

Klasifikasi Potensi Penyakit Jantung Menggunakan Algoritma C4.5

Muhammad Santoso¹, Bilal Al-Akbar², Hilman Nurjaya³, Syaiban Ahmad Ramadhan⁴,
Naufal Ahmad Rizky⁵, Achmad Fadillah⁶

^{1,2,3,4,5,6,7} Sistem Informasi, Teknik dan Informatika, Universitas Bina Sarana Informatika
Jl. Kramat Raya No.98 - Jakarta 10450, Indonesia

e-mail: ¹msantosomuhammad686@gmail.com, ²Bilalakbar1806@gmail.com, ³Hilmannurjaya123@gmail.com,
⁴ibanrama29@gmail.com, ⁵naufalrizky150@gmail.com,
⁶myfadil869@gmail.com

Abstrak - Dengan tren yang mengkhawatirkan dalam beberapa tahun terakhir, penyakit jantung masih menjadi penyebab kematian paling umum di Indonesia dan di seluruh dunia. Tujuan dari penelitian ini adalah untuk meningkatkan pemahaman dan prediksi penyakit jantung koroner dengan menggunakan metode pohon keputusan dengan algoritma C4.5. Algoritma ini unggul karena mudah digunakan dan dapat digambarkan sebagai pohon keputusan berdasarkan faktor risiko pasien. Usia, jenis kelamin, nyeri dada, tekanan darah sistolik, kadar kolesterol total, denyut jantung maksimal, dan potensi adalah beberapa faktor risiko penting dalam studi ini. Algoritma C4.5 bertujuan untuk membuat model prediktif yang membantu dalam diagnosis dan intervensi dini penyakit jantung koroner dengan memanfaatkan elemen-elemen ini. Hasil model pohon keputusan dapat dikomunikasikan dengan mudah kepada profesional karena sederhana dan dapat dipahami.

Kata Kunci: Penyakit Jantung, Klasifikasi, Data Mining, Algoritma C4.5.

Abstract - The paper's abstract templates are used as a guide for writing paper for publications in journals Bina Sarana Informatika. The paper writers must follow the instructions given in this paper. Abstract is restricted from 100-250 words in Indonesian Language and English Language (in italics), should not contain equations, pictures, and table. The letter size for abstract, keywords, and body papers is 10 points using Times New Roman font. In the abstract they contain prologue of the raised problem, the proposed method and the desired results (hypothesis). The number of pages in the manuscript is at least 6 pages, a maximum of 10 pages, with an even number of pages using A4 paper size with 1 (single) spacing.

Keywords: Heart Disease, classification, Data Mining, C4.5 Algorithm

PENDAHULUAN

Penyakit Jantung merupakan salah satu masalah Kesehatan yang serius dan menjadi penyebab utama kematian di seluruh dunia. Klasifikasi penyakit jantung adalah salah satu upaya untuk mengelompokkan jenis-jenis penyakit jantung berdasarkan berbagai karakteristik penyebab, gejala dan perubahan structural atau fungsional pada jantung. Penyakit jantung dapat mencakup berbagai kondisi mulai dari penyakit jantung koroner, gagal jantung, aritmia, hingga penyakit jantung bawaan. Klasifikasi ini sangat penting untuk membantu para ahli Kesehatan dalam mendiagnosis, merencanakan pengobatan, dan memberikan perawatan yang tepat dan sesuai dengan jenis penyakit jantung yang dialami oleh pasien (Agus Oka Gunawan et al., 2023).

Diabetes adalah salah satu kondisi yang dapat menyebabkan komplikasi penyakit berat lainnya seperti darah tinggi, jantung, dll. Hal ini disebabkan oleh masalah metabolisme yang menghalangi pankreas untuk menghasilkan insulin (Aris Faizal & Benyamin, 2019). Jumlah orang yang menderita diabetes meningkat di setiap negara, tetapi penyakit ini tersebar di seluruh dunia. Di Indonesia, prevalensi diabetes adalah 10,9 persen, dan terus meningkat. Penyakit ini menyerang semua orang, bukan hanya yang miskin. Pola hidup yang tidak sehat dianggap sebagai faktor utama penyebabnya. Akibatnya, penting untuk mengetahui tanda-tanda awal diabetes agar dapat mencegah penyakit sebelum terlambat. Gejala awal biasanya muncul pada tahap. Umumnya gejala awal terjadi bagi penderita tahap awal diabetes diantaranya sering merasa haus, sering merasa lapar, gatal-gatal dan juga lemas (Yusnaeni, 2022).



