

Peningkatan Akurasi Prediksi Penyakit Jantung dengan Teknik SMOTEENN pada Algoritma Random Forest

Alfin Rahmada¹, Erliyan Redy Susanto^{*2}

^{1,2}Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Indonesia
Email: ¹alfin_rahmada@teknokrat.ac.id, ²erliyan.redy@teknokrat.ac.id

Abstrak

Penyakit jantung merupakan salah satu penyebab utama kematian di dunia yang dapat mengancam nyawa jika tidak ditangani dengan tepat. Bahaya penyakit ini meliputi serangan jantung mendadak, gagal jantung, hingga stroke akibat komplikasi pembuluh darah. Masalah utama yang dihadapi dalam studi ini adalah akurasi prediksi penyakit jantung saat ini masih cukup rendah sehingga belum layak untuk diimplementasikan dalam bentuk program komputer. Penelitian ini bertujuan untuk meningkatkan akurasi prediksi penyakit jantung dengan menggunakan metode *machine learning*. Dalam studi ini kami menggunakan algoritma *Random Forest* dan mengintegrasikan teknik SMOTEENN (*Synthetic Minority Oversampling Technique and Edited Nearest Neighbor*) ke dalam algoritma. Hasil studi menunjukkan bahwa performa metode yang kami gunakan lebih baik dari studi sebelumnya, dari 86% menjadi 94%. Penerapan teknik SMOTEENN pada algoritma *Random Forest* berhasil meningkatkan akurasi prediksi penyakit jantung secara signifikan. Penelitian ini berkontribusi dalam meningkatkan keakuratan model prediksi medis berbasis *machine learning* untuk mendukung proses diagnosa klinis penyakit jantung.

Kata kunci: Ketidakseimbangan Data, Machine Learning, Penyakit Jantung, Random Forest, SMOTEENN

Improving Heart Disease Prediction Accuracy with SMOTEENN Technique on Random Forest Algorithm

Abstract

Heart disease is one of the leading causes of death in the world that can be life-threatening if not treated properly. The dangers of this disease include sudden heart attack, heart failure, and stroke due to blood vessel complications. The main problem faced in this study is that the current accuracy of heart disease prediction is still quite low so it is not feasible to be implemented in the form of a computer program. This study aims to improve the accuracy of heart disease prediction using machine learning methods. In this study we used Random Forest algorithm and integrated SMOTEENN (*Synthetic Minority Oversampling Technique and Edited Nearest Neighbor*) technique into the algorithm. The results show that the performance of our method is better than previous studies, from 86% to 94%. The application of the SMOTEENN technique to the Random Forest algorithm has significantly improved the accuracy of heart disease prediction. This research contributes to improving the accuracy of machine learning-based medical prediction models to support the clinical diagnosis process of heart disease.

Keywords: Data Imbalance, Machine Learning, Heart Disease, Random Forest, SMOTEENN

1. PENDAHULUAN

Penyakit jantung merupakan salah satu penyebab utama kematian di dunia yang dapat mengancam nyawa jika tidak ditangani dengan tepat. Bahaya penyakit ini meliputi serangan jantung mendadak, gagal jantung, hingga stroke akibat komplikasi pembuluh darah[1][2][3]. Dalam diagnosa penyakit jantung, dokter menggunakan berbagai parameter seperti tekanan darah, kadar kolesterol, riwayat keluarga, berat badan, serta pemeriksaan lanjutan seperti elektrokardiogram (EKG) dan angiografi untuk mendeteksi gangguan pada fungsi jantung[4].

Artificial Intelligence (AI) memiliki peran signifikan dalam bidang medis, khususnya dalam memprediksi penyakit jantung[5][6]. Dengan AI, data medis yang kompleks dapat dianalisis untuk mengidentifikasi pola-pola yang sulit ditemukan secara manual[7][8]. Sebelum mengembangkan aplikasi medis berbasis AI, langkah awal yang dilakukan adalah membangun model prediktif guna memastikan akurasi dan keandalannya. Salah satu

tantangan utama dalam pengembangan model AI adalah rendahnya akurasi akibat ketidakseimbangan data, di mana jumlah kasus positif jauh lebih sedikit dibandingkan kasus negatif[9][10]. Penelitian ini bertujuan untuk meningkatkan akurasi prediksi penyakit jantung dengan menggunakan teknik *Synthetic Minority Oversampling Technique and Edited Nearest Neighbor* (SMOTEENN), yang mengombinasikan metode *oversampling* dan *undersampling* untuk mengatasi ketidakseimbangan data[11].

