

Evaluation of SMOTE Technique in the Comparison of XGBoost and Random Forest Algorithms for Liver Disease Prediction

Wahyutri Nur Rohman^{1*}, I Made Artha Agastya^{2*}

*Informatika, Universitas Amikom Yogyakarta

wahyutrinurrohman@students.amikom.ac.id¹, artha.agastya@amikom.ac.id²

Article Info

Article history:

Received 2025-07-14

Revised 2025-10-11

Accepted 2025-11-05

Keyword:

Data Mining,
Liver Disease Prediction,
Random Forest,
XGBoost,
Model Evaluation.

ABSTRACT

In many countries, including Indonesia, liver disease remains a major cause of morbidity and mortality. Early detection plays a crucial role in improving treatment outcomes. This study evaluates the performance of two widely used machine learning models Random Forest and XGBoost for predicting liver disease, employing the SMOTE balancing technique to address class imbalance. The primary objectives are to enhance model fairness, reduce overfitting, and improve sensitivity toward the minority class. Model performance is assessed using accuracy, precision, recall, and F1-score metrics. The XGBoost model achieved an average accuracy of 99.74%, precision of 99.77%, recall of 99.75%, and F1-score of 99.72%, while the Random Forest model attained an average accuracy of 99.82%, precision of 99.89%, recall of 99.75%, and F1-score of 99.75%. Both models demonstrated excellent predictive capability, with Random Forest slightly outperforming XGBoost. These results highlight the importance of data balancing and robust model validation in developing reliable machine learning models for healthcare decision-making.



This is an open access article under the [CC-BY-SA](#) license.

I. PENDAHULUAN

Di banyak negara, termasuk Indonesia, penyakit hati merupakan salah satu penyebab utama kesakitan dan kematian. oleh karena itu, identifikasi dini sangat penting untuk meningkatkan keberhasilan terapi dan menghindari dampak buruk lainnya. Setiap tahun, penyakit hati merenggut nyawa sekitar dua juta orang, atau setara dengan 4% dari seluruh kematian di dunia—artinya, satu dari setiap 25 orang yang meninggal, penyebabnya adalah penyakit hati. Sekitar dua pertiga dari angka kematian ini terjadi pada laki-laki. Sebagian besar kematian disebabkan oleh komplikasi sirosis dan kanker hati (karsinoma hepatoseluler), sementara hepatitis akut hanya menyumbang sebagian kecil kasus. Di tingkat global, penyebab utama sirosis adalah infeksi hepatitis virus, konsumsi alkohol berlebihan, serta penyakit hati berlemak non-alkohol (NAFLD)[1].

Seiring dengan kemajuan teknologi digital, berbagai metode berbasis machine learning telah diimplementasikan untuk mengotomatisasi proses diagnosis penyakit hati berdasarkan data rekam medis pasien. Berbagai algoritma telah diterapkan, termasuk Multilayer Perceptron[2],

Artificial Neural Network[3], serta model pohon keputusan seperti Random Forest [4], dan diakui sebagai metode yang efektif dalam pengklasifikasian data tabular. Seiring berjalanannya waktu, pendekatan ensemble seperti XGBoost semakin diminati karena kemampuannya dalam mengelola data kompleks dan menghasilkan prediksi yang lebih akurat [5]. Namun, akurasi model tetap sangat tergantung pada kualitas preprocessing dan tantangan umum berupa ketidakseimbangan distribusi kelas yang sering dijumpai dalam dataset medis.

Berbagai penelitian sebelumnya menunjukkan bahwa mayoritas studi hanya menerapkan satu jenis algoritma tanpa membandingkan kinerjanya dengan metode lain yang lebih kompleks atau stabil [6]. Masalah ketidakseimbangan kelas sering diabaikan, meskipun model yang dilatih tanpa penyeimbangan cenderung bias terhadap kelas mayoritas dan kurang responsif terhadap kelas minoritas. Beberapa penelitian juga belum secara eksplisit menerapkan teknik penyeimbangan seperti SMOTE[7], serta belum melaksanakan validasi berlapis atau pengujian stabilitas model dalam konteks data nyata. Selain itu, mayoritas

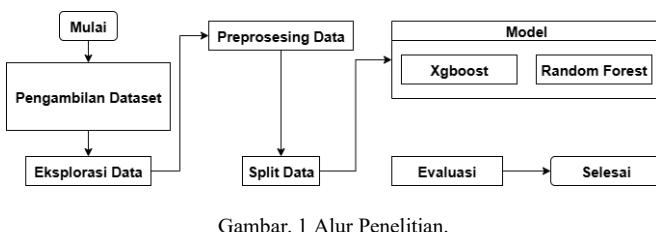
penelitian terbatas pada satu sumber dataset, tanpa pengujian ekstensif terhadap data eksternal yang lebih beragam dan representatif [8]. Ini menciptakan kebutuhan mendesak untuk merancang model prediksi penyakit hati yang tidak hanya akurat, tetapi juga stabil dan adil terhadap kedua kelas prediksi.

Penelitian ini berusaha untuk mengatasi kesenjangan ini dengan mengembangkan dan membandingkan kemampuan dua algoritma ensembel terkemuka, Random Forest dan XGBoost, dalam memprediksi penyakit hati, dengan menggunakan metodologi penyeimbangan data melalui SMOTE[5], [8]. Strategi penyeimbangan ini dianggap dapat meningkatkan sensitivitas model terhadap kelas minoritas dan meningkatkan keadilan prediksi, sekaligus mengurangi overfitting yang terkadang terkait dengan data yang tidak seimbang. Selain itu, berbagai metode validasi, termasuk train-test split dan k-fold cross-validation, digunakan untuk menilai stabilitas model dan meningkatkan kepercayaan pada temuan yang dicapai. Metrik yang komprehensif, seperti akurasi, presisi, recall, dan F1-score, serta pengujian cross-validation digunakan untuk mengevaluasi model.

Oleh karena itu, hasil dari penelitian ini diharapkan dapat memberikan kontribusi yang besar dalam pengembangan sistem pendukung keputusan berbasis pembelajaran mesin yang lebih handal dan praktis di sektor kesehatan. Melalui penelitian ini, diharapkan dapat diperoleh model prediksi penyakit liver dengan performa yang lebih baik dan hasil yang lebih stabil dibandingkan penelitian sebelumnya. Penelitian ini, memberikan konteks dalam pengembangan metode interpretasi model untuk aplikasi klinis serta pengujian model pada dataset yang lebih beragam di masa mendatang.

II. METODE

Penelitian ini menyajikan alur penelitian yang digambarkan dalam sebuah diagram alir, seperti yang diilustrasikan pada Gambar 1.



Gambar. 1 Alur Penelitian.

A. Pengambilan Dataset

Dataset Pasien Penyakit Hati[9], yang bersumber dari Kaggle, terdiri dari 30.691 baris informasi mengenai pasien dengan gangguan penyakit hati dan memiliki 11 fitur. Label dataset muncul dalam variabel hasil, yang memiliki dua nilai: 1 (Pasien Liver), yang mewakili 21.917 data, dan 2 (Pasien Non Liver), yang mewakili 8.774 data. Berikut adalah Informasi detail dari dataset, yang dapat dilihat pada table 1.

