JUSIFOR: Jurnal Sistem Informasi dan Informatika
Vol. 3, No. 2, Desember 2024, hlmn. 160-165
E-ISSN: 2830-3393 | P-ISSN: 2830-2443
DOI: https://doi.org/10.70609/jusifor.v3i2.5729

Implementasi Algoritma Decision Tree Untuk Pendeteksian Penyakit Jantung

Lailin Nafi'ah[1*], Zaehol Fatah[2]

[1] Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Ibrahimy, Indonesia
[2] Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Ibrahimy, Indonesia

Info Artikel

Riwayat Artikel:
Diterima : **29-10-2024**
Direvisi : **31-10-2024**
Diterbitkan : **15-12-2024**

Kata Kunci:

Decision Tree,
Penyakit Jantung,
Data Mining,

ABSTRAK

Penyakit jantung merupakan salah satu penyebab utama kematian di dunia, sehingga penting untuk mengembangkan metode yang efektif dalam mendeteksi dan memprediksi risiko penyakit ini. Algoritma *Decision Tree* telah banyak digunakan dalam bidang kesehatan karena kemampuannya untuk menangani data yang kompleks dan memberikan hasil yang mudah diinterpretasikan. Penelitian ini membahas implementasi algoritma *Decision Tree* untuk prediksi penyakit jantung dengan menggunakan dataset yang berisi berbagai fitur klinis, seperti tekanan darah, kadar kolesterol, usia, jenis kelamin, dan lain-lain. Dalam proses implementasi, data diolah melalui beberapa tahapan, termasuk pembersihan data, pemilihan fitur, serta pembagian data menjadi subset pelatihan dan pengujian. Hasil dari model menunjukkan bahwa algoritma *Decision Tree* mampu mencapai tingkat akurasi yang baik dalam memprediksi risiko penyakit jantung. Selain itu, analisis lebih lanjut terhadap fitur-fitur yang paling berpengaruh dalam prediksi membantu memberikan wawasan yang lebih mendalam tentang faktor risiko utama penyakit jantung. Studi ini menunjukkan bahwa algoritma *Decision Tree* merupakan alat yang efektif untuk digunakan dalam prediksi medis, khususnya penyakit jantung, dan memiliki potensi untuk diintegrasikan ke dalam sistem pendukung keputusan klinis.

***Keywords*:**

Decision Tree,
Heart disease,
Data Mining,

ABSTRACT

Heart disease is one of the main causes of death in the world, so it is important to develop effective methods for detecting and predicting the risk of this disease. The Decision Tree algorithm has been widely used in the health sector because of its ability to handle complex data and provide results that are easy to interpret. This research discusses the implementation of the Decision Tree algorithm for predicting heart disease using a dataset containing various clinical features, such as blood pressure, cholesterol levels, age, gender, etc. In the implementation process, data is processed through several stages, including data cleaning, feature selection, and dividing data into training and testing subsets. The results of the model show that the Decision Tree algorithm is able to achieve a good level of accuracy in predicting the risk of heart disease. In addition, further analysis of the features that are most influential in predictions helps provide deeper insight into the main risk factors for heart disease. This study shows that the Decision Tree algorithm is an effective tool for use in medical prediction, especially heart disease, and has the potential to be integrated into clinical decision support systems.

Penulis Korespondensi:
Lailin Nafi'ah,
Teknologi Informasi,
Universitas Ibrahimy, Situbundo, Jawa Timur, Indonesia
Email: lailinnafiahnafiah@gmail.com

This is an open access article under the CC BY-SA license

1. PENDAHULUAN

Penyakit jantung merupakan salah satu penyakit yang paling sering ditemui di masyarakat dan dapat menginfeksi siapa saja tanpa memandang usia, jenis kelamin, atau gaya hidup [1]. Jantung adalah organ otot yang berfungsi untuk memompa darah melalui pembuluh darah dengan ritme yang teratur dan berkelanjutan. Organ ini sangat penting dalam sistem sirkulasi, karena darah mendistribusikan oksigen dan nutrisi ke seluruh tubuh serta membantu mengeluarkan sisa-sisa metabolisme. Jantung terletak di sisi kiri rongga dada. Sebagai salah satu organ yang paling krusial, jantung berpotensi menyebabkan kematian jika terjadi masalah. Gangguan yang terkait dengan jantung dapat dikategorikan menjadi dua tipe: penyakit jantung dan serangan jantung. [2].

