PENERAPAN ALGORITMA DECISION TREE BERBASIS POHON KEPUTUSAN DALAM KLASIFIKASI PENYAKIT JANTUNG

Rindi Antika, Ahmad Rifa'I, Fatihanursari Dikananda, Dendy Indriya Efendi, Riri Narasati
Teknik Informatika, STMIK IKMI Cirebon
Jln Perjuangan No 10 B Karyamulya Kesambi Kota Cirebon
rindian2107@gmail.com

ABSTRAK

Deteksi dini penyakit jantung merupakan tantangan besar dalam dunia medis karena seringkali terkait dengan rendahnya akurasi dalam mengklasifikasikan kondisi jantung. Banyak orang baru menyadari bahwa mereka menderita penyakit jantung ketika telah mencapai tahap yang sangat parah. Kondisi ini mengakibatkan penanganan medis terlambat dan berpotensi membahayakan nyawa. Rendahnya akurasi dalam mengklasifikasikan penyakit jantung menciptakan kesulitan, dimana model atau algoritma klasifikasi mungkin kesulitan membedakan dengan tepat antara jenis-jenis penyakit jantung yang berbeda. Pengembangan teknologi diagnostik merupakan faktor penting dalam meningkatkan kemampuan untuk mendeteksi kondisi jantung lebih awal. Oleh karena itu, upaya untuk meningkatkan akurasi dalam mengklasifikasikan penyakit jantung memegang peranan besar dalam penanganan yang lebih efektif. Studi ini akan menerapkan pendekatan *Knowledge Discovery in Databases (KDD)* yang melibatkan serangkaian langkah sistematis dalam melakukan analisis data. Dari hasil penerapan algoritma *Decision Tree* dalam klasifikasi penyakit jantung diperoleh *Accuracy* sebesar 93.44%, *Recall* sebesar 75%, *Precision* sebesar 90% dan *F1-Score* sebesar 81.81%. Dalam penelitian selanjutnya diharapkan untuk mencoba eksplorasi metode *tuning hyperparameter* pada algoritma *decision tree* untuk melihat apakah peningkatan performa lebih lanjut dapat dicapai. Dalam penelitian selanjutnya diharapkan mencoba eksplorasi penggunaan algoritma lain seperti *Random Forest, Support Vector Machines (SVM),* atau *Neural Networks* untuk melihat apakah terdapat peningkatan performa dalam klasifikasi penyakit jantung.

Kata kunci: *Decision tree*, Klasifikasi, Penyakit jantung, *Cleveland Clinic Foundation*, *Knowledge Discovery in Databases (KDD)*

1. PENDAHULUAN

Penyakit jantung merupakan faktor utama yang menyebabkan kematian di berbagai belahan dunia, termasuk di Indonesia. Statistik tahun 2020 dari PERKI menunjukkan bahwa penyakit jantung menyumbang sekitar 36% dari total kematian global, melebihi angka kematian akibat kanker dengan tingkat dua kali lipat lebih tinggi. Di Indonesia sendiri, penyakit jantung menjadi penyebab utama kematian sebanyak 26,4% [1]. Upaya pencegahan dan perawatan yang tepat sangat diperlukan untuk mengurangi dampak dan insiden penyakit jantung di Indonesia.

Deteksi dini penyakit jantung adalah tantangan medis besar karena rendahnya akurasi klasifikasi. Banyak orang menyadari penyakit ini pada tahap yang parah, menyebabkan penanganan terlambat dan risiko nyawa. Rendahnya akurasi klasifikasi menciptakan kesulitan dalam membedakan jenis-jenis penyakit jantung. Pengembangan teknologi diagnostik penting untuk deteksi dini. Meningkatkan akurasi klasifikasi penyakit jantung berperan besar dalam penanganan yang lebih efektif.

Penelitian oleh Agus Oka Gunawan tahun 2023 menggunakan data set jantung dengan 11 atribut fitur dan 1 atribut label dalam klasifikasi penyakit jantung [2]. Algoritma yang digunakan pada penelitian yang dilakukan oleh Jannah tahun 2023 adalah algoritam Decision Tree. Algoritma ini digunakan untuk membuat model yang mudah dipahami, fleksibel, dan dapat divisualisasikan [3]. Hasil penelitian Nurmansyah tahun 2022 menunjukkan bahwa algoritma decision tree mampu secara efisien meramalkan keberadaan penyakit jantung dengan tingkat ketepatan sebesar 75%. Meski demikian, terdapat potensi untuk meningkatkan akurasi dengan menggunakan metode yang lebih canggih dan dataset yang lebih luas [4].