TABEL I
DATASET

No	Fitur	Deskripsi	Kategori/nilai
1	Age of the patient	Umur pasien saat awal perawatan.	-
2	Gender of the patient	Jenis Kelamin Pasien	Male = 0, Female = 1.
3	TB	Total Bilirubin	-
4	DB	Direct Bilirubin	-
5	Alkphos	Alkaline Phosphotase	-
6	Sgpt	Alamine Aminotransferase	-
7	Sgot	Aspartate Aminotransferase	-
8	TP	Total Proteins	-
9	ALB	Albumin	-
10	A/G Ratio	Albumin and Globulin Ratio	-
11	Result	Status pasien: Liver Patient or Non Liver Patient	Liver Patient = 1, Non Liver Patient = 2.

Berbagai karakteristik yang ditemukan dalam dataset yang digunakan untuk studi prediksi penyakit hati tercantum dalam Tabel 1. Terdapat 11 fitur utama dalam dataset, yang dikelompokkan ke dalam berbagai kategori penting. Data tersebut mencakup atribut seperti usia pasien pada awal perawatan, jenis kelamin yang dinyatakan sebagai nilai numerik, dan hasil tes laboratorium yang terkait dengan diagnosis penyakit hati.

Fitur Total Bilirubin (TB) dan Direct Bilirubin (DB) berhubungan dengan kadar bilirubin dalam darah, yang dapat mengindikasikan adanya gangguan fungsi hati. Alkaline Phosphotase (Alkphos), Alamine Aminotransferase (Sgpt), dan Aspartate Aminotransferase (Sgot) adalah enzim yang berkaitan dengan metabolisme hati, di mana kadar yang tinggi dapat menandakan kerusakan atau peradangan hati. Selain itu, Total Proteins (TP), Albumin (ALB), dan Albumin and Globulin Ratio (A/G Ratio) juga penting dalam evaluasi kondisi liver karena berhubungan langsung dengan kemampuan hati dalam memproduksi protein.

Adapun fitur Result digunakan sebagai variabel target yang menentukan status pasien apakah Liver Patient (nilai 1) atau Non-Liver Patient (nilai 2). Seluruh data ini menjadi dasar dalam proses analisis dan pembuatan model prediksi untuk mendeteksi penyakit liver secara lebih akurat.

B. Exploratory Data Analysis (EDA)

Pada tahapan ini peneliti menerapkan beberapa metode explorasi data analisis seperti struktur dataset, pengecekan missing values, dan jumlah dari baris kolom result.

C. Preprocessing Data

Preprocessing data adalah tahapan yang perlu dilakukan untuk mengevaluasi data dalam dataset yang akan digunakan pada penelitian ini, sehingga memungkinkan kelangsungan proses berikutnya[10]. Teknik yang digunakan dalam penelitian yaitu penanganan missing value pada atribut dataset dilakukan karena terdapat Informasi yang hilang dan dapat menurunkan hasil dari akurasi saat proses klasifikasi[11].

Selanjutnya, dilakukan pengkodean label menggunakan metode Label Encoding, yaitu teknik sederhana yang menetapkan nilai integer untuk setiap nilai unik dari variabel kategorikal. Meskipun sederhana, metode ini efektif untuk model berbasis pohon keputusan seperti Random Forest dan XGBoost yang tidak sensitif terhadap hubungan ordinal antar kategori[12]. Selain itu, untuk mengatasi permasalahan ketidakseimbangan kelas pada dataset, penelitian ini juga menerapkan teknik balancing data menggunakan Synthetic Minority Over-sampling Technique (SMOTE), yang bertujuan untuk meningkatkan proporsi data minoritas sehingga model tidak bias terhadap kelas mayoritas dan performa klasifikasi menjadi lebih seimbang [13].

Normalisasi adalah langkah penting berikutnya dalam pra-pemrosesan data setelah menangani nilai yang hilang, pengkodean label, dan menyeimbangkan data. Normalisasi adalah cara untuk mengubah nilai atribut numerik sehingga semuanya berada dalam rentang yang ditentukan. Dengan cara ini, semua karakteristik memiliki bobot yang sama dalam proses pemodelan. Hal ini sangat penting karena algoritma pembelajaran mesin dapat menjadi bias terhadap fitur dengan skala yang lebih tinggi ketika karakteristik memiliki rentang nilai yang sangat berbeda. Hal ini membuat model menjadi kurang akurat. Normalisasi meningkatkan konsistensi dan kualitas data, yang membuat proses klasifikasi menjadi lebih akurat dan lebih cepat[14]

D. Split Data

Data pelatihan dan data pengujian adalah dua kategori utama yang menjadi dasar pembagian dataset prapemrosesan yang digunakan dalam investigasi ini. Pelatihan menggunakan 80% data, sedangkan pengujian menggunakan 20% sisanya. Data pelatihan sangat penting bagi model pembelajaran mesin untuk mengenali dan mempelajari karakteristik dan pola dari set data yang tersedia. Sementara itu, data pengujian digunakan untuk mengevaluasi kinerja model yang telah dilatih dan memastikan seberapa baik model tersebut dapat menghasilkan prediksi yang akurat pada data yang belum pernah dilihat sebelumnya. Pemisahan ini sangat penting untuk menjamin bahwa model bekerja dengan baik pada data yang telah diketahui dan dapat menggeneralisasi data yang belum pernah diproses selama pelatihan. Akurasi dan kemampuan generalisasi model diukur dengan menggunakan data uji untuk mengevaluasi seberapa baik kinerjanya pada data yang belum pernah dilihat sebelumnya[15]

E. Modelling

Langkah penting dalam proses penelitian ini adalah pemodelan, di mana algoritma atau model dikembangkan untuk meramalkan klasifikasi penyakit hati dengan menggunakan data dari langkah prapemrosesan sebelumnya. Pada tahap ini, para peneliti menggunakan dua algoritma machine learning yang dikenal memiliki kinerja yang baik dalam menangani masalah klasifikasi Random Forest dan Extreme Gradient Boosting (XGBoost).

XGBoost merupakan algoritma berbasis pohon keputusan yang menerapkan teknik boosting untuk meningkatkan akurasi model dengan meminimalkan kesalahan secara bertahap pada iterasi sebelumnya. Sementara itu, Random Forest merupakan metode ensemble learning yang membentuk sejumlah pohon keputusan pada berbagai subset data, kemudian menggabungkan hasilnya untuk meningkatkan akurasi dan mengurangi overfitting.

Kedua algoritma ini dipilih karena telah banyak digunakan dalam penelitian di bidang kesehatan dan terbukti efektif dalam memproses data dengan fitur numerik maupun kategorikal serta mampu menangani ketidakseimbangan data[16], [17].

F. Evaluasi

Pada tahap evaluasi model klasifikasi, Confusion Matrix memainkan peran penting dalam memahami kesalahan prediksi secara lebih detail. Menurut Sathyaranayanan dan Tantri, Confusion Matrix tidak hanya memungkinkan perhitungan akurasi, tapi juga metrik lain seperti precision, recall, specificity, dan F1-score serta pengujian cross-validation yang sangat dibutuhkan dalam analisis data medis, terutama untuk dataset dengan ketidakseimbangan kelas [18].

Hal ini menunjukkan bahwa Confusion Matrix memberikan gambaran visual dan numerik yang komprehensif tentang kekuatan dan kelemahan masing-masing algoritma dalam tugas klasifikasi[19].

III. HASIL DAN PEMBAHASAN

Pada bagian ini akan menyajikan temuan penelitian dan analisis data yang diperoleh, mulai dari langkah pertama (pengambilan set data) hingga evaluasi dari hasil penelitian. Tujuan dari penelitian ini adalah untuk mengidentifikasi kelebihan dan kekurangan masing-masing model serta variabel-variabel yang mempengaruhi efektivitasnya. Dalam upaya untuk mendapatkan pemahaman yang lebih baik tentang keefektifan teknik yang digunakan, hasilnya akan dibandingkan dengan penelitian sebelumnya untuk menentukan apakah teknik tersebut sesuai atau berbeda.

