

Prediksi Risiko Penyakit Jantung Berdasarkan Kebiasaan Hidup Dengan Menggunakan Model *Machine Learning*

Ahmad Maulana¹

¹Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri

E-mail: [*1molanaahmad011@gmail.com](mailto:1molanaahmad011@gmail.com)

Abstrak – Penyakit jantung merupakan salah satu penyebab utama kematian di dunia dan banyak dipengaruhi oleh kebiasaan hidup seseorang. Oleh karena itu, diperlukan upaya prediksi risiko penyakit jantung secara dini untuk mendukung pencegahan terkena penyakit jantung. Penelitian ini bertujuan untuk memprediksi risiko penyakit jantung berdasarkan kebiasaan hidup menggunakan pendekatan machine learning. Dataset yang digunakan berasal dari Kaggle dan memuat data kesehatan individu, seperti usia, jenis kelamin, indeks massa tubuh, tekanan darah, kebiasaan merokok, konsumsi alkohol, serta tingkat aktivitas fisik. Tahapan penelitian meliputi pra-pengolahan data, penanganan outlier menggunakan metode Interquartile Range (IQR), pembagian data latih dan data uji, serta pelatihan model Random Forest dengan optimasi hyperparameter menggunakan Grid Search. Evaluasi model dilakukan menggunakan metrik akurasi, confusion matrix, dan classification report. Hasil penelitian menunjukkan bahwa model Random Forest mampu mencapai tingkat akurasi sebesar 73% dalam memprediksi risiko penyakit jantung. Hasil ini menunjukkan bahwa pendekatan berbasis kebiasaan hidup cukup efektif untuk digunakan sebagai alat skrining awal dalam mengidentifikasi risiko penyakit jantung, sehingga dapat membantu meningkatkan kesadaran dan upaya pencegahan dini berbasis data.

Kata Kunci — Kebiasaan Hidup, Machine Learning, Penyakit Jantung, Random Forest

1. PENDAHULUAN

Kesehatan jantung memiliki peranan penting dalam menjaga kualitas hidup seseorang, terutama karena penyakit kardiovaskular terus menjadi ancaman kesehatan global [1]. Laporan WHO menunjukkan bahwa pada tahun 2022, penyakit kardiovaskular menyebabkan sekitar 19,8 juta kematian atau sekitar 32% dari seluruh kematian dunia, dengan sebagian besar kasus berkaitan dengan serangan jantung dan stroke [2]. Ketika fungsi jantung tidak bekerja secara optimal, berbagai aspek kehidupan seperti kemampuan beraktivitas fisik, produktivitas [3], dan kesejahteraan psikologis akan ikut terdampak [4]. Oleh karena itu, mengenali serta memprediksi risiko penyakit jantung menjadi langkah awal yang penting untuk memahami kondisi kesehatan seseorang secara lebih komprehensif. Faktor-faktor seperti usia, riwayat penyakit keluarga, serta pola hidup tetap menjadi penentu utama yang berkontribusi terhadap meningkatnya risiko penyakit jantung [5].

Penyakit jantung umumnya disebabkan oleh kombinasi faktor risiko yang berkaitan dengan gaya hidup [6] dan kondisi kesehatan seseorang. Faktor seperti kebiasaan merokok [7], konsumsi alkohol berlebihan [8], kurangnya aktivitas fisik [9], pola makan tinggi lemak jenuh serta stres terbukti berkontribusi terhadap meningkatnya kemungkinan terjadinya penyakit jantung [10]. Faktor gaya hidup berperan hingga lebih dari 70% terhadap total risiko penyakit kardiovaskular global [11]. Perubahan gaya hidup yang sederhana, seperti peningkatan aktivitas fisik dan pengendalian berat badan, mampu menurunkan risiko hingga hingga 30% [12]. Berdasarkan fakta ini, penelitian mengenai prediksi risiko penyakit jantung berbasis kebiasaan hidup dengan pendekatan *machine learning* menjadi relevan untuk dilakukan, karena dapat membantu dalam mengidentifikasi individu berisiko tinggi secara dini dan memberikan rekomendasi pencegahan yang lebih personal serta efisien [13].

