



Comparison of K-Nearest Neighbors and Random Forest Algorithms for Recommendations for a Healthy Lifestyle in Prevent Heart Disease

Perbandingan Algoritma K-Nearest Neighbors dan Random Forest untuk Rekomendasi Gaya Hidup Sehat dalam Mencegah Penyakit Jantung

**Elza Sahelvi^{1*}, Putri Cikita², Riska Mela Sapitri³,
Rahmaddeni⁴, Lusiana Efrizoni⁵**

^{1,2,3,4,5}Program Studi Teknik Informatika, Universitas Sains dan Teknologi Indonesia, Indonesia

E-Mail: ¹2210031802016@sar.ac.id, ²2210031802020@sar.ac.id,
³2210031802095@sar.ac.id, ⁴rahmaddeni@usti.ac.id, ⁵lusiana@stmik-amik-riau.ac.id

Received Feb 12th 2025; Revised May 21th 2025; Accepted May 30th 2025; Available Online Jun 23th 2025, Published Jun 23th 2025

Corresponding Author: Elza Sahelvi

Copyright © 2025 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

Heart disease is a leading cause of mortality, primarily due to unhealthy lifestyle factors. This study compares the performance of K-Nearest Neighbors (KNN) and Random Forest (RF) algorithms in providing recommendations for a healthy lifestyle to prevent heart disease. A dataset comprising 1,025 entries with 14 features underwent preprocessing steps, including normalization, feature selection, and data splitting at ratios of 80:20 and 70:30. Model evaluations were conducted using accuracy, precision, recall, and F1-score metrics. The RF algorithm achieved higher accuracy (99% for the 80:20 split and 98% for the 70:30 split) compared to KNN (83% and 86%, respectively), demonstrating greater stability in classifying heart disease risk. Feature analysis identified chest pain type (CP) as the most influential factor. Based on these findings, it is recommended to adopt a healthy diet, engage in regular physical activity, manage stress effectively, and undergo routine health check-ups. In conclusion, the RF algorithm proved more effective in delivering healthy lifestyle recommendations. Future research may consider utilizing larger datasets with additional variables to enhance prediction accuracy.

Keywords: Classification, Healthy Lifestyle Recommendations, Heart Disease, K-Nearest Neighbors, Random Forest

Abstrak

Penyakit jantung merupakan salah satu penyebab utama kematian yang disebabkan oleh faktor gaya hidup tidak sehat. Untuk mengatasi permasalahan ini, penelitian ini membandingkan algoritma *K-Nearest Neighbors* (KNN) dan *Random Forest* (RF) dalam memberikan rekomendasi gaya hidup sehat guna mencegah penyakit jantung. Dataset yang digunakan terdiri dari 1.025 entri dengan 14 fitur, yang telah melalui tahap *preprocessing*, termasuk normalisasi, seleksi fitur, dan pembagian data 80:20 serta 70:30. Evaluasi model dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa *Random Forest* memiliki akurasi lebih tinggi (99% pada skenario 80:20 dan 98% pada skenario 70:30) dibandingkan KNN (83% dan 86%), serta lebih stabil dalam mengklasifikasikan risiko penyakit jantung. Analisis fitur menunjukkan bahwa *Chest Pain Type* (CP) atau nyeri dada merupakan faktor paling berpengaruh. Berdasarkan hasil ini, direkomendasikan pola makan sehat, aktivitas fisik teratur, manajemen stres, serta pemeriksaan kesehatan rutin. Kesimpulannya, *Random Forest* lebih efektif dalam sistem rekomendasi gaya hidup sehat, dan penelitian selanjutnya dapat menggunakan dataset lebih besar dengan variabel tambahan guna meningkatkan akurasi prediksi.

Kata Kunci: Klasifikasi, K-Nearest Neighbors, Penyakit Jantung, Random Forest, Rekomendasi Gaya Hidup Sehat

1. PENDAHULUAN

Penyakit jantung adalah salah satu penyebab utama kematian di dunia dan telah menjadi tantangan besar dalam sistem kesehatan global. Berdasarkan laporan dari Organisasi Kesehatan Dunia (WHO), lebih dari 17,9 juta kematian per tahun disebabkan oleh penyakit *kardiovaskular*, yang mencakup 31% dari total kematian global [1]. Tingginya angka kematian akibat penyakit jantung yang menjadi alasan penelitian ini dilakukan [2].

Penyakit kronis, seperti *diabetes*, *hipertensi*, dan penyakit jantung, menjadi beban kesehatan masyarakat yang signifikan. Program pencegahan menjadi krusial dalam mengurangi prevalensi dan dampak penyakit ini [3]. Penyakit jantung sering kali disebabkan oleh faktor gaya hidup yang tidak sehat, seperti pola makan yang buruk, kurangnya aktivitas fisik, konsumsi alkohol yang berlebihan, merokok, serta stres yang tinggi. Penyakit jantung adalah kondisi yang melibatkan gangguan pada organ jantung, yang dapat menunjukkan beberapa gejala yang perlu diwaspadai. Gejala-gejala tersebut mencakup rasa mual, muntah, keringat dingin, mudah lelah, sakit kepala, nyeri dada sebelah kiri, sesak napas, lemas, jantung berdebar, dan sensasi dada seperti diremas-remas [4][5][6].

Penelitian ini sejalan dengan Tujuan Pembangunan Berkelanjutan (*Sustainable Development Goals/SDGs*), khususnya SDG 3: "*Good Health and Well-being*", yang bertujuan untuk memastikan kehidupan yang sehat dan meningkatkan kesejahteraan bagi semua orang di segala usia. Penyakit jantung merupakan salah satu penyebab utama kematian di dunia, dan gaya hidup sehat menjadi faktor kunci dalam upaya pencegahannya. Dengan membandingkan algoritma K-Nearest Neighbor (KNN) dan Random Forest (RF) dalam memberikan rekomendasi gaya hidup sehat, penelitian ini berkontribusi pada pengembangan teknologi berbasis kecerdasan buatan yang dapat membantu individu dalam mengadopsi kebiasaan hidup yang lebih sehat. Implementasi sistem rekomendasi ini diharapkan dapat meningkatkan kesadaran masyarakat akan pentingnya pencegahan penyakit jantung serta mendukung pencapaian target SDGs dalam mengurangi angka kematian akibat penyakit tidak menular melalui pendekatan berbasis data dan teknologi.