A. Pengambilan dan Analisa dataset

Dataset yang digunakan dalam penelitian ini memiliki 11 fitur. Dimana proses pemanggilan dan menampilkan dataset menggunakan perintah read dan print dari library pandas.

No	Fitur	Nilai	Nilai	Nilai
	Data Baris Ke-1		Data baris 2	Data baris 8
1	Age of the patient	65	62	55
2	Gender of the patient	Female	Male	Male
3	Total Bilirubin	0.7	10.9	0.9
4	Direct Bilirubin	0.1	5.5	0.3
5	Alkphos Alkaline Phosphotase	187.0	699.0	202.0

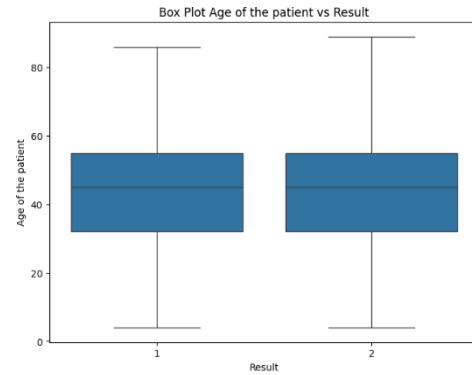
6	Sgpt Alamine Aminotransferase	16.0	64.0	22.0
7	Sgot Aspartate Aminotransferase	18.0	100.0	19.0
8	Total Proteins	6.8	7.5	7.4
9	ALB Albumin	3.3	3.2	4.1
10	A/G Ratio	0.90	0.74	1.20
11	Result	1	1	2

Tabel 2 menyajikan data penelitian yang terdiri dari informasi demografis dan hasil pemeriksaan laboratorium pasien yang digunakan untuk memprediksi penyakit hati. Data ini mencakup usia (Age of the patient) dan jenis kelamin (Gender of the patient) pasien, yang penting karena prevalensi penyakit hati dapat bervariasi berdasarkan kelompok usia dan jenis kelamin. Selanjutnya, beberapa parameter biokimia yang berkaitan langsung dengan fungsi hati juga disertakan, seperti Total Bilirubin dan Direct Bilirubin yang mengukur kadar bilirubin dalam darah, kadar yang tinggi dapat mengindikasikan kerusakan hati atau gangguan saluran empedu. Alkphos (Alkaline Phosphatase) adalah enzim yang biasanya meningkat saat terjadi masalah pada saluran empedu atau tulang.

Selain itu, Sgpt (Alamine Aminotransferase) dan Sgot (Aspartate Aminotransferase) merupakan enzim yang dilepaskan ke aliran darah ketika sel hati atau otot mengalami kerusakan, sehingga menjadi indikator penting dalam diagnosis penyakit hati. Parameter Total Proteins, Albumin (ALB), dan A/G Ratio digunakan untuk menilai kemampuan hati dalam memproduksi protein; nilai albumin yang rendah atau rasio albumin terhadap globulin yang abnormal dapat menunjukkan adanya gangguan hati kronis. Kolom terakhir, Result, berfungsi sebagai label klasifikasi dengan nilai 0 menunjukkan pasien tidak terindikasi penyakit hati dan nilai 1 menunjukkan pasien terindikasi penyakit hati. Kombinasi dari seluruh parameter ini memberikan gambaran menyeluruh mengenai kondisi pasien, sehingga dapat dimanfaatkan oleh algoritma pembelajaran mesin seperti Random Forest dan XGBoost untuk membedakan antara pasien sehat dan pasien yang berisiko secara lebih akurat.

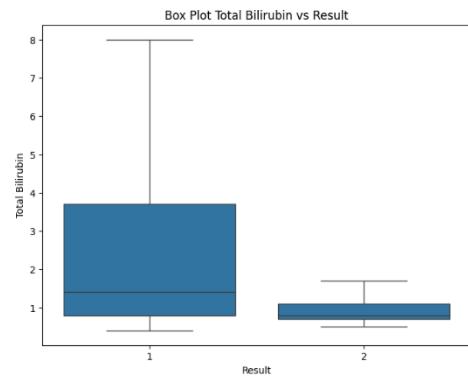
B. Exploratory Data Analysis (EDA)

Tahap Exploratory Data Analysis (EDA) dilakukan untuk memahami karakteristik awal dari dataset yang digunakan sebelum masuk ke tahap pemodelan. Proses ini bertujuan untuk menggali pola, distribusi, serta hubungan antar variabel sehingga dapat memberikan gambaran menyeluruh mengenai data. Melalui visualisasi seperti box plot, histogram, maupun heatmap, EDA membantu dalam mengidentifikasi adanya perbedaan signifikan antara pasien sakit dan tidak sakit, mendeteksi nilai ekstrem (outlier), serta mengevaluasi apakah suatu fitur berpotensi kuat dalam membedakan kondisi kesehatan pasien. Dengan demikian, EDA menjadi langkah penting untuk memastikan kualitas data dan menemukan insight yang relevan bagi analisis lebih lanjut.



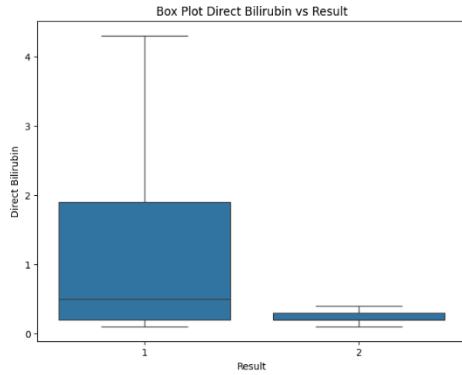
Gambar. 2 Box Plot Age vs Result.

Gambar .2 menunjukkan bahwa distribusi umur antara pasien sakit (Result = 1) dan pasien sehat (Result = 2) relatif serupa. Median keduanya berada di kisaran 45 tahun, dengan rentang data yang cukup lebar mulai dari usia muda hingga lanjut usia. Hal ini mengindikasikan bahwa usia bukanlah faktor pembeda utama antara kondisi pasien sakit dan sehat dalam dataset ini. Dengan kata lain, penyakit yang diteliti tidak hanya menyerang kelompok usia tertentu, melainkan dapat terjadi pada berbagai rentang umur.



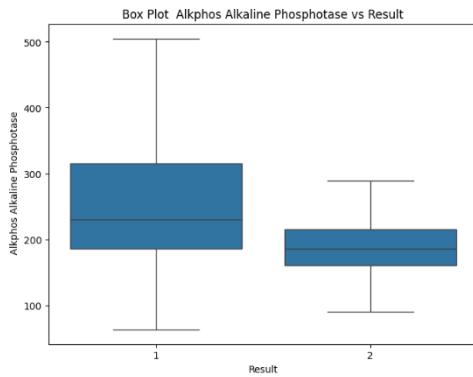
Gambar. 3 Box Plot Total Bilirubin vs Result.

