Analisis Kinerja XGBoost pada Data Seimbang dan tidak Seimbang dalam Klasifikasi Tingkat Keparahan Pasien Hemodialisis

Mochammad Thoriq Wafa[1], Eva Yulia Puspaningrum[2], Ani Dijah Rahajoe[3]
[1,2,3]Universitas Pembangunan Nasional "Veteran" Jawa Timur, Indonesia
E-mail: 21081010309@student.upnjatim.ac.id, evapuspaningrum.if@upnjatim.ac.id, *anidijah.if@upnjatim.ac.id*

Article Info
Article History
Received: 2025-10-07
Revised: 2025-11-13
Published: 2025-12-02

Keywords:
XGBoost;
SMOTE;
Data Imbalance;
Classification;
Hemodialysis.

Abstract
This study evaluates the impact of class imbalance on multiclass classification of hemodialysis severity (mild, moderate, severe) using Extreme Gradient Boosting (XGBoost) and the effectiveness of the Synthetic Minority Over-sampling Technique (SMOTE) in improving model performance. The dataset comprises 5,000 patient records with 23 clinical and laboratory features; preprocessing included median/mode imputation, one-hot encoding, and standardization. As per the study design, SMOTE was applied prior to data partitioning, followed by stratified 60:40 and 70:30 train:test splits. Performance was assessed using accuracy, precision, recall, and F1-score (reported as macro and weighted), alongside confusion-matrix analysis. Results indicate consistent improvements after SMOTE, with accuracy increasing from 0.7055 to 0.8566 (60:40) and from 0.7000 to 0.8705 (70:30), accompanied by better recognition of minority classes. These findings underscore the importance of data balancing for fair and reliable classification to support clinical decision-making, while acknowledging the limitation that applying SMOTE before splitting may yield overly optimistic performance estimates.

Artikel Info
Sejarah Artikel
Diterima: 2025-10-07
Direvisi: 2025-11-13
Dipublikasi: 2025-12-02

Kata kunci:
XGBoost;
SMOTE;
Ketidakseimbangan kelas;
Klasifikasi;
Hemodialisis.

Abstrak
Penelitian ini mengevaluasi pengaruh ketidakseimbangan kelas pada klasifikasi tingkat keparahan hemodialisis (mild, moderate, severe) menggunakan Extreme Gradient Boosting (XGBoost) serta efektivitas Synthetic Minority Oversampling Technique (SMOTE) dalam meningkatkan kinerja model. Dataset mencakup 5.000 rekam pasien dengan 23 fitur klinis–laboratorium; pra-pemrosesan meliputi imputasi median/modus, one-hot encoding, dan standardisasi. Sesuai rancangan penelitian, SMOTE diterapkan sebelum pembagian data, kemudian dilakukan stratified split 60:40 dan 70:30 (latih:uji). Evaluasi menggunakan akurasi, presisi, recall, dan F1-score (macro dan weighted), disertai analisis confusion matrix. Hasil menunjukkan perbaikan konsisten setelah SMOTE, dengan akurasi meningkat dari 0,7055 → 0,8566 (60:40) dan 0,7000 → 0,8705 (70:30), diikuti peningkatan pengenalan kelas minoritas. Temuan ini menegaskan bahwa penyeimbangan data penting untuk klasifikasi yang lebih adil dan andal dalam mendukung keputusan klinis, dengan catatan keterbatasan bahwa penerapan SMOTE sebelum split berpotensi memberikan estimasi kinerja yang lebih optimistis.

I. PENDAHULUAN

Penyakit ginjal kronik (PGK) merupakan salah satu masalah kesehatan global yang prevalensi-nya terus meningkat setiap tahunnya (Kim et al., 2022). Pada stadium lanjut, hemodialisis menjadi terapi utama melalui proses penyaringan darah buatan yang dilakukan secara rutin untuk menggantikan fungsi ginjal. Dalam praktik klinis, pasien hemodialisis menunjukkan tingkat keparahan yang bervariasi, mulai dari ringan (*mild*), sedang (*moderate*), hingga berat (*severe*). Penentuan tingkat keparahan ini krusial untuk mendukung pengambilan keputusan medis, menetapkan prioritas perawatan, serta memperkirakan prognosis pasien (Zhang et al., 2023).

Seiring meningkatnya jumlah pasien dan kompleksitas data medis, *machine learning* (ML) menawarkan pendekatan kuantitatif yang potensial untuk membantu klasifikasi tingkat keparahan pasien hemodialisis (Hassanzadeh et al., 2023). Di antara berbagai algoritma yang tersedia, Extreme Gradient Boosting (XGBoost), sebuah metode *ensemble* berbasis *gradient boosting*, dikenal memiliki performa yang kuat dan stabil untuk data tabular yang kompleks (As'an Hamid and Subhiyakto, 2025). Sejumlah kajian juga melaporkan efektivitas XGBoost pada beragam permasalahan medis, seperti klasifikasi diabetes dan penyakit jantung, sehingga relevan dipertimbangkan pada konteks hemodialisis

(Ratantja Kusumajati et al., 2024; Ubai Dullah et al., 2025).