Gaya hidup sehat adalah sebuah pilihan gaya hidup jangka panjang dengan pola hidup yang baik secara komitmen untuk menjalankan beberapa hal guna mendukung fungsi tubuh, sehingga memberikan efek kesehatan bagi tubuh. Terdapat berbagai upaya yang dapat dilakukan untuk menerapkan gaya hidup sehat dan aktif adalah dengan bersikap menjaga pola asupan makanan sehat dengan diet dan nutrisi, berolahraga secara rutin, memilih nutrisi pendukung yang tepat dan mengikuti komunitas guna mendapat dukungan dari orang yang sama [7]. Terdapat berbagai upaya yang dapat dilakukan untuk menerapkan gaya hidup sehat dan aktif adalah dengan bersikap menjaga pola asupan makanan sehat dengan diet dan nutrisi, berolahraga secara rutin, memilih nutrisi pendukung yang tepat [8]. Gaya hidup sehat meliputi berbagai aspek seperti aktivitas fisik teratur, pola makan seimbang, tidak merokok, dan menjaga berat badan ideal, yang secara signifikan dapat menurunkan risiko penyakit jantung karena itu, sistem rekomendasi yang tepat sangat penting untuk membantu individu memilih dan menerapkan kebiasaan hidup sehat secara personal.

Perkembangan teknologi yang kian pesat sangat membantu dunia kedokteran dalam menangani berbagai macam penyakit. Saat ini telah ada teknologi yang biasa disebut dengan kecerdasan buatan. Salah satu manfaat dari teknologi ini yaitu dapat memprediksi penyakit dengan tingkat akurasi tertentu [9]. Salah satu cara dari kemajuan teknologi yang memudahkan dalam kegiatan medis adalah metode klasifikasi. Klasifikasi merupakan bentuk metode atau teknik yang digunakan untuk mengumpulkan beberapa objek menjadi beberapa kelompok [2][10]. Algoritma seperti KNN dan RF telah menunjukkan potensi besar dalam berbagai aplikasi, termasuk klasifikasi data kesehatan dan sistem rekomendasi berbasis data.

Algoritma KNN merupakan algoritma klasifikasi yang sederhana namun efektif dalam pengolahan data. Algoritma ini dapat digunakan untuk mengklasifikasikan data berdasarkan pola dan kemiripan dengan data [10]. Algoritma KNN dibangun dengan menggunakan beberapa nilai k yang umum digunakan, kemudian dipilih nilai k yang menghasilkan performa terbaik. Ukuran jarak yang digunakan adalah *Euclidean Distance* dan *Manhattan Distance*, kemudian dibandingkan dan diambil serta dilihat model yang akan memberikan performa terbaik dalam kasus data penyakit jantung [11]. Lalu Pada algoritma random forest dimana di dalam algoritma ini terdiri dari algoritma *decision tree* yang dikenal sebagai pohon prediksi dimana setiap pohon keputusan bergantung kepada nilai *random vector* dan di jadikan sebagai sampel secara bebas pada semua pohon (tree) dalam forest tersebut. Random Forest juga memiliki keunggulan dalam menangani data dengan dimensi tinggi dan menangkap interaksi antara variable [12].

Penelitian sebelumnya yang telah membahas implementasi KNN dan RF di berbagai bidang kesehatan, seperti deteksi penyakit, analisis pola makan, dan rekomendasi diet. Sebagai contoh, penelitian sebelumnya yang di lakukan oleh Hamid Rahman, dkk. Menunjukkan bahwa KNN efektif dalam mengklasifikasikan data pasien untuk mendeteksi risiko penyakit tertentu, terutama pada dataset yang relatif kecil [13]. Dan ada juga penelitian yang di lakukan oleh Fitriana Sholekhah, dkk. Menunjukan bahwa KNN juga efektif dalam mengklasifikasi data untuk sindrom metabolik [14]. Sementara itu, penelitian lain mengungkapkan bahwa RF mampu memprediksi faktor risiko penyakit dengan akurasi yang lebih tinggi dibandingkan metode tradisional, terutama pada dataset dengan banyak variabel independen yang saling berinteraksi. Selain itu, penerapan algoritma ini dalam rekomendasi gaya hidup sehat juga mulai mendapatkan perhatian, terutama dalam membantu individu mengadopsi kebiasaan yang dapat menurunkan risiko penyakit kronis [15]. Dan ada juga penelitian sebelumnya yang menggunakan metode RF untuk melakukan deteksi pasien gagal jantung dan menghasilkan tingkat akurasi hingga 91,45% [16][17].

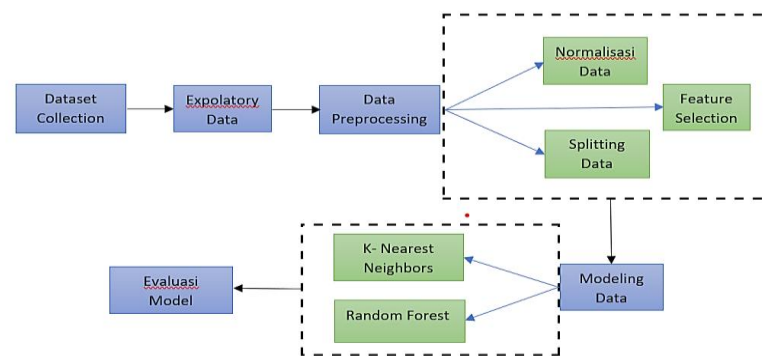
Penelitian ini memiliki signifikansi tersendiri dibandingkan dengan penelitian-penelitian terdahulu, karena secara khusus melakukan komparasi antara algoritma KNN dan RF dalam sistem rekomendasi gaya hidup sehat untuk pencegahan penyakit jantung. Sebagian besar studi sebelumnya cenderung hanya

mengaplikasikan satu algoritma tertentu tanpa membandingkan performa secara menyeluruh dalam konteks sistem rekomendasi berbasis machine learning untuk kesehatan preventif. Penelitian ini tidak hanya mengevaluasi tingkat akurasi, tetapi juga mempertimbangkan efisiensi dari masing-masing algoritma dalam menganalisis faktor-faktor risiko penyakit jantung. Dengan demikian, penelitian ini memberikan kontribusi yang berbeda dan lebih komprehensif, baik secara teoritis dalam pengembangan literatur terkait penerapan machine learning di bidang kesehatan, maupun secara praktis dalam mendukung pemilihan algoritma yang lebih tepat guna untuk pengembangan sistem rekomendasi digital yang efektif, akurat, dan aplikatif dalam meningkatkan kualitas hidup serta mencegah penyakit jantung. Tidak seperti penelitian sebelumnya yang hanya fokus pada deteksi penyakit, studi ini secara khusus membandingkan dua algoritma machine learning dalam konteks sistem rekomendasi berbasis gaya hidup, yang bertujuan mendukung tindakan preventif berbasis data. Hal ini menjadikan penelitian ini unik dalam pendekatan serta kontribusinya terhadap pencegahan penyakit jantung.