Box plot total bilirubin memperlihatkan perbedaan yang cukup jelas antara pasien sakit (Result = 1) dan pasien sehat (Result = 2). Pasien sakit memiliki median total bilirubin yang lebih tinggi, dengan sebaran data yang jauh lebih lebar hingga mencapai nilai ekstrem sekitar 7–8. Sebaliknya, pasien sehat cenderung memiliki kadar total bilirubin yang lebih stabil dan rendah, dengan rentang sempit di sekitar 1. Temuan ini menunjukkan bahwa peningkatan kadar total bilirubin dapat menjadi indikator penting untuk membedakan pasien sakit dari yang sehat.



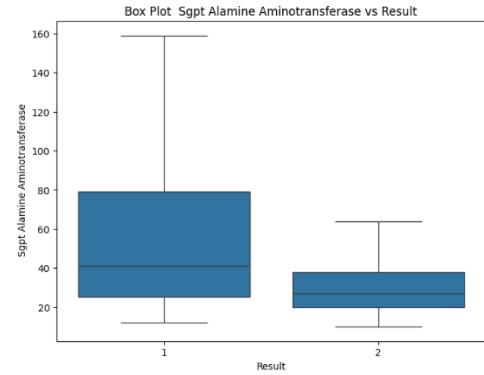
Gambar. 4 Box Plot Direct Bilirubin vs Result.

Gambar 4 juga terdapat perbedaan yang sangat signifikan antara kedua kelompok. Tingginya kadar Bilirubin Langsung pada kelompok yang sakit (Result 1) dibandingkan dengan kelompok yang sehat (Result 2) menunjukkan bahwa Bilirubin Langsung adalah penanda yang kuat (marker) untuk membedakan orang sakit dari orang sehat, dan kemungkinan besar merupakan indikator adanya penyakit yang memengaruhi fungsi hati atau sistem empedu, di mana kadar bilirubin terkonjugasi/langsung akan meningkat.



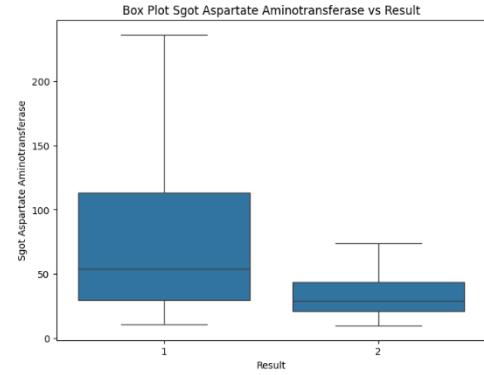
Gambar. 5 Box Plot Alkphos Alkaline Phosphotase vs Result.

Berdasarkan box plot pada gambar 5, terlihat jelas ada perbedaan signifikan pada kadar Alkaline Phosphatase (Alkphos) antara kedua kelompok. Kelompok Result 1 secara konsisten menunjukkan kadar Alkphos yang lebih tinggi (median ≈ 230 U/L) dengan variasi data yang sangat luas, mulai dari nilai terendah ≈ 50 U/L hingga tertinggi ≈ 500 U/L. Sebaliknya, kelompok Result 2 memiliki kadar Alkphos yang lebih rendah (median ≈ 185 U/L) dan sebaran data yang jauh lebih stabil dan sempit (berkisar antara ≈ 90 U/L hingga ≈ 290 U/L). Hal ini mengindikasikan bahwa kategori atau kondisi yang diwakili oleh Result 1 cenderung memiliki kadar Alkphos yang meningkat dan lebih tidak terkontrol dibandingkan dengan Result 2.



Gambar. 6 Box Plot Sgpt Alamine Aminotransferase vs Result.

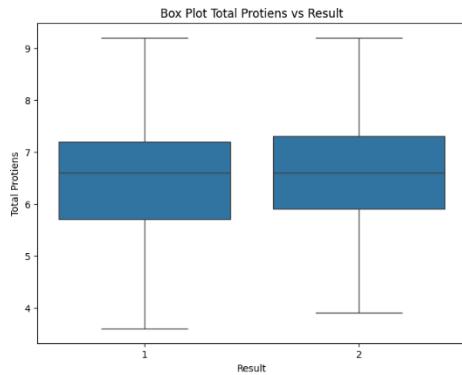
Berdasarkan box plot ini, terdapat perbedaan yang sangat jelas pada kadar enzim SGPT (Alkaline Aminotransferase) antara kedua kelompok. Kelompok Result 1 menunjukkan kadar SGPT yang jauh lebih tinggi dan bervariasi luas. Nilai tengah (median) SGPT pada kelompok 1 berada di sekitar 42 U/L, dengan rentang sebaran data (dari whisker bawah ke atas) yang sangat besar, yaitu dari ≈ 10 U/L hingga mencapai puncaknya di ≈ 160 U/L. Sebaliknya, kelompok Result 2 memiliki kadar SGPT yang secara signifikan lebih rendah dan stabil, dengan nilai tengah (median) hanya sekitar 28 U/L dan rentang sebaran data keseluruhan yang jauh lebih sempit dan rendah, yakni dari ≈ 10 U/L hingga ≈ 64 U/L. Perbedaan ini menunjukkan bahwa kondisi atau status yang diwakili oleh Result 1 cenderung dikaitkan dengan peningkatan SGPT yang merupakan indikator potensial kerusakan atau peradangan hati.



Gambar. 7 Box Plot Sgot Aspartate Aminotransferase vs Result.

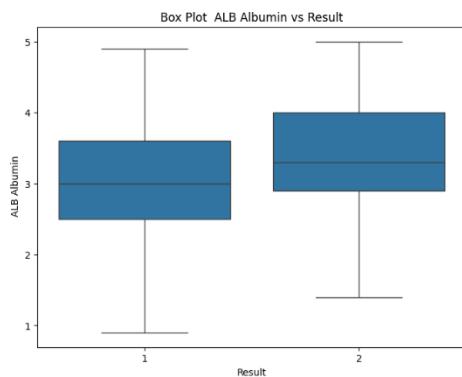
Berdasarkan box plot ini, terdapat perbedaan signifikan pada kadar enzim SGOT (Aspartate Aminotransferase) antara kedua kelompok. Kelompok Result 1 menunjukkan kadar SGOT yang secara substansial lebih tinggi dan sangat bervariasi. Nilai tengah (median) SGOT pada kelompok 1 berada di sekitar 55 U/L, dengan rentang sebaran data (dari whisker bawah ke atas) yang sangat besar, yaitu dari ≈ 10 U/L hingga mencapai puncaknya di ≈ 230 U/L. Sebaliknya, kelompok Result 2 memiliki kadar SGOT yang jauh lebih rendah dan stabil, dengan nilai tengah (median) hanya sekitar 30 U/L dan rentang sebaran data keseluruhan yang jauh lebih sempit, yakni dari ≈ 10 U/L hingga ≈ 75 U/L. Perbedaan drastis

ini mengindikasikan bahwa kondisi atau status yang diwakili oleh Result 1 sangat terkait dengan peningkatan SGOT, yang sering menjadi penanda kuat adanya kerusakan pada sel hati atau jantung.



Gambar. 8 Box Plot Total Proteins vs Result.

