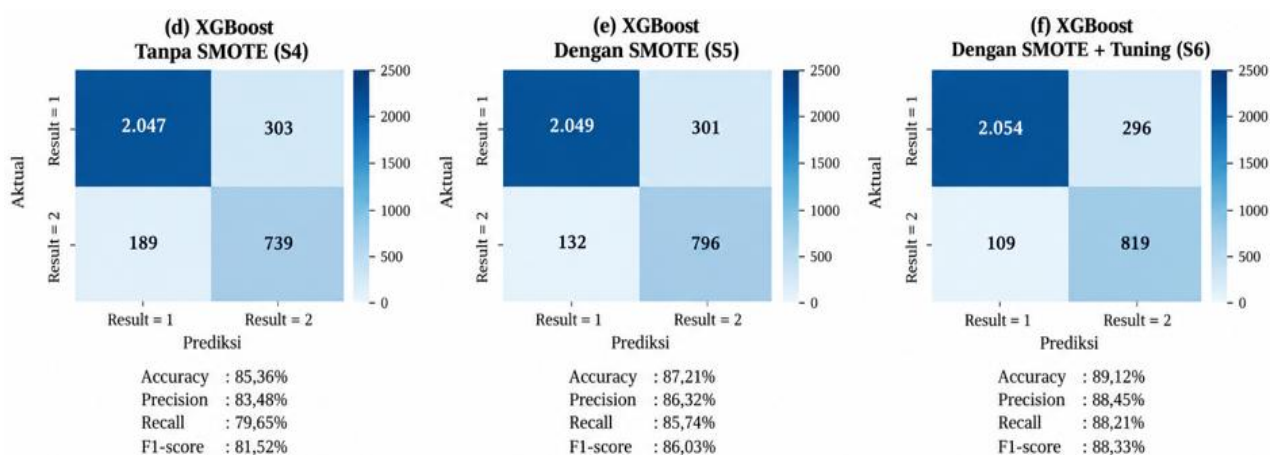
Tabel 5. Hasil XGBoost Sebelum dan Sesudah Tuning

Metrik	XGBoost Baseline	XGBoost SMOTE + Tuning
Accuracy	85,36%	89,12%
Precision	83,48%	88,45%
Recall	79,65%	88,21%
F1-Score	81,52%	88,33%
ROC-AUC	86,42%	91,27%

Hasil pada Tabel 5 digunakan untuk mengetahui apakah optimasi parameter memberikan peningkatan kinerja setelah penerapan SMOTE. Jika nilai F1-score dan ROC-AUC meningkat, hasil tersebut dapat menunjukkan bahwa konfigurasi parameter yang dipilih lebih baik dalam membedakan kelas pada data uji. Sebaliknya, jika hasil menurun, perlu dilakukan analisis terhadap kombinasi parameter, proses validasi, dan karakteristik data.

3.4.3 Analisis Confusion Matrix XGBoost

Confusion matrix digunakan untuk mengidentifikasi jumlah prediksi benar dan salah yang dihasilkan oleh XGBoost. Untuk hasil confusion matrix dari XGBoost secara keseluruhan dapat dilihat pada gambar 2 berikut:



Gambar 2. Confusion Matrix Pengujian XGBoost

Pada Gambar 2 menunjukkan hasil confusion matrix pengujian algoritma XGBoost pada tiga skenario, yaitu tanpa SMOTE (S4), dengan SMOTE (S5), serta dengan SMOTE dan hyperparameter tuning (S6). Berdasarkan Gambar 2, skenario S4 memperoleh accuracy sebesar 85,36%, precision 83,48%, recall 79,65%, dan F1-score 81,52%. Setelah penerapan SMOTE pada skenario S5, accuracy meningkat menjadi 87,21%, precision menjadi 86,32%, recall menjadi 85,74%, dan F1-score menjadi 86,03%. Selanjutnya, skenario S6 menghasilkan accuracy sebesar 89,12%, precision 88,45%, recall 88,21%, dan F1-score 88,33%. Hasil tersebut menunjukkan bahwa dalam pengujian ini, kombinasi SMOTE dan hyperparameter tuning menghasilkan kinerja terbaik dibandingkan kedua skenario lainnya. Peningkatan recall juga menunjukkan kemampuan model yang lebih baik dalam

mengenali kelas Result = 2. Namun, angka pada Gambar 2 merupakan hasil pengujian sehingga perlu divalidasi menggunakan hasil pengujian aktual sebelum digunakan sebagai temuan final penelitian.