Berbagai metode telah digunakan untuk memprediksi penyakit jantung, seperti *Logistic Regression*, *Support Vector Machine* (SVM), dan *Neural Networks*, dengan hasil akurasi yang beragam[12][13]. *Logistic Regression* biasanya menunjukkan akurasi yang cukup baik, sedangkan SVM dan *Neural Networks* cenderung memberikan performa lebih tinggi namun membutuhkan waktu komputasi lebih panjang[14][15]. Dalam penelitian ini, algoritma *Random Forest* dipilih karena kemampuannya mengolah dataset yang kompleks, menghasilkan akurasi tinggi, dan memberikan interpretasi yang lebih mudah dibandingkan metode lainnya[16].

Riset ini menawarkan pendekatan baru dengan mengintegrasikan teknik SMOTEENN ke dalam algoritma *Random Forest* untuk meningkatkan akurasi prediksi penyakit jantung[17][18]. Pendekatan ini berbeda dari penelitian sebelumnya yang umumnya hanya menggunakan kombinasi berbagai rasio[19]. Dengan menggabungkan kedua teknik tersebut, model tidak hanya mampu menangani ketidakseimbangan data, tetapi juga memperbaiki kualitas data dengan menghilangkan *outlier* yang dapat menurunkan performa model[20]. Strategi ini menghasilkan prediksi yang lebih akurat dan tangguh dibandingkan dengan metode tradisional. Studi ini berhasil meningkatkan akurasi model prediksi dari 86% menjadi 94%, melebihi performa metode lain seperti *Logistic Regression*, *Support Vector Machine*, dan *Artificial Neural Networks* yang sebelumnya dilaporkan[21].

2. METODE PENELITIAN

Penelitian ini menggunakan satu metode prediksi pembelajaran mesin dalam prediksi penyakit jantung yaitu metode *Random Forest* dengan tahapan penelitian seperti pada Gambar 1.



Gambar 1 Tahapan Penelitian

a) Pengumpulan Data

Dataset jantung diunduh dan dimuat ke dalam program.

b) Pembagian Fitur dan Target

Dataset yang dimuat dipisahkan menjadi dua bagian utama, fitur X yang berisi informasi atau variabel penentu dan target y yang menjadi fokus prediksi.

c) Penanganan Ketidakseimbangan Data

Data diseimbangkan menggunakan SMOTE untuk *oversampling* kelas minoritas, dan ENN diterapkan untuk mengurangi noise dalam data.

d) Pembagian Dataset

Dataset hasil *resampling* dibagi menjadi data latih dan data uji dengan rasio 80:20.

e) Normalisasi Fitur

Data dinormalisasi menggunakan *Standard Scaler* untuk memastikan semua fitur memiliki skala yang serupa.

f) Pelatihan Model

Model *Random Forest* digunakan untuk proses pelatihan. Model ini dipilih karena keandalannya dalam menangani data dengan banyak fitur serta memberikan hasil yang stabil. Data latih yang telah dinormalisasi dan diseimbangkan digunakan dalam tahap ini.

g) Evaluasi Model

Model diuji menggunakan data uji, menghasilkan metrik seperti akurasi, presisi, recall, F1-score, dan ROC-AUC.

h) Visualisasi Hasil

Confusion matrix dan ROC curve divisualisasikan untuk memberikan pemahaman mendalam tentang kinerja model.

2.1. Random Forest

Random Forest adalah salah satu algoritma pembelajaran mesin berbasis *ensemble learning* yang digunakan untuk tugas klasifikasi dan regresi[22]. Algoritma ini merupakan pengembangan dari metode *Decision Tree*, di mana alih-alih menggunakan satu pohon keputusan (*Decision Tree*), *Random Forest* membangun banyak pohon (hutan) untuk membuat prediksi yang lebih akurat dan robust. Algoritma *Random Forest* dipilih karena kemampuannya mengolah dataset yang kompleks, menghasilkan akurasi tinggi, dan memberikan interpretasi yang lebih mudah dibandingkan metode lainnya[16]. Dalam studi ini beberapa parameter penting dapat disesuaikan untuk mengoptimalkan kinerja model. $n_estimators$, parameter ini menentukan jumlah pohon dalam hutan (*forest*). Defaultnya adalah 100, tetapi jumlah ini dapat ditingkatkan untuk menghasilkan model yang lebih stabil dengan biaya waktu pelatihan tambahan. max_depth , parameter ini membatasi kedalaman maksimum setiap pohon. Kedalaman yang terlalu besar dapat menyebabkan model *overfitting* terhadap data latih. Secara *default*, max_depth tidak diatur sehingga setiap pohon tumbuh hingga sempurna, tetapi ini dapat disesuaikan berdasarkan kebutuhan untuk mengurangi kompleksitas model. $random_state$, parameter ini memastikan reproduktibilitas eksperimen dengan mengatur seed acak yang tetap. Di sini, $random_state$ diatur ke 42 untuk memastikan bahwa hasil pelatihan tetap konsisten di setiap eksekusi[23].

$$\hat{y} = mode(T1(X), T2(X), \dots, TN(X)) \quad (1)$$

Keterangan:

$T_i(X)$: Prediksi dari pohon keputusan i .

mode: Prediksi yang paling sering muncul.