Penelitian ini dimaksudkan untuk meningkatkan tingkat ketepatan dalam mengklasifikasikan penyakit jantung dengan menguji performa algoritma Decision Tree pada dataset khusus. Evaluasi eksperimental akan dilakukan untuk menganalisis hasil percobaan dan meningkatkan akurasi identifikasi pasien berisiko terkena penyakit jantung. Tujuan utamanya adalah berkontribusi pada peningkatan diagnosis awal dan manajemen penyakit jantung.

Penelitian menggunakan algoritma Decision Tree, yang merupakan model prediktif dalam data mining. Decision Tree memanfaatkan struktur pohon untuk membuat keputusan dari analisis data variabel. Dalam eksplorasi dan analisis data, model ini membantu mengidentifikasi pola, hubungan, dan informasi penting. Keunggulan Decision Tree adalah kemampuannya memvisualisasikan keterkaitan variabel, memudahkan pengambilan keputusan dalam kompleksitas data. Model ini dimulai dari simpul tunggal dan bercabang sesuai pilihan yang ada, kunci dalam memahami dan mengoptimalkan hasil analisis penelitian ini.

Hasil penelitian ditujukan kepada komunitas ilmiah dan praktisi kesehatan untuk memperluas pengetahuan dalam klasifikasi penyakit jantung. Informasi ini dapat meningkatkan metode diagnostik, perawatan, dan manajemen penyakit, serta mendorong pengembangan teknologi kesehatan yang lebih canggih untuk deteksi dan penanganan dini penyakit jantung. Diharapkan juga meningkatkan kesadaran masyarakat terhadap signifikansi deteksi dini penyakit jantung serta mendorong pengembangan teknologi diagnostik yang lebih unggul di waktu yang akan datang dalam usaha pencegahan dan penanganan penyakit jantung.

2. TINJAUAN PUSTAKA

Pada tahap ini, akan dilakukan pemaparan mengenai literatur ilmiah yang memiliki relevansi tinggi dengan fokus penelitian yang sedang dilakukan. Proses ini mencakup analisis mendalam terhadap berbagai studi, artikel, serta sumber-sumber teoritis terkait yang bertujuan untuk mendukung, memperluas, dan mengonfirmasi arah penelitian yang sedang dijalankan. Tujuan utama dari tahap ini adalah untuk menguraikan dengan seksama landasan teoritis yang dapat menambah wawasan dan memberikan dukungan yang kuat bagi perjalanan penelitian yang sedang berlangsung.

2.1. Data Mining

Data mining merupakan rangkaian langkah untuk mengeksplorasi dan menganalisis data dalam skala besar dengan tujuan menemukan pola atau Teknik, metode, atau algoritma yang sesuai sangat tergantung pada sasaran dan keseluruhan proses penemuan pengetahuan dalam basis data. Proses *data mining* melibatkan pencarian atau penggalian informasi menarik dalam data yang besar dengan menggunakan teknik atau metode tertentu. [5] [6].

2.2. Decision Tree

Pohon keputusan merupakan suatu struktur diagram alur berbentuk pohon, di mana setiap simpul internal (node non-daun) merepresentasikan pengujian terhadap atribut tertentu. Setiap cabang pada pohon menggambarkan hasil dari pengujian tersebut, sementara setiap simpul daun (simpul terminal) menyimpan informasi label kelas. Simpul teratas dalam struktur pohon merupakan simpul akar [7].

2.3. Klasifikasi

Klasifikasi adalah metode untuk mendefinisikan sekumpulan pola atau properti yang mendeskripsikan dan menunjukkan perbedaan satu kategori informasi dari yang lain guna menentukan apakah suatu objek perilaku dan atribut kelompok yang ditentukan. Klasifikasi juga menunjukkan hubungan antar atribut [8].

2.4. *Confusion Matrix*

Confusion matrix merupakan metode pengujian yang digunakan untuk menilai performa suatu algoritma pembelajaran mesin. Hal ini dilakukan dengan membandingkan prediksi yang dihasilkan oleh algoritma terhadap data sebenarnya yang telah diketahui sebelumnya. *Confusion matrix* memberikan gambaran yang jelas tentang sejauh mana algoritma dapat memprediksi dengan benar setiap kelas atau label pada dataset, serta membantu dalam mengidentifikasi sejumlah metrik evaluasi seperti akurasi, presisi, recall, dan lainnya yang mendalam memahami kinerja algoritma dalam melakukan klasifikasi [9].