Meskipun telah banyak studi yang membahas KNN dan RF secara individual, perbandingan langsung antara kedua algoritma ini dalam konteks rekomendasi gaya hidup sehat untuk mencegah penyakit jantung masih jarang dilakukan. Padahal, penelitian komparatif semacam ini penting untuk mengidentifikasi algoritma yang lebih efektif dan efisien dalam mendukung pencegahan penyakit *kardiovaskular*. Oleh karena itu, penelitian ini bertujuan untuk menganalisis kinerja kedua algoritma dalam memberikan rekomendasi berdasarkan data kesehatan pengguna, serta mengidentifikasi keunggulan dan kelemahan masing-masing algoritma. Evaluasi dilakukan menggunakan metrik-metrik seperti akurasi, presisi, *recall*, dan *F1-score*.

2. METODE PENELITIAN

Proses penelitian dan metode diilustrasikan dengan bagan yang ditunjukkan pada Gambar 1.



Gambar 1. Methodologi Penelitian

2.1. Dataset Collection

Pada penelitian ini, data yang digunakan diperoleh dari dataset penyakit jantung yang tersedia secara publik yang berasal dari kaggle. Dataset ini berisi berbagai parameter kesehatan yang berhubungan dengan risiko penyakit jantung, seperti usia, jenis kelamin, tekanan darah, kadar kolesterol, serta beberapa variabel lainnya yang relevan.

Proses pengumpulan data dilakukan dengan memastikan bahwa data yang digunakan memiliki kelengkapan dan kualitas yang memadai untuk analisis lebih lanjut. Sumber data dipilih berdasarkan ketersediaan informasi yang akurat dan telah digunakan dalam penelitian sebelumnya. Setiap atribut dalam dataset memiliki peran penting dalam membantu model machine learning memahami pola yang terkait dengan kondisi kesehatan pengguna. Tabel 1 merupakan fitur dari dataset penyakit jantung.

Tabel 1. Fitur dalam Dataset

No.	Fitur	Deskripsi
1.	Age	Usia pasien (dalam tahun).
2.	Sex	Jenis kelamin pasien (1 = Laki-laki, 0 = Perempuan).
3.	Chest Pain Type (cp)	Jenis nyeri dada (1-4, dengan kategori yang berbeda).
4.	Resting Blood Pressure (restbps)	Tekanan darah saat istirahat (mm Hg).
5.	Serum Cholesterol (chol)	Kadar kolesterol dalam darah (mg/dl).
6.	Fasting Blood Sugar (fbs)	Gula darah puasa (> 120 mg/dl, 1 = benar, 0 = salah).
7.	Resting ECG (restecg)	Hasil elektrokardiografi saat istirahat (0-2).
8.	Max Heart Rate (thalach)	Denyut jantung maksimum yang dicapai.
9.	Exercise Induced Angina (exang)	Nyeri dada akibat olahraga (1 = Ya, 0 = Tidak).
10.	Oldpeak	Depresi segmen ST yang diinduksi oleh olahraga relatif terhadap istirahat.

No.	Fitur	Deskripsi
11.	Slope of ST Segment (slope)	Kemiringan segmen ST latihan (0-2).
12.	Number of Major Vessels (ca)	Jumlah pembuluh darah utama yang diwarnai dengan fluoroskopi (0-3).
13.	Thalassemia (thal)	Jenis kelainan darah thalassemia (0-3).
14.	Target	Diagnosis penyakit jantung (1 = Mengidap penyakit jantung, 0 = Tidak mengidap penyakit jantung).

2.2. Exploratory Data

Exploratory Data Analysis (EDA) adalah suatu pendekatan awal yang penting dalam analisis data yang bertujuan untuk menggali dan memahami struktur serta karakteristik data tanpa membuat asumsi sebelumnya. Melalui proses EDA (*Exploratory Data Analysis*), peneliti mengeksplorasi data untuk mengidentifikasi pola, menemukan nilai-nilai yang berbeda dari biasanya (*outlier*), menguji hipotesis, dan memverifikasi asumsi yang mendasari analisis data. Tujuan utama dari EDA adalah untuk mendapatkan wawasan yang berharga dari data yang tersedia, sehingga memungkinkan pengguna untuk membuat keputusan yang lebih tepat dan akurat dalam analisis statistik [18].

2.3. Data Preprocessing

Pada tahapan preprosesing data ini digunakan untuk melakukan penyesuaian data terhadap algoritma ataupun metode yang akan digunakan, pada penelitian tahap preprocessing akan dilakukan dalam tiga cara yaitu normalisasi data, *feature selection* dan *splitting data*[12]. Tujuan dari melakukan preprocessing data adalah untuk *machine learning* dapat mengolah data dengan lancar. Dengan melakukan preprocessing data maka dapat menyaring dan memilah data dengan tujuan analisis pada prediksi penyakit jantung [19].

1. Data Normalization

Normalisasi Data merupakan proses dilakukan transformasi sebuah atribut numerik yang di skalakan ke dalam sebuah bentuk lebih sederhana seperti 0 sampai 1. Data yang telah dilakukan proses normalisasi akan berkisar antara 0 sampai dengan 1. Dengan begitu, metode klasifikasi dapat membacanya dengan range yang sama pada setiap atributnya [20].

2. Feature Selection

Langkah selanjutnya ada Feature selection (seleksi fitur) bertujuan untuk memilih feature yang berpengaruh dan mengesampingkan feature yang tidak berpengaruh dalam suatu kegiatan pemodelan atau penganalisaan data. Menggunakan teknik seleksi fitur mengurangi jumlah fitur digunakan untuk belajar dan memilih fitur diskriminasi tinggi dalam proses seleksi fitur. Selain itu, seleksi fitur membantu meningkatkan akurasi dengan memilih fitur yang optimal [21].