Salah satu tantangan utama ML pada data medis adalah ketidakseimbangan kelas, ketika jumlah sampel kelas mayoritas jauh melampaui kelas minoritas (Syukron et al., 2020). Kondisi ini dapat mendorong model menjadi bias terhadap kelas mayoritas dan menghasilkan indikator kinerja yang tidak mencerminkan performa secara menyeluruh di seluruh kelas target (Fadhlullah and Wahyono, 2024). Untuk mengatasinya, berbagai pendekatan telah dikembangkan, mulai dari penyesuaian bobot kelas hingga teknik *resampling*. Di antara teknik *oversampling*, Synthetic Minority Oversampling Technique (SMOTE) banyak digunakan karena membangkitkan sampel sintetis di sekitar tetangga terdekat kelas minoritas, sehingga distribusi kelas menjadi lebih seimbang dan pengenalan kelas minoritas dapat meningkat (Pramesti et al., 2025). Temuan-temuan sebelumnya menunjukkan bahwa SMOTE berpotensi memperbaiki *recall* kelas minoritas dan meningkatkan skor F1 secara keseluruhan (Wulandari and Badieah, 2025; Sarkar et al., 2023).

Meskipun penelitian mengenai penerapan ML untuk penyakit ginjal dan berbagai kasus klinis telah berkembang, sebagian besar studi masih berfokus pada pemilihan algoritma terbaik tanpa mengevaluasi secara sistematis pengaruh penyeimbangan data terhadap kinerja model pada tugas klasifikasi tingkat keparahan hemodialisis. Selain itu, masih terbatas kajian yang secara eksplisit membandingkan kinerja XGBoost pada variasi rasio pembagian data latih dan uji, misalnya 60:40 dan 70:30, dalam konteks data tidak seimbang versus seimbang (Pramesti et al., 2025). Celah ini penting karena pilihan strategi penyeimbangan dan proporsi data latih dapat memengaruhi stabilitas metrik dan keadilan performa antarkelas.

Berdasarkan celah tersebut, penelitian ini menganalisis kinerja XGBoost pada dua kondisi data, yaitu tidak seimbang (*imbalanced data*) dan seimbang hasil penerapan SMOTE (*balanced data*), serta dua skenario pembagian data terstratifikasi, yaitu 60:40 dan 70:30 untuk data latih dan data uji. Secara singkat, dataset yang dikaji berskala menengah, sekitar 5.000 rekam pasien, dengan 23 fitur klinis dan laboratorium. Detail variabel, tahapan pra-pemrosesan, dan prosedur penyeimbangan dipaparkan pada bagian Metode. Evaluasi kinerja meliputi akurasi, presisi, *recall*, dan F1-score, dengan pelaporan *macro* dan *weighted*, serta interpretasi *confusion matrix* untuk membaca pola salah klasifikasi dan dampaknya pada konteks klinis.

Kontribusi utama penelitian ini adalah menunjukkan dampak praktis penyeimbangan data berbasis SMOTE terhadap kinerja XGBoost pada tugas klasifikasi tingkat keparahan hemodialisis, membandingkan dua rasio pembagian data latih dan uji untuk menilai pengaruh ukuran data latih terhadap stabilitas metrik, serta menyajikan kerangka kerja yang transparan dan dapat direplikasi yang meliputi pra-pemrosesan, penyeimbangan, pelatihan, dan evaluasi, dengan penekanan pada keadilan antarkelas. Bagian berikutnya menguraikan dataset, pra-pemrosesan, strategi penyeimbangan kelas, serta rancangan evaluasi yang digunakan.

II. METODE PENELITIAN

1. Desain Penelitian

   Penelitian ini menggunakan pendekatan kuantitatif dengan eksperimen komparatif berbasis machine learning untuk menganalisis kinerja Extreme Gradient Boosting (XGBoost) dalam mengklasifikasikan tingkat keparahan pasien hemodialisis pada dua kondisi data: data tidak seimbang (imbalanced) dan data seimbang hasil Synthetic Minority Oversampling Technique (SMOTE). Tahapan dilakukan berurutan, meliputi pra-pemrosesan, penyeimbangan data, pembagian data menjadi data latih dan data uji, pelatihan model XGBoost, serta evaluasi performa menggunakan metrik klasifikasi. Desain ini bertujuan menilai dampak keseimbangan data terhadap akurasi, presisi, serta kemampuan model mengenali kelas minoritas pada klasifikasi tingkat keparahan.