2.2. SMOTEENN

SMOTEENN (*Synthetic Minority Oversampling Technique and Edited Nearest Neighbor*) adalah teknik pengolahan data yang digunakan untuk mengatasi ketidakseimbangan data (*imbalanced data*), yang sering terjadi ketika jumlah sampel dalam kelas minoritas jauh lebih sedikit dibandingkan kelas mayoritas. SMOTEENN merupakan gabungan dari dua metode yang saling melengkapi, SMOTE (*Synthetic Minority Oversampling Technique*) adalah metode *oversampling* yang menghasilkan data sintesis untuk kelas minoritas[24][25]. Teknik ini bekerja dengan membuat sampel baru di sepanjang garis antara data minoritas yang ada, alih-alih hanya menduplikasi data. Hal ini membantu meningkatkan jumlah data di kelas minoritas secara realistis tanpa mengubah distribusi data secara signifikan. dan ENN (*Edited Nearest Neighbor*) adalah teknik *undersampling* yang digunakan setelah proses SMOTE. ENN berfungsi untuk membersihkan data dengan menghapus sampel dari kelas mayoritas yang salah diklasifikasikan oleh algoritma *K-Nearest Neighbor* (KNN) berdasarkan tetangganya[26]. Ini membantu mengurangi noise dan memastikan bahwa data yang digunakan untuk pelatihan lebih berkualitas. Setelah proses ini, dataset menjadi lebih seimbang, meningkatkan kemampuan model dalam mendeteksi kasus pada kelas minoritas.

$$x_{new} = x_i + \lambda(x_k - x_i) \quad (2)$$

Keterangan:

x_i : Sampel minoritas asli

x_k : Tetangga terdekat x_i

λ : Bilangan acak antara 0 dan 1.

2.3. Confusion Matrix

Confusion Matrix adalah tabel yang digunakan untuk mengevaluasi kinerja model klasifikasi dengan membandingkan hasil prediksi model dengan nilai aktual (data sebenarnya). *Confusion matrix* memberikan informasi tentang prediksi benar dan salah yang dilakukan model untuk setiap kelas[27].

$$\begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix} \quad (3)$$

Keterangan:

TP : True Positives.

FP : False Positives.

FN : False Negatives.

TN : True Negatives.

Confusion matrix terdiri dari beberapa perhitungan.

- a) Akurasi
Mengukur seberapa sering model membuat prediksi yang benar.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

- b) Presisi
Mengukur proporsi prediksi positif yang benar-benar positif.

$$Presisi = \frac{TP}{TP+FP} \quad (5)$$

- c) Recall
Mengukur kemampuan model untuk mendeteksi semua kasus positif yang sebenarnya.

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

- d) Specifycity
Mengukur kemampuan model untuk mengenali semua kasus negatif yang sebenarnya.

$$Specifycity = \frac{TN}{TN+FP} \quad (7)$$

- e) F1-Score
Menggabungkan Precision dan Recall dalam satu metrik dengan mengambil rata-rata harmonisnya.

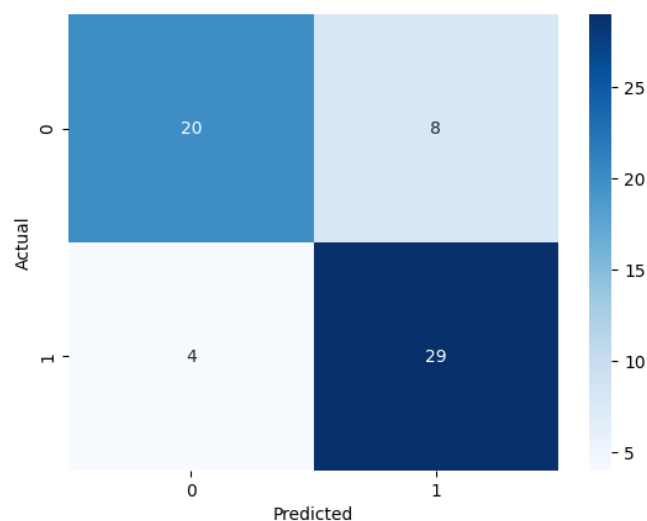
$$F1 = 2 \times \frac{Presisi \times Recall}{Presisi + Recall} \quad (8)$$

3. HASIL DAN PEMBAHASAN

Dalam studi ini kami menggunakan metode *Random Forest*. Ada 2 strategi yang kami gunakan yaitu menggunakan metode *Random Forest* dan *Random Forest* dengan teknik SMOTEENN.

