

# Analisis Kinerja dan Akurasi Algoritma K-Nearest Neighbor Pada Dataset Heart Disease

Zaehol Fatah<sup>1</sup>, Mohammad Hadi Sujatmiko<sup>2</sup>

<sup>1,2</sup> Prodi Teknologi Informasi, Fak. SAINTEK, Universitas Ibrahimy, Situbondo, Jawa Timur, Indonesia

Email : <sup>1</sup>zaeholfatah@gmail.com, <sup>2</sup>m.hadisujatmiko060607@gmail.com

## Abstrak

*Since heart disease is the world's largest cause of mortality, early identification is crucial. This research assesses how effectively the KNN algorithm classifies data on heart disease. With  $k = 3$  and an 80:20 split between training and test data, supervised learning was employed. The target function served as the classification label for the Heart Disease Dataset, which includes characteristics associated with heart disease. RapidMiner software was used for the analytical procedure, which included data splitting, normalization, application of the KNN model, and making use of a confusion matrix to assess the model's performance. According to the findings, the KNN algorithm produced a weighted mean recall of 95.00%, a weighted mean precision of 96.15%, a classification error of 4.55%, and an accuracy of 95.45%. These findings support the notion that the KNN algorithm is highly accurate and dependable in identifying possible cardiac problems.*

**Keywords:** heart disease, K-Nearest Neighbor, data mining, accuracy, classification.

## Abstrak

Penyakit jantung, penyebab kematian tertinggi di dunia, memerlukan diagnosis cepat. Tujuan dari studi ini adalah untuk mengevaluasi tingkat akurasi algoritma KNN dalam mengklasifikasikan data yang berkaitan dengan penyakit jantung. Metode pengajaran supervisi digunakan, dengan parameter  $k = 3$ , dan rasio pembagian data pengajaran dan pengujian sebesar 80:20. Dataset yang digunakan berasal dari Dataset Penyakit Jantung, yang mengandung atribut terkait diagnosis penyakit jantung dengan label klasifikasi target. Proses analisis dilakukan melalui perangkat lunak RapidMiner. Ini mencakup proses split data, normalisasi fase, penerapan model KNN, dan evaluasi kinerja model confusion matriks kecacauan. Menurut hasil penelitian, algoritma KNN memiliki nilai akurasi 95,45%. kesalahan klasifikasi sebesar 4,55%, keakuratan rata-rata rata-rata sebesar 95,00%, dan keakuratan rata-rata rata-rata sebesar 96,15%. Hasil ini menunjukkan algoritma KNN memiliki tingkat akurasi hampir sempurna dan cukup andal untuk deteksi potensi dari dataset yang digunakan.

**Kata Kunci:** klasifikasi, penyakit jantung, data mining, K-Nearest Neighbor, akurasi.

## PENDAHULUAN

Perkembangan teknologi informasi dalam beberapa dekade terakhir telah menghasilkan transformasi signifikan dalam sektor kesehatan (Fatah 2025). Pemanfaatan teknologi tersebut memungkinkan pengolahan data medis dilakukan dengan lebih cepat, akurat, dan sistematis sehingga dapat meningkatkan

kualitas proses pengambilan keputusan klinis. Penggabungan konsep big data, kecerdasan buatan, dan metode data mining turut memperkaya analisis dalam bidang kesehatan, khususnya dalam mendukung deteksi dini berbagai penyakit yang memiliki risiko tinggi terhadap keselamatan pasien. Seiring dengan meningkatnya volume data kesehatan,

penerapan analisis berbasis algoritma menjadi semakin penting untuk mengidentifikasi pola serta indikator penyakit secara lebih tepat dan terukur(Wu et al. 2021).