Banyak pasien dengan penyakit jantung tidak menyadari gejala-gejala awal yang mereka alami, dan sejumlah besar penderita penyakit jantung meninggal akibat serangan jantung [3]. Penyakit jantung sering kali tidak terdeteksi, yang menyebabkan banyak orang meninggal secara tiba-tiba tanpa menyadari bahwa mereka menderita kondisi ini atau mengalami serangan jantung. Penyakit jantung merupakan salah satu penyebab utama kematian global. Menurut laporan dari World Health Organization (2017), diperkirakan sekitar 17,9 juta orang meninggal akibat penyakit ini. Oleh karena itu, penting untuk memberikan penanganan yang tepat bagi penderita jantung secara akurat dan dengan biaya yang terjangkau. [4].

Data mining dapat diartikan sebagai proses pengambilan informasi penting yang bersifat implisit dan belum diketahui sebelumnya dari suatu kumpulan data. Proses ini melibatkan pengumpulan serta analisis data historis untuk menemukan keteraturan, pola, dan hubungan dalam set data yang besar [5]. Data mining merupakan metode untuk menganalisis sekumpulan data guna menemukan hubungan yang tak terduga dan menyajikan informasi dengan cara yang baru, sehingga lebih mudah dimengerti dan bermanfaat bagi pemilik data. Disiplin ini menyatukan berbagai bidang ilmu, seperti pembelajaran mesin, pengenalan pola, statistik, manajemen basis data, dan visualisasi, untuk mengatasi tantangan dalam mengekstrak informasi dari database besar [6].

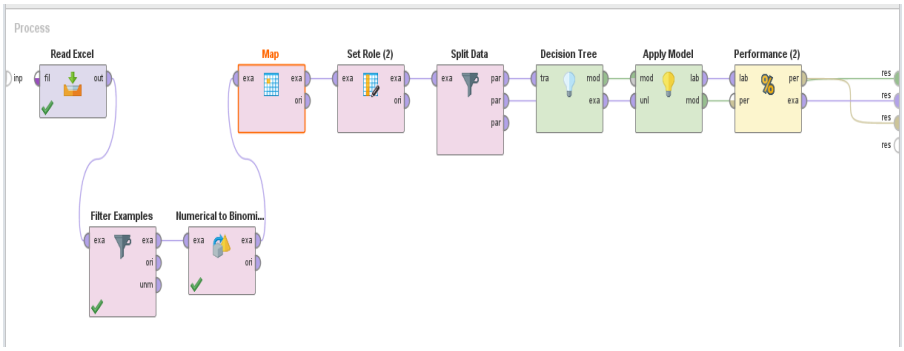
Penelitian ini memanfaatkan algoritma Decision Tree, yang berfungsi sebagai model prediktif dalam data mining. Algoritma ini menggunakan struktur pohon untuk membuat keputusan berdasarkan analisis data dari berbagai variabel. Dalam proses eksplorasi dan analisis data, model ini membantu mengungkap pola, hubungan, dan informasi penting. Keunggulan utama Decision Tree adalah kemampuannya untuk memvisualisasikan keterkaitan antara variabel, sehingga mempermudah pengambilan keputusan dalam situasi yang kompleks [7]. Decision Tree adalah sebuah diagram alir yang mirip dengan struktur pohon. Setiap node internal mewakili pengujian terhadap suatu atribut, sedangkan setiap cabang menandakan hasil dari pengujian tersebut, dan leaf node menunjukkan kelas atau distribusi kelas. Node teratas dikenal sebagai root node, yang berfungsi sebagai classifier untuk mengorganisir kelas-kelas dalam data, dengan representasi kelas-kelas tersebut dalam bentuk pohon. Decision Tree adalah metode machine learning yang dapat menghasilkan model prediktif berdasarkan serangkaian keputusan logis. Dalam bidang kesehatan, penggunaan algoritma Decision Tree dapat membantu dalam mengklasifikasikan risiko seseorang untuk terkena penyakit jantung berdasarkan data yang relevan. [8].

2. METODE PENELITIAN

Data tentang penyakit jantung yang dikumpulkan melalui platform Kaggle (www.kaggle.com) dianalisis menggunakan metode decision tree. Dataset ini terdiri dari 1025 data subjek penelitian ini terdiri dari 14 atribut sekunder dari *dataset* "Klasifikasi Penyakit Jantung". Sebelum menjalankan Decision Tree, dataset melakukan tahap *Data Filter Examples*, selanjutnya *Numerical to Binominal* mengonversi atribut numerik menjadi biner agar sesuai dengan algoritma Decision Tree, kemudian, Map digunakan untuk mengubah atribut sesuai yang ditentukan. *Sat Role* menetapkan peran kolom sebagai target, *Split Data* membagi data menjadi *subset* pelatihan dan pengujian. Langkah terakhir, Decision Tree untuk melatih model dan *Apply Model* menerapkan model ke data pengujian atau data baru, serta Performance mengevaluasi model dengan menghitung metrik seperti akurasi untuk menilai kinerjanya.