Berbagai penelitian sebelumnya menunjukkan bahwa metode *machine learning* memiliki potensi besar dalam meningkatkan ketepatan prediksi risiko penyakit jantung. Algoritma *Random Forest* dilaporkan mampu mencapai tingkat akurasi yang tinggi dan memberikan hasil klasifikasi yang stabil dibandingkan metode lainnya [14]. Penelitian berikutnya juga menegaskan keunggulan *Random Forest* dibandingkan *KNN*, dengan tingkat akurasi mencapai 98–99% dalam memberikan rekomendasi gaya hidup pencegahan [15]. Selain itu, algoritma *Linear Discriminant Analysis* (LDA) terbukti efektif dalam mengklasifikasikan penyakit jantung dengan akurasi sebesar 81,22% menggunakan dataset UCI [16]. Pendekatan lain melalui Carta Prediksi Faktor Risiko WHO (SEAR B) menunjukkan bahwa sebagian besar responden berada pada kategori risiko rendah, dengan usia, tekanan darah, kadar kolesterol, dan status merokok tetap menjadi indikator utama dalam estimasi risiko penyakit jantung [17].

Berdasarkan berbagai temuan tersebut, dapat disimpulkan bahwa pemanfaatan algoritma *machine learning*, khususnya *Random Forest* dan metode klasifikasi lainnya, memiliki potensi yang kuat dalam memprediksi risiko penyakit jantung secara akurat. Namun, perbedaan karakteristik data, variabel gaya hidup yang digunakan, serta konteks populasi penelitian menunjukkan perlunya pengembangan dan evaluasi model prediksi yang lebih spesifik dan kontekstual. Oleh karena itu, penelitian ini difokuskan pada penerapan pendekatan *machine learning* untuk memprediksi risiko penyakit jantung berdasarkan kebiasaan hidup, dengan harapan dapat memberikan kontribusi empiris dalam mendukung upaya pencegahan penyakit kardiovaskular secara lebih efektif dan berbasis data.

2. METODE PENELITIAN

2.1 Dataset

Penelitian ini memanfaatkan *dataset* penyakit kardiovaskular yang tersimpan dalam berkas *cardio_train.csv*. *Dataset* tersebut berisi informasi kesehatan individu yang digunakan dalam analisis risiko penyakit jantung, meliputi usia, jenis kelamin, tinggi badan, berat badan, tekanan darah sistolik (*ap_hi*), tekanan darah diastolik (*ap_lo*), kadar glukosa, kebiasaan merokok, konsumsi alkohol, tingkat aktivitas fisik, serta variabel target berupa status penyakit kardiovaskular (*cardio*).

Pada tahap awal, atribut *id* dihapus karena tidak memiliki pengaruh terhadap proses klasifikasi. Selain itu, dilakukan rekayasa fitur dengan menambahkan indeks massa tubuh (*Body Mass Index / BMI*) sebagai indikator kondisi berat badan individu. *BMI* dihitung menggunakan Persamaan 1:

$$BMI = \frac{weight}{height/100^2} \dots \dots \dots (1)$$

Dengan *weight* menyatakan berat badan dalam kilogram dan *height* menyatakan tinggi badan dalam sentimeter. Selain itu, variabel usia yang awal dinyatakan dalam satuan hari dikonversi menjadi tahun untuk meningkatkan keterbacaan dan interpretasi data.

2.2 Pra-Pengolahan Data

2.2.1 Pemeriksaan dan penanganan nilai hilang

Pemeriksaan nilai hilang dilakukan untuk memastikan kelengkapan data sebelum proses pemodelan. Tahap ini bertujuan untuk mengidentifikasi adanya nilai kosong (*missing values*) yang dapat memengaruhi kinerja model. Hasil pemeriksaan menunjukkan bahwa seluruh variabel pada *dataset* tidak mengandung nilai hilang, sehingga tidak diperlukan proses imputasi maupun penghapusan data. Dengan demikian, seluruh data dapat digunakan secara langsung pada tahap analisis berikutnya.