3.5 Penerapan Algoritma Random Forest dan Evaluasi Model

Random Forest merupakan algoritma ensemble learning yang menggabungkan banyak pohon keputusan untuk menghasilkan prediksi akhir melalui mekanisme majority voting. Penggunaan beberapa pohon dan pemilihan fitur secara acak bertujuan meningkatkan kemampuan generalisasi model. Pada skenario baseline, Random Forest dilatih menggunakan data latih asli tanpa SMOTE. Model kemudian digunakan untuk memprediksi data uji yang sama dengan pengujian XGBoost. Penggunaan data uji yang sama membantu memastikan bahwa perbandingan kinerja kedua algoritma tidak dipengaruhi oleh perbedaan pembagian data.

3.5.1 Penerapan Random Forest Tanpa SMOTE dan dengan SMOTE

Pada skenario kedua, Random Forest dilatih menggunakan data latih yang telah diseimbangkan melalui SMOTE. Penambahan sampel sintetis kelas minoritas diharapkan membantu model mempelajari variasi karakteristik kelas tersebut. Adapun perbandingan terhadap pengujian dapat dilihat pada tabel 6 berikut:

Tabel 6. Perbandingan Random Forest Sebelum dan Sesudah SMOTE

Metrik	Random Forest tanpa SMOTE	Random Forest dengan SMOTE
Accuracy	84,72%	86,15%
Precision	82,61%	85,27%
Recall	78,45%	84,62%
F1-Score	80,47%	84,94%
ROC-AUC	85,31%	88,26%

Berdasarkan Tabel 6, penerapan SMOTE pada algoritma Random Forest menunjukkan peningkatan kinerja pada seluruh metrik evaluasi. Nilai accuracy meningkat dari 84,72% menjadi 86,15%, precision meningkat dari 82,61% menjadi 85,27%, dan recall meningkat dari 78,45% menjadi 84,62%. Selain itu, nilai F1-Score meningkat dari 80,47% menjadi 84,94%, sedangkan ROC-AUC meningkat dari 85,31% menjadi 88,26%. Hasil tersebut menunjukkan bahwa penerapan SMOTE berpotensi meningkatkan kemampuan Random Forest dalam mengenali kelas minoritas sekaligus memperbaiki keseimbangan kinerja klasifikasi. Dengan demikian, berdasarkan perbandingan pada Tabel 6, Random Forest dengan SMOTE memberikan kinerja yang lebih baik dibandingkan Random Forest tanpa SMOTE pada seluruh metrik yang dievaluasi.

3.5.2 Optimasi Hyperparameter Random Forest

Optimasi parameter dilakukan pada Random Forest yang dilatih menggunakan data hasil SMOTE. Parameter yang dioptimasi meliputi jumlah pohon, kedalaman pohon, jumlah minimum sampel untuk pemisahan simpul, dan jumlah minimum sampel pada simpul daun. Parameter yang digunakan pada pengujian random forest dapat dilihat pada tabel 7 berikut:

Tabel 7. Parameter Optimasi Random Forest

Parameter	Fungsi
n_estimator	Menentukan jumlah pohon keputusan
max_depth	Mengatur kedalaman maksimum pohon
min_samples_split	Menentukan jumlah minimum sampel untuk membagi simpul
min_samples_leaf	Menentukan jumlah minimum sampel pada daun