Tabel 1. Confusion Matrix

		Kelas Sebenarnya	
		True	*False*
Kelas Prediksi	*True*	TP	FP
	False	FN	TN

Tabel 1 merupakan *Confusion Matrix* yang dapat digunakan dalam klasifikasi suatu model atau algoritma. *Confusion Matrix* memiliki empat elemen utama yaitu, Positif Benar (TP), Negatif Benar (TN), Positif Salah (FP), dan Negatif Salah (FN). Empat elemen yang dapat dijelaskan sebagai yang berikut:

a. Positif Benar (TP): Jumlah data dengan nilai Positif yang diprediksi dengan benar sebagai Positif.
b. Negatif Salah (FP): Jumlah data dengan nilai Negatif yang salah diprediksi sebagai Positif.
c. Negatif Benar (TN): Jumlah data dengan nilai Negatif yang diprediksi dengan benar sebagai Negatif.
d. Positif Salah (FN): Jumlah data dengan nilai Positif yang salah diprediksi sebagai Negatif.

Penilaian menggunakan *Confusion Matrix* memberikan nilai Akurasi, *Recall*, Presisi, dan *F1-Score*. Perhitungan dapat dijelaskan sebagai berikut:

a. *Accuracy* adalah sebuah ukuran yang mendefinisikan seberapa dekatnya nilai prediksi dari suatu model dengan nilai aktual atau sebenarnya dari data yang diamati. Ini mencerminkan sejauh mana model atau algoritma mampu mengidentifikasi dengan tepat dan akurat kelas atau nilai yang sebenarnya dalam suatu dataset. Ukuran ini memberikan gambaran mengenai tingkat keakuratan model dalam melakukan prediksi, dengan membandingkan seberapa dekat prediksi yang dihasilkan dengan fakta yang sebenarnya dalam data.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

b. Recall adalah model untuk mengidentifikasi secara akurat semua instance yang benar positif dalam dataset. Secara spesifik, *Recall* dapat dianggap sebagai rasio antara jumlah item relevan yang berhasil diidentifikasi oleh model (*True Positive*)

dan jumlah keseluruhan item relevan yang sebenarnya ada dalam dataset (*True Positive* dan *False Negative*). Semakin tinggi nilai *Recall*, semakin sedikit item yang relevan yang terlewatkan oleh model.

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

c. *Precision* diartikan sebagai perbandingan antara jumlah jumlah item yang relevan yang terpilih dibandingkan dengan total item yang terpilih. *Precision* adalah indikator seberapa tepat suatu sistem dalam mengidentifikasi item yang relevan dari keseluruhan item yang dipilih. Metrik ini mencerminkan sejauh mana sistem memberikan jawaban yang akurat dan relevan terhadap kebutuhan informasi yang diminta.

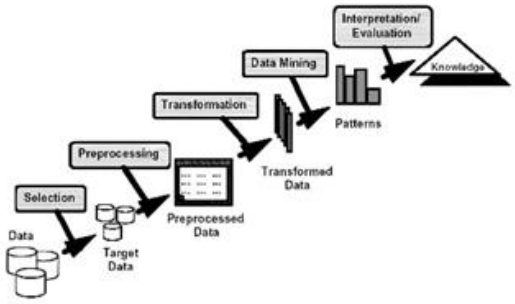
$$Precision = \frac{TP}{TP+FP} \quad (3)$$

d. *F1-Score* merupakan metrik yang mengkombinasikan *Precision* dan *Recall* dalam sebuah nilai tunggal. Hal ini dilakukan dengan mengambil *harmonic mean* dari *Precision* dan *Recall*. *Harmonic mean* memberikan penekanan pada nilai terkecil di antara keduanya. *F1-Score* berguna untuk memberikan gambaran yang seimbang antara *Precision* dan *Recall*, semakin tinggi nilai *F1-Score*, semakin baik model dalam melakukan klasifikasi yang seimbang antara *Precision* dan *Recall*.