Kematian adalah takdir tuhan yang telah ditetapkan, penyakit jantung salah satunya, sehingga diagnosis penyakit jantung harus dideteksi dengan cepat(Putri, Ristyawan, and Muzaki 2022). Dengan jumlah data medis yang semakin meningkat, data mining telah menjadi salah satu metode yang berguna untuk membantu proses pengambilan keputusan medis(Sahelvi et al. 2025). K-Nearest Neighbor (KNN)[3] adalah algoritma yang paling banyak digunakan untuk mengklasifikasikan data medis. karena memiliki keunggulan dalam kesederhanaan dan efektivitas pada data non-linear(Rahmansyah et al. 2024).

Namun, efektivitas KNN bergantung pada pemilihan nilai k, jarak pengukuran, dan distribusi data. Beberapa penelitian sebelumnya, seperti oleh Amril et al. (2021) dan Putri et al. (2022)(Samosir et al. 2021), menunjukkan bahwa KNN dapat mencapai tingkat akurasi tinggi dalam klasifikasi penyakit jantung, namun hasilnya masih bervariasi tergantung pada konfigurasi data dan parameter yang digunakan(Fredilio et al. 2023). Berdasarkan fenomena tersebut, penelitian ini berfokus pada analisis akurasi algoritma KNN dalam mengklasifikasikan penyakit jantung menggunakan dataset Heart Disease Dataset dengan nilai k = 3 dan rasio pembagian data training dan testing sebesar 0.8:0.2.

enelitian ini bertujuan untuk mengkaji tingkat akurasi algoritma K-Nearest Neighbor (KNN) dalam melakukan klasifikasi pada dataset penyakit jantung. Selain itu, juga diarahkan untuk mengevaluasi kinerja model klasifikasi melalui pengukuran metrik precision, recall, dan error rate sebagai indikator objektif performa sistem. Lebih lanjut, bertujuan menilai efektivitas algoritma KNN sebagai pendekatan prediktif dalam mendukung penentuan diagnosis penyakit

jantung secara komputasional (Sistem et al. 2025).

## METODE

Algoritma KNN adalah teknik pembelajaran instance-based yang menentukan kelas data berdasarkan mayoritas label dari sejumlah k tetangga terdekat yang memiliki tingkat kemiripan tertinggi(Lestari 2014). KNN dinilai efektif dalam klasifikasi penyakit jantung karena mampu mengenali pola hubungan non-linear antar karakteristik klinis(Zahri 2025). Namun, kualitas data, proses normalisasi, dan pemilihan nilai k yang tepat sangat memengaruhi kinerja algoritma(Tukan, Weking, and Watomakin 2025). Oleh karena itu, untuk memastikan kemampuan KNN untuk membuat prediksi yang akurat dan dapat diandalkan, evaluasi kinerja model dilakukan dengan menggunakan metrik akurasi, ketepatan, recall, dan tingkat kesalahan.

Sehubungan regresi dan klasifikasi, algoritma KNN digunakan(Nikmatun et al. 2019). Dengan menggunakan fungsi jarak geometris seperti Euclidean Distance, KNN menentukan kelas data uji berdasarkan mayoritas kelas k tetangga terdekatnya.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Semakin kecil jarak, semakin besar peluang data tersebut memiliki kelas yang sama. Usia, jenis kelamin, tekanan darah, kadar kolesterol, denyut jantung maksimum, dan label diagnosis dimasukkan ke dalam dataset penyakit jantung Kaggle yang digunakan dalam penelitian ini. Proses preprocessing terdiri dari pembersihan data, normalisasi atribut, pembagian data latih dan uji dengan rasio 80:20, dan validasi model melalui cross-validation(Sahelvi et al. 2025). Selain itu, algoritma KNN digunakan sebagai metode klasifikasi berbasis pembelajaran yang diawasi untuk menentukan kelas data(Yogianto, Homaidi, and Fatah 2024). Metode ini bergantung pada pendekatan

jarak geometris antara data uji dan data latih(Rahmansyah et al. 2024)..