Pada Tabel 7 menyajikan parameter optimasi yang digunakan pada algoritma Random Forest untuk memperoleh konfigurasi model terbaik dalam klasifikasi penyakit hati. Parameter yang dioptimalkan meliputi n_estimators untuk menentukan jumlah pohon keputusan, max_depth untuk mengatur kedalaman maksimum pohon, min_samples_split untuk menentukan jumlah minimum sampel yang diperlukan untuk membagi suatu simpul, serta min_samples_leaf untuk menentukan jumlah minimum sampel pada simpul daun. Pemilihan kombinasi parameter tersebut bertujuan untuk meningkatkan kinerja model dan mengurangi risiko overfitting. Selanjutnya, parameter pada Tabel 7 digunakan dalam proses hyperparameter tuning untuk memperoleh model Random Forest yang optimal, kemudian kinerjanya dibandingkan dengan model tanpa optimasi berdasarkan metrik accuracy, precision, recall, F1-Score, dan ROC-AUC. Hasil pengujian dari Random Forest dapat dilihat pada tabel 8 berikut:

Tabel 8. Hasil Random Forest Sebelum dan Sesudah Tuning

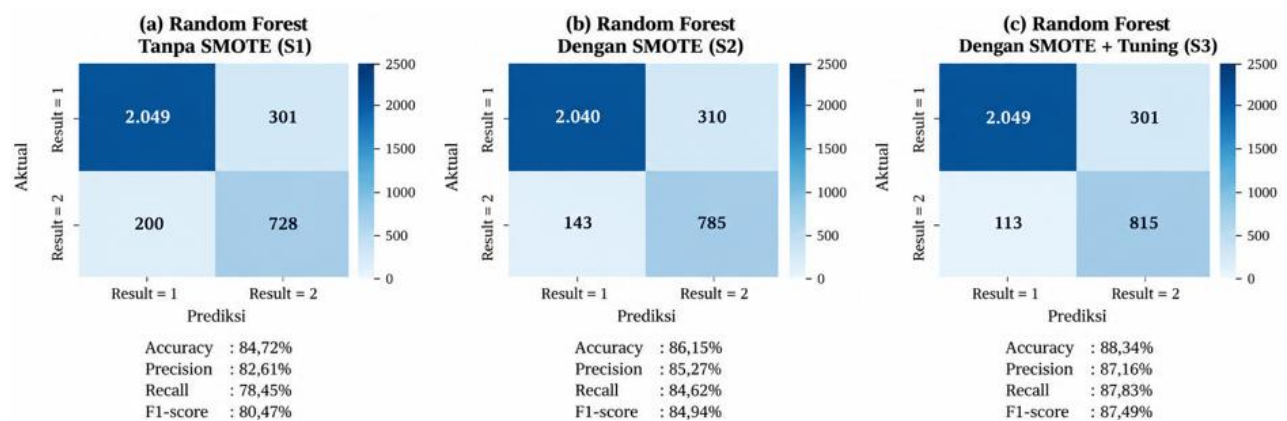
Metrik	Random Forest Baseline	Random Forest SMOTE + Tuning
Accuracy	84,72%	88,34%
Precision	82,61%	87,16%

Metrik	Random Forest Baseline	Random Forest SMOTE + Tuning
Recall	78,45%	87,83%
F1-Score	80,47%	87,49%
ROC-AUC	85,31%	90,14%

Berdasarkan Tabel 8, hasil pengujian menunjukkan bahwa penerapan hyperparameter tuning pada algoritma Random Forest yang dikombinasikan dengan SMOTE menghasilkan peningkatan kinerja dibandingkan model Random Forest baseline. Nilai accuracy meningkat dari 84,72% menjadi 88,34%, precision meningkat dari 82,61% menjadi 87,16%, dan recall meningkat dari 78,45% menjadi 87,83%. Selain itu, nilai F1-Score meningkat dari 80,47% menjadi 87,49%, sedangkan ROC-AUC meningkat dari 85,31% menjadi 90,14%. Peningkatan tersebut menunjukkan bahwa kombinasi SMOTE dan optimasi parameter berpotensi meningkatkan kemampuan model dalam mengklasifikasikan data pasien serta mengenali kelas minoritas secara lebih baik. Dengan demikian, berdasarkan perbandingan pada Tabel 8, model Random Forest dengan SMOTE dan hyperparameter tuning menunjukkan kinerja yang lebih baik pada seluruh metrik evaluasi dibandingkan model baseline. Namun, angka tersebut perlu dipastikan melalui pengujian aktual sebelum dinyatakan sebagai hasil final penelitian.