2. Dataset Penelitian

   Dataset bersumber dari platform Kaggle (GSK RD Solutions, 2023). Data berisi rekam pasien hemodialisis yang mencakup parameter klinis dan laboratorium yang lazim digunakan, misalnya kreatinin, hemoglobin, urea, tekanan darah, dan variabel relevan lain. Variabel target adalah Disease Severity dengan tiga kelas, yaitu *Mild*, *Moderate*, dan *Severe*. Secara ringkas, dataset berukuran 5.000 rekam dan, setelah pengodean kategorikal, menghasilkan 23 fitur. Distribusi awal menunjukkan ketidakseimbangan kelas, dengan Severe paling dominan dan Mild paling sedikit; oleh karena itu, penelitian membandingkan kondisi data asli yang tidak seimbang dengan data hasil penyeimbangan

SMOTE untuk menilai pengaruhnya terhadap kinerja XGBoost.

3. Pra-pemrosesan

Pra-pemrosesan dilakukan untuk memastikan kualitas dan konsistensi data sebelum pelatihan model. Nilai hilang pada fitur numerik diimputasi menggunakan median karena lebih *robust* terhadap pencilan, sedangkan pada fitur kategorikal diimputasi menggunakan modus. Fitur kategorikal kemudian direpresentasikan ke bentuk numerik melalui one-hot encoding agar sesuai dengan kebutuhan model berbasis pohon. Seluruh fitur selanjutnya dinormalisasi menggunakan StandardScaler untuk menyeragamkan skala fitur dan menstabilkan proses pembelajaran.

4. Penyeimbangan Data Menggunakan SMOTE

Setelah pra-pemrosesan, data diseimbangkan menggunakan SMOTE, yaitu teknik oversampling yang membangkitkan sampel sintetis pada kelas minoritas berdasarkan kedekatan tetangga dalam ruang fitur. Sesuai rancangan penelitian, SMOTE diterapkan sebelum pembagian data guna menghasilkan working dataset yang seimbang, sehingga proporsi kelas pada subset yang dihasilkan menjadi lebih seragam. Parameter yang digunakan mengikuti praktik umum, yaitu k_neighbors = 5 dan sampling_strategy = 'auto' sehingga setiap kelas diseimbangkan terhadap kelas dominan global. Pendekatan ini ditujukan untuk mengurangi bias kelas mayoritas dan meningkatkan pengenalan kelas minoritas; konsekuensi metodologis dan mitigasinya dibahas pada bagian pembahasan.

5. Pembagian Data

Setelah penyeimbangan, dataset dibagi menjadi data latih dan data uji dengan dua skenario rasio 60:40 dan 70:30, menggunakan stratified sampling agar proporsi kelas tetap terjaga pada kedua subset. Dua rasio dipilih untuk menilai sensitivitas kinerja terhadap ukuran data latih. Data latih digunakan untuk membangun model XGBoost, sedangkan data uji digunakan untuk mengukur kemampuan generalisasi model.

6. Model Klasifikasi Menggunakan XGBoost

Algoritma yang digunakan adalah XGBoost, metode ensemble berbasis gradient boosting yang menyusun sejumlah pohon keputusan secara bertahap untuk meminimalkan kesalahan (boosting). XGBoost dikenal efektif untuk data tabular yang kompleks dan efisien secara komputasi. Hiperparameter disetel secara manual dan dijaga konsisten lintas skenario, meliputi learning_rate 0,1, max_depth 6, n_estimators 200, subsample 0,9, dan colsample_bytree 0,9. Model dilatih pada data latih dan dievaluasi pada data uji yang tidak digunakan selama pelatihan.

7. Evaluasi Kinerja Model

Evaluasi kinerja menggunakan akurasi, presisi, sensitivitas (recall), dan F1-score. Untuk klasifikasi multikelas, metrik dilaporkan dalam dua skema perata-rataan, yaitu macro average (rata per kelas, menekankan keadilan antarkelas) dan weighted average (mempertimbangkan proporsi kelas). Selain itu, confusion matrix digunakan untuk menelaah pola salah klasifikasi per kelas dan implikasi klinisnya. Perbandingan dilakukan antara kondisi tidak seimbang dan seimbang pada kedua skenario split guna menilai dampak SMOTE terhadap kinerja XGBoost.

III. HASIL DAN PEMBAHASAN

A. Hasil Penelitian

Bagian ini menyajikan perbandingan hasil evaluasi XGBoost pada dua kondisi data, yaitu tidak seimbang dan seimbang dengan SMOTE, serta dua skenario split 60:40 dan 70:30. Fokus analisis diarahkan pada perubahan metrik multikelas (akurasi, presisi, recall, F1 tertimbang) dan pola kesalahan pada confusion matrix, sehingga pembacaan hasil tidak sekadar mendeskripsikan proses, tetapi juga menekankan arti praktisnya bagi klasifikasi klinis.

1. Hasil Pra-pemrosesan Data

Setelah imputasi, one-hot encoding menghasilkan total 23 fitur; seluruh fitur kemudian dinormalisasi dengan StandardScaler. Distribusi target sebelum penyeimbangan sangat timpang (Severe dominan, Mild paling kecil). Dengan SMOTE, tiap kelas menjadi 3.671 sampel (total 11.013), sehingga dataset kerja seimbang dan siap dievaluasi pada dua kondisi (imbalanced vs balanced).