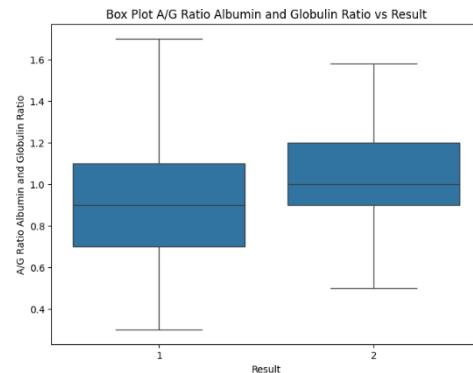
Berdasarkan box plot ini, terlihat bahwa tidak ada perbedaan yang signifikan pada kadar Total Protein antara kelompok Result 1 dan Result 2. Kedua kelompok menunjukkan nilai median yang hampir identik, berada di sekitar 6.6 g/dL. Selain itu, rentang nilai (sebaran) 50% data tengah (kotak) dan rentang keseluruhan data (whisker) juga sangat mirip. Untuk Result 1, nilai berkisar antara ≈ 3.5 hingga ≈ 9.2 g/dL, sementara untuk Result 2, nilainya juga berkisar dari ≈ 3.8 hingga ≈ 9.2 g/dL. Kesamaan ini menunjukkan bahwa Total Protein bukanlah penanda yang efektif untuk membedakan kondisi atau kategori yang diwakili oleh Result 1 dan Result 2, karena kadarnya serupa di kedua kelompok.



Gambar. 9 Box Plot ALB Albumin vs Result.

Box plot albumin menunjukkan bahwa pasien sehat (Result = 2) cenderung memiliki kadar albumin yang lebih tinggi dibandingkan pasien sakit (Result = 1). Median albumin pada pasien sehat berada di atas 3, sementara pada pasien sakit lebih rendah, meskipun distribusi keduanya masih cukup beririsan. Selain itu, sebaran kadar albumin pada pasien sakit lebih bervariasi dengan adanya nilai rendah hingga mendekati 1, sedangkan pasien sehat memiliki rentang lebih stabil. Temuan ini mengindikasikan bahwa kadar

albumin yang lebih rendah dapat berkaitan dengan kondisi penyakit pada pasien.



Gambar. 10 Box Plot A/G Ratio vs Result.

Box plot rasio A/G memperlihatkan bahwa pasien sehat (Result = 2) memiliki median rasio A/G yang sedikit lebih tinggi dibandingkan pasien sakit (Result = 1). Selain itu, distribusi nilai pada pasien sehat lebih terkonsentrasi di sekitar nilai 1, sedangkan pasien sakit menunjukkan variasi yang lebih besar dengan adanya nilai yang sangat rendah hingga mendekati 0.3. Hal ini menunjukkan bahwa rasio A/G yang lebih rendah berpotensi berkaitan dengan kondisi penyakit, sedangkan rasio yang lebih tinggi cenderung ditemukan pada pasien sehat.

Hasil eksplorasi data menunjukkan bahwa tidak semua variabel dalam dataset memiliki perbedaan signifikan antara pasien sakit dan sehat. Faktor usia dan total protein, misalnya, tidak memberikan perbedaan berarti sehingga kurang efektif sebagai penanda penyakit. Namun, beberapa variabel laboratorium lain seperti total bilirubin, bilirubin langsung, alkaline phosphatase (Alkphos), SGPT, SGOT, albumin, serta rasio A/G memperlihatkan perbedaan distribusi yang sangat jelas antara kelompok sakit (Result = 1) dan sehat (Result = 2). Pasien sakit cenderung memiliki nilai yang lebih tinggi pada indikator enzim hati (bilirubin, SGOT, SGPT, Alkphos) dan nilai yang lebih rendah pada albumin maupun rasio A/G. Pola distribusi yang konsisten ini memperkuat dugaan bahwa indikator-indikator tersebut dapat menjadi variabel prediktif penting dalam mendeteksi kondisi penyakit.

C. Penanganan missing values

Pada tahap ini, data yang memiliki nilai kosong tidak dihapus, melainkan diisi kembali menggunakan teknik imputasi. Untuk data numerik, seperti Total Bilirubin, Direct Bilirubin, Alkphos Alkaline Phosphotase, dan parameter laboratorium lainnya, nilai yang hilang diganti dengan nilai median. Median dipilih karena lebih tahan terhadap pengaruh nilai ekstrem (outlier), sehingga distribusi data tetap terjaga. Sementara itu, untuk data kategorikal seperti Gender of the patient, nilai yang hilang diisi dengan nilai modus, yaitu kategori yang paling sering muncul. Dengan cara ini, semua data tetap dapat dimanfaatkan untuk analisis tanpa mengurangi jumlah sampel, sehingga model memiliki

informasi yang lengkap dan representatif. Hasil perbandingan sebelum dan sesudah proses imputasi dapat dilihat pada Tabel 4 dan Tabel 5.

TABEL 3
MISSING VALUES

Fitur	Jumlah
Age of the patient	2
Gender of the patient	902
Total Bilirubin	648
Direct Bilirubin	561
Alkphos Alkaline Phosphotase	796
Sgpt Alamine Aminotransferase	538
Sgot Aspartate Aminotransferase	462
Total Protiens	463
ALB Albumin	494
A/G Ratio	559
Result	0

TABEL 4
PENANGANAN MISSING VALUES

Fitur	Jumlah
Age of the patient	0
Gender of the patient	0
Total Bilirubin	0
Direct Bilirubin	0
Alkphos Alkaline Phosphotase	0
Sgpt Alamine Aminotransferase	0
Sgot Aspartate Aminotransferase	0
Total Protiens	0
ALB Albumin	0
A/G Ratio	0
Result	0

D. Encoding

Pada tahap encoding, peneliti menerapkan teknik one-hot encoding untuk mengonversi variabel kategorikal, seperti Gender of the patient menjadi format yang dapat digunakan dalam model machine learning. One-hot encoding mengubah kategori Female dan Male menjadi dua kolom terpisah dengan nilai 0 dan 1, yang mewakili keberadaan masing-masing kategori pada setiap baris data. Pada kolom Result, terjadi perubahan label dari 1 menjadi 0 dan 2 menjadi 1. Hal ini dilakukan agar format data sesuai dengan kebutuhan algoritma XGBoost, yang secara default mengenali label kelas dimulai dari angka 0. Perubahan ini bersifat teknis dan tidak mengubah makna data, melainkan hanya menyesuaikan struktur label agar model dapat memproses data dengan benar serta menghasilkan prediksi yang optimal.

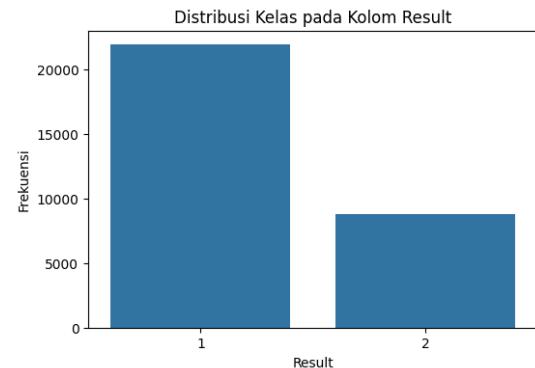
Selain itu, variabel numerik lainnya, seperti Total Bilirubin, Direct Bilirubin, dan berbagai indikator medis lainnya, dipertahankan dalam bentuk aslinya menggunakan teknik passthrough. Teknik ini memungkinkan model untuk memproses data numerik tanpa perubahan, memastikan bahwa informasi yang terkandung dalam variabel numerik tetap utuh. Dengan demikian, tahap encoding ini memungkinkan data yang terstruktur menjadi siap digunakan untuk analisis lebih lanjut tanpa kehilangan informasi penting dari fitur kategorikal dan numerik. Hasil encoding dapat dilihat pada Tabel 5.