3.1. Metode Random Forest

Berikut ini optimasi model *Random Forest* untuk prediksi penyakit jantung menggunakan teknik SMOTEENN dan tanpa menggunakan teknik SMOTEENN. *Confusion Matrix* algoritma *Random Forest* disajikan pada Gambar 2.



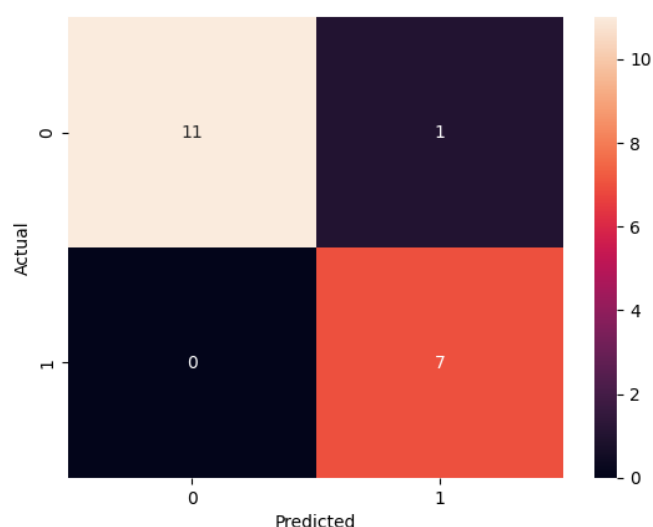
Gambar 2 *Confusion Matrix Random Forest*

Gambar 2 menampilkan *Confusion Matrix*, sebuah alat yang umum digunakan dalam pembelajaran mesin untuk mengevaluasi kinerja model klasifikasi. Matriks ini secara khusus mengevaluasi model *Random Forest*,

yang merupakan salah satu metode pembelajaran ensemble yang populer. Baris pada matriks mewakili label kelas aktual dari data, sedangkan kolom mewakili label yang diprediksi oleh model *Random Forest*. Setiap sel menunjukkan jumlah kejadian di mana model memprediksi kelas tertentu (kolom) saat kelas aktualnya berbeda (baris).

Sel diagonal mewakili prediksi yang benar, menunjukkan bahwa model berhasil mengklasifikasikan data dengan benar. Sebagai contoh, sel pada perpotongan baris 0 dan kolom 0 menunjukkan 20 kejadian di mana model dengan benar memprediksi kelas 0. Sementara itu, sel di luar diagonal mewakili prediksi yang salah atau kesalahan klasifikasi. Misalnya, sel pada baris 1 dan kolom 0 menunjukkan 4 kejadian di mana model salah memprediksi kelas 0 padahal kelas aktualnya adalah 1.

Dengan menganalisis *Confusion Matrix*, dapat dipahami seberapa baik model *Random Forest* mengklasifikasikan berbagai kelas, mengidentifikasi potensi bias, dan memahami jenis kesalahan yang dibuat oleh model. Informasi ini dapat digunakan lebih lanjut untuk meningkatkan akurasi model dengan menangani area tertentu di mana model berkinerja buruk. *Confusion Matrix* algoritma *Random Forest* menggunakan teknik SMOTEENN disajikan pada Gambar 3.



Gambar 3 *Confusion Matrix Random Forest* dengan teknik SMOTEENN

Gambar 3 menggambarkan *Confusion Matrix* menggunakan teknik SMOTEENN untuk meningkatkan akurasi prediksi penyakit jantung, sebuah alat visualisasi yang digunakan untuk mengevaluasi kinerja model klasifikasi. Matriks ini menunjukkan jumlah *instance* yang diklasifikasikan dengan benar dan salah untuk setiap kelas. Baris dalam matriks mewakili label kelas aktual, sedangkan kolom mewakili label kelas yang diprediksi. Elemen diagonal (11 dan 7) menunjukkan jumlah *instance* yang diklasifikasikan dengan benar untuk setiap kelas, sedangkan elemen di luar diagonal (1 dan 0) menunjukkan jumlah *instance* yang diklasifikasikan secara salah.

Sebagai contoh, sel di kanan atas (1) menunjukkan bahwa satu *instance* dari kelas 0 salah diprediksi sebagai kelas 1. Demikian pula, sel di kiri bawah (0) menunjukkan bahwa tidak ada *instance* dari kelas 1 yang salah diklasifikasikan sebagai kelas 0. *Confusion Matrix* dapat digunakan untuk menghitung berbagai metrik kinerja, seperti akurasi, presisi, recall, dan F1-score. Metrik-metrik ini membantu mengevaluasi kinerja keseluruhan model dan mengidentifikasi area yang memerlukan perbaikan[28][29].