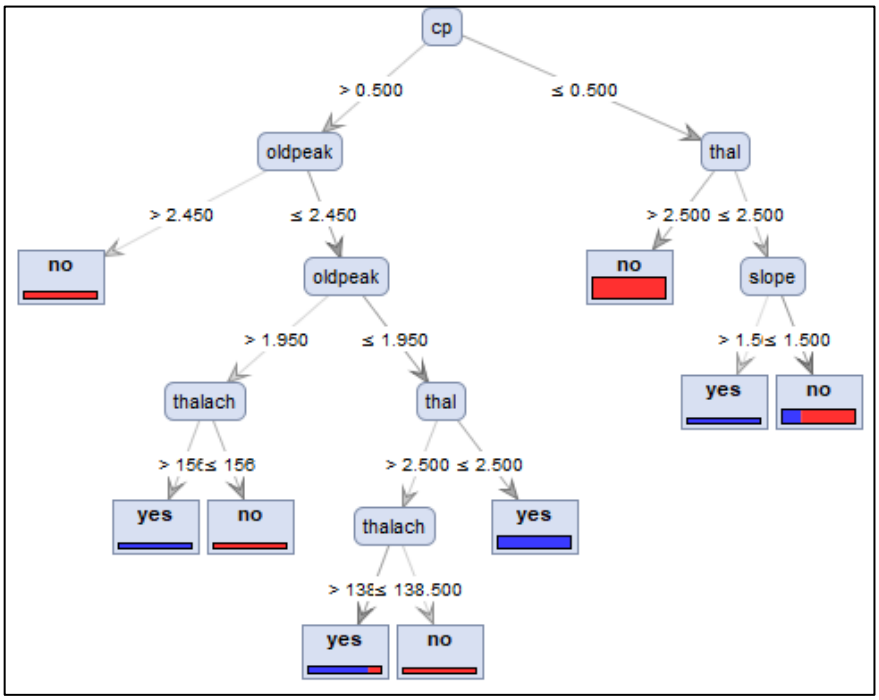
3. HASIL DAN ANALISIS

Gambar 1 ini sebuah alur proses dalam platform RapidMiner, yang biasanya digunakan untuk analisis data dan pembelajaran mesin, digambarkan di sini. Proses dimulai dengan Read Excel, yang mengekstrak data sebagai *input* dari *file* Excel. Setelah itu, contoh-contoh difilter untuk menghilangkan contoh yang tidak diperlukan, dan nilai data diubah dari numerik menjadi *binominal* dengan mengubah dari numerik menjadi *binominal*. Selanjutnya, blok Peta mungkin digunakan untuk memetakan atau mengubah nilai data sebelum disesuaikan perannya melalui Set *Role*. Selanjutnya, blok *split* data digunakan untuk memisahkan data pelatihan dan pengujian. Selanjutnya, model data pelatihan dibuat dengan menggunakan Algoritma Decision Tree. Setelah model dibuat, blok *Apply Model* digunakan untuk menerapkannya pada data eksperimen. Blok terakhir, Performansi, menilai kinerja model dengan menggunakan sejumlah metrik evaluasi



Gambar 1. Model Klasifikasi

Gambar 2 merupakan diagram pohon keputusan yang digunakan untuk membuat keputusan klasifikasi berdasarkan sejumlah variabel. Kita mulai dengan mengukur variabel cp (kemungkinan menyebabkan sakit dada atau nyeri dada) pada puncak pohon dengan batasan tertentu. Setelah itu, data didistribusikan ke percabangan lain. Nilai variabel seperti *oldpeak, thalach, thal,* dan *slope* digunakan oleh setiap percabangan untuk membuat keputusan. Di sebelah kiri diagram, jika nilai cp lebih dari 0.500, percabangan berikutnya didasarkan pada *oldpeak* (yang mungkin merujuk pada depresi segmen ST di elektrokardiogram), tetapi jika *oldpeak* kurang dari 2.450, pohon membagi lebih jauh berdasarkan nilai *thalach* (detak jantung maksimum yang mungkin). Jika nilai ini lebih dari 156, itu diklasifikasikan *yes*, yang menunjukkan indikasi positif; sebaliknya, jika nilainya kurang dari 156, itu diklasifikasikan no. Untuk membuat keputusan, berbagai variabel dipertimbangkan.