TABEL 5
LABEL ENCODING

No	Fitur	Nilai Sebelum Encoding	Nilai Sesudah Encoding
Data baris Ke-1			
1	Age of the patient	65.0	65.0
2	Gender of the patient	Female	0
3	Total Bilirubin	0.7	0.7
4	Direct Bilirubin	0.1	0.1
5	Alkphos Alkaline Phosphotase	187.0	187.0
6	Sgpt Alamine Aminotransferase	16.0	16.0
7	Sgot Aspartate Aminotransferase	18.0	18.0
8	Total Protiens	6.8	6.8
9	ALB Albumin	3.3	3.3
10	A/G Ratio	0.90	0.90
11	Result	1	0

E. Balancing data

Data ini awalnya mengalami ketidakseimbangan dataset, dimana terdapat pada class 1 dan 2 pada dataset penelitian ini. Dari gambar 11 terlihat ketidakseimbangan dari dataset.



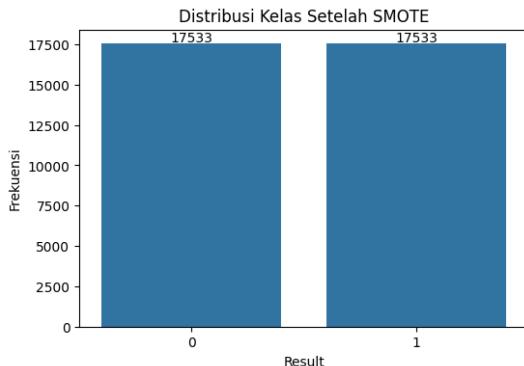
Gambar 11. Sebelum dilakukan Balancing data.

Untuk mengatasi ketidakseimbangan kelas pada data, digunakan metode SMOTE (Synthetic Minority Over-sampling Technique). Setelah penerapan SMOTE, jumlah data pelatihan meningkat dari 24.552 menjadi 35.066 sampel, dengan distribusi yang kini seimbang antara kedua kelas — masing-masing terdiri dari 17.533 data untuk kelas 1 dan 2. Proses ini memastikan model memperoleh pembelajaran yang adil dari kedua kelas, sehingga mampu mengurangi bias terhadap kelas mayoritas dan meningkatkan kemampuan generalisasi terhadap data baru.

Dengan dilakukannya balancing data menggunakan SMOTE, diharapkan performa model klasifikasi meningkat, terutama pada metrik recall dan F1-score untuk kelas minoritas, yang sangat penting agar kasus Liver Disease tidak terlewatkan.

Proses SMOTE (Synthetic Minority Over-sampling Technique) hanya diterapkan pada data training, bukan pada keseluruhan dataset. Hal ini dilakukan untuk mencegah kebocoran data (data leakage) antara data pelatihan dan data pengujian. Dengan kata lain, data dibagi terlebih dahulu

menjadi data training dan testing menggunakan proporsi 80:20, kemudian SMOTE diterapkan hanya pada data training agar distribusi kelas menjadi seimbang sebelum model dilatih. Sementara itu, data testing tetap dibiarakan dalam kondisi aslinya agar dapat digunakan untuk menguji kemampuan model pada data yang merepresentasikan kondisi nyata di dunia sebenarnya.



Gambar 12. Setelah dilakukan Balancing data.

F. Normalisasi data

Pada tahap normalisasi data, teknik MinMaxScaler digunakan untuk mengubah nilai variabel numerik menjadi kisaran antara 0 dan 1. Pendekatan ini mengubah setiap nilai fitur sehingga nilai minimumnya adalah 0 dan nilai terbesarnya adalah 1. Teknik ini sangat penting untuk memberikan penskalaan yang seragam pada semua fitur, sehingga tidak ada satu pun fitur yang mendominasi model karena kisaran nilai yang lebih luas. Tabel 6 menunjukkan hasil sebelum dan sesudah normalisasi data yang menggambarkan perubahan distribusi nilai fitur setelah proses normalisasi.

TABEL 6
NORMALISASI DATA

No	Fitur	Nilai Sebelum Normalisasi Data	Nilai Sesudah Normalisasi Data
Data baris Ke-1			
1	Age of the patient	65.0	0.709302
2	Gender of the patient	0	0.0
3	Total Bilirubin	0.7	0.004021
4	Direct Bilirubin	0.1	0.000000
5	Alkphos Alkaline Phosphotase	187.0	0.060576
6	Sgt Alamine Aminotransferase	16.0	0.003015
7	Sgot Aspartate Aminotransferase	18.0	0.001626
8	Total Protiens	6.8	0.594203
9	ALB Albumin	3.3	0.521739
10	A/G Ratio	0.90	0.24
11	Result	1	1

G. Split data

Pada tahap ini, dilakukan pemisahan dataset menjadi dua variabel utama, yaitu X sebagai kumpulan fitur independen (seluruh atribut kecuali kolom result) dan y sebagai variabel target (result). Proses pembagian data dilakukan

menggunakan teknik train-test split dengan proporsi 80% untuk data latih (train) dan 20% untuk data uji (test). Untuk menjaga konsistensi dalam pengacakan data setiap kali kode dijalankan, digunakan parameter random_state 42. Tabel 7 menunjukkan hasil pemisahan data tersebut, yang menggambarkan distribusi antara data latih dan data uji setelah proses pembagian dilakukan.

TABEL 7
SPLIT DATA

Keterangan	Data Training	Data Testing	Jumlah
	Dengan Smote		
Proporsi	80%	20%	100%
Jumlah	24552	6139	30691

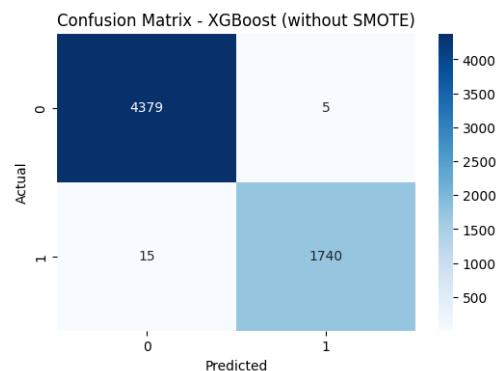
H. Modelling

Penelitian ini menggunakan dua algoritma pembelajaran mesin, yaitu Extreme Gradient Boosting (XGBoost) dan Random Forest, sebagai model klasifikasi utama. Kedua model dijalankan menggunakan parameter bawaan (default) dari pustaka Scikit-learn, tanpa dilakukan optimasi hiperparameter. Pendekatan ini dipilih untuk memberikan gambaran awal kinerja masing-masing algoritma secara adil dan konsisten pada kondisi standar.

Pada tahap pemodelan, kinerja kedua model akan dievaluasi menggunakan cross-validation, sehingga hasil yang diperoleh lebih reliabel dan tidak hanya bergantung pada satu pembagian data.

I. Evaluasi

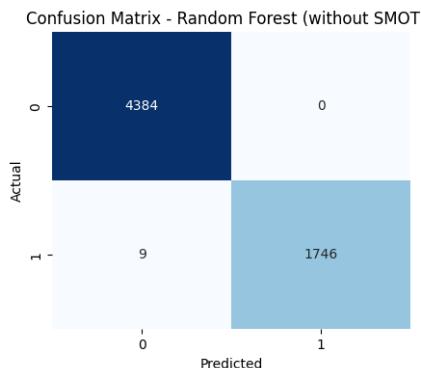
Selama tahap evaluasi, akurasi prediksi model dibandingkan dengan data aktual. Untuk mengevaluasi kinerja model, pendekatan ini membutuhkan matriks yang disebut confusion matrix. Dalam evaluasi yang dilakukan tanpa menggunakan smote pada data train maupun data test menghasilkan confusion matrix sebagai berikut.