Dalam penelitian sebelumnya menggunakan 3 metode pembelajaran mesin dalam prediksi penyakit jantung yang terdiri dari *Support Vector Machine* (SVM), *Logistic Regression* dan *Artificial Neural Network* (ANN) dengan pembagian data dan uji 80% : 20% didapatkan hasil seperti pada Tabel 1.

Tabel 1 Pembagian Data 80% : 20%

Confusion Matrix	Precision	Recall	F1 Score	Accuracy
SVM	0.67	0.88	0.76	0.70
Logistic Regression	0.88	0.88	0.88	0.86
ANN	0.87	0.84	0.86	0.85

Berdasarkan pada Tabel 1, metode *Logistic Regression* memiliki nilai presisi tertinggi sebesar 0.88. Baik SVM maupun *Logistic Regression* sama-sama mencapai nilai recall tertinggi sebesar 0.88. Selain itu, *Logistic Regression* juga mencatat nilai F1-score tertinggi sebesar 0.88 dan akurasi tertinggi sebesar 0.86[19]. Pembagian data dengan uji 80 : 20 algoritma *Random Forest* didapatkan hasil seperti Tabel 2.

Tabel 2 Pembagian Data 80% : 20%

Confusion Matrix	Precision	Recall	F1 Score	Accuracy
Random Forest	0.76	0.87	0.88	0.78

Tabel 2 menunjukkan hasil evaluasi model *Random Forest* yang dilakukan dengan pembagian data sebesar 80% untuk pelatihan dan 20% untuk pengujian. Model memiliki nilai presisi sebesar 0,76, yang menunjukkan tingkat keakuratan prediksi positif dari keseluruhan prediksi positif yang dilakukan oleh model, nilai recall mencapai 0,87, menunjukkan kemampuan model dalam mendeteksi semua data positif yang benar, model memiliki nilai F1 Score sebesar 0,88, yang merupakan rata-rata harmonis dari presisi dan recall, menandakan keseimbangan yang baik antara keduanya, tingkat akurasi model sebesar 0,78, yang menunjukkan persentase prediksi model yang benar terhadap keseluruhan data uji. Pembagian data dengan uji 80 : 20 algoritma *Random Forest* menggunakan teknik SMOTEENN didapatkan hasil seperti Tabel 3.

Tabel 3 Pembagian Data 80% : 20%

Confusion Matrix	Precision	Recall	F1 Score	Accuracy
Random Forest	0.87	1.0	0.93	0.94

Tabel 3 menampilkan hasil evaluasi model *Random Forest* dengan pembagian data sebesar 80% untuk pelatihan dan 20% pengujian dengan menerapkan teknik SMOTEENN untuk meningkatkan akurasi prediksi penyakit jantung. Hasil evaluasi menunjukkan metrik kinerja yang tinggi, dengan nilai Precision sebesar 0,87, yang mencerminkan tingkat keakuratan prediksi positif. Recall mencapai nilai sempurna 1,0, menunjukkan bahwa model mampu mengidentifikasi semua data positif secara benar. F1 Score sebesar 0,93 menunjukkan keseimbangan yang sangat baik antara presisi dan recall. Selain itu, Accuracy model tercatat sebesar 0,94, yang berarti 94% dari total prediksi yang dibuat model adalah benar. Secara keseluruhan, hasil ini menunjukkan performa yang sangat baik dari model *Random Forest* dalam melakukan klasifikasi.

3.2. Diskusi

Peningkatan akurasi dalam prediksi penyakit jantung dengan penggunaan algoritma *Random Forest* yang dikombinasikan dengan teknik SMOTEENN memiliki implikasi signifikan untuk aplikasi klinis. Model ini, dengan akurasi mencapai 94%, menunjukkan potensi besar untuk digunakan sebagai alat bantu diagnostik dalam menentukan risiko penyakit jantung. Dalam praktik klinis, prediksi yang lebih akurat memungkinkan identifikasi dini pasien berisiko tinggi, sehingga dapat dilakukan intervensi yang lebih cepat dan tepat, mengurangi kemungkinan komplikasi serius.