Penyakit jantung koroner (PJK) disebabkan oleh plak yang menumpuk di dalam arteri koroner, yang menyediakan oksigen untuk otot jantung. Statistik global menunjukkan 9,4 juta kematian per tahun. Diabetes adalah salah satu kondisi yang dapat menyebabkan komplikasi penyakit berat lainnya seperti darah tinggi, jantung, dll. Hal ini disebabkan oleh masalah metabolisme yang menghalangi pankreas untuk menghasilkan insulin (Riska et al., 2021).

Salah satu jenis penyakit yang dapat membahayakan hidup seseorang adalah hipertensi. Di Negara maju, prevalensi hipertensi mencapai 37,3%. Sekitar 95% kasus hipertensi disebabkan oleh penyakit lain seperti stroke, penyakit jantung koroner, gangguan fungsi ginjal, atau gangguan fungsi kognitif. Hipertensi cenderung meningkat di Indonesia. Di Indonesia, rata-rata 25,8 persen orang berusia 25 hingga 70 tahun menderita hipertensi. Namun, hanya 36,8 persen penderita yang diobati oleh tenaga medis, dan sisanya tidak terdiagnosis dengan baik (Azwanti & Elisa, 2019).

Klasifikasi penyakit jantung sangat penting dalam penelitian medis, pengembangan obat, dan perancangan intervensi kesehatan masyarakat. Dengan memahami perbedaan antar jenis penyakit jantung, para ahli kesehatan dapat merancang strategi pengobatan yang lebih efektif dan memberikan perawatan yang lebih spesifik sesuai dengan kebutuhan masing-masing pasien (Amalia Hilda & Evicienna, n.d.) . Dengan demikian, upaya klasifikasi ini merupakan langkah penting dalam meningkatkan deteksi dini, pengobatan, dan pencegahan penyakit jantung untuk meningkatkan kesehatan jantung secara keseluruhan, penelitian ini dibuat agar dapat memecahkan masalah dalam menentukan penyebab utama penyakit jantung (Novitasari et al., 2023).

Penyakit jantung merujuk pada berbagai kondisi yang memengaruhi jantung, organ utama dalam sistem kardiovaskular yang bertanggung jawab memompa darah ke seluruh tubuh. Latar belakang potensi penyakit jantung melibatkan berbagai faktor risiko dan penyebab yang dapat meningkatkan kemungkinan seseorang mengalami masalah jantung. Berdasarkan data yang kami miliki ada beberapa faktor utama penyebab penyakit jantung seperti Tekanan darah tinggi, Kadar kolesterol, Hr max, dan nyeri dada.

Untuk itu, kami mengolah data ini untuk mengetahui faktor apa yang paling berpotensi dalam penyakit jantung , dengan adanya hasil dari olah data ini kami berharap dapat menjadi pengingat bagi siapapun agar tetap menjaga pola hidup yang sehat dan mengantisipasi potensi penyakit jantung sejak dini.