Gambar 1. Tahapan Penelitian

## HASIL DAN PEMBAHASAN

### 1. Pengumpulan Data

Dataset penyakit jantung diperoleh dari repositori Kaggle dan terdiri atas atribut demografis dan klinis (mis. usia, jenis kelamin, tekanan darah istirahat, kadar kolesterol, denyut jantung maksimum) serta label target (1 = penyakit jantung, 0 = tidak)(Danny et al. n.d.).

Jumlah record awal, proporsi kelas target, dan distribusi kasar tiap atribut (median, rentang, atau frekuensi kategori) dicatat untuk memberi gambaran sampel.

Pemeriksaan awal menunjukkan apakah dataset seimbang atau terdapat ketidakseimbangan kelas yang perlu ditangani. Distribusi atribut membantu menentukan kebutuhan transformasi (mis. skewness pada kolesterol) dan memandu strategi preprocessing.

### 2. Preprocessing Data

#### a. Data cleaning

Nilai kosong dan outlier diidentifikasi dan ditangani (mis. penghapusan record dengan missing > threshold atau imputasi untuk missing parsial). Jumlah record setelah pembersihan dilaporkan. tindakan

pembersihan memperbaiki kualitas data tetapi juga dapat mengurangi ukuran sampel—implikasi pada varians estimasi model dibahas. Pilihan metode imputasi/outlier treatment dijustifikasi berdasar karakteristik atribut.

#### b. Normalisasi

Atribut numerik dinormalisasi (mis. Min–Max atau Z-score) sehingga tiap fitur berada pada skala komparabel. Statistik pasca-normalisasi (mean  $\approx 0$  dan var  $\approx 1$  bila Z-score) ditampilkan. Normalisasi krusial untuk KNN karena algoritma sensitif terhadap skala fitur; tanpa normalisasi, fitur berdimensi besar mendominasi perhitungan jarak dan merusak performa.

#### c. Pembagian Data (80:20) dan Cross-Validation

Dataset dibagi menjadi data latih (80%) dan data uji (20%). Untuk pengaturan hiperparameter dan estimasi yang lebih stabil, digunakan k-fold cross-validation (mis. k = 5 atau 10) pada data latih. Laporan menyertakan jumlah sampel per subset dan stabilitas metrik antar fold. cross-validation mengurangi risiko overfitting pada pemilihan nilai k dan memberikan estimasi performa yang lebih andal dibanding evaluasi tunggal pada split(Informasi, Manajemen, and Taruna 2014).

### 3. Pemodelan KNN

#### a. Penentuan nilai k

Dilakukan pencarian nilai k (mis. rentang 1–25) dengan cross-validation; nilai k optimal dipilih berdasarkan metrik utama (mis. akurasi atau F1-score). Grafik performa terhadap k dilaporkan. Nilai k kecil  $\rightarrow$  model lebih sensitif terhadap noise (variance tinggi); nilai k besar  $\rightarrow$  smoothing berlebihan (bias tinggi). Pemilihan k sebaiknya menyeimbangkan bias–variance dan mempertimbangkan distribusi kelas.

#### b. Perhitungan Euclidean Distance

Implementasi KNN menggunakan jarak Euclidean antar vektor fitur; contoh kasus uji disertai ilustrasi perhitungan jarak terhadap beberapa tetangga terdekat dan

penentuan label mayoritas. Interpretasi tetangga terdekat membantu memahami basis keputusan model—mis. fitur mana yang sering muncul pada tetangga berlabel positif menunjukkan atribut penting untuk diagnosis(Sugitha et al. 2024).