$$F1-Score = \frac{2(Precision*Recall)}{Precision+Recall} \quad (4)$$

3. METODE PENELITIAN

Studi ini akan menerapkan pendekatan *Knowledge Discovery in Databases* (KDD), sebuah metode komprehensif yang melibatkan serangkaian langkah dalam menggali, menganalisis, dan mengeksplorasi data yang besar dan kompleks. Merupakan proses yang sistematis dan terstruktur dalam mengekstrak data berharga dimana sebelumnya tidak terdeteksi atau tidak diketahui secara langsung dari dalam kumpulan data besar. Tujuan utamanya adalah untuk menemukan pola, tren, atau pengetahuan yang dapat memberikan wawasan baru yang berguna dan potensial untuk diaplikasikan dalam berbagai bidang [10]. Berikut adalah langkah-langkah yang akan diambil dalam proses penerapan metode KDD ini:



Gambar 1. Tahapan KDD [11]

Gambar 1 menggambarkan langkah-langkah dalam *Knowledge Discovery in Databases* (KDD), yang mencakup Pemilihan, Pra-pemrosesan, Transformasi, Pertambangan Data, dan Interpretasi / Evaluasi.

3.1. Selection

Pada langkah ini, proses pemilihan atribut dataset menjadi fokus utama. Seleksi data menjadi suatu kebutuhan karena tidak semua atribut yang tersedia dalam dataset digunakan dalam proses pengolahan data mining. Langkah ini mempertimbangkan relevansi, kegunaan, dan kontribusi setiap atribut terhadap analisis yang sedang dilakukan, sehingga hanya atribut yang paling signifikan dan relevan yang dipertahankan untuk proses selanjutnya. Hal ini bertujuan untuk mengoptimalkan kualitas dan efisiensi analisis *data mining* yang akan diterapkan.

3.2. Preprocessing

Pada tahap ini, dilakukan penanganan terhadap nilai-nilai yang hilang (missing value) dan seleksi atribut pada data. Tujuannya adalah untuk memperbaiki kualitas data dengan mengisi atau menghapus nilai yang hilang serta memilih atribut yang paling relevan dan signifikan. Proses ini bertujuan untuk meningkatkan kualitas data dan memastikan bahwa atribut yang dipilih secara tepat dapat membantu meningkatkan nilai akurasi dalam analisis dan pemodelan yang akan dilakukan selanjutnya.

3.3. Transformation

Pada langkah ini, dilakukan proses penting yang melibatkan transformasi data ke dalam format yang tepat dan cocok untuk proses eksplorasi dan analisis data yang disebut data mining. Proses ini bertujuan untuk menyesuaikan struktur dan karakteristik data agar dapat diaplikasikan secara optimal dalam algoritma dan teknik analisis yang diterapkan dalam data mining.

3.4. Data Mining

Pada proses ini, terjadi pembagian dataset yang telah disiapkan menjadi dua bagian, yakni data training dan data testing, dengan proporsi pembagian sebesar 8:2. Setelah itu, dilakukan penerapan algoritma

Decision Tree pada dataset yang berisi informasi tentang penyakit jantung. Langkah ini bertujuan untuk melakukan analisis dan pembelajaran pada data latih (training data) menggunakan algoritma Decision Tree guna membuat model yang dapat memprediksi atau mengklasifikasikan data uji (testing data) dengan akurat.

3.5. Interpretation/Evaluation

Pada langkah ini, dilakukan evaluasi menyeluruh terhadap hasil penerapan algoritma Decision Tree pada dataset yang relevan. Evaluasi ini mencakup analisis mendalam terhadap kinerja algoritma, seperti akurasi, presisi, recall, dan f1-score guna memahami seberapa baik algoritma tersebut berfungsi dalam konteks pemodelan dan klasifikasi data.

4. HASIL DAN PEMBAHASAN

4.1. Selection

Pada tahap ini, dilakukan penggunaan dataset yang berasal dari situs web https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset. Dataset ini terdiri dari data mengenai penyakit jantung yang mencakup sejumlah 303 entri dengan 14 atribut yang tersedia dalam format Comma Separated Values (CSV). Salah satu atribut yang menjadi fokus sebagai label dalam penelitian ini adalah 'Target'. Dataset ini menjadi dasar pengembangan dan evaluasi model dalam analisis klasifikasi penyakit jantung yang dilakukan dalam penelitian.