Membandingkan hasil penelitian ini dengan studi terdahulu membantu memberikan pemahaman yang lebih mendalam mengenai kontribusi yang dihasilkan oleh penelitian ini. Penelitian terdahulu melakukan komparasi antara tiga algoritma pembelajaran mesin (*Support Vector Machine*, *Logistic Regression*, dan *Artificial Neural Network*) dalam memprediksi penyakit jantung. Hasil menunjukkan bahwa *Logistic Regression* memberikan akurasi terbaik sebesar 86% pada skenario pembagian data 80:20. Sebaliknya, ANN memberikan hasil terbaik dengan akurasi 82% pada pembagian 70:30, sementara SVM menunjukkan performa terbaik dalam hal recall pada pembagian data lainnya[19]. Sedangkan penelitian ini membahas pendekatan dalam mengoptimalkan model *Random Forest* untuk prediksi penyakit jantung menggunakan teknik SMOTEENN[30][31][32]. Teknik ini menggabungkan *oversampling* (SMOTE) dan *undersampling* (ENN) untuk menangani masalah ketidakseimbangan data, di mana kelas minoritas sering kali diabaikan oleh model prediksi. Pendekatan ini bertujuan untuk meningkatkan akurasi pada kasus minoritas (pasien yang terdiagnosis penyakit jantung) tanpa mengorbankan informasi dari kelas mayoritas. SMOTEENN memberikan peningkatan signifikan dibandingkan teknik *oversampling* atau *undersampling* lainnya karena kombinasi dari dua pendekatan tersebut secara sinergis mengatasi ketidakseimbangan data.

Hasil penelitian menunjukkan bahwa model *Random Forest* yang dioptimalkan menggunakan SMOTEENN menghasilkan performa yang baik berdasarkan metrik-metrik evaluasi seperti akurasi, presisi, recall, F1-score, dan AUC-ROC. ROC-AUC sebesar 0,99 menunjukkan kemampuan model yang hampir sempurna dalam membedakan pasien dengan dan tanpa penyakit jantung. Evaluasi ini mempertegas bahwa

penggunaan SMOTEENN dapat meningkatkan kemampuan model menangani ketidakseimbangan data, memberikan prediksi yang lebih andal untuk data medis.

4. KESIMPULAN

Studi ini memberikan pendekatan baru yang efektif untuk mengatasi ketidakseimbangan data dalam prediksi medis. Hasil studi menunjukkan bahwa algoritma *Random Forest* dengan memanfaatkan teknik SMOTEENN dapat meningkatkan performa prediksi penyakit jantung. Pendekatan ini berhasil meningkatkan akurasi prediksi secara signifikan menjadi 94%, dengan nilai ROC-AUC mencapai 99%, menandakan kemampuan model yang hampir sempurna. Pendekatan ini memberikan kontribusi signifikan dalam mengatasi tantangan dalam prediksi penyakit jantung. Kontribusi ini menjadi langkah penting dalam mendukung pengembangan sistem pendukung keputusan berbasis *machine learning* yang lebih andal untuk aplikasi medis. Dengan hasil studi ini maka model dapat dikembangkan dalam bentuk aplikasi program komputer. Penelitian di masa mendatang dapat diarahkan pada pengintegrasian teknik lain untuk meningkatkan interpretabilitas model, selain itu, perlu dilakukan eksplorasi terhadap berbagai kombinasi algoritma dan teknik penyeimbangan data lainnya untuk menguji generalisasi metode pada dataset medis yang lebih beragam.