3. Splitting Data

Langkah selanjutnya adalah membagi data menjadi data latih dan data uji. Dengan pemisahan data 70:30 dan 80:20. Data yang belum pernah digunakan dalam suatu penelitian, tetapi juga berguna untuk mengevaluasi keberhasilan atau kegagalan suatu penelitian, disebut data pengujian, sedangkan data pelatihan adalah data yang digunakan untuk melakukan penelitian [21].

2.4. Modeling Data

Pada tahap pemodelan data, penelitian ini membandingkan performa algoritma K-Nearest Neighbors (KNN) dan Random Forest (RF) dalam memberikan rekomendasi gaya hidup sehat untuk mencegah penyakit jantung. Pemilihan kedua algoritma ini didasarkan pada keunggulannya dalam tugas klasifikasi, terutama dalam analisis data kesehatan.

2.4.1. K-Nearest Neighbors (K-NN)

K-NN adalah teknik untuk mengklasifikasikan dataset yang telah diklasifikasi sebelumnya. Keakuratan algoritma K-NN sangat dipengaruhi oleh fitur yang berbeda ketika nilainya tidak sesuai dengan nilai perkiraan. Beberapa studi yang menggunakan algoritma K-NN menangani hampir secara eksklusif pemilihan fitur dan pembobotan untuk meningkatkan efisiensi algoritma dalam klasifikasi [22]. Algoritma K-NN memiliki keuntungan dalam menghasilkan data yang kuat atau tidak ambigu dan efektif bila digunakan pada data yang cukup besar [23]. Tujuan dari algoritma K-NN adalah untuk mengklasifikasikan objek baru berdasarkan atribut dan data pelatihannya. K-NN menggunakan klasifikasi pada tetangga-tetangga sebagai nilai prediksi dari *query instance* yang baru [24]. Proses metode K-NN dilakukan dengan mencari kelompok k objek pada data latih yang paling dekat dengan objek pada data baru atau data uji [25].

Langkah pertama dalam proses ini adalah menentukan parameter K, yang merupakan jumlah tetangga terdekat yang akan dipertimbangkan dalam pengambilan keputusan. Pemilihan nilai K yang tepat sangat penting, karena dapat mempengaruhi akurasi model. Nilai K yang kecil dapat membuat model terlalu sensitif terhadap noise, sedangkan nilai K yang terlalu besar dapat mengakibatkan model yang terlalu umum dan kehilangan detail penting. Setelah menentukan nilai K, langkah berikutnya adalah menghitung jarak antara

data training dan data testing. Jarak yang paling umum digunakan dalam K-NN adalah jarak Euclidean. Persamaan Euclidean Distance ditunjukkan pada persamaan 1.

$$euc = \sqrt{(\sum_{i=1}^n (p_i - q_i)^2)} \quad (1)$$

di mana p_i adalah nilai dari data training, q_i adalah nilai dari data testing, dan n adalah dimensi data. Proses ini melibatkan pengukuran seberapa jauh setiap data testing dari semua data training, sehingga kita dapat mengidentifikasi tetangga terdekat.

Setelah jarak dihitung, langkah selanjutnya adalah mengurutkan jarak yang terbentuk dari yang terdekat hingga yang terjauh. Pengurutan ini memungkinkan kita untuk dengan mudah mengidentifikasi K tetangga terdekat. Setelah itu, kita menentukan K tetangga terdekat berdasarkan urutan jarak yang telah dihitung. Setiap tetangga terdekat ini memiliki kelas yang bersesuaian, yang merupakan label atau kategori dari data training.

Langkah terakhir adalah menghitung jumlah kelas dari tetangga terdekat tersebut. Dengan menganalisis kelas-kelas yang dimiliki oleh K tetangga terdekat, kita dapat menetapkan kelas yang paling sering muncul sebagai kelas untuk data testing yang sedang dievaluasi. Proses ini memungkinkan kita untuk mengklasifikasikan data baru berdasarkan pola yang telah dipelajari dari data training, sehingga K-NN menjadi metode yang sederhana namun efektif dalam pengklasifikasian data [26].

2.4.2. Random Forest (RF)

RF adalah salah satu algoritma klasifikasi dan regresi yang populer dalam data mining dan *machine learning*. Random forest menggabungkan konsep dari metode *ensemble* dan *decision tree* [27]. Random forest adalah suatu algoritma yang digunakan pada klasifikasi data jumlah besar. Klasifikasi Random Forest dilakukan melalui penggabungan pohon dengan melakukan training yang dilakukan membangkitkan pohon klasifikasi dengan banyak versi yang kemudian mengkombinasikannya untuk memperoleh prediksi akhir, maka dalam Random Forest proses pengacakan untuk membentuk pohon klasifikasi tidak hanya dilakukan untuk data sampel saja melainkan juga pada pengambilan variabel prediktor [28].

2.5. Evaluasi Model

Evaluasi model adalah proses untuk menilai kinerja suatu algoritma dalam melakukan prediksi atau klasifikasi berdasarkan data yang diberikan. Dalam penelitian ini, evaluasi model bertujuan untuk mengukur efektivitas algoritma yang digunakan dalam memberikan rekomendasi gaya hidup sehat guna mencegah penyakit jantung. Proses evaluasi dilakukan dengan menggunakan berbagai metrik penilaian, seperti akurasi, presisi, recall, F1-score, dan waktu komputasi. Metrik-metrik ini digunakan untuk memahami sejauh mana model dapat menghasilkan prediksi yang benar serta bagaimana model menangani ketidak seimbangan data dan variasi karakteristik dalam dataset.

Untuk memahami performa model klasifikasi kami menggunakan *Confusion Matrix*, Table 2 merupakan tabel yang menggambarkan hasil prediksi model dibandingkan dengan nilai sebenarnya.

Table 2. *Confusion Matrix*

Actual/ Predicted	Positive (1)	Negative (0)
Positive (1)	True Positive (TP)	False Negative (FN)
Negative (0)	False Positive (FP)	True Negative (TN)

Dari *Confusion Matrix* ini, beberapa metrik evaluasi dapat dihitung dengan persamaan 2 – 5.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

Evaluasi model sangat penting dalam machine learning karena membantu menentukan apakah suatu algoritma bekerja dengan baik, apakah perlu dilakukan penyesuaian parameter, atau apakah model yang berbeda lebih cocok untuk dataset yang digunakan.