2.2.2 Deteksi *Outlier*

Deteksi *outlier* dilakukan untuk mengidentifikasi nilai ekstrem yang menyimpang dari pola umum distribusi data. Nilai ekstrem dapat menyebabkan bias pada proses pelatihan model dan menurunkan performa klasifikasi. Deteksi *outlier* dilakukan pada atribut numerik utama, yaitu tinggi badan, berat badan, tekanan darah sistolik, tekanan darah diastolik, kadar glukosa, kebiasaan merokok, konsumsi alkohol, aktivitas fisik, serta BMI. Proses ini dilakukan secara visual menggunakan *boxplot* untuk mengamati sebaran data dan keberadaan nilai yang berada di luar rentang normal.

2.2.3 Penanganan *Outlier* dengan Metode IQR (*Interquartile Range*)

Penanganan *outlier* pada penelitian ini menggunakan metode *Interquartile Range* (IQR). Metode ini menghitung rentang antar kuartil untuk menentukan batas bawah dan batas atas nilai yang dianggap normal. IQR dihitung menggunakan Persamaan 2:

Nilai batas bawah dan batas atas *outlier* ditentukan menggunakan Persamaan 3 dan Persamaan 4:

Nilai yang berada di luar batas tersebut tidak dihapus, melainkan dikoreksi menggunakan metode capping, yaitu menggantinya dengan nilai batas terdekat. Pendekatan ini dipilih untuk mempertahankan jumlah data tetap utuh sekaligus mengurangi pengaruh nilai ekstrem terhadap proses pembelajaran model.

2.3 Pembagian Data

Dataset yang telah melalui tahap pra-pengolahan data kemudian dibagi menjadi data latih dan data uji. Pembagian data dilakukan dengan 80% sebagai data latih dan 20% sebagai data uji menggunakan metode *train-test split*. Proses pembagian ini dilakukan secara stratifikasi berdasarkan variable target (*cardio*) agar proporsi antara data latih dan data uji tetap seimbang. Selain itu *random state* ditetapkan untuk memastikan hasil pembagian data konsisten dan dapat diulang.

2.4 Hyperparameter Tuning

Model klasifikasi yang digunakan dalam penelitian ini adalah *Random Forest*. Untuk memperoleh kinerja model yang optimal, dilakukan proses *hyperparameter tuning* menggunakan metode *Grid Search Cross-Validation*. Beberapa parameter yang dioptimasi meliputi jumlah pohon (*n_estimators*), kedalaman maksimum pohon (*max_depth*), serta jumlah minimum sampel untuk pemisahan node (*min_samples_split*).

Random Forest bekerja dengan membangun sejumlah pohon keputusan dan menentukan kelas akhir berdasarkan mekanisme *majority voting*. Secara konseptual, proses prediksi *Random Forest* dapat dinyatakan pada Persamaan 5:

Dengan $h_n(x)$ menyatakan hasil prediksi dari masing-masing pohon keputusan. Proses pencarian parameter terbaik dilakukan menggunakan validasi silang sebanyak tiga lipatan (*3-fold cross-validation*) dengan metrik evaluasi berupa akurasi.

2.5 Evaluasi Model

2.5.1 Akurasi

Akurasi digunakan sebagai metrik utama untuk mengevaluasi kinerja model klasifikasi. Matriks ini mengukur proporsi prediksi yang benar terhadap seluruh data uji. Perhitungan akurasi dinyatakan menggunakan persamaan 6 :

Dengan *TP* menyatakan *true positive*, *TN* menyatakan *true negative*, *FP* menyatakan *false positive*, dan *FN* menyatakan *false negative*.