4.2. Preprocessing

Pada tahap ini akan dilakukan penanganan *missing value*, namun pada dataset yang digunakan tidak ditemukan adanya *missing value*, kemudian selanjutnya akan dilakukan pemilihan atribut, atribut pada dataset penyakit jantung sebanyak 14 atribut yaitu *Age*, *Sex*, *Cp*, *Trestbps*, *Chol*, *Bps*, *Restecg*, *Thalach*, *Exang*, *Oldepak*, *Slope*, *Ca*, *Thal*, dan *Target*, namun atribut yang akan digunakan pada penelitian ini hanya 6 atribut yaitu *chol*, *exang*, *slope*, *thalach*, *trestbps* dan *target*. Atribut yang diplih adalah atribut yang berpengaruh terhadap gejala-gejala yang paling dialami oleh penderita penyakit jantung.

Tabel 2. Preprocessing Penyakit Jantung

Chol	*Exang*	*Slope*	*Thalach*	*Trestbps*	*Target*
233	0	3	150	145	0
286	1	2	108	160	1
229	1	2	129	120	0
250	0	3	187	130	0
204	0	1	172	130	0
236	0	1	178	120	0
268	0	3	160	140	1
354	1	1	163	120	0
254	0	2	147	130	1
203	1	3	155	140	0
192	0	2	148	140	0
294	0	2	153	140	0
256	1	2	142	130	1
265	0	1	173	120	0
199	0	1	162	172	0

Tabel 2 merupakan data penyakit jantung setelah dilakukan *preprocessing* pemilihan 6 atribut.

4.3. Transformation

Transformation data merupakan tahapan yang dilakukan untuk mengubah data yang belum sesuai ke dalam bentuk yang sesuai dan siap untuk diproses data mining. Pada dataset penyakit jantung dilakukan transformasi pada atribut *target* yaitu dari bentuk angka 1 & 0 diubah menjadi *yes* & *no*.

Tabel 3. Transformation Penyakit Jantung

Chol	*Exang*	*Slope*	*Thalach*	*Trestbps*	*Target*
233	0	3	150	145	*No*
286	1	2	108	160	*Yes*
229	1	2	129	120	*No*
250	0	3	187	130	*No*
204	0	1	172	130	*No*
236	0	1	178	120	*No*
268	0	3	160	140	*Yes*
354	1	1	163	120	*No*
254	0	2	147	130	*Yes*
203	1	3	155	140	*No*
192	0	2	148	140	*No*
294	0	2	153	140	*No*
256	1	2	142	130	*Yes*
265	0	1	173	120	*No*
199	0	1	162	172	*No*

Tabel 3 merupakan data penyakit jantung setelah dilakukan transformasi data.

4.4. Data Mining

Setelah melakukan beberapa tahapan yang diantaranya data selection, preprocessing, dan transformation maka tahap selanjutnya adalah data mining. Teknik pemodelan data mining yang dipilih adalah klasifikasi menggunakan algoritam Decision Tree. Klasifikasi algoritma decision tree digunakan untuk mencapai tujuan awal yaitu meningkatkan akurasi dalam klasifikasi penyakit jantung. Berikut penerapan data mining dalam klasifikasi penyakit jantung menggunakan algoritma decision tree.

4.5. Interpretation/Evaluation

Pada tahap ini dilakukan evaluasi terhadap model yang telah dibuat, bentuk evaluasi yang dilakukan adalah dengan menghitung nilai *accuracy*, *precision*, *recall* dan *f1-scroe* dari algoritma *Decision Tree* menggunakan *Confusin Matrix*.

Tabel 4. Confusion Matrix Decision Tree

	True No	**True Yes**
Pred. No	48	3
Pred. Yes	1	9

Berdasarkan hasil *Confusion Matrix* pada tabel 4 dapat dilihat bahwa sebanyak 9 data diklasifikasikan "*Yes*" dengan *valid*, kemudian sebanyak 48 data diklasifikasikan "*No*" dengan *valid*.

Setelah diketahui hasil klasifikasi penyakit jantung menggunakan algoritma Decision Tree, maka dapat diketahui nilai Accuracy, Recall, Precision dan F1-Score.