Tabel 1. Literatur Penelitian

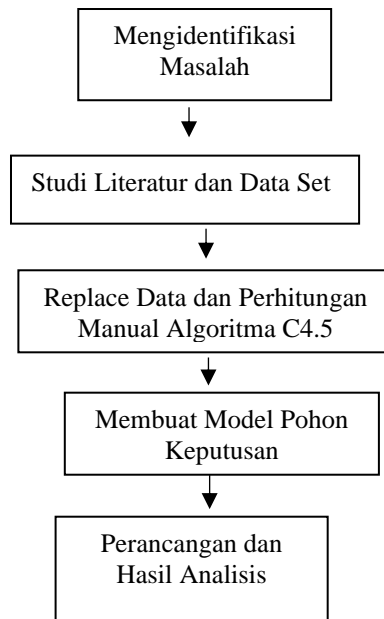
<i>Research Problem (RP)</i>	<i>Literatur Support</i>
RP1 Mengklasifikasi penyakit jantung	Klasifikasi adalah proses mengajar dan mendidik fungsi target dengan memetakan setiap set karakteristik (atribut) ke satu jumlah label kelas yang tersedia. Dengan menggunakan 909 kasus dan 6 karakteristik, memprediksi kelangsungan hidup penyakit jantung. Hasil penelitian menunjukkan bahwa metode pohon keputusan memiliki nilai terbaik. Selain itu, S.B. Kotsiantis menjelaskan dalam paper rivew bahwa metode decision tree memiliki kelebihan dalam pengolahan kumpulan data penyakit jantung, seperti transparansi pengetahuan atau klasifikasi, kecepatan klasifikasi, dan sifat diskrit, binari, atau kontinu (Anggraini et al., 2022) , (Komarudin et al., 2021).
RP2 Menentukan Atribut atau dataset yang menjadi penyebab penyakit jantung	Pada tahap ini, tipe data akan diubah untuk menyesuaikan dengan set data jantung yang akan diproses menjadi informasi atau lebih sederhana dilakukan encoding atau konversi data untuk membuatnya dapat digunakan dalam format tertentu, seperti yang ditunjukkan di bawah ini: 1. Usia (Fitur) = Integer 3. Jenis Kelamin (Fitur) = Binominal 4. Tipe Nyeri Dada (Fitur) = Polynominal 5. RestingBP (Fitur) = Integer 6. Kolesterol (Fitur) = Integer 7. EKG istirahat (Fitur) = Polynominal 8. MaxHR (Fitur) = Integer 9. Latihan Angina (Fitur) = Binominal 10. Puncak Tua (Fitur) = Real 11. ST_Lereng (Fitur) = Poly Penyakit Jantung (Label) adalah biner (Sepharni et al., 2022)

Sumber: Hasil Penelitian (2023)

METODE PENELITIAN

Dalam penelitian ini, data diperoleh melalui sumber Kaggle, terdiri dari 918 dataset, dan metode yang diterapkan adalah algoritma C4.5. Langkah-langkah penyelesaian masalah dapat ditemukan dalam Gambar 1, yang mencakup identifikasi masalah, studi literatur, penggantian data, perhitungan manual algoritma C4.5, serta perancangan dan analisis.

Algoritma decision tree series C4.5 diimplementasikan untuk mengklasifikasikan penyakit jantung. C4.5 merupakan perkembangan dari algoritma ID3, digunakan untuk membangun pohon keputusan dalam metode klasifikasi data mining yang bersifat prediktif. Elemen-elemen kunci dalam algoritma C4.5 melibatkan nilai Entropy, Gain, Split info, dan Gain ratio. Pembentukan simpul akar model tree didasarkan pada Gain ratio tertinggi.



Sumber: Hasil Penelitian (2023)

Gambar 1. Langkah-Langkah Penyelesaian Masalah

1. Mengidentifikasi Masalah

Pada tahap awal diperlukan untuk mengetahui permasalahan yang dihadapi subjek penelitian dan menentukan ruang lingkup penelitian.

2. Studi Literatur dan Data Set

Studi literatur dilakukan untuk mengumpulkan dan memproses data terkait penyakit jantung dan faktor gejala yang terdiagnosa. Sumber data berasal dari buku, jurnal, dan artikel ilmiah. Data set tentang penyakit jantung diambil dari situs web Kaggle.

3. Replace Data dan Perhitungan Manual Algoritma C4.5

Pada tahap ini, proses penggantian data mencakup transformasi bahasa (dari Inggris ke Indonesia) dan konversi nilai 0 menjadi nominal dan 1 menjadi penyakit jantung. Tujuannya adalah agar data terlihat lebih intuitif. Perhitungan manual melibatkan 918 dataset yang dipilih secara acak, disesuaikan dengan formula algoritma decision tree series C4.5.