3. HASIL DAN PEMBAHASAN

3.1. Dataset Collection

Dataset yang digunakan dalam penelitian ini merupakan dataset penyakit jantung yang terdiri dari 1025 entri dengan 14 fitur, termasuk variabel target yang menunjukkan apakah seseorang mengidap penyakit jantung (1) atau tidak (0). Dataset ini telah melalui tahap eksplorasi awal untuk memastikan kualitas data. Data bisa dilihat pada Table 3.

Table 3. Dataset Penyakit Jantung

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	52	1	0	125	212	0	1	168	0	1	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3	61	1	0	148	203	0	1	161	0	0	2	1	3	0
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0
...
1020	59	1	1	140	221	0	1	164	1	0	2	0	2	1
1021	60	1	0	125	258	0	0	141	1	2.8	1	1	3	0
1022	47	1	0	110	275	0	0	118	1	1	1	1	2	0
1023	50	0	0	110	254	0	0	159	0	0	2	0	2	1
1024	54	1	0	120	188	0	1	113	0	1.4	1	1	3	0

Distribusi data target menunjukkan bahwa terdapat 526 individu dengan penyakit jantung dan 499 individu yang tidak mengidap penyakit jantung. Data ini cukup seimbang untuk digunakan dalam model klasifikasi.

3.2. Preprocessing

Tahapan preprocessing data sangat penting dalam memastikan bahwa data yang digunakan dalam pelatihan model berada dalam kondisi optimal. Preprocessing bertujuan untuk meningkatkan kualitas data, menghilangkan noise, serta memastikan bahwa data memiliki format yang sesuai agar algoritma dapat bekerja secara efisien.

1. Data Normalization

Data Normalization (Normalisasi data) dilakukan untuk memastikan bahwa setiap fitur dalam dataset memiliki skala yang seragam sehingga dapat meningkatkan performa model dalam proses klasifikasi. Teknik yang digunakan dalam penelitian ini adalah *Min-Max Scaling*, di mana nilai setiap fitur dikonversi ke rentang antara 0 hingga 1. Normalisasi ini penting karena model berbasis jarak seperti KNN sangat bergantung pada skala fitur. Setelah proses normalisasi, distribusi data menjadi lebih seragam dan mengurangi kemungkinan fitur tertentu mendominasi prediksi model. Table 4 merupakan hasil normalisasi pada *dataset* :

Table 4. Hasil Normalisasi (bagian 1)

age	sex	cp	trestbps	chol	fbs	restecg
-0.268437	0.661504	-0.915755	-0.377636	-0.659332	-0.418878	0.891255
-0.158157	0.661504	-0.915755	0.479107	-0.833861	2.387330	-1.004049
1.716595	0.661504	-0.915755	0.764688	-1.396233	-0.418878	0.891255
0.724079	0.661504	-0.915755	0.936037	-0.833861	-0.418878	0.891255
0.834359	-1.511706	-0.915755	0.364875	0.930822	2.387330	0.891255

Tabel 4 menunjukkan hasil normalisasi untuk fitur-fitur demografis dan hasil pemeriksaan awal pasien.

Tabel 5. Hasil Normalisasi (bagian 2)

thalach	exang	oldpeak	slope	ca	thal	target
0.821321	-0.712287	-0.060888	0.995433	1.209221	1.089852	0
0.255968	1.403928	1.727137	-2.243675	-0.731971	1.089852	0
-1.048692	1.403928	1.301417	-2.243675	-0.731971	1.089852	0
0.516900	-0.712287	-0.912329	0.995433	0.238625	1.089852	0
-1.874977	-0.712287	0.705408	-0.624121	2.179817	-0.522122	0

Tabel 5 mencakup fitur yang berhubungan dengan aktivitas jantung, gejala, dan diagnosis akhir (target).

2. Feature Selection

Saya menggunakan Metode *SelectKBest* dengan *ANOVA F-test (f_classif)* untuk melakukan seleksi fitur. Metode ini digunakan untuk mengukur pengaruh masing-masing fitur terhadap variabel target berdasarkan skor *F-test*. Dengan pendekatan ini, saya dapat mengidentifikasi fitur yang paling berpengaruh serta mengeliminasi fitur yang kurang signifikan, sehingga model yang dihasilkan lebih optimal dan tidak terbebani oleh fitur yang tidak relevan. Table 6 merupakan hasil *Feature Selection*.

Table 6. Hasil *Feature Selection* (Fitur yang paling berpengaruh)

No	Feature	Score
9	oldpeak	208.0028
7	thalach	194.2581
2	cp	188.1611
8	exang	181.3255
11	ca	151.5469

Table 7. Hasil *Feature Selection* (Fitur yang paling tidak berpengaruh)

No	Feature	Score
0	age	49.20551
3	trestbps	15.70774
4	chol	14.66663
6	restecg	12.21142
5	fbs	1.86193

3. Splitting Data

Data dibagi menjadi dua bagian, yaitu data latih (training set) dan data uji (testing set), dengan dua skenario pembagian yaitu, 80:20 (80% data latih, 20% data uji) dan 70:30 (70% data latih, 30% data uji). Tujuan dari pembagian ini adalah untuk memastikan model memiliki cukup data untuk dilatih sekaligus menyisakan data yang cukup untuk pengujian guna mengukur performa model secara akurat. Pemisahan data dilakukan secara acak untuk menghindari bias dalam pelatihan model. Dengan tahapan preprocessing yang baik, model dapat memperoleh performa yang lebih optimal dalam mempelajari pola dari dataset, serta dapat menghasilkan prediksi yang lebih akurat dalam mendeteksi risiko penyakit jantung.

3.3. Modeling Data

Pada tahap pemodelan data, dilakukan eksperimen menggunakan Algoritma KNN dan RF dengan dua skenario pembagian data, yaitu 80:20 dan 70:30. Evaluasi dilakukan dengan menggunakan metrik akurasi, presisi, *recall*, dan *F1-score* untuk menilai kinerja masing-masing model.