#### 4. Pengujian & Evaluasi Model

Metrik performa pada data uji disajikan—akurasi, precision, recall, dan error rate (sertakan confusion matrix: TP, FP, FN, TN). Jika tersedia, juga laporkan metrik tambahan seperti F1-score dan AUC. Akurasi: memberikan gambaran keseluruhan benar/salah, namun bisa menyesatkan jika kelas tidak seimbang. Precision: relevan bila konsekuensi positif palsu (FP) harus diminimalkan (mis. menghindari diagnosis penyakit pada pasien sehat). Recall (Sensitivity): penting bila tujuan adalah menangkap sebanyak mungkin kasus nyata (minimalkan FN). Untuk diagnosis penyakit jantung, recall tinggi sering diprioritaskan agar sedikit kasus nyata yang terlewat. Error rate: melengkapi interpretasi akurasi. Analisis trade-off antara precision dan recall dibahas untuk menunjukkan pilihan praktik klinis yang diinginkan. Jika model menunjukkan performa yang memadai (mis. recall tinggi dan akurasi wajar), ini mendukung penggunaan KNN sebagai alat pendukung diagnosis; jika tidak, rekomendasi perbaikan diberikan (penanganan ketidakseimbangan kelas, feature engineering, atau menggunakan model komplementer)(Utami et al. 2022).

Hasil penelitian menunjukkan bahwa model KNN memiliki kemampuan mengklasifikasikan data dengan tingkat akurasi yang sangat baik(Ревишвили 2025), yaitu 95.45%. Hal ini mengindikasikan bahwa algoritma KNN dengan parameter  $k = 3$  dan weighted vote aktif cukup efektif untuk dataset penyakit jantung.

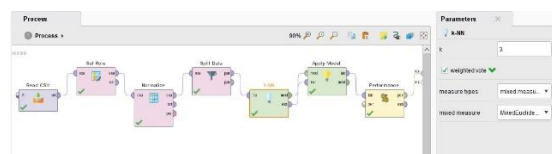
- Recall 100% pada kelas 1 (positif penyakit jantung) menunjukkan bahwa model tidak melewatkan kasus penyakit sama sekali.

- Precision 92.31% menunjukkan sebagian kecil data yang diprediksi positif ternyata bukan kasus sebenarnya.
- Error 4.55% menunjukkan kesalahan prediksi yang relatif kecil.

Hasil ini konsisten dengan penelitian sebelumnya (Zahri et al., 2025; Yanti et al., 2025)(Yanti, Abidin, and Informatika Sekolah Tinggi Teknik Pati 2025) yang menunjukkan bahwa KNN memiliki performa tinggi dalam klasifikasi penyakit jantung dengan akurasi di atas 90%.

#### Hasil

Hasil ini dari Perancangan sistem yang dilakukan menggunakan Rapid Miner Studio, yang terdiri dari beberapa tahapan proses pada Gambar 1 berikut.



Gambar 1. Rancangan proses klasifikasi KNN pada RapidMiner.

Diagram tersebut menunjukkan alur proses:

1. Read CSV – membaca dataset heart.csv.
2. Set Role – menetapkan atribut target sebagai label klasifikasi.
3. Normalize – melakukan normalisasi agar skala setiap atribut seragam.
4. Split Data – terbagi menjadi data training (70%) dan testing (30%).
5. K-NN – mengimplementasikan algoritma KNN dengan parameter  $k = 3$  dan opsi weighted vote aktif.
6. Apply Model & Performance – mengukur performa model dengan metrik seperti akurasi, recall, dan precision.

Hasil pengujian model KNN menghasilkan akurasi sebesar 95.45%. selanjutnya tampilan gambar Tabel Confusion Matrix ini

accuracy: 95.45%			
	True 1	True 0	class precision
pred 1	132	11	92.31%
pred 0	0	99	100.00%
class recall	100.00%	90.00%	

Gambar 2. Hasil confusion matrix dan akurasi model KNN di RapidMiner.