4. Membuat Pohon Keputusan

Tahapan pemodelan menggunakan algoritma C4.5 [14] meliputi:

1. Menyiapkan data training. Pada tahap ini dataset akan digunakan sebagai data training, dataset pada proses sebelumnya telah dikelompokkan pada kelas-kelas tertentu.
2. Menentukan akar pohon. Pada tahap ini dilakukan perhitungan nilai entropi guna penentuan akar pohon berdasarkan dari nilai information gain yang paling tinggi. Entropi digunakan untuk mengukur tingkat keberagaman kumpulan data. Rumus dari entropi sebagai berikut [15]:

$$Entropi (S) = \sum p_i * \log_2 p_i$$

Keterangan:

S: himpunan kasus

n: jumlah partisi

pi: jumlah proporsi sampel untuk kelas i

3. Menentukan nilai information gain

Information gain merupakan salah satu cara menentukan pemilihan atribut pada tiap node di tree. Perhitungan information gain menggunakan rumus sebagai berikut:

$$Gain(S, A) = Entropi(S) - \sum_{i=0}^n |S_i| * Entropi(S_i)$$

Keterangan:

Entropi(S_i): entropi untuk kasus yang memiliki nilai i

S: himpunan kasus

A: atribut penentu

N: jumlah partisi atribut A

|S_i|: jumlah kasus nilai i

|S|: jumlah seluruh kasus dalam S

4. Mengulangi langkah kedua

Pada tahap ini mengulang langkah kedua sampai semua record terpartisi. Tahap ini akan berhenti apabila semua record dalam simpul N mendapat kelas yang sama, tidak terdapat atribut dalam record yang bisa dipartisi kembali, dan tidak ada record dalam cabang pohon yang kosong.

5. Perancangan dan Hasil Analisis

Tahap ini melibatkan pengolahan data set jantung dalam RapidMiner dan analisis hasilnya. Metode K-Fold Cross Validation digunakan untuk pembentukan data training dan data testing dari keseluruhan dataset. Nilai K sebesar 10 direkomendasikan untuk memastikan iterasi pohon keputusan sebanyak 10.

HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data dan Atribut

Kebutuhan data training pada penelitian ini bersumber dari data potensi penyakit jantung pada web Kaggle. Dari data tersebut kemudian akan digunakan sebagai data operasional, seperti yang ditampilkan dalam Tabel 2. Pada table 2 diketahui hasil dari dataset potensi penyakit jantung yang telah melewati proses data selection, pre processing dan transformation. Terdapat enam atribut dengan menentukan atribut potensi sebagai atribut target.

Tabel 2. Dataset Potensi Penyakit Jantung

Umur	Jenis Kelamin	Nyeri Dada	Tekanan Darah Sistolik	Kadar kolesterol	Denyut Jantung Maximal	Potensi
Dewasa	Laki-laki	Atypical Angina	Tinggi	Tinggi	Diatas normal	YA
Pra Lanjut Usia	Perempuan	Non Anginal pain	Tinggi	Normal	Normal	YA
Dewasa	Perempuan	Non Anginal Pain	Normal	Ambang batas	Normal	TIDAK
Dewasa	Laki-laki	Asymptonic	Normal	Normal	Normal	TIDAK
Lanjut Usia	Laki-laki	Asymptonic	Tinggi	Tinggi	Normal	YA
Pra Lanjut Usia	Laki-laki	Atypical Angina	Tinggi	Tinggi	Normal	YA
Lanjut Usia	Perempuan	Atypical Angina	Tinggi	Normal	Normal	YA
Dewasa	Laki-Laki	Non Anginal pain	Normal	Normal	Normal	TIDAK

Sumber: Hasil Penelitian (2023)

Dalam penentuan klasifikasi potensi penyakit jantung dengan atribut umur, jenis kelamin, nyeri dada, tekanan darah sistolik, kadar kolesterol, dan denyut jantung maksimal memiliki beberapa kriteria seperti yang terlihat didalam Tabel 3.