Gambar 2. *Decision Tree*

Gambar 3 menunjukkan *confusion matrix* dari hasil model klasifikasi dengan akurasi 87,50%. Tabel tersebut membagi hasil prediksi ke dalam empat kategori utama: prediksi ya (prediksi positif), prediksi no (prediksi negatif), prediksi benar ya (prediksi benar positif), dan prediksi benar tidak (prediksi benar negatif). Dalam hal metrik, ketepatan untuk kelas "*yes*" adalah 77,78%, yang menunjukkan bahwa sekitar 77,78% dari semua prediksi "*yes*" benar-benar positif. Ketepatan untuk kelas "*no*" lebih tinggi, yaitu 93,33 persen.

accuracy: 87.50%

	true yes	true no	class precision
pred. yes	7	2	77.78%
pred. no	1	14	93.33%
class recall	87.50%	87.50%	

Gambar 3. Akurasi

Gambar 4 ini adalah matriks *confusion* untuk model klasifikasi yang menggunakan metrik *recall* dan ketepatan untuk dua kelas: "*yes*" dan "*no*". Dari matriks ini, kita dapat melihat bahwa dari delapan kasus yang sebenarnya adalah "*yes*", model memprediksi 7 di antaranya dengan benar (true positive) dan membuat satu kesalahan dengan memprediksi "*no*" ketika sebenarnya adalah "*yes*" (false negative). Untuk kelas " Dalam hal metrik kinerja, tingkat keakuratan untuk kelas "*yes*" adalah 77,78%, yang menunjukkan bahwa dari semua prediksi "*yes*", 77,78% adalah benar. Sementara tingkat keakuratan untuk kelas "*no*" adalah 93.33%, yang menunjukkan bahwa prediksi "*no*" lebih akurat. *Recall* kedua kelas ini.

recall: 87.50% (positive class: no)

	true yes	true no	class precision
pred. yes	7	2	77.78%
pred. no	1	14	93.33%
class recall	87.50%	87.50%	

Gambar 4. *Recall*

Confusion matrix dengan metrik ketepatan dan *recall* untuk dua kelas, yaitu "*yes*" dan "*no*". Dari 16 kasus yang sebenarnya "*yes*", model memprediksi 14 kasus dengan benar (true negative) dan salah memprediksi 2 sebagai "*yes*" (false positive). Dari perspektif metrik, akurasi untuk kelas "*yes*" adalah 77,78%, yang menunjukkan bahwa dari semua prediksi "*yes*", 77,78% benar-benar merupakan kasus "*yes*". Akurasi untuk kelas "*no*" adalah 93,33%, yang menunjukkan bahwa prediksi "*no*" lebih akurat. *Recall* untuk kedua kelas ini adalah 87,5%, yang menunjukkan bahwa akurasi yang lebih tinggi untuk kelas "

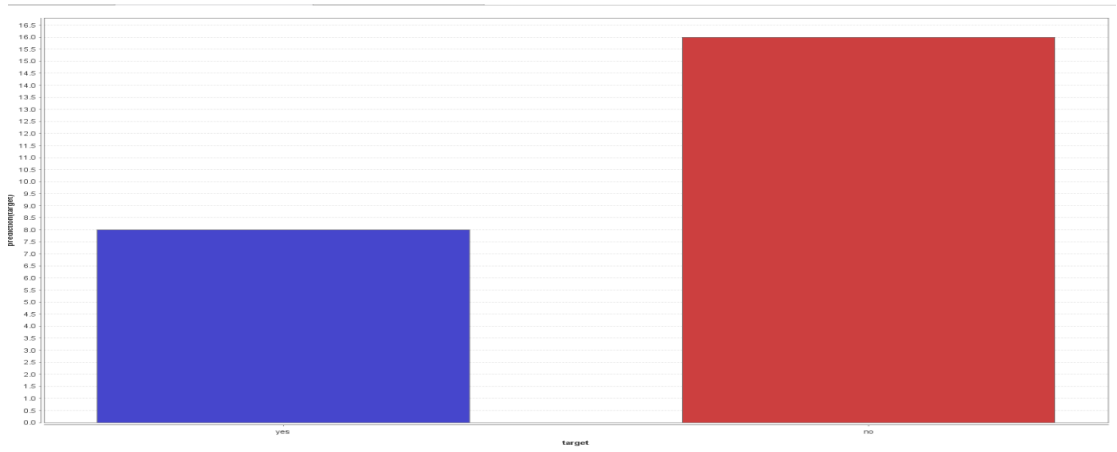
precision: 93.33% (positive class: no)

	true yes	true no	class precision
pred. yes	7	2	77.78%
pred. no	1	14	93.33%
class recall	87.50%	87.50%	

