

Implementasi Teknik Data Mining untuk Prediksi Kanker Paru – Paru Menggunakan Algoritma Decision Tree C4.5

Angela Putri Rudhiyanto^{1*}, Najla Damayanti², Esthi Ayu Anggita³, Ananda Putri Maharani⁴, Amanda Chickita Aprilia Juanda⁵

¹⁻⁵Teknologi Informasi, Universitas Bina Sarana Informatika, Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta
E-mail: 19232161@bsi.ac.id

* Corresponding Author

 <https://doi.org/10.31004/jerkin.v4i3.4976>

ARTICLE INFO

Article history

Received: 23 Nov 2025

Revised: 05 Dec 2025

Accepted: 30 Dec 2025

Kata Kunci:

Machine Learning, Regresi Linear, Prediksi Nilai Akhir, Kehadiran, Nilai Tugas

Keywords:

Machine Learning, Linear Regression, Final Grade Prediction, Attendance, Assignment Score

ABSTRACT

Penelitian ini bertujuan untuk menerapkan teknik data mining dalam memprediksi penyakit kanker paru-paru menggunakan algoritma Decision Tree C4.5. Dataset yang digunakan adalah *Lung Cancer Dataset* dengan jumlah 309 data dan 16 atribut yang mencakup faktor demografi, gaya hidup, serta gejala klinis pasien. Proses penelitian dilakukan menggunakan perangkat lunak RapidMiner melalui beberapa tahapan, yaitu pengumpulan data, persiapan data, pembuatan model, serta pengujian dan evaluasi model. Algoritma C4.5 membangun pohon keputusan berdasarkan nilai *Gain Ratio* untuk menentukan atribut yang paling berpengaruh terhadap klasifikasi kanker paru-paru. Hasil penelitian menunjukkan bahwa algoritma Decision Tree C4.5 mampu menghasilkan performa klasifikasi yang sangat baik dengan nilai akurasi sebesar 96,76%, *classification error* sebesar 3,24%, nilai Kappa sebesar 0,844, *weighted mean recall* sebesar 89,86%, dan *weighted mean precision* sebesar 94,93%. Pohon keputusan yang dihasilkan menunjukkan bahwa atribut ALLERGY menjadi faktor dominan dalam proses klasifikasi, diikuti oleh atribut lain seperti *Yellow_Finger*, *Peer_Pressure*, dan *Swallowing_Difficulty*. Berdasarkan hasil tersebut, dapat disimpulkan bahwa algoritma Decision Tree C4.5 efektif dan interpretatif dalam memprediksi kanker paru-paru serta berpotensi digunakan sebagai alat bantu analisis awal dalam mendukung pengambilan keputusan medis.

This study aims to implement data mining techniques to predict lung cancer using the Decision Tree C4.5 algorithm. The dataset used in this research is the Lung Cancer Dataset, consisting of 309 records and 16 attributes covering demographic factors, lifestyle habits, and clinical symptoms of patients. The research process was conducted using RapidMiner software through several stages, including data collection, data preprocessing, model construction, and model evaluation. The C4.5 algorithm generates a decision tree based on the Gain Ratio criterion to identify the most influential attributes in lung cancer classification. The results show that the Decision Tree C4.5 algorithm achieved excellent classification performance with an accuracy of 96.76%, a classification error of 3.24%, a Kappa value of 0.844, a weighted mean recall of 89.86%, and a weighted mean precision of 94.93%. The generated decision tree indicates that the ALLERGY attribute is the most dominant factor in classifying lung cancer, followed by other attributes such as Yellow_Finger, Peer_Pressure, and Swallowing_Difficulty. These findings indicate that the Decision Tree C4.5 algorithm is effective and highly interpretable for lung cancer prediction and has strong potential as an early decision-support tool in medical diagnosis.