3.1 Pengumpulan Data dan Atribut

Kebutuhan data training pada penelitian ini bersumber dari data potensi penyakit jantung pada web Kaggle. Dari data tersebut kemudian akan digunakan sebagai data operasional, seperti yang ditampilkan dalam Tabel 2. Pada table 2 diketahui hasil dari dataset potensi penyakit jantung yang telah melewati proses data selection, pre processing dan transformation. Terdapat enam atribut dengan menentukan atribut potensi sebagai atribut target.

Tabel 2. Dataset Potensi Penyakit Jantung

Umur	Jenis Kelamin	Nyeri Dada	Tekanan Darah Sistolik	Kadar kolesterol	Denyut Jantung Maximal	Potensi
Dewasa	Laki-laki	Atypical Angina	Tinggi	Tinggi	Diatas normal	YA
Pra Lanjut Usia	Perempuan	Non Anginal pain	Tinggi	Normal	Normal	YA
Dewasa	Perempuan	Non Anginal Pain	Normal	Ambang batas	Normal	TIDAK
Dewasa	Laki-laki	Asymptonic	Normal	Normal	Normal	TIDAK
Lanjut Usia	Laki-laki	Asymptonic	Tinggi	Tinggi	Normal	YA
Pra Lanjut Usia	Laki-laki	Atypical Angina	Tinggi	Tinggi	Normal	YA
Lanjut Usia	Perempuan	Atypical Angina	Tinggi	Normal	Normal	YA
Dewasa	Laki-Laki	Non Anginal pain	Normal	Normal	Normal	TIDAK

Sumber: Hasil Penelitian (2023)

Dalam penentuan klasifikasi potensi penyakit jantung dengan atribut umur, jenis kelamin, nyeri dada, tekanan darah sistolik, kadar kolesterol, dan denyut jantung maksimal memiliki beberapa kriteria seperti yang terlihat didalam Tabel 3.

Tabel 3. Atribut Data Set Potensi Penyakit Jantung

NO	Kriteria	Keterangan
1	Umur	Anak (0-9), Remaja (10-18), Dewasa (19-44), Pra Lanjut Usia (45-59), Lanjut Usia (60-185)
2	Jenis Kelamin	Laki-laki, Perempuan
3	Nyeri Dada	Atypical Angina (ATA), Non Anginal Pain (NAP), Asymptonic (ASY), Typical Angina (TA)
4	Tekanan Darah Sistolik	Rendah (0-89), Normal (90-130), Tinggi (131-250)
5	Kadar kolesterol	Normal (0-199), Ambang Batas (200-239), Tinggi (240-500)
6	Denyut Jantung Maximal	Dibawah Normal (0-49), Normal (50-220), Diatas normal (221-400)

Sumber: Hasil Penelitian (2023)

3.2 Pengolahan Data Menggunakan Algoritma C4.5

Dalam tahap ini dilakukan perhitungan untuk menentukan akar pohon berdasarkan rumus entropi dengan menggunakan dataset pada Tabel 2 sehingga dihasilkan nilai entropi dari masing- masing atribut seperti yang terlihat dalam Tabel 4.

Tabel 4. Entropy Potensi Penyakit Jantung

Atribut	Keterangan	Jumlah	Iya	Tidak	Entropy
Umur	Dewasa	178	137	41	0,778598
	Pra Lanjut Usia	487	352	135	0,851618
	Lanjut Usia	253	199	54	0,748005
	Atypical Angina	173	173	0	0

Nyeri Dada	Non Anginal Pain	203	139	64	0,899164
	Asymptonic	496	338	158	0,902796
	Typical Angina	46	38	8	0,666578
Jenis Kelamin	Laki-laki	725	528	197	0,843928
	Perempuan	193	160	33	0,659952
Tekanan Darah Sistolik	Tekanan Darah Rendah	2	0	2	0
	Tekanan Darah Normal	491	263	228	0,996332
	Tekanan Darah Tinggi	425	425	0	0
Kadar Kolesterol	Kolesterol Normal	318	174	144	0,99357
	Kolesterol Ambang Batas	237	151	86	0,945039
	Kolesterol Tinggi	363	363	0	0
Denyut Jantung Maximal	Denyut Jantung Dibawah Normal	0	0	0	0
	Denyut Jantung Normal	834	604	230	0,849636
	Denyut Jantung Diatas Normal	84	84	0	0

Sumber: Hasil Penelitian (2023)

Dari hasil perhitungan entropi dari setiap node pada Tabel 4 akan digunakan untuk menentukan nilai *information gain*.