2.5.2 Confusion Matrix

Confusion Matrix digunakan untuk menggambarkan performa model secara lebih rinci dengan membandingkan hasil prediksi dan label aktual. Matriks ini menunjukkan distribusi prediksi benar dan salah pada masing-masing kelas, sehingga dapat dianalisis jenis kesalahan klasifikasi yang terjadi. Matriks ini terdiri dari empat komponen utama, yaitu *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN), yang masing-masing menunjukkan jumlah prediksi benar dan salah pada setiap kelas.

2.5.3 Classification Report

Classification report digunakan untuk melengkapi evaluasi model dengan menampilkan nilai *precision*, *recall*, dan *F1-score* pada setiap kelas. Evaluasi ini memberikan gambaran yang lebih komprehensif mengenai kemampuan model dalam mengklasifikasikan individu dengan dan tanpa risiko penyakit kardiovaskular, khususnya pada kondisi distribusi kelas yang tidak seimbang.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Dataset yang digunakan dalam penelitian ini bersumber dari Kaggle berupa data penyakit kardiovaskular yang disimpan dalam berkas format CSV, yaitu *cardio_train.csv*. *Dataset* ini berisi data Kesehatan yang digunakan untuk memprediksi risiko penyakit jantung berdasarkan kebiasaan hidup.

id	age	gender	height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	cardio	
0	0	18393	2	168	62.0	110	80	1	1	0	0	1	0
1	1	20228	1	156	85.0	140	90	3	1	0	0	1	1
2	2	18857	1	165	64.0	130	70	3	1	0	0	0	1
3	3	17623	2	169	82.0	150	100	1	1	0	0	1	1
4	4	17474	1	156	56.0	100	60	1	1	0	0	0	0

Gambar 1. Dataset *cardio_train.csv*

Berdasarkan Gambar 1, *dataset* memiliki beberapa atribut yang menggambarkan karakteristik individu, seperti usia, jenis kelamin, tinggi badan, berat badan, tekanan darah, kadar kolesterol, kadar glukosa, serta variabel target *cardio* yang menunjukkan status penyakit kardiovaskular.