Gambar 13. Extreme Gradient Boost Confusion Matrix without smote.

Gambar 13 menampilkan confusion matrix untuk model XGBoost tanpa penerapan SMOTE. Model ini menunjukkan hasil yang cukup baik dengan 5 kesalahan prediksi pada kelas 0 dan 15 kesalahan pada kelas 1. Sebagian besar data berhasil diklasifikasikan dengan benar (4.379 dan 1.740 sampel), yang menandakan bahwa XGBoost tetap mampu memberikan

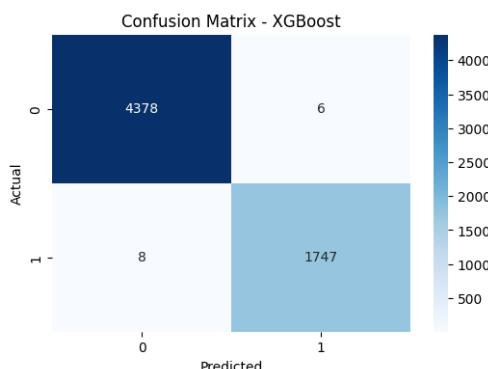
kinerja tinggi meskipun data belum seimbang. Namun, perbedaan jumlah kesalahan antar kelas menunjukkan bahwa tanpa SMOTE, model masih sedikit bias terhadap kelas mayoritas. Selanjutnya pada model random forest tanpa smote.



Gambar 14. Random Forest Confusion Matrix without Smote.

Gambar 14 menampilkan confusion matrix untuk model Random Forest tanpa penerapan SMOTE. Hasilnya menunjukkan performa yang sangat baik, dengan tidak ada kesalahan prediksi pada kelas No Disease dan hanya 9 kesalahan pada kelas Liver Disease. Hal ini mengindikasikan bahwa Random Forest mampu mempelajari pola data dengan sangat baik meskipun tanpa proses penyeimbangan kelas, serta memiliki akurasi dan ketepatan tinggi dalam membedakan pasien sehat dan yang berisiko penyakit liver.

Tetapi dalam Penerapan SMOTE (Synthetic Minority Over-sampling Technique) sangat penting dalam penelitian ini karena dataset awal menunjukkan ketidakseimbangan kelas yang signifikan antara pasien sehat dan pasien dengan penyakit liver. Ketidakseimbangan ini dapat menyebabkan model cenderung bias terhadap kelas mayoritas, sehingga kurang mampu mengenali kasus penyakit yang sebenarnya (kelas minoritas). Dengan menerapkan SMOTE, dilakukan peningkatan jumlah sampel sintetis pada kelas minoritas berdasarkan kedekatan data, tanpa mengubah distribusi aslinya.

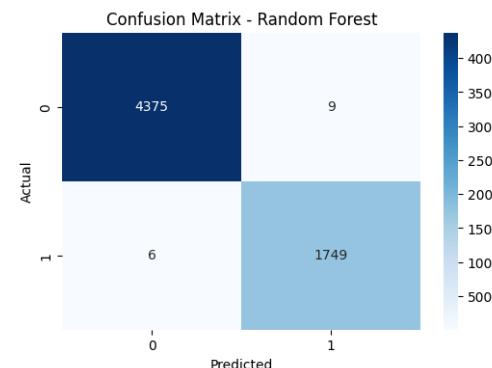


Gambar 15. Extreme Gradient Boost Confusion Matrix

Langkah ini membantu model untuk belajar pola secara lebih seimbang, meningkatkan sensitivitas (recall) terhadap kasus penyakit, serta menghasilkan prediksi yang lebih adil

dan akurat, khususnya dalam konteks deteksi penyakit liver yang membutuhkan identifikasi dini terhadap kasus positif.

Gambar 15 menunjukkan confusion matrix untuk model XGBoost. Terlihat bahwa model ini hanya membuat 6 kesalahan prediksi pada kelas 0 dan 8 kesalahan pada kelas 1, sementara sebagian besar data berhasil diklasifikasikan dengan benar (4.378 dan 1.747 sampel). Hal ini menunjukkan bahwa model XGBoost memiliki akurasi dan konsistensi yang sangat tinggi, serta mampu membedakan kedua kelas dengan baik.



Gambar 16. Random Forest Confusion Matrix

Gambar 16 menampilkan confusion matrix untuk model Random Forest. Model ini menghasilkan 9 kesalahan prediksi pada kelas No Disease dan 6 kesalahan pada kelas Liver Disease, dengan sebagian besar data berhasil diklasifikasikan dengan benar (4.375 dan 1.749 sampel). Hasil ini menunjukkan bahwa Random Forest memiliki performa yang sangat baik dan seimbang, dengan tingkat kesalahan yang rendah pada kedua kelas, sehingga mampu mendeteksi kondisi penyakit liver dengan tingkat akurasi tinggi.

Dengan jumlah kesalahan yang lebih sedikit pada kedua kelas, model Random Forest tampak memiliki kinerja yang lebih unggul dalam mengklasifikasikan data pada dataset ini. Untuk memastikan bahwa perbedaan performa antara kedua model bukan sekadar kebetulan akibat pembagian data, dilakukan pengujian cross-validation sebanyak 30 kali lipatan (30-fold cross-validation) pada masing-masing model. Pengujian ini mengevaluasi metrik akurasi, presisi, recall, dan F1-score.

TABEL 8
PERFORMA ALGORITMA XGBOOST

Percobaan ke	Accuracy	Precision	Recall	F1-Score
1	99.82%	99.82%	99.82%	99.82%
2	99.31%	99.48%	99.14%	99.31%
3	100%	100%	100%	100%
4	99.57%	99.82%	99.31%	99.57%
5	100%	100%	100%	100%
...
26	99.82%	99.82%	99.82%	99.82%
27	99.91%	99.82%	100%	99.91%
28	99.65%	99.65%	99.65%	99.65%
29	99.82%	99.82%	99.82%	99.82%
30	99.74%	99.48%	100%	99.74%
Rata-rata	99.74%	99.76%	99.72%	99.74%

Tabel 8 menunjukkan hasil pengujian algoritma XGBoost selama 30 kali percobaan menggunakan metode cross-validation. Nilai akurasi, precision, recall, dan F1-score yang diperoleh sangat tinggi, dengan sebagian besar percobaan menghasilkan performa di atas 99%. Rata-rata performa XGBoost mencapai accuracy 99,74%, precision 99,76%, recall 99,72%, dan F1-score 99,74%, menunjukkan kemampuan model dalam mengklasifikasikan data secara konsisten dan akurat. Hasil ini memperlihatkan bahwa XGBoost mampu menjaga keseimbangan antara tingkat kesalahan yang rendah dan prediksi yang tepat pada setiap kategori.