1. Menentukan Nilai *Information Gain*. Dengan menggunakan rumus *information gain* pada Tabel 4 didapatkan nilai *information gain* yang nantinya akan digunakan untuk menentukan akar pohon yang paling atas. Hasil dari perhitungan nilai *information gain* dapat dilihat pada Tabel 5.

Tabel 5. Hasil Perhitungan Information Gain

Atribut	Information Gain	Gain Ratio
Umur	0,003237	0,002222
Nyeri Dada	0,083232	0,05882
Jenis Kelamin	0,006891	0,009288
Tekanan Darah Sistolik	0,279244	0,274715
Kadar kolesterol	0,223982	0,14326
Denyut Jantung Maximal	0,040249	0

Sumber: Hasil Penelitian (2023)

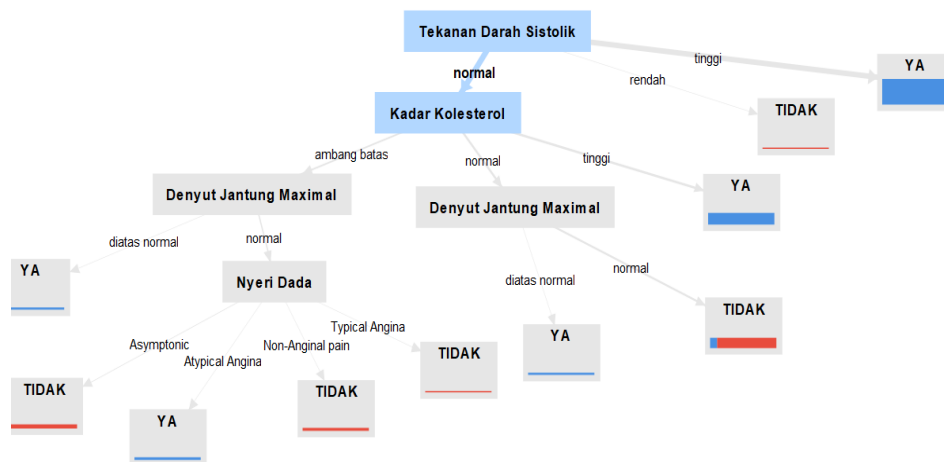
Berdasarkan Tabel 5 diketahui atribut yang memiliki nilai *information gain* tertinggi adalah atribut Tekanan Darah Sistolik yang akan dijadikan sebagai akar pohon.

2. Mengulangi langkah ke 2 sampai record dapat terpatisi

Pada tahap ini diulang kembali langkah ke dua, namun perhitungan entropi hanya berdasarkan dari atribut jenis buku dikarenakan sebagai akar pohon. Setelah didapat nilai masing-masing entropi kemudian ditentukan nilai *information gain* atributnya.

3.3 Pembuatan Model Pohon Keputusan

Berdasarkan hasil perhitungan nilai *information gain* didapatkan pohon keputusan seperti yang terlihat dalam Gambar 2.



Sumber: Hasil Penelitian (2023)

Gambar 2. Hasil Decision Tree

Berdasarkan hasil pohon keputusan pada Gambar 2 diketahui atribut yang paling atas dan menempati akar pohon adalah tekanan darah sistolik yang diikuti oleh kadar kolesterol, denyut jantung maximal dan nyeri dada.