2. Hasil Penyeimbangan Data Menggunakan SMOTE

Distribusi awal kelas target menunjukkan ketimpangan yang cukup

besar antara kelas Severe, Moderate, dan Mild, sehingga dilakukan proses Synthetic Minority Oversampling Technique (SMOTE) untuk menyeimbangkan jumlah sampel, dan perbandingan jumlah sampel sebelum dan sesudah proses tersebut dapat dilihat pada Tabel 1.

Tabel 1. Distribusi Sebelum dan Sesudah SMOTE

Kelas	Sebelum SMOTE	Sesudah SMOTE
Severe	3.671	3.671
Moderate	1.184	3.671
Mild	145	3.671
Total	5.000	11.013

SMOTE meningkatkan kepadatan di sekitar sampel minoritas sehingga membantu *recall* dan menurunkan ketimpangan antarkelas. Namun, ada risiko overfitting lokal, pergeseran distribusi fitur dan overlap antarkelas (*covariate shift*), serta *trade-off* presisi–*recall*. Pada studi ini kami memakai setelan konservatif (*k_neighbors* = 5, *sampling_strategy* = 'auto') dan memantau risiko melalui uji distribusi sebelum–sesudah (KS/proporsi kategori), proyeksi PCA/UMAP, serta kurva *precision–recall* per kelas; karena SMOTE diterapkan sebelum pembagian data, kami mencatat kemungkinan estimasi yang sedikit lebih optimistis.

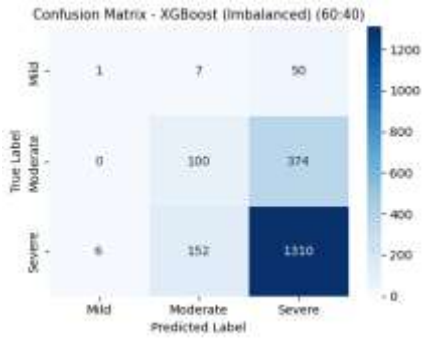
3. Hasil Klasifikasi XGBoost pada Data Tidak Seimbang dan Seimbang (Split 60:40)

Pada kondisi data asli yang tidak seimbang, prediksi cenderung berpihak pada kelas Severe. Setelah penyeimbangan dengan SMOTE, kinerja meningkat pada seluruh metrik; rincian nilai disajikan pada Tabel 2. Analisis selanjutnya menyoroti pola salah-klasifikasi untuk menilai perubahan pemerataan pengenalan antarkelas.

Tabel 2. Hasil Perbandingan Kinerja XGBoost (Split 60:40)

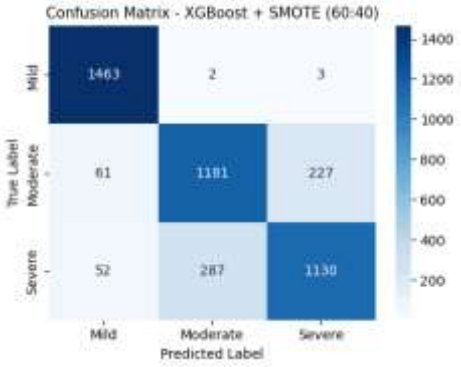
Kondisi Data	Akurasi	Precision	Recall	F1-Score
Tidak Seimbang	0.7055	0.65	0.71	0.67
Seimbang	0.8566	0.85	0.86	0.85

Perubahan distribusi kesalahan divisualisasikan pada Gambar 1–2: setelah SMOTE, diagonal *confusion matrix* menguat dan kesalahan pada Mild serta Moderate berkurang nyata.



Gambar 1. Confusion Matrix XGBoost pada Data Tidak Seimbang (Split 60:40)

Sebelum penyeimbangan, kesalahan didominasi oleh pergeseran prediksi ke Severe: lebih dari 85% sampel Mild diprediksi sebagai Severe (1,7% benar sebagai Mild; 86,2% salah ke Severe), dan sekitar 79% Moderate juga bergeser ke Severe. Dampaknya, recall Mild dan Moderate sangat rendah, sedangkan Severe tampak tinggi karena dominasi jumlah kasus.



Gambar 2. Confusion Matrix XGBoost pada Data Seimbang (Split 60:40)

Setelah penyeimbangan, diagonal *confusion matrix* menguat. *Recall* Mild nyaris sempurna (≈ 99,7% benar sebagai Mild), Moderate meningkat nyata (≈ 80,4% benar; sisa ≈ 15,5% masih salah ke Severe), sementara Severe menjadi lebih seimbang (≈ 76,9% benar; sebagian bergeser ke Moderate). Pola ini menunjukkan bahwa pemerataan data mengurangi bias ke kelas mayoritas dan memperbaiki pengenalan kelas yang semula minoritas, dengan sedikit penurunan *recall* pada Severe yang diimbangi peningkatan keadilan antar kelas.