3.2 Prapengolahan Data

3.2.1 Pengecekan Nilai Hilang

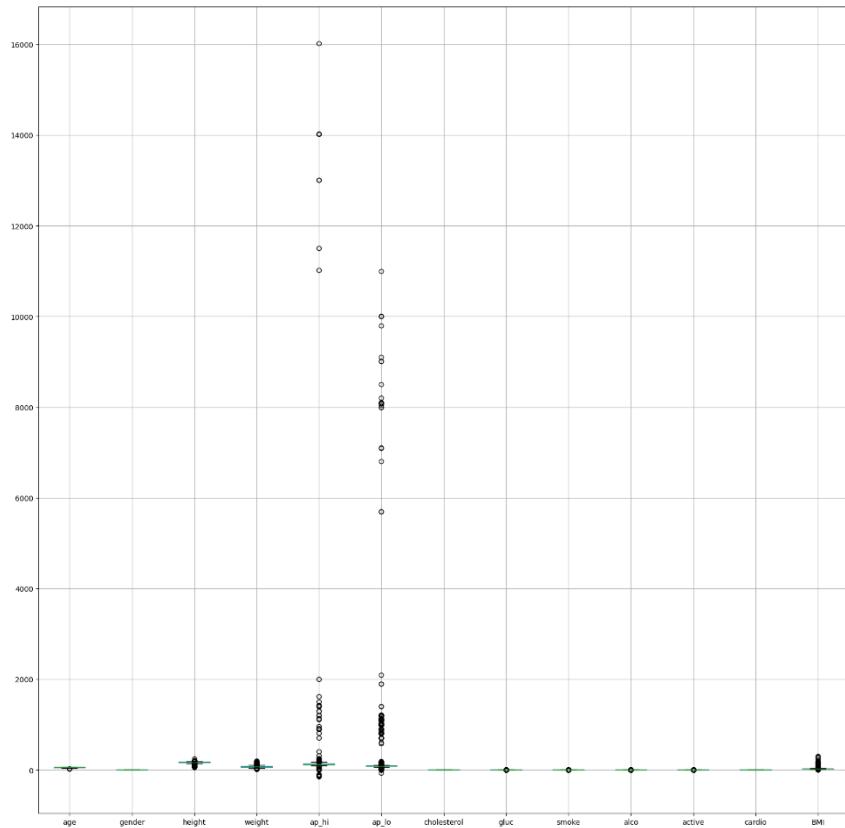
Tabel 1. Cek Nilai Hilang

Variabel	Jumlah Nilai Hilang
<i>age</i>	0
<i>gender</i>	0
<i>height</i>	0
<i>weight</i>	0
<i>ap_hi</i>	0
<i>ap_lo</i>	0
<i>cholesterol</i>	0
<i>gluc</i>	0
<i>smoke</i>	0
<i>alco</i>	0
<i>active</i>	0
<i>BMI</i>	0
<i>cardio</i>	0

Tabel 1, menggambarkan pencarian nilai yang hilang dalam proses prapengolahan data.

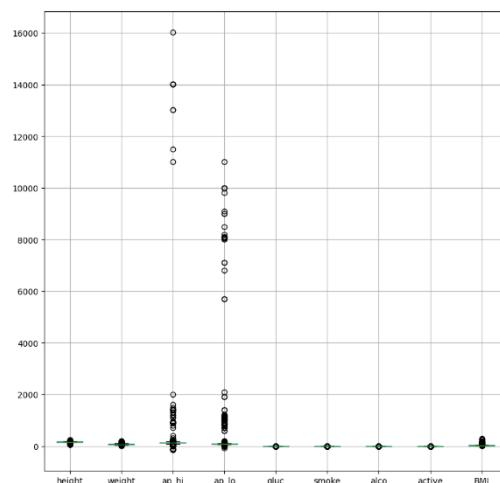
3.2.2 Deteksi *Outlier*

Untuk mendekripsi *Outlier* dilakukan visualisasi menggunakan *boxplot* pada variabel numerik.



Gambar 2. *Boxplot* Data Mendekripsi *Outlier*

Gambar 2, menampilkan sebaran *outlier* yang ditampilkan menggunakan *Boxplot*, gambar tersebut menampilkan keseluruhan variabel yang ada *outlier* (nilai ekstrem). Variabel-variabel tersebut kemudian dipilih dan dibersihkan supaya data lebih realistik dan bersih.

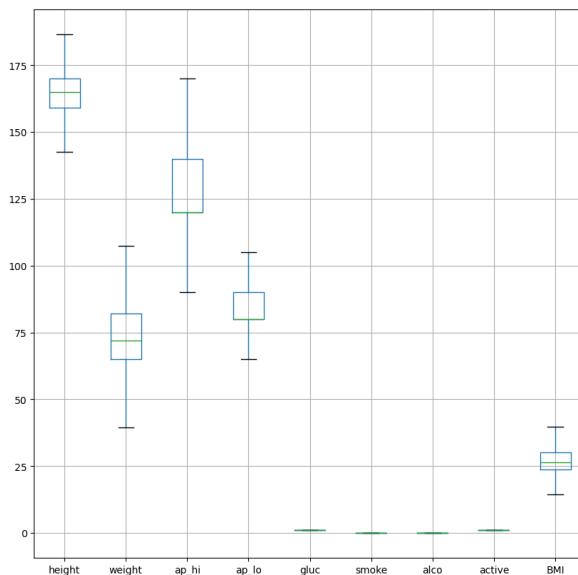


Gambar 3. *Boxplot* Menampilkan Fitur yang Memiliki *Outlier*

Gambar 3 menampilkan variable yang hanya terdapat *outlier* di dalamnya. Variabel tersebut dipilih yang ada *outlier*-nya saja karena mempermudah visualisasi untuk membersihkan *outlier* tersebut.