3.4 Hasil Keputusan

Hasil evaluasi dengan menggunakan metode cross validation dan 10 fold validation pada dataset dapat dilihat dari gambar 3. Berdasarkan gambar 3, hasil akurasi Algoritma C.45 menunjukkan performa 96,74%.

accuracy: 96.74% +/- 2.46% (micro average: 96.73%)

	true YA	true TIDAK	class precision
pred. YA	661	3	99.55%
pred. TIDAK	27	227	89.37%
class recall	96.08%	98.70%	

Sumber: Hasil Penelitian (2023)

Gambar 3 Hasil Keputusan

KESIMPULAN

Berdasarkan penelitian, kami menyimpulkan dalam mengklasifikasi potensi penyakit jantung pada data yang kami ambil dari Kaggle bahwa klasifikasi berdasarkan tekanan darah tinggi sistolik, jenis kelamin, kadar kolesterol, max hr, kadar gula darah didapatkan bahwa tekanan darah tinggi merupakan faktor yang paling berpotensi dalam penyakit jantung, selanjutnya diikuti oleh kadar kolesterol yang tinggi, max hr tinggi, dan kadar gula darah yang tinggi.

Kami menyarankan untuk penelitian berikutnya menggunakan metode penelitian yang lain seperti K Means atau metode Data Mining lainnya untuk menentukan prediksi potensi penyakit jantung atau dilakukan dengan menggabungkan beberapa metode agar mendapatkan hasil yang lebih akurat dan lebih baik.

REFERENSI

- Agus Oka Gunawan, I. M., Indah Saraswati, I. D. A., Riswana Agung, I. D. G., & Eka Putra, I. P. (2023). Klasifikasi Penyakit Jantung Menggunakan Algoritma Decision Tree Series C4.5 Dengan Rapidminer. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 5(2), 73–83. <https://doi.org/10.47233/jteksis.v5i2.775>
- Amalia Hilda, & Evicienna. (n.d.). *227517-sistem-penunjang-keputusan-kesehatan-unt-2d22c05d*.
- Angraini, R., Haerani, E., Jasril, J., & Afrianty, I. (2022). Pengelompokan Penyakit Pasien Menggunakan Algoritma K-Means. *JURIKOM (Jurnal Riset Komputer)*, 9(6), 1840. <https://doi.org/10.30865/jurikom.v9i6.5145>
- Aris Faizal, & Benyamin. (2019). *Router Research Penerapan Data Mining untuk Identifikasi Penyakit Diabetes Melitus dengan Menggunakan Metode Klasifikasi* (Vol. 1, Issue 1). <http://ejournal.stipwunaraha.ac.id/index.php/router>
- Azwanti, N., & Elisa, E. (2019). Analisis Pola Penyakit Hipertensi Menggunakan Algoritma C4.5. *InfoTekJar (Jurnal Nasional Informatika Dan Teknologi Jaringan)*, 3(2), 116–123. <https://doi.org/10.30743/infotekjar.v3i2.944>
- Komarudin, R., Vicenna, P., Yudha, B., Maulana, Y. I., Afni, N., Salim, A., & Carolina, I. (2021). Bianglala Informatika Penerapan Metode Algoritma C4.5 Dalam Klasifikasi Diagnosa Penyakit Umum Menggunakan WEKA. *Bianglala Informatika*, 9(2).
- Novitasari, F., Haerani, E., Nazir, A., Jasril, J., & Insani, F. (2023). Sistem Klasifikasi Penyakit Jantung Menggunakan Teknik Pendekatan SMOTE Pada Algoritma Modified K-Nearest Neighbor. *Building of Informatics, Technology and Science (BITS)*, 5(1). <https://doi.org/10.47065/bits.v5i1.3610>
- Riska, S., Alham, J. I., Yosrita, E., & Cahyaningtyas, ; Rizqia. (2021). *Sistem Diagnosis Penyakit Jantung Koroner Dengan Menggunakan Algoritma C4.5 Berbasis Website (Studi Kasus: RSUD Dr. Soedarso Pontianak)*. 14(2). <https://doi.org/10.33322/petir.v14i2.1388>
- Sepharni, A., Hendrawan, I. E., Rozikin, C., & Karawang, S. (2022). *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*. <https://www.kaggle.com/fedesoriano/heart->
- Yusnaeni, W. (2022). This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License Penerapan Algoritma C4.5 Dalam Prediksi Resiko Diabetes Tahap Awal (Early Stage Diabetes). *Jurnal Teknik Komputer AMIK BSI*. <https://doi.org/10.31294/jtk.v4i2>