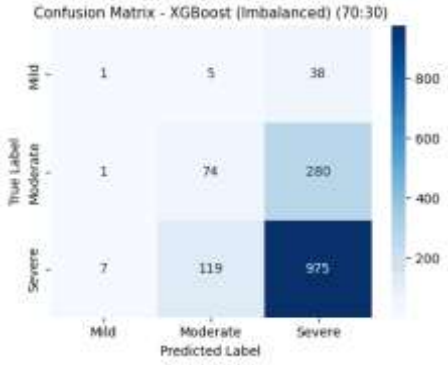
4. Hasil Klasifikasi XGBoost pada Data Tidak Seimbang dan Seimbang (Split 70:30)

Pada rasio 70:30, pola peningkatan kinerja setelah SMOTE konsisten dengan skenario 60:40; rincian nilai numerik disajikan pada Tabel 3. Secara ringkas, terjadi kenaikan menyeluruh pada metrik multikelas (akurasi, presisi, recall, dan F1 tertimbang), menandakan pemerataan pengenalan antar-kelas.

Tabel 3. Hasil Perbandingan Kinerja XGBoost (Split 70:30)

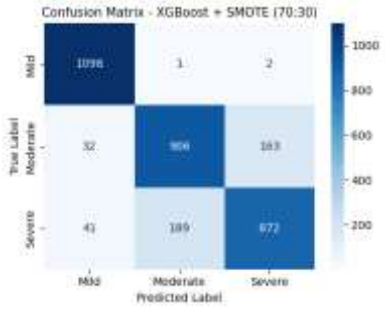
Kondisi Data	Akurasi	Precision	Recall	F1-Score
Tidak Seimbang	0.7000	0.65	0.70	0.66
Seimbang	0.8705	0.87	0.87	0.87

Kinerja model pada data seimbang kembali menunjukkan peningkatan yang konsisten pada seluruh metrik evaluasi. Peningkatan tertinggi terjadi pada kelas Mild yang sebelumnya hampir tidak terdeteksi, kini mencapai Recall 1.00 dan F1-score 0.97, menunjukkan bahwa model menjadi jauh lebih adil terhadap semua kelas target.



Gambar 3. Confusion Matrix XGBoost pada Data Tidak Seimbang (Split 70:30)

Pada kondisi tidak seimbang, prediksi terpusat pada Severe. Sampel Mild dan Moderate banyak salah diklasifikasikan ke Severe sehingga *recall* keduanya rendah, sedangkan *recall* Severe tampak tinggi karena dominasi jumlah kasus. Pola ini menegaskan bias mayoritas sebelum penyeimbangan.



Gambar 4. Confusion Matrix XGBoost pada Data Seimbang (Split 70:30)

Setelah penyeimbangan dengan SMOTE, diagonal confusion matrix menguat. Recall Mild mendekati 100% (≈ 99,7%) dan Moderate meningkat (≈ 82% benar), sementara Severe tetap baik meski sebagian bergeser ke Moderate (≈ 79% benar). Ini menunjukkan model menjadi lebih adil antarkelas.

Skenario 70:30 menegaskan bahwa penyeimbangan dengan SMOTE efektif memperbaiki kinerja dan pemerataan pengenalan kelas, dengan visual pada Gambar 3–4 memperlihatkan penguatan diagonal dan penurunan salah klasifikasi pada kelas yang sebelumnya minoritas. Detail angka lengkap tersaji pada Tabel 3.

5. Ringkasan Kinerja XGBoost pada Seluruh Skenario

Seluruh hasil pada empat skenario (imbalanced vs balanced; 60:40 dan 70:30) dirangkum pada Gambar 5, sehingga pola peningkatan kinerja lebih mudah dibaca tanpa mengulang nilai dari tabel. Secara umum, penerapan SMOTE memberikan kenaikan konsisten pada semua metrik evaluasi di kedua rasio pembagian data.

Gambar 5. Ringkasan metrik

Secara agregat, peningkatan absolut rata-rata di seluruh metrik adalah ≈ +0,18 poin. Jika diuraikan per metrik, akurasi ≈ +0,16, presisi ≈ +0,21, recall ≈ +0,16, dan

F1-score ≈ +0,20 (rata-rata dari skenario 60:40 dan 70:30). Tren ini menegaskan bahwa penyeimbangan data bukan hanya menaikkan akurasi global, tetapi juga memeratakan pengenalan antarkelas yang sebelumnya timpang.

Kenaikan pasca-SMOTE serupa pada 60:40 dan 70:30. Dengan porsi data latih lebih besar, skenario 70:30 cenderung memberi peningkatan yang sedikit lebih menonjol pada kelas minoritas; namun karena ukuran uji lebih kecil, estimasi dapat sedikit lebih berfluktuasi. Kesimpulan empiris tidak bergantung pada satu rasio tertentu karena pola peningkatan konsisten di keduanya.

Secara keseluruhan, SMOTE terbukti meningkatkan kinerja XGBoost dan keadilan antarkelas dengan biaya waktu komputasi yang relatif kecil. Bagian Pembahasan berikut menafsirkan implikasi metodologis dan klinis dari temuan ini, termasuk potensi *trade-off* serta batasan penyeimbangan sebelum pembagian data.