DAFTAR PUSTAKA

- [1] L. A. Sposato *et al.*, “Post-Stroke Cardiovascular Complications and Neurogenic Cardiac Injury,” *J. Am. Coll. Cardiol.*, vol. 76, no. 23, pp. 2768–2785, Dec. 2020, doi: 10.1016/j.jacc.2020.10.009.
- [2] A. Kumari, A. P. Singh, and C. V. Obiajulu, “Exploring the Relationship Between Muscle-Enhancing Steroids and Heart Attacks: A Comprehensive Overview,” *Int. J. Pharm. Res. Allied Sci.*, vol. 13, no. 3, pp. 19–34, 2024, doi: 10.51847/yzw5XWDJwZ.
- [3] F. Yulidasari, H. Salsabila, G. A. Sedar, G. A. N. Putri, H. Putery, and H. A. Zawary, “PENYULUHAN METODE ”CERDIK” SEBAGAI UPAYA PENGENDALIAN HIPERTENSI DI DESA WONOREJO RT 7 DAN 9 KECAMATAN SATUI,” *SELAPARANG J. Pengabd. Masy. Berkemajuan*, vol. 7, no. 2, p. 1300, Jun. 2023, doi: 10.31764/jpmb.v7i2.15214.
- [4] C. Shao, J. Wang, J. Tian, and Y. Tang, “Coronary Artery Disease: From Mechanism to Clinical Practice,” Springer, Singapore, 2020, pp. 1–36. doi: 10.1007/978-981-15-2517-9_1.
- [5] H. Haryanto, “Analisis Big Data dan Artificial Intelligence (AI): dalam Industri Khususnya Prediksi Penyakit Jantung dengan Python,” *Go Infotech J. Ilm. STMIK AUB*, vol. 30, no. 1, pp. 76–86, Jun. 2024, doi: 10.36309/goi.v30i1.262.
- [6] P. Mathur, S. Srivastava, X. Xu, and J. L. Mehta, “Artificial Intelligence, Machine Learning, and Cardiovascular Disease,” *Clin. Med. Insights Cardiol.*, vol. 14, p. 117954682092740, Jan. 2020, doi: 10.1177/1179546820927404.
- [7] H. M. Iskandar, “ANALISIS KOMUNIKASI PASIEN DENGAN DOKTER PADA APLIKASI KONSULTASI KESEHATAN BERBASIS ARTIFICIAL INTELEGENCE AICARE,” *J. Pariwisata*, vol. 4, no. 1, pp. 9–19, Jan. 2024, doi: 10.31294/pariwara.v4i1.5641.
- [8] I. Castiglioni *et al.*, “AI applications to medical images: From machine learning to deep learning,” *Phys. Medica*, vol. 83, pp. 9–24, Mar. 2021, doi: 10.1016/j.ejmp.2021.02.006.
- [9] M. Tiara Triani Br Sirait, N. Siti Fathonah, and M. Nurkamal Fauzan, “PEMANFAATAN ALGORITMA ADASYN DAN SUPPORT VECTOR MACHINE DALAM MENINGKATKAN AKURASI PREDIKSI KANKER PARU-PARU,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 5, pp. 8773–8778, Sep. 2024, doi: 10.36040/jati.v8i5.10752.
- [10] F. Thabtah, S. Hammoud, F. Kamalov, and A. Gonsalves, “Data imbalance in classification: Experimental evaluation,” *Inf. Sci. (Ny.)*, vol. 513, pp. 429–441, Mar. 2020, doi: 10.1016/j.ins.2019.11.004.
- [11] A. A. G. W. S. Erlangga, I. G. A. Gunadi, and I. M. G. Sunarya, “Kombinasi Oversampling dan Undersampling dalam Menangani Class Imbalanced dan Overlapping pada Klasifikasi Data Bank Marketing,” *J. Resist. (Rekayasa Sist. Komputer)*, vol. 7, no. 1, pp. 32–42, Apr. 2024, doi: 10.31598/jurnalresistor.v7i1.1515.
- [12] N. F. Sahamony, T. Terttiaavini, and H. Rianto, “Analisis Perbandingan Kinerja Model Machine Learning untuk Memprediksi Risiko Stunting pada Pertumbuhan Anak,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 413–422, Feb. 2024, doi: 10.57152/malcom.v4i2.1210.