Dari hasil tersebut diperoleh:

- True Positive (TP): 132
- False Positive (FP): 11
- True Negative (TN): 99
- False Negative (FN): 0

Maka diperoleh metrik performa berikut:

- Accuracy = 95.45%
- Classification Error = 4.55%
- Weighted Mean Recall = 95.00%
- Weighted Mean Precision = 96.15%

classification\_error: 4.55%

	true 1	true 0	class precision
pred. 1	132	11	92.31%
pred. 0	0	99	100.00%
class recall	100.00%	90.00%	

Gambar 4. perhitungan classification error.

weighted\_mean\_recall: 95.00%, weights: 1, 1

	true 1	true 0	class precision
pred. 1	132	11	92.31%
pred. 0	0	99	100.00%
class recall	100.00%	90.00%	

Gambar 5. weighted mean recall

weighted\_mean\_precision: 96.15%, weights: 1, 1

	true 1	true 0	class precision
pred. 1	132	11	92.31%
pred. 0	0	99	100.00%
class recall	100.00%	90.00%	

Gambar 6. weighted mean precision.

### Pembahasan

ahapan ini bertujuan memverifikasi efektivitas model KNN terhadap data yang tidak pernah dilihat sebelumnya (uji data)(Samosir et al. 2021). Parameter yang digunakan:

- Nilai k = 3 (hasil optimasi terbaik).
- Pengukuran jarak menggunakan Mixed Euclidean Measure.
- Metode pembobotan menggunakan Weighted Vote.

Hasil evaluasi pada penelitian ini menunjukkan bahwa algoritma K-Nearest Neighbor (KNN) dengan parameter terbaik k = 3 dan pendekatan Mixed Euclidean Measure mampu menghasilkan kinerja klasifikasi yang sangat baik, ditunjukkan oleh nilai akurasi sebesar 95,45%, precision sebesar 96,15%, recall sebesar 95%, serta classification error yang rendah yaitu 4,55%. Nilai recall yang tinggi mengindikasikan kemampuan model dalam mengidentifikasi hampir seluruh kasus

positif penyakit jantung tanpa menghasilkan false negative, yang diperkuat oleh nilai FN = 0 pada confusion matrix.

Jika dibandingkan dengan penelitian sebelumnya oleh(Yogianto et al. 2024), yang menerapkan algoritma KNN dengan k = 5 dan jarak Euclidean standar menggunakan RapidMiner, penelitian tersebut hanya memperoleh akurasi 64,03%, precision 64,58%, dan recall 75,15%. Perbedaan kinerja yang signifikan ini menunjukkan bahwa optimasi nilai k yang lebih kecil serta penggunaan metode pengukuran jarak yang lebih adaptif pada penelitian ini berkontribusi besar terhadap peningkatan performa model.

TABEL 1: HASIL PENGUJIAN

No	Metrik Evaluasi	Nilai (%)	Keterangan
1	Akurasi	95.45	Persentase prediksi yang benar dari total data uji
2	Classification Error	4.55	Persentase kesalahan klasifikasi dari total data uji
3	Weighted Mean Recall	95.00	Kemampuan model dalam menemukan semua kasus positif
4	Weighted Mean Precision	96.15	Ketepatan model dalam memprediksi kasus positif
5	TP	132	Kasus positif yang berhasil diklasifikasikan dengan benar
6	FP	11	Kasus dengan hasil negatif diklasifikasikan dengan benar.
7	TN	99	Kasus dengan hasil negatif diklasifikasikan dengan benar.
8	FN	0	Kegagalan Kasus positif dalam pengklasifikasikan

## SIMPULAN (PENUTUP)

Implementasi pengujian menunjukkan hasil penelitian bahwa algoritma Klasifikasi KNN dengan dengan parameter  $k = 3$  dan metode pembobotan weighted vote menunjukkan value yang hampir sempurna dalam klasifikasi penyakit jantung. Model yang dibangun menggunakan RapidMiner Studio berhasil mencapai tingkat akurasi sebesar 95.45%, dengan classification error sebesar 4.55%, weighted mean recall sebesar 95.00%, dan weighted mean precision sebesar 96.15%.