B. Pembahasan

1. Analisis Hasil Pra-pemrosesan Data.

   Tahap pra-pemrosesan data bertujuan memastikan kualitas dan konsistensi dataset sebelum pelatihan model, yang meliputi imputasi nilai hilang, transformasi data kategorikal menggunakan One-hot encoding, dan normalisasi dengan StandardScaler. Setelah proses ini, seluruh atribut dipastikan bebas dari nilai kosong dan memiliki distribusi nilai yang lebih seragam. Normalisasi berperan penting untuk mencegah dominasi fitur dengan skala besar, seperti tekanan darah atau berat badan, terhadap fitur lain yang bernilai kecil, sehingga model XGBoost dapat mempelajari hubungan antarfitur tanpa bias skala. Secara keseluruhan, pra-pemrosesan ini memberikan dasar yang kuat bagi tahap pelatihan dan evaluasi model berikutnya.

2. Penyeimbangan Data Menggunakan SMOTE

   Distribusi awal kelas target menunjukkan ketimpangan yang signifikan, di mana kelas Severe mendominasi dataset dengan 3.671 sampel sementara kelas Mild hanya memiliki 145 sampel, sehingga model sulit mengenali pola pada kelas minoritas. Setelah penerapan Synthetic Minority Oversampling Technique (SMOTE), jumlah sampel pada setiap kelas menjadi seimbang, masing-masing sebanyak 3.671 data. Penyeimbangan ini terbukti meningkatkan kemampuan model dalam mengenali seluruh kelas secara proporsional. Teknik SMOTE menghasilkan sampel sintetis baru berdasarkan *nearest neighbors* dari data minoritas, bukan sekadar menduplikasi data yang ada. Hasil ini sejalan dengan temuan Syukron et al. (2020) serta Mubarok et al. (2022) yang menyatakan bahwa penggunaan SMOTE dapat mengurangi bias terhadap kelas mayoritas dan meningkatkan sensitivitas model pada data medis dengan ketidakseimbangan kelas.

3. Analisis Kinerja XGBoost pada Data Tidak Seimbang (Split 60:40).

   Pada kondisi data tidak seimbang, model XGBoost menghasilkan akurasi sebesar 0.7055 dengan nilai F1-score tertimbang 0.67. Kinerja tertinggi dicapai pada kelas Severe dengan F1-score 0.82, sedangkan kelas Mild hampir tidak terdeteksi dengan F1-score hanya 0.03. Kondisi ini menunjukkan bahwa model terlalu berfokus pada kelas mayoritas dan gagal mengenali kelas minoritas, yang merupakan karakteristik umum pada kasus *imbalanced data*. Kelemahan ini menegaskan bahwa tanpa penyeimbangan, model cenderung memiliki akurasi tinggi secara keseluruhan namun rendah dalam ketepatan klasifikasi pada kelas yang penting secara klinis, sehingga secara diagnostik masih berisiko karena dapat mengabaikan pasien dengan tingkat keparahan ringan.

4. Pengaruh SMOTE terhadap Kinerja XGBoost (Split 60:40).

   Setelah penerapan SMOTE, performa model meningkat secara signifikan dengan akurasi mencapai 0.8566 dan F1-score sebesar 0.85, di mana seluruh kelas berhasil dikenali dengan lebih seimbang. Kelas Mild yang sebelumnya hanya memiliki recall 0.02 kini meningkat menjadi 1.00, menunjukkan bahwa semua data pada kelas tersebut berhasil diidentifikasi dengan benar. Hasil ini menegaskan bahwa SMOTE efektif dalam mengurangi bias klasifikasi dan meningkatkan kemampuan generalisasi

model. Peningkatan kinerja tidak hanya terlihat pada akurasi, tetapi juga pada nilai recall dan precision, sehingga model menjadi lebih andal dalam memprediksi pasien pada setiap tingkat keparahan.

5. Analisis Kinerja XGBoost pada Data Tidak Seimbang (Split 70:30).

   Hasil pengujian pada pembagian data 70:30 menunjukkan pola yang serupa dengan skenario sebelumnya, di mana model XGBoost yang dilatih pada data tidak seimbang menghasilkan akurasi sebesar 0.7000 dengan F1-score tertimbang 0.66. Meskipun proporsi data latih lebih besar, model tetap menunjukkan bias terhadap kelas mayoritas karena distribusi data tidak berubah. Temuan ini menegaskan bahwa peningkatan jumlah data latih tidak selalu menjamin peningkatan performa model apabila distribusi kelas tetap tidak seimbang.