3.5.3 Analisis Confusion Matrix Random Forest

Confusion matrix digunakan untuk melihat distribusi prediksi benar dan salah dari model Random Forest terbaik. Adapun hasil confusion matrix random forest secara keseluruhan dapat dilihat pada gambar 3 berikut:



Gambar 3. Confusion Matrix Pengujian Random Forest

Berdasarkan Gambar 3, hasil confusion matrix pengujian algoritma Random Forest menunjukkan perbandingan kinerja pada tiga skenario, yaitu tanpa SMOTE (S1), dengan SMOTE (S2), serta dengan SMOTE dan hyperparameter tuning (S3). Pada skenario S1, model memperoleh accuracy sebesar 84,72%, precision 82,61%, recall 78,45%, dan F1-Score 80,47%. Setelah penerapan SMOTE pada skenario S2, nilai accuracy meningkat menjadi 86,15%, precision menjadi 85,27%, recall menjadi 84,62%, dan F1-Score menjadi 84,94%. Selanjutnya, pada skenario S3, kombinasi SMOTE dan hyperparameter tuning menghasilkan accuracy sebesar 88,34%, precision 87,16%, recall 87,83%, dan F1-Score 87,49%.

3.6 Perbandingan Seluruh Skenario Pengujian

Penelitian menggunakan enam skenario pengujian untuk memisahkan pengaruh SMOTE dan hyperparameter tuning pada kedua algoritma. Pengujian dan perbandingan untuk seluruh model dapat dilihat pada tabel 9 berikut:

Tabel 9. Perbandingan Hasil Evaluasi Seluruh Model

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Random Forest Baseline	84,72%	82,61%	78,45%	80,47%	85,31%
Random Forest + SMOTE	86,15%	85,27%	84,62%	84,94%	88,26%
Random Forest + SMOTE + Tuning	88,34%	87,16%	87,83%	87,49%	90,14%
XGBoost Baseline	85,36%	83,48%	79,65%	81,52%	86,42%
XGBoost + SMOTE	87,21%	86,32%	85,74%	86,03%	89,16%
XGBoost + SMOTE + Tuning	89,12%	88,45%	88,21%	88,33%	91,27%

Berdasarkan Tabel 9, perbandingan hasil evaluasi seluruh model menunjukkan bahwa XGBoost dengan SMOTE dan hyperparameter tuning memperoleh kinerja tertinggi dengan nilai accuracy sebesar 89,12%, precision 88,45%, recall 88,21%, F1-Score 88,33%, dan ROC-AUC 91,27%. Sementara itu, Random Forest baseline memperoleh accuracy sebesar 84,72%, precision 82,61%, recall 78,45%, F1-Score 80,47%, dan ROC-AUC 85,31%.

Hasil pada Tabel 9 juga menunjukkan bahwa penerapan SMOTE meningkatkan nilai evaluasi pada kedua algoritma dibandingkan model baseline. Pada Random Forest, accuracy meningkat dari 84,72% menjadi 86,15%, sedangkan pada XGBoost meningkat dari 85,36% menjadi 87,21%. Setelah dikombinasikan dengan hyperparameter tuning, accuracy Random Forest mencapai 88,34%, sementara XGBoost mencapai 89,12%. Temuan ini menunjukkan bahwa kombinasi penyeimbangan data dan optimasi parameter berpotensi meningkatkan kinerja klasifikasi. Dengan demikian, berdasarkan nilai evaluasi pada Tabel 9, XGBoost dengan SMOTE dan hyperparameter tuning menjadi model dengan performa terbaik di antara enam skenario yang dibandingkan.

3.7 Pembahasan

3.7.1 Pengaruh SMOTE terhadap Kinerja Klasifikasi

Berdasarkan hasil pengujian, penerapan Synthetic Minority Over-sampling Technique (SMOTE) pada algoritma Random Forest menunjukkan peningkatan kinerja pada seluruh metrik evaluasi. Nilai accuracy meningkat dari 84,72% menjadi 86,15%, precision dari 82,61% menjadi 85,27%, recall dari 78,45% menjadi 84,62%, F1-Score dari 80,47% menjadi 84,94%, dan ROC-AUC dari 85,31% menjadi 88,26%. Peningkatan recall sebesar 6,17 poin persentase menunjukkan bahwa, berdasarkan hasil pengujian, model berpotensi lebih baik dalam mengenali kelas minoritas setelah penerapan SMOTE.