-
- [13] A. S. Prabowo and F. I. Kurniadi, "Analisis Perbandingan Kinerja Algoritma Klasifikasi dalam Mendeteksi Penyakit Jantung," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 7, no. 1, pp. 56–61, Oct. 2023, doi: 10.47970/siskom-kb.v7i1.468.
- [14] A. Kurani, P. Doshi, A. Vakharia, and M. Shah, "A Comprehensive Comparative Study of Artificial Neural Network (ANN) and Support Vector Machines (SVM) on Stock Forecasting," *Ann. Data Sci.*, vol. 10, no. 1, pp. 183–208, Feb. 2023, doi: 10.1007/s40745-021-00344-x.
- [15] A. P. Rodrigues *et al.*, "Real-Time Twitter Spam Detection and Sentiment Analysis using Machine Learning and Deep Learning Techniques," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–14, Apr. 2022, doi: 10.1155/2022/5211949.
- [16] M. Raka Sujono, A. Bahtiar, and B. Irawan, "ANALISIS MODEL MACHINE LEARNING UNTUK JENIS ASPAL DI JAWA BARAT MENGGUNAKAN ALGORITMA DECISION TREE DAN RANDOM FOREST," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 6, pp. 3886–3891, Feb. 2024, doi: 10.36040/jati.v7i6.8275.
- [17] J. Yang and J. Guan, "A Heart Disease Prediction Model Based on Feature Optimization and Smote-Xgboost Algorithm," *Information*, vol. 13, no. 10, p. 475, Oct. 2022, doi: 10.3390/info13100475.
- [18] A. Seva, S. N. Tirumala Rao, and M. Sireesha, "Prediction of Liver Disease with Random Forest Classifier Through SMOTE-ENN Balancing," in *2024 IEEE 13th International Conference on Communication Systems and Network Technologies (CSNT)*, IEEE, Apr. 2024, pp. 928–933. doi: 10.1109/CSNT60213.2024.10546170.
- [19] F. Handayani, "Komparasi Support Vector Machine, Logistic Regression Dan Artificial Neural Network Dalam Prediksi Penyakit Jantung," *J. Edukasi dan Penelit. Inform.*, vol. 7, no. 3, p. 329, Dec. 2021, doi: 10.26418/jp.v7i3.48053.
- [20] W. Nugraha, R. Sabaruddin, and S. Murni, "Teknik Scaling Menggunakan Robust Scaler Untuk Mengatasi Outlier Data Pada Model Prediksi Serangan Jantung," *Techno.Com*, vol. 23, no. 2, pp. 319–327, May 2024, doi: 10.62411/tc.v23i2.10463.
- [21] C. Shao, J. Wang, J. Tian, and Y. Tang, "Coronary Artery Disease: From Mechanism to Clinical Practice," Springer Nature Singapore Pte Ltd., 2020, pp. 1–36. doi: 10.1007/978-981-15-2517-9_1.
- [22] O. O. Palit *et al.*, "Akurasi Metode Mesin Pembelajaran dalam Analisis Variabel Penting Faktor Risiko Sindrom Down," *Indones. J. Comput. Sci.*, vol. 13, no. 5, Oct. 2024, doi: 10.33022/ijcs.v13i5.4354.
- [23] A. Nautiyal and A. K. Mishra, "Drill Bit Selection and Drilling Parameter Optimization using Machine Learning," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1261, no. 1, p. 012027, Dec. 2023, doi: 10.1088/1755-1315/1261/1/012027.
- [24] J. Li, Q. Zhu, Q. Wu, and Z. Fan, "A novel oversampling technique for class-imbalanced learning based on SMOTE and natural neighbors," *Inf. Sci. (Ny.)*, vol. 565, pp. 438–455, Jul. 2021, doi: 10.1016/j.ins.2021.03.041.
- [25] M. H. Jamal, N. Naz, M. A. K. Khattak, F. Saeed, S. N. Altamimi, and S. N. Qasem, "A Comparison of Re-Sampling Techniques for Detection of Multi-Step Attacks on Deep Learning Models," *IEEE Access*, vol. 11, pp. 127446–127457, 2023, doi: 10.1109/ACCESS.2023.3332512.
- [26] T. S. Albalawi, "Heart-SecureCloud: A Secure Cloud-Based Hybrid DL System for Diagnosis of Heart Disease Through Transformer-Recurrent Neural Network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 9, 2024, doi: 10.14569/IJACSA.2024.0150916.
- [27] R. Hayami, Soni, and I. Gunawan, "Klasifikasi Jamur Menggunakan Algoritma Naïve Bayes," *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 3, no. 1, pp. 28–33, May 2022, doi: 10.37859/coscitech.v3i1.3685.
- [28] M. R. Maulana, A. Sucipto, and H. M. Mulyo, "OPTIMISASI PARAMETER SUPPORT VECTOR MACHINE DENGAN PARTICLE SWARM OPTIMIZATION UNTUK PENINGKATAN KLASIFIKASI DIABETES," *J. Inform. Teknol. dan Sains*, vol. 6, no. 4, pp. 802–812, Nov. 2024, doi: 10.51401/jinteks.v6i4.4784.
- [29] Y. K. Bintang and H. Imaduddin, "PENGEMBANGAN MODEL DEEP LEARNING UNTUK DETEKSI RETINOPATI DIABETIK MENGGUNAKAN METODE TRANSFER LEARNING," *JIPi (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 9, no. 3, pp. 1442–1455, Aug. 2024, doi: 10.29100/jipi.v9i3.5588.
- [30] I. D. Mienye and Y. Sun, "Effective Feature Selection for Improved Prediction of Heart Disease," 2022,

- pp. 94–107. doi: 10.1007/978-3-030-93314-2_6.
- [31] N. L. Fitriyani, M. Syafrudin, G. Alfian, C. Yang, J. Rhee, and S. M. Ulyah, “Chronic Disease Prediction Model Using Integration of DBSCAN, SMOTE-ENN, and Random Forest,” in *2022 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETISIS)*, IEEE, Jun. 2022, pp. 289–294. doi: 10.1109/ICETISIS55481.2022.9888806.
- [32] A. U. Nandhini and K. Dharmarajan, “Random forest and genetic algorithm united with hyperparameter for diabetes prediction by using WBSMOTE, wrapper approach,” *Int. J. Syst. Syst. Eng.*, vol. 13, no. 2, pp. 207–227, 2023, doi: 10.1504/IJSSE.2023.131226.