6. Pengaruh SMOTE terhadap Kinerja XGBoost (Split 70:30).

   Penerapan SMOTE pada skenario pembagian data 70:30 menghasilkan peningkatan performa yang konsisten, dengan akurasi mencapai 0.8705 dan F1-score sebesar 0.87. Kelas Mild kembali menunjukkan hasil terbaik dengan recall 1.00 dan F1-score 0.97, menandakan bahwa penyeimbangan data membuat model lebih adil dalam mengenali seluruh kelas target. Peningkatan ini juga menunjukkan bahwa proporsi data latih yang lebih besar memberikan kemampuan generalisasi yang lebih baik karena model mempelajari lebih banyak variasi data. Dengan demikian, kombinasi antara rasio data latih yang tinggi dan penerapan SMOTE menjadi konfigurasi paling optimal dalam penelitian ini.

7. Perbandingan dan Interpretasi Keseluruhan.

   Jika dibandingkan antar semua skenario, peningkatan kinerja model terlihat konsisten, di mana model dengan penerapan SMOTE selalu mengungguli model tanpa SMOTE pada rasio 60:40 maupun 70:30. Rata-rata peningkatan akurasi mencapai sekitar +0.15 poin dan F1-score naik sekitar +0.18 poin. Kombinasi XGBoost dengan SMOTE pada pembagian data 70:30 menghasilkan performa terbaik dengan akurasi 0.8705 dan F1-score 0.8689. Hasil ini menegaskan bahwa penyeimbangan data sebelum pelatihan merupakan tahap penting untuk meningkatkan kinerja algoritma pembelajaran mesin pada data medis yang tidak seimbang. Temuan ini sejalan dengan studi Pramesti et al. (2025) yang juga melaporkan peningkatan signifikan pada nilai recall dan AUC setelah penerapan teknik SMOTE dalam klasifikasi penyakit medis, menunjukkan bahwa pendekatan SMOTE + XGBoost tidak hanya meningkatkan akurasi tetapi juga reliabilitas model dalam mendukung pengambilan keputusan klinis.

8. Implikasi Hasil Penelitian.

   Hasil penelitian ini memberikan implikasi penting bagi penerapan machine learning di bidang kesehatan, terutama dalam sistem pendukung keputusan klinis. Model yang dibangun dengan data seimbang memungkinkan proses klasifikasi tingkat keparahan pasien dilakukan secara lebih adil dan konsisten. Temuan ini juga menegaskan bahwa tahapan pra-pemrosesan dan penyeimbangan data memiliki peran yang sama pentingnya dengan pemilihan algoritma dalam membentuk model yang andal. Oleh karena itu, kombinasi XGBoost dan SMOTE dapat direkomendasikan sebagai pendekatan unggulan untuk menangani dataset medis dengan tingkat ketidakseimbangan kelas yang tinggi, seperti pada kasus pasien hemodialisis.

IV. SIMPULAN DAN SARAN

A. Simpulan

Penelitian ini menegaskan bahwa ketidakseimbangan data berdampak besar pada kinerja sistem klasifikasi berbasis XGBoost untuk penentuan tingkat keparahan pasien hemodialisis. Pada kondisi tidak seimbang, prediksi cenderung condong ke kelas mayoritas; setelah penyeimbangan dengan SMOTE, seluruh metrik utama meningkat dan pengenalan antarkelas menjadi lebih merata pada skenario 60:40 maupun 70:30 (lihat ringkasan pada tabel dan gambar hasil). Temuan ini menunjukkan bahwa pengendalian distribusi kelas merupakan langkah kunci untuk memperoleh performa yang adil, bukan sekadar akurasi agregat yang tinggi. Secara praktis, teknik

penyeimbangan data perlu menjadi tahap baku dalam pengembangan sistem pendukung keputusan klinis berbasis pembelajaran mesin, khususnya ketika label minoritas merepresentasikan kondisi kritis, dengan pelaporan macro-F1/weighted-F1 dan pembacaan confusion matrix per kelas agar mutu klasifikasi tidak didominasi kelas mayoritas.

Konfigurasi XGBoost + SMOTE pada rasio data latih yang lebih besar (70:30) menunjukkan kinerja yang stabil tanpa biaya komputasi berlebih, sehingga layak dijadikan rujukan awal untuk studi sejenis. Dengan demikian, penelitian ini memperkuat bukti empiris bahwa pendekatan SMOTE + XGBoost efektif untuk menangani ketidakseimbangan pada data medis multikelas dan menghasilkan klasifikasi yang lebih adil di seluruh kelas target, sekaligus menyajikan pipeline yang transparan dan dapat direplikasi serta menegaskan pentingnya dimensi keadilan performa dalam evaluasi. Ke depan, disarankan validasi tambahan, misalnya resampling yang dibatasi pada data latih, cross-validation, atau pengujian eksternal lintas rumah sakit dan periode waktu, untuk memastikan generalisasi yang lebih kuat.