1. K-Nearest Neighbors (K-NN)

Table 3. Hasil *Splitting Data* 80:20 KNN

	Precision	Recall	F1-Score	Support
0	0.88	0.77	0.82	102
1	0.80	0.89	0.84	103
Accuracy	-	-	0.83	205
Macro avg	0.84	0.83	0.83	205
Weighted	0.84	0.83	0.83	205

Table 4. Hasil *Splitting Data* 70:30 KNN

	Precision	Recall	F1-Score	Support
0	0.87	0.85	0.86	159
1	0.84	0.87	0.85	149
Accuracy	-	-	0.86	308
Macro avg	0.86	0.86	0.86	308
Weighted	0.86	0.86	0.86	308

Dari hasil Table 3 dan Table 4, dapat dilihat bahwa akurasi model KNN pada skenario 80:20 adalah 83%, sedangkan pada skenario 70:30 meningkat menjadi 86%. Peningkatan akurasi ini dapat dijelaskan oleh jumlah data latih yang lebih besar pada skenario 70:30, yang memungkinkan model untuk mempelajari lebih banyak pola dari dataset sebelum diuji. Selain itu, nilai presisi, *recall*, dan *F1-score*

juga mengalami peningkatan pada skenario 70:30, yang menunjukkan bahwa model menjadi lebih baik dalam membedakan antara individu yang berisiko terkena penyakit jantung dan yang tidak.

2. Random Forest

Table 5. Hasil *Splitting Data 80:20 Random Forest*

	Precision	Recall	F1-Score	Support
0	0.97	1.00	0.99	102
1	1.00	0.97	0.99	103
Accuracy	-	-	0.99	205
Macro avg	0.99	0.99	0.99	205
Weighted	0.99	0.99	0.99	205

Table 6. Hasil *Splitting Data 70:30 Random Forest*

	Precision	Recall	F1-Score	Support
0	0.96	1.00	0.98	159
1	1.00	0.96	0.98	149
Accuracy	-	-	0.98	308
Macro avg	0.98	0.98	0.98	308
Weighted	0.98	0.98	0.98	308

Dari hasil table 5 dan table 6 di atas, dapat dilihat bahwa RF memiliki performa yang jauh lebih unggul dibandingkan KNN, dengan akurasi mencapai 99% pada skenario 80:20 dan 98% pada skenario 70:30. Meskipun terjadi sedikit penurunan akurasi pada skenario 70:30, nilai presisi, *recall*, dan *F1-score* tetap sangat tinggi, menunjukkan bahwa model tetap mampu melakukan klasifikasi dengan sangat baik. Hal ini menunjukkan bahwa RF memiliki kemampuan generalisasi yang lebih baik dibandingkan KNN.

3.4. Analisis Evaluasi Model

Table 7 menunjukkan bahwa RF memiliki performa lebih unggul dibandingkan KNN dalam semua metrik evaluasi. Pada skenario 80:20, Random Forest mencapai akurasi 99%, sedangkan KNN hanya 83%. Begitu juga pada skenario 70:30, dimana RF memperoleh 98%, lebih tinggi dibandingkan 86% pada KNN.

Table 7. Perbandingan KNN dan RF

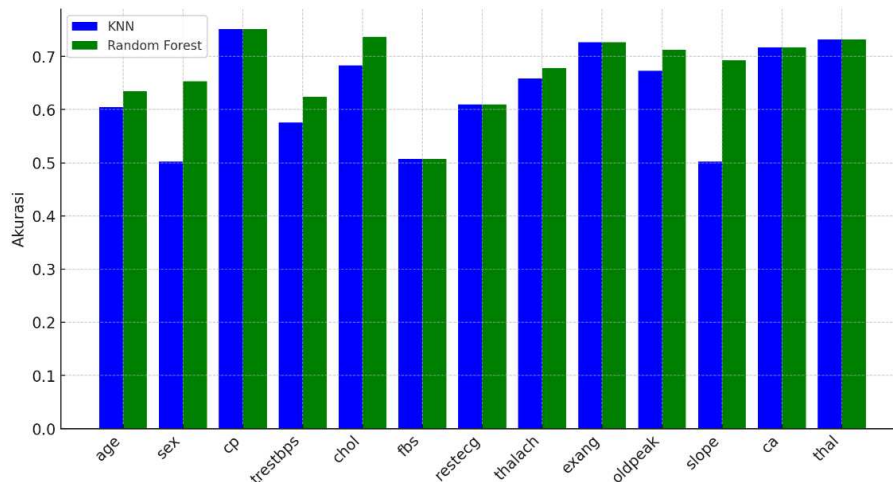
		KNN				Random Forest					
		Precision	Recall	F1 Score	Support			Precision	Recall	F1 Score	Support
80:20	0	0.88	0.77	0.82	102	0	0.97	1.00	0.99	102	
	1	0.80	0.89	0.84	103	1	1.00	0.97	0.99	103	
	Accuracy	-	-	0.83	205	Accuracy	-	-	0.99	205	
	Macro avg	0.84	0.83	0.83	205	Macro avg	0.99	0.99	0.99	205	
	Weighted	0.84	0.83	0.83	205	Weighted	0.99	0.99	0.99	205	
70:30	0	0.87	0.85	0.86	159	0	0.96	1.00	0.98	159	
	1	0.84	0.87	0.85	149	1	1.00	0.96	0.98	149	
	Accuracy	-	-	0.86	308	Accuracy	-	-	0.98	308	
	Macro avg	0.86	0.86	0.86	308	Macro avg	0.98	0.98	0.98	308	
	Weighted	0.86	0.86	0.86	308	Weighted	0.98	0.98	0.98	308	

Keunggulan RF terletak pada kemampuannya dalam menangani data berdimensi tinggi dan menangkap hubungan kompleks antar fitur, sehingga lebih stabil dalam prediksi. Sementara itu, KNN lebih sensitif terhadap skala data dan bergantung pada pemilihan parameter K. Berdasarkan hasil evaluasi ini, RF lebih direkomendasikan untuk sistem rekomendasi gaya hidup sehat karena akurasinya yang lebih tinggi dan stabilitasnya dalam klasifikasi risiko penyakit jantung.

Untuk memvisualisasikan perbedaan performa antara KNN dan RF, hasil evaluasi juga ditampilkan dalam bentuk diagram batang. Diagram batang menunjukkan bahwa Random Forest secara konsisten memiliki akurasi, presisi, *recall*, dan *F1-score* yang lebih tinggi dibandingkan KNN dalam semua skenario pembagian data.

Dari hasil akurasi pada gambar diagram batang di atas, dapat dilihat bahwa *Chest Pain Type* (CP) atau nyeri dada merupakan faktor penyebab tertinggi seseorang berisiko terkena penyakit jantung. Dengan hasil uji ini, kami merekomendasikan penerapan gaya hidup sehat guna mengurangi risiko terkena penyakit jantung. Salah satu langkah utama adalah menjaga pola makan sehat, seperti mengonsumsi makanan tinggi serat, vitamin, dan mineral, termasuk buah, sayuran, serta biji-bijian, yang dapat membantu menjaga kesehatan *kardiovaskular*. Sebaliknya, makanan tinggi lemak jenuh, garam, dan gula perlu dibatasi untuk mencegah peningkatan tekanan darah dan kadar kolesterol. Menggunakan minyak sehat seperti minyak zaitun serta

memilih sumber protein sehat seperti ikan dan kacang-kacangan lebih dianjurkan dibandingkan daging merah dan makanan olahan.