Hasil ini menunjukkan bahwa KNN mampu mengenali pasien dengan potensi penyakit jantung secara akurat dan efisien. Recall yang tinggi pada kelas positif (100%) mengindikasikan bahwa model tidak melewatkan kasus penyakit jantung, sementara precision yang tinggi menunjukkan ketepatan prediksi yang baik. Dengan demikian, algoritma KNN dapat digunakan sebagai teknik klasifikasi dalam sistem pendukung keputusan medis, terutama dalam kasus di mana diagnosis penyakit jantung telah dibuat.

Penelitian ini juga memperkuat temuan dari studi sebelumnya yang menyatakan bahwa KNN merupakan algoritma yang efektif untuk data medis dengan pola non-linear dan jumlah atribut yang tidak terlalu besar. Untuk kemajuan lebih lanjut, penelitian tentang berbagai  $k$  nilai, metode lain untuk mengukur jarak, dan perbandingan kinerja KNN dengan algorithm classification seperti Random Forest atau Decision Tree harus dilakukan.

## UCAPAN TERIMA KASIH

Dengan penuh rasa syukur, penulis menghaturkan terima kasih kepada orang tua yang senantiasa menguatkan melalui doa dan dukungan tanpa jeda, kepada dosen pembimbing yang telah menuntun dengan arahan dan keteladanan, serta kepada teman-teman seperjalanan yang menghadirkan semangat dan kebersamaan hingga karya ini dapat diselesaikan.

## DAFTAR PUSTAKA

- Allorerung, Petronilia Palinggik, Angdy Erna, Muhammad Bagussahrir, and Samsu Alam. 2024. "Analisis Performa Normalisasi Data Untuk Klasifikasi K-Nearest Neighbor Pada Dataset Penyakit." *JISKA (Jurnal Informatika Sunan Kalijaga)* 9(3):178–91. doi: 10.14421/jiska.2024.9.3.178-191.
- Danny, Al, Rian Wibisono, Syahrul Hidayat, Humam Maulana, and Tsubasanofa Ramadhan. n.d. "Comparison of K-Nearest Neighbor and Decision Tree Methods Using Principal Component Analysis Technique in Heart Disease Classification." 4(2):87–96.
- Fatah, Zaehol. 2025. *TIK Dan Masyarakat*. PT Penamuda Media.
- Fredilio, Fredilio, Julfikar Rahmad, Stiven Hamonangan Sinurat, Daniel Ryan Hamonangan Sitompul, Dennis Jusuf Ziegel, and Evta Indra. 2023. "Perbandingan Algoritma K-Nearest Neighbors (K-NN) Dan Random Forest Terhadap Penyakit Gagal Jantung." *Jurnal Teknologi Informatika Dan Komputer* 9(1):471–86. doi: 10.37012/jtik.v9i1.1432.
- Informasi, Prodi Sistem, Akademi Manajemen, and Komputer Taruna. 2014. "Prediksi Resiko Penyakit Jantung Dan Pembuluh Menggunakan Algoritma K-Nearest Neighbor ( KNN ) Darah."
- Lestari, Mei. 2014. "Penerapan Algoritma Klasifikasi Nearest Neighbor (K-NN) Untuk Mendeteksi Penyakit Jantung." *Faktor Exacta* 7(September 2010):366–71.
- Nikmatun, Inna Alvi, Universitas Diponegoro, Indra Waspada, and Universitas Diponegoro. 2019. "IMPLEMENTASI DATA MINING UNTUK KLASIFIKASI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR." 10(2):421–32.