3.2.3 Penanganan *Outlier* Metode IQR

Penanganan *outlier* dilakukan menggunakan metode *Interquartile Range* (IQR) dengan pendekatan *capping*. Metode ini tidak menghapus data, tetapi menyesuaikan nilai ekstrem agar berada dalam batas bawah dan batas atas yang ditentukan.



Gambar 4. Boxplot Setelah *Outlier* Dibersihkan

Gambar 4 menggambarkan variabel yang terdapat *outlier* sudah dibersihkan dan ditangani. Setelah dilakukan penanganan *outlier*, distribusi data menjadi lebih terkendali dan nilai ekstrem berhasil diminimalkan tanpa mengurangi jumlah data.

Tabel 2. Ringkasan Penanganan *Outlier* dengan Metode IQR

Variable	Q1	Q3	IQR	Batas Bawah	Batas Atas	Jumlah <i>Outlier</i>
height	159	170	11	142,5	186,5	519
weight	65	82	17	39,5	107,5	1.819
ap_hi	120	140	20	90	170	1.435
ap_lo	80	90	10	65	105	4.632
gluc	1	1	0	1	1	10.521
smoke	0	0	0	0	0	6.169
alco	0	0	0	0	0	3.764
active	1	1	0	1	1	13.739
BMI	23,88	30,22	6,35	14,35	39,74	1.995

Tabel 2 menunjukkan batas bawah dan batas atas hasil perhitungan metode IQR yang digunakan sebagai acuan dalam proses *capping* pada setiap variabel.

3.3 Pembagian Data

Dataset yang telah melalui tahap prapengolahan selanjutnya dibagi menjadi data latih dan data uji dengan perbandingan 80% dan 20%. Pembagian dilakukan menggunakan metode *train-test split*.

Tabel 3. Pembagian Data Latih dan Data Uji

Jenis Data	Jumlah Data	Percentase
Data Latih	56.000	80%
Data Uji	14.000	20%
Total	70.000	100%

3.4 Hyperparameter Tuning

Tabel 4. Parameter untuk *Random Forest*

Parameter	Nilai
Jumlah Pohon (<i>n_estimators</i>)	200
Kedalaman Maksimum (<i>max_depth</i>)	15
Minimum Sample untuk Split (<i>min_samples_split</i>)	2
Bobot Kelas (<i>class_weight</i>)	Balanced
Random state	42

Proses pencarian parameter terbaik dilakukan menggunakan metode *Grid Search* dengan validasi silang sebanyak tiga lipatan (*3-fold cross validation*). Parameter yang diuji meliputi jumlah pohon, kedalaman maksimum pohon, dan jumlah minimum sampel untuk pemisahan node. Berdasarkan hasil pengujian, diperoleh kombinasi parameter terbaik sebagaimana ditunjukkan pada Tabel 4, yang selanjutnya digunakan dalam proses pelatihan model *Random Forest*.

3.5 Evaluasi

3.5.1 Akurasi model

Tabel 5. Nilai Akurasi Model *Random Forest*

Model	Jumlah Data Uji	Akurasi
<i>Random Forest</i>	14.000	0,73

Model *Random Forest* menghasilkan nilai akurasi sebesar 73% pada data uji. Hasil ini menunjukkan bahwa model mampu mengklasifikasikan risiko penyakit jantung dengan cukup baik. Model ini tidak menggunakan data medis lanjutan seperti hasil elektrokardiogram, pemeriksaan laboratorium klinis lengkap, maupun riwayat medis mendalam, yang umumnya dapat meningkatkan tingkat akurasi prediksi.

Dengan keterbatasan tersebut, capaian akurasi yang diperoleh dapat dikatakan cukup sebagai alat prediksi risiko awal (*early risk screening*). Model ini mampu memberikan gambaran awal mengenai potensi risiko penyakit jantung berdasarkan pola gaya hidup dan indikator kesehatan umum, sehingga dapat dimanfaatkan sebagai sistem untuk meningkatkan kesadaran dan pencegahan dini.