B. Saran

1. Saran untuk penelitian selanjutnya
   a) Uji SMOTE–ENN atau ADASYN untuk mereduksi noise dan memperbaiki batas keputusan kelas minoritas.
   b) Lakukan Bayesian Optimization dengan stratified cross-validation untuk penalaan hiperparameter yang efisien dan replikabel.
   c) Tambahkan kalibrasi probabilitas (Platt atau isotonic) dan analisis ambang per kelas untuk menyelaraskan keputusan klinis.
   d) Lakukan validasi eksternal lintas rumah sakit/periode guna menilai generalisasi dan ketahanan terhadap *dataset shift*.
   e) Laporkan indikator keadilan multikelas (macro-F1, per-class recall, balanced accuracy) secara konsisten.
2. Saran untuk implementasi dan penerapan
   a) Integrasikan ke sistem pendukung keputusan klinis dengan antarmuka sederhana dan penjelasan fitur utama.
   b) Terapkan pemantauan pascadeploy: *data-drift*, *performance-drift*, dan audit *precision–recall* per kelas.
   c) Siapkan tata kelola: SOP validasi periodik, dokumentasi versi model, dan mekanisme *rollback*.
   d) Lakukan pelatihan pengguna serta tetapkan ambang tindak lanjut jelas untuk kasus berisiko tinggi.
   e) Pastikan privasi dan keamanan data: anonymisasi, kontrol akses, dan enkripsi.

DAFTAR RUJUKAN

As'an Hamid, M. and Subhiyakto, R. (2025) 'Performance comparison of Random Forest, SVM, and XGBoost algorithms with SMOTE for stunting prediction', Journal of Applied Informatics and Computing, volume(issue), Journal of Applied Informatics and Computing (JAIC). http://jurnal.polibatam.ac.id/index.php/JAIC

Fadhlullah, M.K.T.Q., Wahyono, 2024. Classification of Tuberculosis Based on Chest X-Ray Images for Imbalance Data using SMOTE. International Journal of Computing and Digital Systems 16, 981–993. https://doi.org/10.12785/ijcds/160171

GSK RD Solutions. (2023) Hemodialysis Realtime Hospital Dataset [Data set]. Kaggle. Available at: https://www.kaggle.com/datasets/gskrdsolutions/hemodialysis-realtime-hospital-dataset (Accessed 9 November 2025).

Hassanzadeh, R., Farhadian, M., Rafieemehr, H., 2023. Hospital mortality prediction in traumatic injuries patients: comparing different SMOTE-based machine learning algorithms. BMC Med Res Methodol 23. https://doi.org/10.1186/s12874-023-01920-w

Kim, J., Mun, S., Lee, S., Jeong, K., Baek, Y., 2022. Prediction of metabolic and pre-metabolic syndromes using machine learning models with anthropometric, lifestyle, and biochemical factors from a middle-aged population in Korea. BMC Public Health 22. https://doi.org/10.1186/s12889-022-13131-x

Mubarok, M.R., Muliadi, M. and Herteno, R. (2022) 'Hyper-parameter tuning pada XGBoost untuk prediksi keberlangsungan hidup pasien gagal jantung', KLIK –

Kumpulan Jurnal Ilmu Komputer, 9(2). https://doi.org/10.20527/klik.v9i2.484

Pramesti, B.T., Anggraeny, F.T., Syaifullah, W. and Saputra, J. (2025) 'Klasifikasi data tidak seimbang menggunakan SMOTE-ENN dengan optimasi Bayesian' http://Jiip.stkipyapisdompu.ac.id

Ratantja Kusumajati, F., Rahmat, B. and Junaidi, A. (2024) 'Implementation of SMOTETomek in diabetes classification using XGBoost' https://kursorjournal.org/index.php/kursor/article/view/410

Sarkar, S., Zhou, Jing, Scaboo, A., Zhou, Jianfeng, Aloysius, N., Lim, T.T., 2023. Assessment of Soybean Lodging Using UAV Imagery and Machine Learning. Plants 12. https://doi.org/10.3390/plants12162893

Syukron, M., Santoso, R. and Widiharih, T. (2020) 'Perbandingan metode SMOTE-Random Forest dan SMOTE-XGBoost untuk klasifikasi tingkat penyakit hepatitis C pada data tidak seimbang' https://ejournal3.undip.ac.id/index.php/gaussian/

Dullah, A.U., Darmawan, A.Y., Pertiwi, D.A.A. and Unjung, J. (2025) 'Extreme Gradient Boosting Model with SMOTE for Heart Disease Classification', Jurnal Informatika Sunan Kalijaga, 10(1), pp. 48–62. https://doi.org/10.14421/jiska.2025.10.1.48-62

Wulandari, N. and Badieah, N. (2025) 'Implementasi teknik resampling untuk mengatasi ketidakseimbangan data terhadap klasifikasi anemia menggunakan Support Vector Machine', Jurnal Rekayasa Sistem Informasi dan Teknologi, 2(3), pp. 942–951. https://doi.org/10.70248/jrsit.v2i3.1856

Zhang, B., Dong, X., Hu, Y., Jiang, X., Li, G., 2023. Classification and prediction of spinal disease based on the SMOTE-RFEXGBoost model. PeerJ Comput Sci 9, 1–20. https://doi.org/10.7717/PEERJ-CS.1280