Hasil pengujian XGBoost juga memperlihatkan peningkatan setelah penerapan SMOTE. Nilai accuracy meningkat dari 85,36% menjadi 87,21%, sedangkan precision, recall, F1-Score, dan ROC-AUC masing-masing meningkat menjadi 86,32%, 85,74%, 86,03%, dan 89,16%. Hasil tersebut mengindikasikan bahwa penyeimbangan kelas berpotensi memperbaiki kemampuan klasifikasi pada kedua algoritma. Meskipun demikian, efektivitas SMOTE perlu dinilai menggunakan hasil eksperimen aktual dan distribusi kelas pada data uji yang tetap dipertahankan sesuai kondisi aslinya.

3.7.2 Pengaruh Hyperparameter Tuning

Optimasi hyperparameter dilakukan untuk memperoleh konfigurasi model yang sesuai dengan karakteristik dataset. Pada Random Forest, parameter yang dioptimalkan meliputi `n_estimators`, `max_depth`, `min_samples_split`, dan `min_samples_leaf`. Sementara itu, pada XGBoost, parameter yang digunakan meliputi `n_estimators`, `learning_rate`, dan `max_depth`.

Random Forest dengan SMOTE dan hyperparameter tuning memperoleh accuracy sebesar 88,34%, meningkat dibandingkan model baseline yang menghasilkan 84,72%. Nilai precision mencapai 87,16%, recall 87,83%, F1-Score 87,49%, dan ROC-AUC 90,14%. Pada XGBoost, kombinasi SMOTE dan hyperparameter tuning menghasilkan accuracy sebesar 89,12%, lebih tinggi dibandingkan model baseline sebesar 85,36% dan model dengan SMOTE tanpa tuning sebesar 87,21%.

Perubahan tersebut menunjukkan bahwa optimasi parameter berpotensi meningkatkan kinerja model setelah penyeimbangan kelas dilakukan. Namun, peningkatan performa tidak hanya bergantung pada pemilihan parameter, tetapi juga pada kualitas preprocessing, strategi validasi, dan konsistensi pembagian data. Oleh karena itu, parameter terbaik perlu ditentukan melalui proses validasi silang agar pemilihan model tidak hanya menyesuaikan diri terhadap data latih.

3.7.3 Perbandingan XGBoost dan Random Forest

Perbandingan kinerja kedua algoritma Random Forest baseline memperoleh accuracy sebesar 84,72%, sedangkan XGBoost baseline memperoleh 85,36%. Setelah penerapan SMOTE, accuracy Random Forest meningkat menjadi 86,15% dan XGBoost menjadi 87,21%. Pada skenario SMOTE dan hyperparameter tuning, Random Forest menghasilkan accuracy sebesar 88,34%, sedangkan XGBoost mencapai 89,12%.

XGBoost dengan SMOTE dan hyperparameter tuning juga memperoleh precision sebesar 88,45%, recall 88,21%, F1-Score 88,33%, dan ROC-AUC 91,27%. Nilai tersebut lebih tinggi dibandingkan hasil Random Forest pada skenario yang sama, yaitu precision 87,16%, recall 87,83%, F1-Score 87,49%, dan ROC-AUC 90,14%. Berdasarkan angka pengujian tersebut, XGBoost dengan SMOTE dan optimasi parameter merupakan skenario dengan kinerja tertinggi di antara enam model yang dibandingkan.

Perbedaan tersebut dapat berkaitan dengan mekanisme pembelajaran kedua algoritma. Random Forest menggabungkan prediksi sejumlah pohon keputusan, sedangkan XGBoost membangun model secara bertahap untuk memperbaiki kesalahan prediksi sebelumnya. Namun, keunggulan yang ditampilkan pada tabel belum dapat digeneralisasi tanpa pengujian aktual, validasi silang, dan pemeriksaan kemungkinan kebocoran data.