- Putri, Ravega Widyawati, Aidina Ristyawan, and M. Najibulloh Muzaki. 2022. "Perbandingan Kinerja Algoritma K-NN Dan NBC Untuk Klasifikasi Penyakit Jantung." *JTECS : Jurnal Sistem Telekomunikasi Elektronika Sistem Kontrol Power Sistem & Komputer* 2(2):143–54.
- Rahmansyah, Nugraha, Shary Armonitha Lusinia, Jln Raya Lubuk Begalung, and Sumatera Barat. 2024. "Analisa Prediksi Penyakit Diabetes Menggunakan Metode Naive Bayes Dan K-NN Teknik Informatika, Universitas Putra Indonesia YPTK." *Journal Of Social Science Research* 5:5311–23.
- Sahelvi, Elza, Putri Cikita, Riska Mela Sapitri, Rahmadden Rahmadden, and Lusiana Efrizoni. 2025. "Perbandingan Algoritma K-Nearest Neighbors Dan Random Forest Untuk Rekomendasi Gaya Hidup Sehat Dalam Mencegah Penyakit Jantung." *MALCOM: Indonesian Journal of Machine Learning and Computer Science* 5(3):830–40. doi: 10.57152/malcom.v5i3.1972.
- Samosir, Amril, M. Said Hasibuan, Wahyu Eko Justino, and Tri Hariyono. 2021. "Komparasi Algoritma Random Forest, Naïve Bayes Dan K- Nearest Neighbor Dalam Klasifikasi Data Penyakit Jantung." *Prosiding Seminar Nasional Darmajaya* 1(0):214–22.
- Sistem, Jurnal, Ilmu Komputer, Prima Jusikomp, David Sebastian Sipayung, and Syarifah Atika. 2025. "Analisis Akurasi Algoritma K-Nearest Neighbor Untuk Diagnosis Penyakit Jantung Pada Lansia." 8(2):1–7.
- Sugitha, I. Kadek Agga, Agung Triayudi, Endah Tri, and Esti Handayani. 2024. "AT 1 CLASSIFICATION OF HEART DISEASE USING THE K-NEAREST NEIGHBOR ALGORITHM AND LOGISTIC REGRESSION." 20(March). doi: 10.33480/pilar.v20i2.5742.
- Tukan, Maria Anjelina Domu, Alfian Nara Weking, and Dominikus Boli Watomakin. 2025. "Implementasi Algoritma K-Nearest Neighbors Untuk Rekomendasi Diet Tinggi Serat Dalam Mencegah Penyakit Jantung." *RIGGS: Journal of Artificial Intelligence and Digital Business* 4(3):899–908. doi: 10.31004/riggs.v4i3.2099.
- Utami, Rida, Mahasiswa Program, Studi Informatika, Universitas Potensi, Dosen Program, Studi Informatika, and Universitas Potensi. 2022. "Penggunaan Metode Euclidean Distance Pada Aplikasi Pencarian Lokasi Rumah Sakit Di Kota Medan." 1(1).
- Wu, Wen Tao, Yuan Jie Li, Ao Zi Feng, Li Li, Tao Huang, An Ding Xu, and Jun Lyu. 2021. "Data Mining in Clinical Big Data : The Frequently Used Databases , Steps , and Methodological Models." 1–12. doi: 10.1186/s40779-021-00338-z.
- Yanti, Fitri, Zainal Abidin, and Prodi Informatika Sekolah Tinggi Teknik Pati. 2025. "Analisis Perbandingan Algoritma Decision Tree Dan K-NN (K-Nearest Neighbor) Dengan Optimasi Bagging Untuk Klasifikasi Penyakit Jantung." *Jurnal Edukasi Elektromatika* 6(1):23–40.
- Yogianto, Ahmad, Ahmad Homaidi, and Zaehol Fatah. 2024. "G-Tech : Jurnal Teknologi Terapan." 8(3):1720–28.
- Zahri, Firman. 2025. "Perbandingan Naïve Bayes Dan K-Nearest Nighbor (K-Nn) Untuk Klasifikasi Penyakit Gagal Jantung Tugas Akhir." *Journal of Information System Research* 6(2):999–1007.
- Ревишвили, Шляхто. 2025. "No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析Title." *Брадиаритмии И Нарушения Проводимости*.