Tabel 5 Klasifikasi Algoritma Decision Tree

Accuracy	0.9344
Recall	0.75
Precision	0.9
F1-Score	0.8181

Dari hasil penerapan algoritma *Decision Tree* dalam klasifikasi penyakit jantung diperoleh *Accuracy* sebesar 93.44%, *Recall* sebesar 75%, *Precision* sebesar 90% dan *F1-Score* sebesar 81%.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian menunjukkan bahwa penerapan algoritma *decision tree* memberikan performa yang baik dalam mengklasifikasikan penyakit jantung. Dengan akurasi sebesar 93.44%, hasil tersebut menunjukkan tingkat keakuratan yang tinggi dalam memprediksi kasus-kasus penyakit jantung dalam dataset yang digunakan. Selain itu, nilai *recall* sebesar 75% menunjukkan kemampuan model dalam mengidentifikasi secara efektif jumlah kasus penyakit jantung yang sebenarnya positif. *Precision* sebesar 90% mengindikasikan bahwa dari kasus yang diprediksi positif oleh model, sebagian besar benar-benar merupakan kasus penyakit jantung. Sementara itu, nilai *F1-score* sebesar 81.81% menggambarkan kombinasi yang seimbang antara *recall* dan *precision*. Kesimpulannya, penerapan algoritma *decision tree* dalam klasifikasi penyakit jantung berhasil memberikan hasil yang menjanjikan dengan performa yang baik dalam memprediksi serta mengidentifikasi kasus penyakit jantung dalam dataset yang digunakan

DAFTAR PUSTAKA

[1] D. Larassati, A. Zaidiah, and S. Afrizal, "Sistem prediksi penyakit jantung koroner menggunakan metode naïve bayes," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 07, no. 02, p. 546, 2022, [Online]. Available: https://www.jurnal.stkippgritulungagung.ac.id/index.php/jipi/article/view/2842

[2] I. M. Agus Oka Gunawan, I. D. A. Indah Saraswati, I. D. G. Riswana Agung, and I. P. Eka Putra, "Klasifikasi Penyakit Jantung Menggunakan Algoritma Decision Tree Series C4.5 Dengan Rapidminer," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 5, no. 2, pp. 73–83, 2023, doi: 10.47233/jteksis.v5i2.775.

[3] S. S. N. Jannah, N. F. Riska, and ..., "Pengklasifikasian Penyakit Jantung dengan Metode Decision Tree," *E-Prosiding …*, 2023, [Online]. Available: https://prosidingsnsa.statistics.unpad.ac.id/?journal=prosidingsnsa&page=article&op=view&path%5B%5D=348

[4] D. Nurmansyah, "Prediksi Penyakit Jantung Menggunakan Algoritma Decision Tree," no. June, pp. 1–4, 2022, [Online]. Available: https://www.researchgate.net/publication/361510915

[5] M. R. Sulistio, N. Suarna, and O. Nurdiawan, "Analisa Penerapan Metode Clustering X-Means Dalam Pengelompokan Penjualan Barang," *J. Teknol. Ilmu Komput.*, vol. 1, no. 2, pp. 37–42, 2023, doi: 10.56854/jtik.v1i2.49.

[6] R. R. Khaerullah, N. Suarna, and O. Nurdiawan, "Analisa Pengelompokan Dataset Komputer Menggunakan Algoritma X-Means," *J. Inform. dan Teknol. Inf.*, vol. 1, no. 2, pp. 125–132, 2023, doi: 10.56854/jt.v1i2.135.

[7] E. E. Fauziah and A. F. Zulfikar, "Penerapan Metode Decision Tree Menggunakan Algoritma Iterative Dichotomiser 3 (ID3) Untuk Klasifikasi Resiko Penyakit Jantung," *OKTAL J. Ilmu Komput. dan Sci.*, vol. 2, no. 4, p. 1219, 2023, [Online]. Available: https://journal.mediapublikasi.id/index.php/oktal/article/view/1302

[8] O. N. Washilaturrizqi, "Implementasi Algoritma C4 . 5 Untuk Menentukan Penerima," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 373–377, 2023.

[9] F. Handayani, "Komparasi Support Vector Machine, Logistic Regression Dan Artificial Neural Network Dalam Prediksi Penyakit Jantung," *JEPIN (Jurnal Edukasi dan Penelit. Inform.*, vol. 7, no. 3, p. 334, 2021, [Online]. Available: https://jurnal.untan.ac.id/index.php/jepin/article/view/48053

[10] F. Firmansyah and O. Nurdiawan, "Penerapan Data Mining Menggunakan Algoritma Frequent Pattern - Growth Untuk Menentukan Pola Pembelian Produk Chemicals," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 547–551, 2023, doi: 10.36040/jati.v7i1.6371.

[11] N. Meilani and O. Nurdiawan, "Data Mining untuk Klasifikasi Penderita Kanker Payudara Menggunakan Algoritma K-Nearest Neighbor," *J. Wahana Inform.*, vol. 2, no. 1, pp. 177–187, 2023, [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer.