

Penerapan Algoritma Random Forest dan Teknik SMOTE untuk Prediksi Kematian Akibat Gagal Jantung Menggunakan RapidMiner

Zaehol Fatah¹, Islamiyatul Addewiyah²
^{1,2,3} Universitas Ibrahimy, Situbondo
Email : islaamiyatuladewiyah34@gmail.com

Abstrak

Penelitian ini mengusulkan pengembangan model klasifikasi yang efektif untuk memprediksi risiko kematian akibat gagal jantung. Pendekatan ini mengintegrasikan Synthetic Minority Oversampling Technique (SMOTE) dengan algoritma Random Forest. Tingkat morbiditas dan mortalitas global sebagian besar diakibatkan oleh gagal jantung, dengan kemampuan prediksi yang akurat menjadi krusial untuk intervensi dini dan peningkatan prognosis pasien. Model prediktif medis menghadapi tantangan signifikan karena ketidakseimbangan kelas data. Hal ini terjadi ketika jumlah pasien bertahan hidup (kelas mayoritas) jauh melebihi jumlah pasien meninggal (kelas minoritas), sehingga memicu bias dalam model. Untuk mengatasi ini, teknik SMOTE diterapkan untuk menyeimbangkan distribusi data dengan menghasilkan sampel sintesis dari kelas minoritas. Dataset yang digunakan berasal dari Kaggle, terdiri dari 299 data rekam medis pasien dengan 13 atribut klinis relevan. Proses penelitian meliputi tahapan preprocessing data, balancing data menggunakan SMOTE, pelatihan model menggunakan Random Forest, dan evaluasi komprehensif. Penilaian performa model dilakukan dengan menggunakan berbagai metrik standar seperti akurasi, presisi, recall, skor F1, dan Area Under the Receiver Operating Characteristic (AUC). Eksperimen yang dilakukan memperlihatkan bahwa model ini sukses meraih nilai akurasi pada angka 84.60%. Di samping itu, model tersebut memperoleh presisi 90.49%, *recall* 86.79%, dan skor *Area Under the Curve* (AUC) mencapai 0.916. Capaian F1-score yang bernilai 88.59% semakin membuktikan adanya ekuilibrium yang kuat antara presisi dan *recall* pada model yang dikembangkan. Bukti yang ada memperlihatkan bahwa penggunaan bersama Random Forest dengan SMOTE menghasilkan peningkatan substansial dalam kemampuan meramalkan risiko kematian akibat gagal jantung, menawarkan alat yang berpotensi berharga bagi tenaga medis dalam mengidentifikasi pasien berisiko tinggi dan merencanakan strategi intervensi yang lebih optimal.

Kata Kunci: classification, heart failure, random forest, rapidminer, smote.

Abstract

The primary goal of this investigation is to formulate an efficacious classification model for predicting the mortality risk linked to heart failure. The methodology leverages the Random Forest algorithm in conjunction with the Synthetic Minority Oversampling Technique (SMOTE). Globally, heart failure is recognized as a primary contributor to both morbidity and mortality, making accurate predictive capability crucial for early intervention and improving patient prognosis. Medical predictive models face a significant challenge posed by class imbalance in the dataset, where the number of minority cases (e.g., deceased patients) substantially outweighs the majority class (surviving patients), leading to model bias. To alleviate this issue, the SMOTE technique is deployed to restore class balance by generating novel synthetic instances for the minority group. The dataset utilized originates from Kaggle, comprising 299 patient medical records with 13 relevant clinical attributes. The research process encompasses data preprocessing, data balancing using SMOTE, model training employing Random Forest, and comprehensive evaluation. Model performance assessment was conducted using various standard metrics such as accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic (AUC). The experiments conducted

demonstrated that the developed model successfully achieved an accuracy value of 84.60%. Moreover, the model attained a precision of 90.49%, a recall of 86.79%, and an AUC (Area Under the Curve) score reaching 0.916. The substantial F1-score of 88.59% further confirms the robust equilibrium between the model's precision and recall metrics. The evidence suggests that the joint utilization of Random Forest and SMOTE results in a substantial improvement in the ability to forecast heart failure mortality risk, offering a potentially valuable tool for medical professionals in identifying high-risk patients and planning more optimal intervention strategies.

Keyword: classification, heart failure, random forest, rapidminer, smote.

PENDAHULUAN

Gagal jantung merupakan kondisi kronis yang terus menjadi masalah kesehatan global, berkontribusi pada angka kematian yang signifikan setiap tahunnya. Kemampuan untuk memprediksi risiko kematian akibat gagal jantung secara akurat dapat memberikan intervensi medis yang lebih cepat dan tepat, sehingga berpotensi menyelamatkan banyak nyawa (Febrian et al., 2025). Namun, dalam konteks data medis, seringkali ditemukan masalah ketidakseimbangan kelas (*imbalanced dataset*), di mana jumlah data untuk satu kelas (misalnya, pasien yang meninggal) jauh lebih sedikit dibandingkan kelas lainnya (pasien yang bertahan hidup). Situasi ketidaksetaraan ini memiliki potensi untuk memicu bias pada model prediktif, sehingga menjadikannya kurang andal untuk mendeteksi kasus-kasus minoritas. Konsekuensinya, mutu prediksi secara keseluruhan menurun (Arisandi, 2023).

Berbagai algoritma klasifikasi telah dikembangkan untuk mengatasi tantangan ini. Dalam tugas klasifikasi, Random Forest diakui sebagai algoritma yang kuat. Prinsip kerjanya adalah mendirikan sejumlah besar pohon keputusan dan kemudian mengagregasi *output* prediksi mereka untuk memperoleh keputusan akhir. Keuntungan utama dari strategi ini adalah mengurangi risiko *overfitting* dan meningkatkan performa akurasi (Hidayat, 2023). Meskipun demikian, performa Random Forest, seperti banyak algoritma pembelajaran mesin lainnya, sangat bergantung pada keseimbangan data (Sidiq et al., 2025). etika data tidak seimbang,

algoritma mungkin cenderung lebih memihak kelas mayoritas, mengabaikan pola penting dalam kelas minoritas (Nugroho & Rilvani, n.d.).

Berbagai teknik diusulkan sebagai respons terhadap masalah ketidakseimbangan dataset. Dalam konteks ini, Synthetic Minority Oversampling Technique (SMOTE) termasuk teknik yang paling sering digunakan dan terbukti ampuh. SMOTE melaksanakan fungsinya dengan memproduksi sampel sintesis baru, yang ditujukan untuk kelas minoritas. Hal ini merupakan kemajuan dibandingkan *oversampling* sederhana, karena memungkinkan kelas minoritas untuk diperkaya tanpa duplikasi langsung dari observasi data yang sudah ada (Erlin et al., 2022). Penerapan SMOTE dapat meningkatkan kualitas prediksi dengan memastikan bahwa model memiliki cukup informasi dari kelas minoritas untuk belajar dan membuat prediksi yang lebih seimbang (Khatib Sulaiman et al., 2023).

Secara menyeluruh, telah terdapat bukti bahwa SMOTE efektif dalam menanggulangi isu ketidakseimbangan kelas dalam konteks gagal jantung kronis. Efeknya adalah peningkatan efikasi diskriminasi risiko yang signifikan, khususnya ketika metode ini digabungkan dengan algoritma *machine learning* yang mutakhir (Wang et al., 2021). Bertolak dari konteks latar belakang di atas, fokus studi ini adalah menciptakan model klasifikasi yang kuat (*robust*) dalam memperkirakan risiko kematian pasien gagal jantung. Penelitian ini akan mengimplementasikan algoritma Random Forest yang dikombinasikan dengan teknik SMOTE

untuk mengatasi masalah ketidakseimbangan kelas pada dataset pasien gagal jantung. Studi-studi sebelumnya telah menunjukkan potensi Random Forest dalam klasifikasi penyakit jantung dan efektivitas SMOTE dalam menangani data tidak seimbang pada berbagai aplikasi medis (Sidiq et al., 2025). Pendekatan kombinasi ini telah diakui sebagai metode yang handal, bahkan dalam pengembangan model hibrida yang mengintegrasikan Random Forest dengan algoritma lain seperti SVM untuk mencapai akurasi prediksi gagal jantung yang superior (Sajid Abdillah et al., 2025). Dataset yang digunakan berasal dari Kaggle, yang berisi rekam medis pasien dengan berbagai atribut klinis. Sasaran fundamental dari studi ini adalah mengaplikasikan algoritma Random Forest guna memperkirakan risiko mortalitas yang disebabkan oleh gagal jantung serta meningkatkan akurasi prediksi dengan mengatasi ketidakseimbangan kelas menggunakan teknik SMOTE.

Hasil penelitian ini diharapkan memberikan kontribusi besar dalam strategi deteksi awal dan manajemen klinis pasien gagal jantung. Wawasan yang berharga juga diberikan mengenai kemandirian kombinasi Random Forest dan SMOTE saat berhadapan dengan data medis yang tidak seimbang. Studi ini juga diharapkan dapat menjadi referensi bagi penelitian lebih lanjut dalam penerapan teknik *machine learning* untuk masalah kesehatan dengan data yang tidak seimbang.

METODE

Atribut Dataset

Sumber data yang dimanfaatkan dalam studi ini berasal dari *Heart Failure Clinical Record*, yang dapat diakses melalui Kaggle. *Dataset* ini meliputi 299 rekam medis pasien dan mencakup 13 atribut prediktor, seperti usia, tekanan darah, tingkat enzim CPK, anemia, dan variabel lainnya. Target klasifikasi untuk model ini adalah *DEATH_EVENT*, yang menunjukkan apakah pasien meninggal (1)

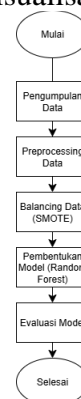
atau tidak (0). Spesifikasi dari atribut-atribut tersebut terlampir pada Tabel 1.

Tabel 1. Rincian dan Penjelasan Variabel Data

No	Variabel	Keterangan	Tipe
1	Usia Subjek Penelitian	Rentang Usia Pasien	Integer
2	anemia	Anemia (0 = tidak, 1 = ya)	Binominal
3	creatinine_phosphokinase	Enzim CPK	Integer
4	diabetes	Diabetes	Binominal
5	ejection_fraction	Persentase ejection fraction	Integer
6	high_blood_pressure	Tekanan darah tinggi	Binominal
7	platelets	Jumlah platelet	Real
8	serum_creatinine	Kadar kreatinin	Real
9	serum_sodium	Kadar sodium	Integer
10	sex	Jenis kelamin	Binominal
11	smoking	Status merokok	Binominal
12	time	Lama follow-up (hari)	Integer
13	DEATH_EVENT	Meninggal (1) atau tidak (0)	Binominal

Alur Penelitian

Untuk mencapai tujuan, serangkaian prosedur sistematis diterapkan dalam penelitian ini. Alur kerja dari keseluruhan proses dapat divisualisasikan dalam Gambar 1.



Gambar 1. Bagan Alir Prosedur Riset Algoritma Random Forest

Pohon keputusan menjadi basis bagi Random Forest, sebuah algoritma yang tergolong dalam *ensemble learning*. Selama periode pelatihan, algoritma ini beroperasi dengan menciptakan banyak pohon keputusan yang bersifat independen. Untuk pembentukan keputusan, setiap *node* pada pohon memilih subset fitur secara acak. Sementara itu, keseluruhan pohon di dalam *forest* mendapatkan pelatihan dari subset data yang dihasilkan melalui metode *bagging* (*bootstrap aggregating*). Mekanisme ini berkontribusi

dalam memitigasi *overfitting* yang umumnya dialami oleh pohon keputusan yang berdiri sendiri (*tunggal*). Selain itu, ia menambah kestabilan dan tingkat ketepatan (*akurasi*) model secara menyeluruh. Dalam proses klasifikasi, keputusan final dicapai melalui suara terbanyak (*mayoritas voting*) dari hasil prediksi tiap pohon. Keunggulan *Random Forest* mencakup kapasitasnya dalam mengolah data berdimensi tinggi, kemampuannya menangani nilai yang hilang (*missing values*) tanpa eksplisit, dan menyediakan penilaian terhadap signifikansi fitur. Namun, seperti yang telah dibahas sebelumnya, performanya sangat bergantung pada keseimbangan data, dan *oversampling* dapat menjadi langkah penting untuk mengoptimalkannya (Darmawan et al., 2024).

SMOTE

Sebagai sebuah teknik *oversampling*, Synthetic Minority Oversampling Technique (SMOTE) memiliki fungsi utama untuk menyesuaikan kesetaraan distribusi kelas pada *dataset*. Teknik ini sangat berguna jika kelas minoritas didominasi oleh jumlah sampel yang minim relatif terhadap kelas mayoritas (Indrawati, 2021). SMOTE beroperasi melalui penciptaan sampel sintesis yang sama sekali baru bagi kelas minoritas, yang berbeda dari sekadar penggandaan sampel yang sudah tersedia. Prosedur ini diawali dengan pemilihan sebuah sampel dari kelas minoritas, diikuti dengan identifikasi sejumlah tetangga terdekat (sebagai contoh, 5 tetangga terdekat). Sampel artifisial kemudian dibentuk di sepanjang ruas garis yang menyatukan sampel data asli dengan sampel-sampel tetangganya. Melalui penciptaan data sintesis tersebut, SMOTE berperan dalam menyetarakan distribusi kelas. Hal ini memfasilitasi model *machine learning* agar dapat menyerap pola dari kelas minoritas dengan tingkat efektivitas yang lebih tinggi, yang berujung pada peningkatan akurasi prediksi, khususnya

pada kelas yang populasinya kecil. Telah terdemonstrasi bahwa penerapan SMOTE secara substansial meningkatkan performa model klasifikasi ketika berhadapan dengan *dataset* medis yang tidak setara (Nugroho & Rilvani, n.d.).

RapidMiner

RapidMiner merupakan platform *data science* yang menyediakan ruang kerja terpadu untuk penyusunan data, pelatihan *machine learning*, dan penerapan (*deployment*) model. Sebagai alat yang digunakan dalam penelitian ini, RapidMiner menawarkan berbagai operator grafis yang memungkinkan pengguna untuk merancang dan menjalankan proses analitik tanpa perlu menulis kode yang kompleks. Fungsi ini memberikan asistensi krusial dalam melaksanakan seluruh fase studi, yang mencakup pra-pemrosesan data (*preprocessing*), penyeimbangan data dengan SMOTE, proses *training* model Random Forest, sampai pada penilaian kinerja menggunakan indikator-indikator metrik yang sesuai. Antarmuka visual RapidMiner mempermudah pengguna untuk memahami alur kerja dan mengimplementasikan berbagai algoritma *machine learning*, menjadikannya pilihan ideal untuk penelitian ini karena kemudahan penggunaan dan fungsionalitas yang komprehensif (Homepage et al., n.d.).

HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil dari implementasi algoritma Random Forest yang dikombinasikan dengan teknik SMOTE untuk prediksi kematian akibat gagal jantung. Hasil ini didapatkan setelah melalui serangkaian tahapan yang sistematis, mulai dari pemodelan hingga evaluasi kinerja model.

Hasil

Sesuai diagram pemodelan yang telah diuraikan, aplikasi teknik ini dieksekusi memanfaatkan RapidMiner. Tahap awal dimulai dengan memuat *dataset* dari berkas Excel menggunakan operator yang bernama 'Read Excel'. Setelah itu, data dialirkan ke operator

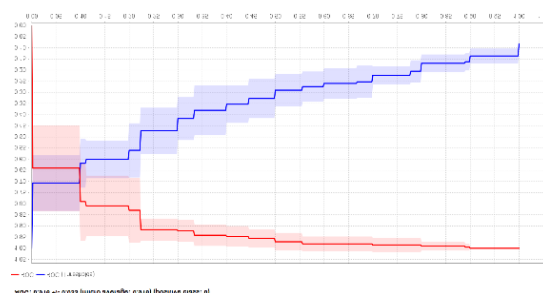
'Cross Validation'. Operator tersebut bertanggung jawab memfragmentasi *dataset* menjadi 10 *fold*, yang kemudian digunakan untuk melatih dan menguji model berulang kali (*iteratif*). Metodologi *cross-validation* ini memainkan peran vital dalam menjamin ketangguhan (*robustness*) model dan menangkal *overfitting* melalui pengujian performa model pada beragam subset data (Rahayu et al., 2024). Di dalam blok "Cross Validation", terdapat dua *sub-proses* utama: "Training" dan "Testing". Pada *sub-proses* "Training", teknik SMOTE Upsampling diterapkan terlebih dahulu untuk menyeimbangkan kelas dalam *dataset*. Pengaturan parameter SMOTE Upsampling meliputi jumlah tetangga (*number of neighbours*) sebanyak 5, normalisasi data (*normalize*) dicentang, dan penyeimbangan kelas (*equalize classes*) juga dicentang. Kelas minoritas secara spesifik diidentifikasi sebagai '1' (yaitu, pasien meninggal), dan opsi "round integers" juga diaktifkan. Pengaturan ini memastikan bahwa sampel sintetis yang dihasilkan relevan dan efektif dalam mengatasi ketidakseimbangan data.

Setelah SMOTE, data yang sudah seimbang kemudian digunakan untuk melatih algoritma Random Forest. Parameter Random Forest diatur dengan jumlah pohon (*number of trees*) sebanyak 200 dan kedalaman maksimum (*maximal depth*) sebesar 5. Kriteria pemisahan (*criterion*) yang digunakan adalah *gini index*. Pemilihan parameter ini didasarkan pada upaya untuk mencapai keseimbangan antara kompleksitas model dan kemampuannya untuk menggeneralisasi dengan baik pada data baru.

Dalam sub-proses 'Testing', model yang telah mendapatkan pelatihan diaplikasikan pada data uji melalui operator bernama 'Apply Model'. Luran prediksi dari 'Apply Model' selanjutnya diteruskan (*diumpankan*) menuju operator 'Performance' untuk mengukur kinerja model. Operator 'Performance' bertugas menghitung beragam indikator evaluasi,

meliputi Akurasi, Presisi, *Recall*, *F1-Score*, dan AUC.

Setelah penerapan SMOTE, distribusi data menjadi lebih seimbang, yang berkontribusi pada peningkatan performa model. Model berhasil mencapai akurasi sebesar 84.60%, presisi 90.49%, recall 86.79%, dan AUC 0.916. Dengan nilai F1-score yang mencapai 88.59%, terindikasi adanya ekuilibrium yang solid antara metrik presisi dan *recall*. Skor tinggi ini mengisyaratkan bahwa model tidak hanya berhasil memprediksi kasus positif dengan tepat, tetapi juga sukses mengenali mayoritas dari kasus positif yang sebenarnya ada. Hasil evaluasi model dan performa secara detail disajikan pada visualisasi kurva ROC dan performa pada Gambar 2 dan 3.



Gambar 2. Kurva ROC Model Prediksi

Performa model dalam memisahkan kelas positif dan negatif ditampilkan melalui Kurva Receiver Operating Characteristic (ROC). Sebuah kurva yang berada dekat dengan pojok kiri atas mengisyaratkan performa model yang memuaskan. Kapabilitas diskriminasi model yang istimewa dikonfirmasi oleh Hasil perhitungan AUC menunjukkan angka 0.916.

```

PerformanceVector
PerformanceVector:
accuracy: 84.60% +/- 6.15% (micro average: 84.62%)
ConfusionMatrix:
True: 1 0
    1: 77 27
    0: 19 176
precision: 90.49% +/- 4.92% (micro average: 90.26%) (positive class: 0)
ConfusionMatrix:
True: 1 0
    1: 77 27
    0: 19 176
recall: 86.79% +/- 7.55% (micro average: 86.70%) (positive class: 0)
ConfusionMatrix:
True: 1 0
    1: 77 27
    0: 19 176
AUC (optimistic): 0.916 +/- 0.033 (micro average: 0.916) (positive class: 0)
AUC: 0.916 +/- 0.033 (micro average: 0.916) (positive class: 0)
AUC (pessimistic): 0.916 +/- 0.033 (micro average: 0.916) (positive class: 0)
    
```

Gambar 3. Hasil evaluasi performa model

Hasil evaluasi menunjukkan akurasi rata-rata $84.60\% \pm 6.19\%$, presisi $90.49\% \pm 4.92\%$, recall $86.79\% \pm 7.55\%$, dan AUC 0.916 ± 0.033 dengan hasil confusion matrix yang stabil. Hasil ini membuktikan konsistensi model dalam prediksi kematian akibat gagal jantung dengan F1-Score sebesar 88.59% yang memperlihatkan ekuilibrium yang optimal antara presisi dan *recall*.

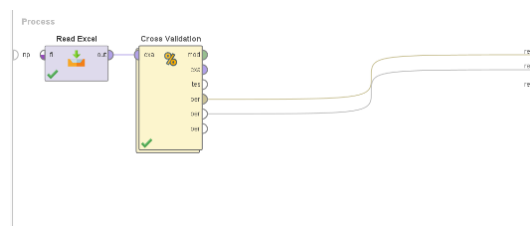
Tabel 2. Standarisasi Nilai AUC (Area Under the Curve)

Nilai AUC	Interpretasi Kualitas Model
0.5 - 0.6	Gagal (Failed)
0.6 - 0.7	Buruk (Poor)
0.7 - 0.8	Rata-rata (Average)
0.8 - 0.9	Baik (Good)
0.9 - 1.0	Sangat Baik (Excellent)

Berdasarkan Tabel 2(Nahm, 2022), Dengan nilai AUC mencapai 0.916, model klasifikasi dianggap termasuk dalam kategori 'Sangat Baik'. Pencapaian ini menunjukkan kapabilitas diskriminasi yang istimewa dalam mengidentifikasi perbedaan antara pasien yang meninggal dan pasien yang bertahan.

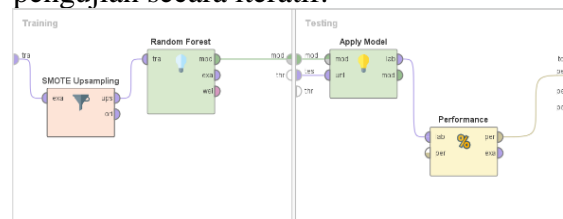
Pembahasan

Algoritma Random Forest, sebagai inti dari pemodelan ini, bekerja dengan membangun banyak pohon keputusan yang membentuk sebuah "hutan". Setiap pohon dilatih secara independen pada *subset* data yang diambil secara acak dengan penggantian (*bagging*), dan pada setiap *node* dalam pohon, *subset* fitur juga dipilih secara acak untuk menentukan pemisahan terbaik. Proses ini berkontribusi pada pengurangan *overfitting* dan peningkatan generalisasi model. Teknik ini selanjutnya disinergikan dengan metode SMOTE guna mengatasi persoalan ketidakseimbangan kelas dalam *dataset*. Langkah-langkah yang dilakukan dalam proses pemodelan ini dapat dideskripsikan dalam gambar berikut, yang dimulai dari proses pembacaan data, validasi silang, hingga pelatihan dan pengujian model:



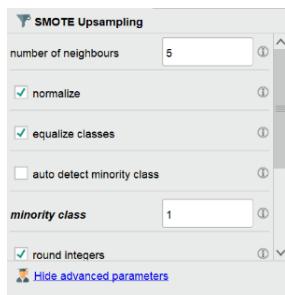
Gambar 4. Proses Baca Data dan Cross Validation

Proses pemodelan diawali dengan alur kerja tingkat atas yang sistematis, sebagaimana diilustrasikan pada Gambar 4. Langkah pertama adalah pembacaan dataset, yang dilakukan menggunakan operator 'Read Excel' untuk memuat data rekam medis pasien. Data yang berhasil dimuat kemudian diarahkan ke operator 'Cross Validation'. Penggunaan validasi silang (dalam penelitian ini 10 fold) merupakan langkah penting untuk memastikan robustness model dan mencegah overfitting, dengan membagi dataset menjadi subset pelatihan dan pengujian secara iteratif.



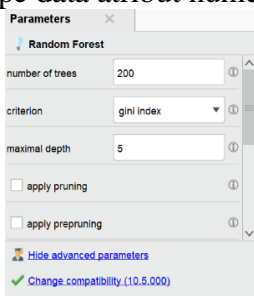
Gambar 5. Proses Training dan Testing Model dalam Cross Validation

Detail implementasi model di dalam blok *Cross Validation* disajikan pada Gambar 5, yang secara eksplisit memisahkan fase *Training* (Pelatihan) dan *Testing* (Pengujian). Pada fase Pelatihan, data diolah melalui dua tahapan kunci: Langkah awal adalah penerapan teknik SMOTE *Upsampling* untuk menyetarakan distribusi kelas. Selanjutnya, data yang sudah seimbang tersebut dimanfaatkan guna melakukan proses *training* pada algoritma Random Forest. Pada fase 'Pengujian' berikutnya, model yang sudah mendapatkan pelatihan diterapkan pada data uji memanfaatkan operator 'Apply Model'. Luaran dari 'Apply Model' langsung diteruskan untuk dinilai oleh operator 'Performance', demi mendapatkan indikator kinerja yang lengkap (*komprehensif*).



Gambar 6. Pengaturan Parameter SMOTE Upsampling

Konfigurasi spesifik untuk operator SMOTE Upsampling disajikan pada Gambar 6, yang menunjukkan strategi yang diadopsi untuk mengatasi ketidakseimbangan data. Parameter kunci diatur dengan number of neighbours sebanyak 5 untuk menentukan jarak dalam penciptaan sampel sintesis. Opsi normalize dan equalize classes diaktifkan untuk menjamin proses normalisasi dan penyeimbangan proporsi kelas. Kelas minoritas, yaitu pasien meninggal ('1'), secara eksplisit ditargetkan untuk oversampling, dengan round integers dicentang untuk mempertahankan integritas tipe data atribut numerik.



Gambar 7. Pengaturan Parameter Random Forest

Detail konfigurasi algoritma Random Forest yang digunakan dalam fase pelatihan model ditunjukkan pada Gambar 7. Untuk mencapai performa ensemble yang optimal, number of trees ditetapkan sebanyak 200. Sebagai kriteria pemisahan (*criterion*) pada setiap *node* pohon, dipilih gini index. Secara krusial, maximal depth dibatasi hingga 5. Pembatasan kedalaman ini merupakan upaya strategis untuk mengontrol kompleksitas model dan meningkatkan kemampuan generalisasi pada data yang belum pernah dilihat,

sehingga mengurangi risiko *overfitting* pada *dataset* medis.

SIMPULAN (PENUTUP)

Mengacu pada temuan dan analisis yang telah diselesaikan, dapat ditarik simpulan bahwa penggabungan algoritma Random Forest dan teknik SMOTE adalah metode yang luar biasa efektif dan andal (*robust*) untuk meramalkan risiko mortalitas yang disebabkan oleh gagal jantung. Penerapan SMOTE berhasil mengatasi masalah fundamental ketidakseimbangan kelas pada *dataset*, sebuah tantangan umum dalam data medis, yang secara krusial meningkatkan kinerja model pada kelas minoritas (pasien yang meninggal). Ini memastikan bahwa model tidak hanya akurat secara keseluruhan tetapi juga sensitif terhadap kasus-kasus yang paling penting untuk dideteksi.

Model yang dihasilkan menunjukkan performa yang sangat kuat dan konsisten, dibuktikan dengan nilai akurasi sebesar 84.60%, presisi 90.49%, recall 86.79%, dan nilai AUC yang tinggi sebesar 0.916. Skor F1 yang mencapai 88.59% semakin menegaskan adanya ekuilibrium unggul antara metrik presisi dan *recall*. Capaian ini memperlihatkan bahwa model tidak hanya unggul dalam mengklasifikasikan secara benar pasien yang berisiko meninggal (presisi yang tinggi), tetapi juga berhasil mendeteksi mayoritas dari kasus positif yang sesungguhnya (tingginya *recall*), seiring dengan upaya meminimalkan kekeliruan identifikasi. Nilai AUC 0.916 menempatkan model ini dalam kategori "Sangat Baik" berdasarkan standarisasi AUC, menegaskan kemampuan diskriminasi model yang luar biasa. Dengan demikian, penelitian ini membuktikan bahwa algoritma Random Forest yang dikombinasikan dengan teknik SMOTE tidak hanya efektif meningkatkan akurasi prediksi kematian akibat gagal jantung tetapi juga memberikan hasil yang stabil dan dapat diandalkan dalam kondisi data yang kompleks. Kontribusi yang dihasilkan diperkirakan dapat menjadi pijakan untuk mengembangkan sistem

pendukung keputusan klinis yang lebih mutakhir. Hal ini bertujuan untuk mengasistensi tenaga medis dalam melakukan identifikasi awal dan merencanakan tindakan intervensi yang lebih terpersonalisasi serta tepat waktu, sehingga pada akhirnya meningkatkan *outcome* (luaran) pasien penderita gagal jantung.

UCAPAN TERIMA KASIH

Kepada Bapak Dosen Pembimbing, Penulis mengekspresikan rasa terima kasih yang setulusnya atas bimbingan, arahan, dan dukungan yang tak ternilai selama proses penelitian ini. Kami juga menghaturkan apresiasi tertinggi kepada semua pihak yang berkepentingan di Universitas Ibrahimy, khususnya Fakultas Sistem Informasi. Dukungan berupa fasilitas dan kesempatan yang mereka berikan telah memungkinkan terlaksananya penelitian ini dengan sukses.

DAFTAR PUSTAKA

- Arisandi, R. (2023). PERBANDINGAN MODEL KLASIFIKASI RANDOM FOREST DENGAN RESAMPLING DAN TANPA RESAMPLING PADA PASIEN PENDERITA GAGAL JANTUNG. *Jurnal Gaussian*, 12(1), 136–145. <https://doi.org/10.14710/j.gauss.12.1.136-145>
- Darmawan, R., Herry Chrisnanto, Y., Abdillah, G., Sains dan Informatika, F., Jenderal Achmad Yani Jln Terusan Jend, U., Cimahi, K., & Barat, J. (2024). KLASIFIKASI DIAGNOSA PENYAKIT TIROID MENGGUNAKAN METODE RANDOM FOREST. In *Jurnal Informatika & Rekayasa Elektronika* (Vol. 7, Issue 2). <http://e-journal.stmiklombok.ac.id/index.php/jireISSN.2620-6900>
- Erlin, E., Desnelita, Y., Nasution, N., Suryati, L., & Zoromi, F. (2022). Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 21(3), 677–690. <https://doi.org/10.30812/matrik.v21i3.1726>
- Febrian, M. R., Saifudin, I., & Suharso, W. (2025). Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma Naive Bayes The Prediction Of Brain Failure Using The Naive Bayes Algorithm. In *Jurnal Smart Teknologi* (Vol. 6, Issue 2). <http://jurnal.unmuhjember.ac.id/index.php/JST>
- Hidayat, A. S. H. A. F. (2023). *Klasifikasi Penyakit Jantung Menggunakan Random Forest Classifier*. Homepage, J., Dewi, S., Kresnawati, A., & Salsabil, S. H. (n.d.). *IJIRSE: Indonesian Journal of Informatic Research and Software Engineering Implementation Of Data Mining In Determining Promotional Strategies At Al-Amah Vocational School Cimanggung Using C4.5 Algorithm Implementasi Data Mining Untuk Menentukan Strategi Promosi Di Smk Al-Amah Cimanggung Menggunakan Algoritma C4.5*.
- Indrawati, A. (2021). PENERAPAN TEKNIK KOMBINASI OVERSAMPLING DAN UNDERSAMPLING UNTUK MENGATASI PERMASALAHAN IMBALANCED DATASET. *Jurnal Informatika Dan Komputer) Akreditasi KEMENRISTEKDIKTI*, 4(1). <https://doi.org/10.33387/jiko>
- Khatib Sulaiman, J., Mizwar Rahim, A. A., Yanuar Risca Pratiwi, I., Ainul Fikri, M., & Amikom Yogyakarta, U. (2023). Klasifikasi Penyakit Jantung Menggunakan Metode Synthetic Minority Over-Sampling Technique Dan Random Forest Clasifier. *Indonesian Journal of Computer Science Attribution*, 12(5), 2023–2995.
- Nahm, F. S. (2022). Receiver operating characteristic curve: overview and practical use for clinicians. *Korean Journal of Anesthesiology*, 75(1), 25–36. <https://doi.org/10.4097/kja.21209>
- Nugroho, A., & Rilvani, E. (n.d.). Penerapan Metode Oversampling SMOTE Pada Algoritma Random Forest Untuk Prediksi Kebangkrutan Perusahaan Application of the SMOTE Oversampling Method to the Random Forest Algorithm for Predicting Company Bankruptcy. In *Februari* (Vol. 22, Issue 1).
- Rahayu, M., Luthfiarta, A., Cahyaningrum, L., & Azzahra, A. N. (2024). Pengaruh Oversampling dan Cross Validation Pada Model Machine Learning Untuk Sentimen Analisis Kebijakan Luaran Kelulusan Mahasiswa. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 8(1), 163. <https://doi.org/10.30865/mib.v8i1.7012>
- Sajid Abdullah, M., Mulyo, H., Wahyu, G., & Wibowo, N. (2025). Heart Failure Classification Using a Hybrid Model Based on SVM and Random Forest. *Journal of Dinda Data Science, Information Technology, and Data Analytics*, 5(2), 208–219. <http://journal.ittelkom-pwt.ac.id/index.php/dinda>
- Sidiq, S., Korespondensi, P., & Shobi Mabur, N. (2025). *Pengembangan Model Prediksi Risiko Diabetes Menggunakan Pendekatan AdaBoost dan Teknik Oversampling SMOTE*. <https://doi.org/10.58602/jima-ilkom.v4i1.41>
- Wang, K., Tian, J., Zheng, C., Yang, H., Ren, J., Li, C., Han, Q., & Zhang, Y. (2021). Improving risk identification of adverse outcomes in chronic heart failure using smote +enn and machine learning. *Risk Management and Healthcare Policy*, 14, 2453–2463. <https://doi.org/10.2147/RMHP.S310295>