Gambar 5. *Precision*

Grafik batang yang ditampilkan pada Gambar 6 memperlihatkan perbandingan antara dua kategori, yaitu "*yes*" dan "*no*", berdasarkan nilai frekuensi atau jumlah. Pada kategori "*yes*", yang direpresentasikan dengan

batang berwarna biru, terlihat bahwa jumlahnya berada di kisaran 80 hingga 90. Sementara itu, kategori "*no*" yang digambarkan dengan batang berwarna merah memiliki jumlah yang jauh lebih besar, yaitu sekitar 150 hingga 160. Dari perbandingan ini, terlihat jelas bahwa kategori "*no*" memiliki frekuensi hampir dua kali lipat dibandingkan kategori "*yes*". Hal ini mengindikasikan bahwa lebih banyak data, responden, atau hasil yang berada dalam kategori "*no*" dibandingkan dengan "*yes*".



Gambar 6. Grafik Batang

4. KESIMPULAN

Berdasarkan hasil pengujian yang telah dilakukan maka dapat diambil Kesimpulan bahwa Algoritma decision tree efektif untuk prediksi penyakit jantung menunjukkan bahwa algoritma ini dapat memprediksi risiko penyakit jantung dengan akurasi yang tinggi berdasarkan tekanan darah, kolesterol, usia, dan jenis kelamin. Pembersihan data, pemilihan fitur yang relevan, dan pembagian data menjadi subset pelatihan dan pengujian adalah semua bagian dari proses pengolahan data. Selain itu, alat seperti tableau public digunakan untuk menampilkan data secara visual. Hasil evaluasi model yang dikembangkan menunjukkan akurasi prediksi sebesar 87,50%, dengan 93.33% lebih presisi untuk kelas "*no*" dibandingkan dengan 77,78% untuk kelas "*yes*".

REFERENSI

[1] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *Jurnal Media Informatika Budidarma*, vol. 4, no. 2, p. 437, Apr. 2020, doi: 10.30865/mib.v4i2.2080.

[2] D. Haganta Depari *et al.*, "Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung," *Jurnal Informatik Edisi ke*, vol. 18, p. 2022.

[3] A. B. Wibisono and A. Fahrurozi, "PERBANDINGAN ALGORITMA KLASIFIKASI DALAM PENGKLASIFIKASIAN DATA PENYAKIT JANTUNG KORONER," *Jurnal Ilmiah Teknologi dan Rekayasa*, vol. 24, no. 3, pp. 161–170, 2019, doi: 10.35760/tr.2019.v24i3.2393.

[4] Petra Valentino and Siska Narulita, "Performansi Algoritma Decision Tree (C4.5) untuk Prediksi Penyakit Jantung," *Jurnal Cakrawala Informasi*, vol. 3, no. 2, pp. 18–24, Dec. 2023, doi: 10.54066/jci.v3i2.349.

[5] Petra Valentino and Siska Narulita, "Performansi Algoritma Decision Tree (C4.5) untuk Prediksi Penyakit Jantung," *Jurnal Cakrawala Informasi*, vol. 3, no. 2, pp. 18–24, Dec. 2023, doi: 10.54066/jci.v3i2.349.

[6] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 4, no. 2, p. 437, Apr. 2020, doi: 10.30865/mib.v4i2.2080.

[7] Petra Valentino and Siska Narulita, "Performansi Algoritma Decision Tree (C4.5) untuk Prediksi Penyakit Jantung," *Jurnal Cakrawala Informasi*, vol. 3, no. 2, pp. 18–24, Dec. 2023, doi: 10.54066/jci.v3i2.349.

[8] A. Rachmat Raharja, A. Pramudianto, and Y. Muchsam, "Penerapan Algoritma Decision Tree dalam Klasifikasi Data 'Framingham' Untuk Menunjukkan Risiko Seseorang Terkena Penyakit Jantung dalam 10 Tahun Mendatang."

[9] M. Wilda Al -Aluf and Z. Fatah, "Klasifikasi Algoritma Decision Tree Untuk Tingkat Kemiskinan Di Indonesia," *J Comput Sci Technol*, vol. 3, pp. 55–62, doi: 10.59435/jocstec.v3i1.404.

[10] A. Yogianto, A. Homaidi, and Z. Fatah, "Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung," *G-Tech: Jurnal Teknologi Terapan*, vol. 8, no. 3, pp. 1720–1728, Jul. 2024, doi: 10.33379/gtech.v8i3.4495.