TABEL 9
PERFORMA ALGORITMA RANDOM FOREST

Percobaan ke	Accuracy	Precision	Recall	F1-Score
1	99.91%	100%	99.82%	99.91%
2	99.82%	99.82%	99.82%	99.82%
3	100%	100%	100%	100%
4	99.74%	100%	99.48%	99.74%
5	99.91%	100%	99.82%	99.91%
...
26	99.82%	99.82%	99.82%	99.82%
27	99.91%	100%	99.82%	99.91%
28	99.82%	99.82%	99.82%	99.82%
29	99.91%	100%	99.82%	99.91%
30	99.65%	99.65%	99.65%	99.65%
Rata-rata	99.82%	99.89%	99.75%	99.82%

Tabel 9 memperlihatkan performa algoritma Random Forest pada 30 kali percobaan cross-validation. Hasil evaluasi menunjukkan nilai akurasi, precision, recall, dan F1-score yang hampir selalu berada di kisaran 99% atau lebih. Secara rata-rata, Random Forest mencatat accuracy 99,82%, precision 99,89%, recall 99,75%, dan F1-score 99,82%, yang menggambarkan kemampuan model ini dalam membuat prediksi yang sangat presisi dengan tingkat kesalahan yang minimal. Performa yang konsisten ini menegaskan keandalan Random Forest dalam mengolah data dan mengklasifikasikan dengan tingkat akurasi yang sangat tinggi.

Berdasarkan perbandingan kedua tabel, baik XGBoost maupun Random Forest menunjukkan performa yang sangat tinggi dan perbedaan nilai yang relatif kecil di semua metrik evaluasi. XGBoost unggul tipis pada konsistensi nilai recall, sementara Random Forest sedikit lebih unggul pada nilai akurasi dan precision rata-rata. Mengingat perbedaan yang sangat kecil ini, kedua algoritma sama-sama layak digunakan untuk klasifikasi pada dataset ini, meskipun Random Forest dapat dianggap sedikit lebih unggul dalam memberikan prediksi yang lebih presisi dan akurat secara keseluruhan.

IV. KESIMPULAN

Model Random Forest dan XGBoost menunjukkan kinerja yang sangat baik dalam klasifikasi data, sebagaimana terlihat dari hasil evaluasi menggunakan validasi silang. Model Random Forest memperoleh akurasi rata-rata sebesar 99,82%,

presisi 99,89%, recall 99,75%, dan F1-score 99,82%. Sementara itu, model XGBoost mencatat akurasi rata-rata 99,74%, presisi 99,77%, recall 99,72%, dan F1-score 99,74%. Random Forest sedikit unggul pada metrik presisi dan F1-score, namun perbedaan kinerjanya dengan XGBoost tergolong sangat kecil.

Secara keseluruhan, hasil ini menunjukkan bahwa kedua model mampu melakukan prediksi dengan tingkat kesalahan yang sangat rendah. Random Forest cenderung lebih stabil pada berbagai metrik evaluasi, sedangkan XGBoost tetap menjadi alternatif yang sangat kompetitif berkat efisiensi waktu pelatihan dan prediksi yang cepat. Kedua model ini sama-sama cocok digunakan pada aplikasi yang memerlukan akurasi tinggi dan kinerja optimal.

DAFTAR PUSTAKA

- [1] H. Devarbhavi, S. K. Asrani, J. P. Arab, Y. A. Nartey, E. Pose, and P. S. Kamath, “Global burden of liver disease: 2023 update,” *J Hepatol*, vol. 79, no. 2, pp. 516–537, Aug. 2023, doi: 10.1016/j.jhep.2023.03.017.
- [2] M. I. Pradipta, Z. Situmorang, and R. W. Sembiring, “Multilayer Perceptron Performance Analysis in Liver Disease Classification,” *Sinkron*, vol. 9, no. 1, pp. 426–434, Jan. 2024, doi: 10.33395/sinkron.v9i1.13202.
- [3] A. S. Afrah, “Sistem Diagnosa Penyakit Liver Menggunakan Metode Artificial Neural Network: Studi Berdasarkan Dataset Indian Liver Patient Dataset,” *Jurnal Informatika Jurnal Pengembangan IT*, vol. 8, pp. 308–312, May 2023, doi: 10.30591/jpit.v8i3.5346.
- [4] R. Amin, R. Yasmin, S. Ruhi, M. H. Rahman, and M. S. Reza, “Prediction of chronic liver disease patients using integrated projection based statistical feature extraction with machine learning algorithms,” *Inform Med Unlocked*, vol. 36, p. 101155, Jan. 2023, doi: 10.1016/J.IIMU.2022.101155.
- [5] S. Noor, S. A. AlQahtani, and S. Khan, “XGBoost-Liver: An Intelligent Integrated Features Approach for Classifying Liver Diseases Using Ensemble XGBoost Training Model,” *Computers, Materials & Continua*, vol. 83, no. 1, pp. 1435–1450, 2025, doi: 10.32604/cmc.2025.061700.
- [6] R. Kashyap and B. Kaur, “Liver Disease Prediction using Machine Learning Algorithms,” *Int J Comput Appl*, vol. 185, no. 27, pp. 36–44, Aug. 2023, doi: 10.5120/ijca2023923022.
- [7] J. Lu, “Research on Prediction of Liver Disease Based on Machine Learning Models,” *Highlights in Science, Engineering and Technology*, vol. 68, pp. 21–28, Oct. 2023, doi: 10.54097/hset.v68i.11926.
- [8] M. A. Nugraha, M. I. Mazdadi, A. Farmadi, Muliadi, and T. H. Saragih, “Penyeimbangan Kelas SMOTE dan Seleksi Fitur Ensemble Filter pada Support Vector Machine untuk Klasifikasi Penyakit Liver,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 6, pp. 1273–1284, Dec. 2023, doi: 10.25126/jtiik.2023107234.
- [9] Abhishek Shrivastava, “Liver Disease Patient Dataset 30K train data,” 2021.
- [10] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, “Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19),” *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, p. 406, Apr. 2021, doi: 10.30865/mib.v5i2.2835.
- [11] M. R. A. Prasetya, A. M. Priyatno, and Nurhaeni, “Penanganan Imputasi Missing Values pada Data Time Series dengan Menggunakan Metode Data Mining,” *Jurnal Informatika dan Teknologi*, pp. 52–62, Jun. 2023, doi: 10.37034/jidt.v5i2.324.

- [12] K. Doctor, T. Mao, and H. Mhaskar, "Encoding of data sets and algorithms," *Applied Numerical Mathematics*, vol. 200, pp. 209–235, Jun. 2024, doi: 10.1016/j.apnum.2023.07.013.
- [13] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.
- [14] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, Jun. 2022, doi: 10.1016/J.GLTP.2022.04.020.
- [15] I. Muraina, "Ideal Dataset Splitting Ratios In Machine Learning Algorithms: General Concerns For Data Scientists and Data Analysis," Jun. 2022.
- [16] T. Inoue *et al.*, "XGBoost, a Machine Learning Method, Predicts Neurological Recovery in Patients with Cervical Spinal Cord Injury," *Neurotrauma Rep*, vol. 1, no. 1, pp. 8–16, Jan. 2020, doi: 10.1089/neur.2020.0009.
- [17] A. Y. Mahmoud, "Novel efficient feature selection: Classification of medical and immunotherapy treatments utilising Random Forest and Decision Trees," *Intell Based Med*, vol. 10, p. 100151, Jan. 2024, doi: 10.1016/J.IBMED.2024.100151.
- [18] S. Sathyaranayanan, "Confusion Matrix-Based Performance Evaluation Metrics," *African Journal of Biomedical Research*, pp. 4023–4031, Nov. 2024, doi: 10.53555/AJBR.v27i4S.4345.
- [19] E. Helmut, E. Helmut, F. Fitriyani, and P. Romadiana, "Classification Comparison Performance of Supervised Machine Learning Random Forest and Decision Tree Algorithms Using Confusion Matrix," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 13, no. 1, pp. 92–97, Feb. 2024, doi: 10.32736/sisfokom.v13i1.1985.