Gambar 5. Diagram Batang Perbandingan Akurasi KNN dan *Random Forest* untuk Setiap Atribut

Selain pola makan, aktivitas fisik secara teratur juga berperan penting dalam menjaga kesehatan jantung. Olahraga aerobik seperti berjalan kaki, berlari, atau bersepeda selama 30 menit per hari, 5 kali seminggu, dapat meningkatkan daya tahan jantung dan memperbaiki sirkulasi darah. Latihan kekuatan seperti angkat beban dan yoga juga membantu menjaga elastisitas pembuluh darah. Mengurangi gaya hidup *sedentary*, seperti duduk terlalu lama, dengan rutin bergerak atau berjalan setiap satu jam, juga disarankan. Selain itu, manajemen stres yang baik sangat diperlukan karena stres yang berlebihan dapat meningkatkan tekanan darah dan risiko penyakit jantung. Teknik relaksasi seperti meditasi, pernapasan dalam, serta tidur yang cukup (7–9 jam per malam) dapat membantu menjaga stabilitas fungsi jantung dan keseimbangan emosional.

Menghindari kebiasaan buruk, seperti merokok dan konsumsi alkohol berlebihan, juga menjadi faktor penting dalam pencegahan. Kandungan zat berbahaya dalam rokok dan alkohol dapat merusak pembuluh darah serta meningkatkan risiko *hipertensi* dan gagal jantung. Oleh karena itu, berhenti merokok dan membatasi konsumsi alkohol sangat disarankan untuk menjaga kesehatan jantung. Selain itu, pemeriksaan kesehatan secara rutin juga diperlukan, terutama bagi individu dengan riwayat keluarga penderita penyakit jantung. Pemantauan tekanan darah, kadar kolesterol, dan gula darah secara berkala dapat membantu mendeteksi risiko lebih awal serta memungkinkan intervensi medis yang lebih cepat dan tepat.

4. KESIMPULAN

Penelitian ini membandingkan Algoritma KNN dan RF dalam memberikan rekomendasi gaya hidup sehat untuk mencegah penyakit jantung. Hasil evaluasi menunjukkan bahwa RF memiliki performa yang lebih unggul, dengan akurasi mencapai 99% pada skenario 80:20 dan 98% pada skenario 70:30, dibandingkan KNN yang hanya mencapai 83% dan 86%. Selain itu, fitur CP teridentifikasi sebagai faktor paling berpengaruh terhadap risiko penyakit jantung, sehingga sistem rekomendasi berbasis data ini dapat menjadi alat bantu yang efektif dalam pencegahan. Meski hasilnya menjanjikan, penelitian ini masih memiliki keterbatasan, seperti ukuran dataset yang kecil dan variabel yang belum mencakup seluruh aspek gaya hidup. Selain itu, metode evaluasi belum menggunakan validasi silang yang dapat meningkatkan keandalan hasil. Untuk penelitian selanjutnya, disarankan menggunakan *dataset* yang lebih besar dan variabel tambahan seperti pola tidur, stres, serta riwayat keluarga, serta menerapkan teknik *cross-validation* dan membandingkan dengan algoritma lain agar sistem rekomendasi yang dihasilkan lebih akurat dan aplikatif.

UCAPAN TERIMA KASIH

Kami mengucapkan terima kasih kepada Universitas Sains dan Teknologi Indonesia atas dukungan fasilitas dan sumber daya dalam penelitian ini. Kami juga berterima kasih kepada seluruh rekan peneliti dan akademisi yang telah memberikan masukan serta bimbingan dalam penyusunan artikel ini. Dukungan dan motivasi dari keluarga serta teman-teman sangat berarti dalam menyelesaikan penelitian ini.

REFERENSI

- [1] WHO, "Cardiovascular diseases. Key Facts 2017.," pp. 1–6, 2019, [Online]. Available: [https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))
- [2] R. Maula and S. Bahri, "Penerapan Algoritma K-Nearest Neighbour pada Penyakit Jantung," vol. 13, no. 1, pp. 42–51, 2024, doi: 10.14421/fourier.2024.131.42-51.