3.5.2 *Confusion Matrix*

Tabel 6. *Confusion Matrix* Hasil Klasifikasi

Aktual/Prediksi	Tidak Berisiko(0)	Berisiko(1)
Tidak Berisiko (0)	5.425	1.579
Berisiko (1)	2.200	4.796

Berdasarkan hasil evaluasi menggunakan *confusion matrix* pada tabel 6, diperoleh bahwa model mampu mengklasifikasikan sebanyak 5.425 data sebagai kelas tidak berisiko secara benar (*true negative*). Sementara itu, terdapat 1.579 data tidak berisiko yang diprediksi sebagai berisiko (*false positive*). Untuk kelas berisiko, model berhasil mengidentifikasi 4.796 data secara benar (*true positive*), sedangkan 2.200 data berisiko masih salah diklasifikasikan sebagai tidak berisiko (*false negative*).

3.5.3 Classification Report

Tabel 7. Hasil Classification Report Model Random Forest

Kelas	Precision	Recall	F1-Score	Support
Tidak Berisiko (0)	0,71	0,77	0,74	7.004
Berisiko(1)	0,75	0,69	0,72	6.996
Akurasi			0,73	14.000

Nilai *precision* dan *recall* pada kedua kelas menunjukkan bahwa model memiliki kemampuan yang relatif seimbang dalam mendeteksi individu berisiko dan tidak berisiko. Nilai *F1-score* yang mendekati nilai akurasi menunjukkan performa model yang stabil.

4. SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan algoritma *Random Forest* mampu memprediksi risiko penyakit jantung berdasarkan kebiasaan hidup dan indikator kesehatan umum dengan tingkat akurasi sebesar 73%. Hasil ini menunjukkan bahwa pendekatan berbasis gaya hidup cukup efektif untuk digunakan sebagai skrining awal dalam mengidentifikasi potensi risiko penyakit jantung. Meskipun demikian, keterbatasan data yang digunakan, yang belum mencakup informasi medis klinis secara mendalam, menyebabkan hasil prediksi belum sepenuhnya optimal. Oleh karena itu, untuk memperoleh prediksi risiko penyakit jantung yang lebih akurat dan andal, diperlukan penggunaan data medis yang lebih lengkap dan terintegrasi.

5. SARAN

Penelitian berikutnya dapat dikembangkan dengan menambahkan dan memperdalam variabel kebiasaan hidup agar model prediksi mampu merepresentasikan tingkat risiko penyakit jantung dengan lebih baik. Selain itu, penggabungan data gaya hidup dengan data medis sebagai informasi pendukung dapat dipertimbangkan untuk meningkatkan akurasi hasil prediksi, tanpa mengubah peran utama model sebagai sarana skrining awal berbasis kebiasaan hidup.