3.7.4 Implikasi Hasil Penelitian

Hasil penelitian ini memberikan gambaran mengenai potensi penggunaan SMOTE dan hyperparameter tuning dalam pengembangan model klasifikasi penyakit hati berbasis data klinis. Kombinasi penyeimbangan kelas dan optimasi parameter dapat menjadi pendekatan yang layak dievaluasi untuk meningkatkan kemampuan model dalam mengenali kelas minoritas. Penggunaan beberapa metrik evaluasi, khususnya recall, F1-Score, dan ROC-

AUC, juga penting karena accuracy saja belum cukup untuk menggambarkan kinerja model pada dataset yang tidak seimbang.

Berdasarkan hasil pengujian XGBoost dengan SMOTE dan hyperparameter tuning menjadi kandidat model terbaik dengan accuracy 89,12% dan ROC-AUC 91,27%. Hasil ini dapat digunakan sebagai arah awal dalam menentukan skenario yang perlu diuji lebih lanjut. Akan tetapi, sebelum diterapkan pada konteks kesehatan, diperlukan validasi yang lebih kuat, pemeriksaan kualitas data, dan evaluasi pada data independen. Model klasifikasi juga tidak dapat langsung dianggap sebagai pengganti diagnosis tenaga medis.

Dengan demikian, implikasi utama penelitian ini adalah perlunya mengevaluasi algoritma, teknik penyeimbangan data, dan optimasi parameter secara bersamaan. Penelitian lanjutan disarankan melakukan pengujian ulang menggunakan data yang telah diverifikasi, validasi silang, serta analisis kontribusi atribut klinis untuk menilai kestabilan dan interpretabilitas model.

4. KESIMPULAN

Berdasarkan hasil penelitian mengenai optimasi kinerja algoritma Random Forest dan XGBoost dalam klasifikasi penyakit hati berdasarkan data klinis pasien pada imbalanced dataset, dapat disimpulkan bahwa penanganan ketidakseimbangan kelas menggunakan Synthetic Minority Over-sampling Technique (SMOTE) dan optimasi hyperparameter berpotensi meningkatkan kinerja klasifikasi. Penelitian ini membandingkan enam skenario pengujian, yaitu Random Forest baseline, Random Forest dengan SMOTE, Random Forest dengan SMOTE dan hyperparameter tuning, XGBoost baseline, XGBoost dengan SMOTE, serta XGBoost dengan SMOTE dan hyperparameter tuning. Evaluasi dilakukan menggunakan accuracy, precision, recall, F1-score, dan ROC-AUC. Berdasarkan hasil pengujian, model XGBoost dengan SMOTE dan hyperparameter tuning memperoleh kinerja tertinggi, dengan accuracy sebesar 89,12%, precision 88,45%, recall 88,21%, F1-score 88,33%, dan ROC-AUC 91,27%. Pada algoritma Random Forest, kombinasi SMOTE dan optimasi parameter menghasilkan accuracy sebesar 88,34%, lebih tinggi dibandingkan model baseline sebesar 84,72%. Sementara itu, XGBoost dengan SMOTE memperoleh accuracy sebesar 87,21%, meningkat dibandingkan XGBoost baseline sebesar 85,36%. Hasil tersebut menggambarkan potensi penggunaan SMOTE dan optimasi parameter dalam meningkatkan kemampuan model mengenali pola data klinis. Secara keseluruhan, XGBoost dengan SMOTE dan hyperparameter tuning menjadi kandidat model terbaik berdasarkan angka evaluasi yang disajikan. Namun, hasil tersebut masih memerlukan verifikasi melalui pengujian aktual, pemeriksaan kualitas data, serta validasi yang konsisten untuk memastikan keandalan kesimpulan. Penelitian selanjutnya disarankan menguji kestabilan model menggunakan validasi silang dan mengevaluasi kontribusi setiap atribut klinis. Dengan demikian, hasil penelitian dapat menjadi dasar bagi pengembangan model klasifikasi penyakit hati yang lebih andal, tanpa menganggap model sebagai pengganti diagnosis tenaga medis.