- [3] M. Fahrurrozi *et al.*, “Evaluasi Efektivitas Program Pencegahan Penyakit Kronis : Fokus Pada Gaya Hidup Sehat Dan Pengelolaan penyakit kronis , masih terdapat kesenjangan pengetahuan mengenai efektivitas program-program memengaruhi gaya hidup sehat dan efektivitas pengelolaan st”.
- [4] Dewi Nasien, S. Sirvan, D. Deny, R. S. Ryan Syahputra, A. Akbar Marunduri, and R. Prawinata See, “Klasifikasi Penyakit Jantung Menggunakan Decision Tree dan KNN Menggunakan Ekstraksi Fitur PCA,” *JEKIN - J. Tek. Inform.*, vol. 4, no. 1, pp. 18–24, 2024, doi: 10.58794/jekin.v4i1.641.
- [5] B. P. dan P. Kesehatan, “Kementerian Kesehatan Republik Indonesia Direktur Kementerian Kesehatan Republik Indonesia,” *Kementeri. Kesehat. RI*, vol. 7, no. 35, p. 1, 2021, [Online]. Available: <https://www.kemkes.go.id/article/view/18052800006/ini-penyebab-stunting-pada-anak.html#:~:text=Stunting merupakan kondisi gagal pertumbuhan,dan memiliki keterlambatan dalam berpikir.%0Ahttps://www.kemkes.go.id/article/view/19093000001/penyakit-jantung-pen>
- [6] S. R. J. I. Alham, “Sistem Diagnosis Penyakit Jantung Koroner Dengan Menggunakan Algoritma C4.5 Berbasis Website (Studi Kasus: RSUD Dr. Soedarso Pontianak),” *Petir*, vol. 14, no. 2, pp. 214–222, 2021, doi: 10.33322/petir.v14i2.1338.
- [7] M. Scientifict and J. Volume, “Perbandingan Algoritma Decision Tree dan Deep Learning dalam Prediksi Masalah Kesehatan berdasarkan Kebiasaan Gaya Hidup Fadhillah Rashidatul A ' la , Zaehol Fatah Universitas Ibrahimy , Indonesia lifestyle habits ; health prediction ; decision tree ; dee,” vol. 2, no. 10, 2024.
- [8] P. P. Tiara and L. Lasnawati, “Makna Gaya Hidup Sehat Dalam Perpektif Teori Interaksionisme Simbolik,” *Humantech J. Ilm. Multidisiplin Indones.*, vol. 1, no. 11, pp. 1627–1638, 2022.
- [9] A. Putranto, N. L. Azizah, and A. I. Ratna Ika, “Sistem Prediksi Penyakit Jantung Berbasis Web Menggunakan Metode SVM dan Framework Streamlit,” *J. Penerapan Sist. Inf. (Komputer Manajemen)*, vol. 4, no. 2, pp. 442–452, 2023, [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/heart+disease>
- [10] M. F. Akbarollah, W. Wiyanto, D. Ardiatma, and A. T. Zy, “Penerapan Algoritma K-Nearest Neighbor Dalam Klasifikasi Penyakit Jantung,” *J. Comput. Syst. Informatics*, vol. 4, no. 4, pp. 850–860, 2023, doi: 10.47065/josyc.v4i4.4071.
- [11] A. Samosir, M. S. Hasibuan, W. E. Justino, and T. Hariyono, “Komparasi Algoritma Random Forest, Naïve Bayes dan K- Nearest Neighbor Dalam klasifikasi Data Penyakit Jantung,” *Pros. Semin. Nas. Darmajaya*, vol. 1, no. 0, pp. 214–222, 2021, [Online]. Available: <https://jurnal.darmajaya.ac.id/index.php/PSND/article/view/2955>
- [12] Z. Susanti, P. Sirait, and E. S. Panjaitan, “Peningkatan Kinerja Random Forest Melalui Seleksi Fitur Secara Pca Untuk Mendeteksi Penyakit Diabetes Tahap Awal,” *Sains dan Teknol.*, vol. 4, no. 3, pp. 51–56, 2023.
- [13] ”ع.ع. ا. احمد جاسم, “نظام تقويمي لمستوى القدرات الحركية لتلاميذ الصفوف (1، 2، 3) الابتدائي بطيئي التعلم and ح. فهمي سليمان *Sport. Cult.*, vol. 15, no. 1, pp. 72–86, 2024, doi: 10.25130/sc.24.1.6.
- [14] F. Sholekhah, A. D. Putri, R. Rahmadden, and L. Efrizoni, “Perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbors untuk Klasifikasi Metabolik Sindrom,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 507–514, 2024, doi: 10.57152/malcom.v4i2.1249.
- [15] N. Nurdiansyah, F. S. Febriyan, Z. Gesit, and D. Amanta, “Mental Health Analysis to Prevent Mental Disorders in Students Using The K-Nearest Neighbor (K-NN) Algorithm and Random Forest Algorithm Analisis Kesehatan Mental untuk Mencegah Gangguan Mental pada Mahasiswa Menggunakan Algoritma K-Nearest Neighbor (K,” vol. 5, no. January, pp. 1–9, 2025.
- [16] 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, 2023, doi: 10.47970/siskom-kb.v7i1.468.
- [17] Y. Yuliani, “Algoritma Random Forest Untuk Prediksi Kelangsungan Hidup Pasien Gagal Jantung Menggunakan Seleksi Fitur Bestfirst,” *Infotek J. Inform. dan Teknol.*, vol. 5, no. 2, pp. 298–306, 2022, doi: 10.29408/jit.v5i2.5896.
- [18] Farhanuddin, Sarah Ennola Karina Sihombing, and Yahfizham, “Komparasi Multiple Linear Regression dan Random Forest Regression Dalam Memprediksi Anggaran Biaya Manajemen Proyek Sistem Informasi,” *J. Comput. Digit. Bus.*, vol. 3, no. 2, pp. 86–97, 2024, doi: 10.56427/jcbd.v3i2.408.
- [19] J. K. Filemon and W. F. Senjaya, “Prediksi Penyakit Jantung Berdasarkan Indikator- Indikator Kesehatan,” vol. 6, no. November, pp. 272–283, 2024.
- [20] M. D. Purbolaksono, M. Irvan Tantowi, A. Imam Hidayat, and A. Adiwijaya, “Perbandingan Support Vector Machine dan Modified Balanced Random Forest dalam Deteksi Pasien Penyakit Diabetes,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 393–399, 2021, doi: 10.29207/resti.v5i2.3008.
- [21] D. Septhya *et al.*, “Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 15–19, 2023, doi: 10.57152/malcom.v3i1.591.

-
- [22] A. Putri *et al.*, “Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 20–26, 2023, doi: 10.57152/malcom.v3i1.610.
 - [23] S. Jasmine Putri, Q. Attaqwa, A. Pratama, and Rahmaddeni, “Klasifikasi Menentukan Jadwal Kerja Data Karyawan Menggunakan Algoritma C4.5 dan K-nearest Neighbor,” *SENTIMAS Semin. Nas. Penelit. dan Pengabd. Masy.*, pp. 215–221, 2022, [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas>
 - [24] R. Wahyudi, M. Orisa, and N. Vendyansyah, “Penerapan Algoritma K-Nearest Neighbors Pada Klasifikasi Penentuan Gizi Balita (Studi Kasus Di Posyandu Desa Bluto),” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 5, no. 2, pp. 750–757, 2021, doi: 10.36040/jati.v5i2.3738.
 - [25] F. T. Admojo and Ahsanawati, “Klasifikasi Aroma Alkohol Menggunakan Metode KNN,” *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 34–38, 2020, doi: 10.33096/ijodas.v1i2.12.
 - [26] F. Putra, H. F. Tahiyat, R. M. Ihsan, R. Rahmaddeni, and L. Efrizoni, “Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, pp. 273–281, 2024, doi: 10.57152/malcom.v4i1.1085.
 - [27] J. Schlenger, “Random Forest,” *Comput. Sci. Sport*, pp. 201–207, 2024, doi: 10.1007/978-3-662-68313-2_24.
 - [28] E. R. B. Sebayang, Y. H. Chrisnanto, and Melina, “Klasifikasi Data Kesehatan Mental di Industri Teknologi Menggunakan Algoritma Random Forest,” *IJESPG J.*, vol. 1, no. 3, pp. 237–253, 2023.