DAFTAR PUSTAKA

- [1] N. A. Khasanah, N. M. A. Wulansari, and N. K. Azhari, “Hubungan Health Anxiety dengan Kualitas Hidup pada Pasien dengan Gangguan Kardiovaskuler di SMC RS Telogorejo Semarang,” *J. Ris. Ilmu Kesehat. Umum dan Farm.*, vol. 3, no. 1, pp. 155–165, 2025.
- [2] World Health Organization, “Cardiovascular diseases (CVDs).” Accessed: Dec. 10, 2025. [Online]. Available: [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))
- [3] A. C. Puspita, S. T. Y. Rahmawati, and S. Suparmi, “Aktivitas Fisik Rutin dalam Produktivitas Harian pada Dewasa Akhir,” *Jenggala J. Ris. Pengemb. Dan Pelayanan Kesehat.*, vol. 3, no. 2, pp. 23–29, 2024.
- [4] N. F. Arifudin and B. Kristinawati, “Dampak Masalah Psikologis Terhadap Kualitas Hidup Pasien Gagal Jantung: Systematic Review,” *Heal. Inf. J. Penelit.*, 2023.
- [5] H. W. Dhany, A. I. Permana, F. Izhari, A. P. Ginting, and Z. G. Pratama, “Aplikasi prediksi serangan jantung untuk warga Kelurahan Pelawi Utara,” *J. Minfo Polgan*, vol. 13, no. 2, pp. 2096–2101, 2024.
- [6] I. S. Nasution, A. D. Rahmadani, W. Audina, D. P. Sari, and N. D. Sari, “Systematic Review: Pengaruh Gaya Hidup dan Pengetahuan Masyarakat terhadap Risiko Penyakit Jantung Koroner,” *Sehat Rakyat J. Kesehat. Masy.*, vol. 4, no. 2, pp. 287–298, 2025.
- [7] S. Syarli and L. Arini, “Faktor Penyebab Hipertensi Pada Lansia: Literatur Review: Faktor Penyebab Hipertensi Pada Lansia: Literatur Review,” *Ahmar metastasis Heal. J.*, vol. 1, no. 3, pp. 112–117, 2021.
- [8] S. M. Dewi, B. Saputra, and M. Daniati, “The Relationship Of Alcohol Consumption And Sleep Quality To The Event Of Hypertension: Hubungan Konsumsi Alkohol Dan Kualitas Tidur Terhadap Kejadian Hipertensi,” *J. Keperawatan Hang Tuah (Hang Tuah Nurs. Journal)*, vol. 2, no. 1, pp. 49–62, 2022.
- [9] I. Pashar and L. Wendikbo, “Hubungan Pola Makan Dan Aktivitas Fisik Dengan Penyakit Jantung Koroner Di Rsud Labuang Baji Makassar,” *J. Pubnursing Sci.*, vol. 2, no. 01, pp. 31–42, 2024.

- [10] D. Triyono, R. Liani, A. W. Utami, S. Tristiyanti, and A. Supriatna, “Penyakit Jantung Koroner Di Indonesia: Peran Faktor Risiko Dan Upaya Pencegahan,” *HUMANIS J. Ilmu-Ilmu Sos. dan Hum.*, vol. 17, no. 1, pp. 86–94, 2025.
- [11] W. Handayani, “Faktor-Faktor Risiko Penyakit Kardiovaskular: Artikel Review,” *Pengemb. Ilmu dan Prakt. Kesehat.*, vol. 4, no. 3, pp. 139–158, 2025.
- [12] G. K. Ghodeshwar, A. Dube, D. Khobragade, and A. H. Dube, “Impact of lifestyle modifications on cardiovascular health: a narrative review,” *Cureus*, vol. 15, no. 7, 2023.
- [13] W. Fauziah *et al.*, “Pemeriksaan Kesehatan Jantung Dalam Rangka Pengembangan Aplikasi Screening Jantung Berbasis Android,” *Budimas J. Pengabdi. Masy.*, vol. 6, no. 3, 2024.
- [14] A. Y. Agusyul, “Prediksi Penyakit Jantung Menggunakan Algoritma Random Forest,” *J. Minfo Polgan*, vol. 12, no. 2, 2023.
- [15] E. Sahelvi, P. Cikita, R. M. Sapitri, R. Rahmaddeni, and L. Efrizoni, “Perbandingan Algoritma K-Nearest Neighbors dan Random Forest untuk Rekomendasi Gaya Hidup Sehat dalam Mencegah Penyakit Jantung: Comparison of K-Nearest Neighbors and Random Forest Algorithms for Recommendations for a Healthy Lifestyle in Prevent Heart Disease,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. 3, pp. 830–840, 2025.
- [16] W. Wardianto, “Prediksi Risiko Penyakit Jantung Menggunakan Algoritma Linear Discriminant Analysis,” *J. Ilm. Sist. Inf. dan Tek. Inform.*, vol. 7, no. 2, pp. 282–293, 2024.
- [17] B. R. Santoso, E. E. M. Gaghauna, and I. Akbar, “Prediksi kejadian penyakit jantung dan pembuluh darah di UPT Puskesmas Rawat Inap Alabio,” *J. Persat. Perawat Nas. Indones.*, vol. 8, no. 1, pp. 1–11, 2023.