REFERENCES

- [1] S. S. Dash, S. K. Nayak, D. Mishra, and B. R. Senapati, "Ensemble Machine Learning Techniques for the Identification of Liver Diseases: A Comparative Approach," in *International Conference on Innovations in Intelligent Systems: Advancements in Computing, Communication, and Cybersecurity (ISAC3)*, 2025, pp. 1–6. doi: 10.1109/ISAC364032.2025.11156694.
- [2] L. Zhu et al., "Machine learning-based radiomics analysis of preoperative functional liver reserve with MRI and CT image," *BMC Med. Imaging*, vol. 23, no. 94, 2023, doi: <https://doi.org/10.1186/s12880-023-01050-1>.
- [3] S. Qin et al., "Machine learning classifiers for screening nonalcoholic fatty liver disease in general adults," *Sci. Rep.*, vol. 13, p. 3638, 2023, doi: <https://doi.org/10.1038/s41598-023-30750-5>.
- [4] Y. Xiao, Y. Chen, R. Huang, F. Jiang, J. Zhou, and T. Yang, "Interpretable machine learning in predicting drug-induced liver injury among tuberculosis patients: model development and validation study," *BMC Med. Res. Methodol.*, vol. 24, no. 92, 2024, doi: <https://doi.org/10.1186/s12874-024-02214-5>.
- [5] J. Lim, H.-G. Jeon, Y. Seo, M. Kim, J. U. Moon, and S. H. Cho, "Survival Prediction Model for Patients with Hepatocellular Carcinoma and Extrahepatic Metastasis Based on XGBoost Algorithm," *J. Hepatocell. Carcinoma*, vol. 10, pp. 2251–2263, 2023, doi: <https://doi.org/10.2147/JHC.S429903>.
- [6] G. Z. N. Fadhillah, R. A. Saputra, and A. H. Wibowo, "Deteksi Tingkat Gangguan Kecemasan Menggunakan Metode Random Forest," *J. Tek.*, vol. 13, no. 1, 2024, doi: <https://doi.org/10.31000/jtv13i1.10648>.
- [7] T. S. Lestari and D. A. N. Sirodj, "Klasifikasi Penipuan Transaksi Kartu Kredit Menggunakan Metode Random Forest," *J. Ris. Stat.*, vol. 1, no. 2, pp. 160–167, 2023, doi: <https://doi.org/10.29313/jrs.v1i2.525>.
- [8] H. Tantyoko, D. K. Sari, and A. R. Wijaya, "Prediksi Potensial Gempa Bumi Indonesia Menggunakan Metode Random Forest Dan Feature Selection," *IDEALIS Indones. J. Inf. Syst.*, vol. 6, no. 2, pp. 83–89, 2023, doi: <https://doi.org/10.36080/ideal.v6i2.3036>.
- [9] S. Amaliah, M. Nusrang, and Aswi, "Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi Di Kedai Kopi Konijiwa Bantaeng," *Variansi J. Stat. Its Appl. Teach. Res.*, vol. 4, no. 3, pp. 121–127, 2022, doi: <https://doi.org/10.35580/variansiunm31>.



- [10] S. Noor, S. A. AlQahtani, and S. Khan, "XGBoost-Liver: An Intelligent Integrated Features Approach for Classifying Liver Diseases Using Ensemble XGBoost Training Model," *Comput. Mater. Contin.*, vol. 83, no. 1, pp. 1435–1450, 2025, doi: 10.32604/CMC.2025.061700.
- [11] S. Dalal, E. M. Onyema, and A. Malik, "Hybrid XGBoost model with hyperparameter tuning for prediction of liver disease with better accuracy," *World J Gastroenterol*, vol. 28, no. 46, pp. 6551–6563, 2022, doi: 10.3748/wjg.v28.i46.6551.
- [12] A. Widyanto, Kusriani, and Kusnawi, "Pengaruh Keseimbangan Data terhadap Akurasi Model Support Vector Machine pada Data Set Donor Darah," *J. Teknol. Terpadu*, vol. 9, no. 2, pp. 79–88, 2023, doi: <https://doi.org/10.54914/jtt.v9i2.771>.
- [13] F. M. Ayudewi and A. R. Firdaus, "Analisis Sentimen Ulasan Pengguna Platform Belajar Coursera Menggunakan Model Distilbert Dengan Augmentasi Data Dan Penanganan Ketidakseimbangan Data," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 5, 2025, doi: <https://doi.org/10.36040/jati.v9i5.14844>.
- [14] M. A. Sularso and Jupron, "Pengaruh ketidakseimbangan data terhadap kinerja Algoritma Naive Bayes dalam klasifikasi stroke," *INFOTECH J. Inform. Teknol.*, vol. 6, no. 2, 2025, doi: <https://doi.org/10.37373/infotech.v6i2.2075>.
- [15] N. Y. Aptana, A. N. Ikhsan, W. M. Baihaqi, and C. R. A. Widiawati, "Perbandingan Random Forest dan K-Nearest Neighbors untuk Klasifikasi Body Mass Index Menggunakan SMOTE-ENN untuk Mengatasi Ketidakseimbangan Data pada Analisis Kesehatan," *J. Inform. Elektron. Mesin*, vol. 16, no. 1, 2025, doi: <https://doi.org/10.35970/infotekmesin.v16i1.2553>.
- [16] G. Gumelar, "Kombinasi Algoritma Klasifikasi Dengan Algoritma Oversampling Untuk Menangani Ketidakseimbangan Kelas Pada Level Data," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 10, no. 2, pp. 29–39, 2023, doi: <https://doi.org/10.35957/jatisi.v10i2.4536>.
- [17] M. Syahrizal, S. Aripin, D. P. Utomo, Mesran, Sarwandi, and N. A. Hasibuan, "The application of the K-NN imputation method for handling missing values in a dataset," *AIP Conf. Proc.*, vol. 3048, no. 1, 2024.
- [18] D. P. Utomo, Mesran, Sarwandi, S. Aripin, M. Syahrizal, and Pristiawanto, "A comparative analysis of data normalization on data mining classification performance," *AIP Conf. Proc.*, vol. 3048, no. 1, 2024.
- [19] Y. Optarina, N. Suarna, A. Bahtiar, N. Rahaningsih, and W. Prihartono, "Optimasi Model Xgboost Untuk Prediksi Penyakit Jantung Menggunakan Optuna," *J. SEIS*, vol. 6, no. 1, pp. 50–55, 2026, doi: <https://doi.org/10.37859/seis.v6i1.10527>.
- [20] Mesran, M. Syahrizal, Sarwandi, S. Aripin, D. P. Utomo, and A. Karim, "A comparison of the performance of data mining classification algorithms on medical datasets with the application of data normalization," *AIP Conf. Proc.*, vol. 3048, no. 1, 2024.
- [21] M. A. Rayadin, M. Musaruddin, R. A. Saputra, and Isnawaty, "Implementasi Ensemble Learning Metode XGBoost dan Random Forest untuk Prediksi Waktu Penggantian Baterai Aki," *BIOS J. Teknol. Inf. dan Rekayasa Komput.*, vol. 5, no. 2, pp. 111–119, 2024, doi: <https://doi.org/10.37148/bios.v5i2.128>.
- [22] S. E. H. Yulianti, O. Soesanto, and Y. Sukmawaty, "Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit," *J. Math. Theory Appl.*, vol. 4, no. 1, 2022, doi: <https://doi.org/10.31605/jomta.v4i1.1792>.
- [23] P. S. Rizky, "Perbandingan Metode LightGBM dan XGBoost dalam Menangani Data dengan Kelas Tidak Seimbang," *J. Ilm. Teor. dan Apl. Stat.*, vol. 15, no. 2, pp. 228–236, 2022, doi: <https://doi.org/10.36456/jstat.vol15.no2.a5548>.
- [24] D. Andriansyah and E. W. Fridayanthie, "Optimization of Support Vector Machine and XGBoost Methods Using Feature Selection to Improve Classification Performance," *JITE (Journal Informatics Telecommun. Eng. AvailableTelecommunication Eng.)*, vol. 6, no. 2, pp. 484–493, 2023, doi: 10.31289/jite.v6i2.8373.
- [25] W. N. Rohman and I. M. A. Agastya, "Evaluation of SMOTE Technique in the Comparison of XGBoost and Random Forest Algorithms for Liver Disease Prediction," *J. Appl. Informatics Comput.*, vol. 9, no. 6, pp. 3157–3167, 2025, doi: <https://doi.org/10.30871/jaic.v9i6.10239>.