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

How to Cite: Angela Putri Rudhiyanto, et al (2025). Implementasi Teknik Data Mining untuk Prediksi Kanker Paru – Paru Menggunakan Algoritma Decision Tree C4.5, 4(3) 16663-16669. <https://doi.org/10.31004/jerkin.v4i3.4976>

PENDAHULUAN

Penyakit kanker paru-paru merupakan masalah kesehatan serius sehingga diperlukan pendekatan berbasis data untuk membantu proses klasifikasi penyakit [1]. Penyakit paru-paru masih menjadi salah satu penyebab utama morbiditas dan mortalitas di Indonesia. Kanker paru-paru merupakan salah satu penyakit yang memiliki angka kematian tinggi, dan keterlambatan dalam diagnosis sering terjadi karena banyak pasien tidak menyadari keberadaan penyakit ini. Oleh karena itu, diperlukan pendekatan berbasis data, seperti teknik data mining, untuk membantu proses prediksi dan deteksi dini secara lebih cepat dan sistematis [2]. Pendekatan data mining, terutama algoritma Decision Tree C4.5, telah terbukti efektif dalam deteksi penyakit paru-paru, dengan C4.5 menunjukkan akurasi tinggi dan rendahnya tingkat false negative, sehingga potensial untuk digunakan sebagai alat pendukung pengambilan keputusan medis di sistem kesehatan [3].

Teknik data mining dengan algoritma Decision Tree banyak digunakan karena menghasilkan model yang mudah dipahami dan efektif dalam klasifikasi penyakit paru-paru menggunakan data sekunder dengan bantuan RapidMiner [4].

Algoritma C4.5 dipilih karena merupakan salah satu algoritma klasifikasi pohon keputusan yang umum digunakan dan efektif dibandingkan algoritma lain seperti CART dan ID3. C4.5 mampu menganalisis data gejala pasien dan menghasilkan model decision tree yang mudah dipahami, sehingga sesuai untuk menganalisis kemungkinan terjadinya kanker paru-paru berdasarkan data medis yang tersedia [5].

Penelitian ini bertujuan untuk menerapkan algoritma Decision Tree C4.5 dalam mengklasifikasikan penyakit kanker paru-paru serta mengevaluasi performa model yang dihasilkan dalam memprediksi risiko kanker paru-paru berdasarkan data medis. Evaluasi dilakukan melalui analisis kuantitatif menggunakan metrik kinerja seperti akurasi, precision, recall, dan nilai Kappa untuk menilai keandalan model dalam proses klasifikasi. [6]

METODE

Penelitian ini menggunakan pendekatan data mining untuk melakukan klasifikasi penyakit kanker paru-paru. Algoritma yang digunakan adalah Decision Tree C4.5, karena mampu menghasilkan model klasifikasi yang mudah dipahami serta efektif dalam pengambilan keputusan medis.

Metode yang Digunakan

Metode penelitian yang diterapkan adalah data mining dengan teknik klasifikasi. Data mining adalah proses untuk mendapatkan data dan mengungkap informasi atau kaidah yang tersembunyi di dalamnya. Istilah data mining muncul pertama kali pada tahun 1990 di bidang ilmu komputer. Data mining juga dikenal dengan berbagai nama lain, seperti knowledge discovery in database (KDD), knowledge extraction, pattern analysis, information harvesting, dan business intelligence [7]. *Data mining can be a valuable technique to improve knowledge management in an organization. It can help identify patterns, perform predictive analytics, discover hidden knowledge, and improve the efficiency of the knowledge management process.* [8]

Dalam konteks data mining, terkandung satu tahap yang dikenal sebagai klasifikasi. Klasifikasi merupakan suatu prosedur yang dapat membantu pengelompokan data [7] yang kemudian berfungsi untuk mendefinisikan kelas data ditemukan. Selama proses klasifikasi, data diorganisasikan dalam bentuk pohon keputusan.

Algoritma Decision Tree C4.5 merupakan salah satu algoritma klasifikasi yang banyak digunakan karena mampu menghasilkan model yang mudah dipahami dan bersifat interpretatif. Algoritma ini bekerja dengan membangun pohon keputusan berdasarkan atribut yang paling berpengaruh terhadap kelas target. Pemilihan atribut terbaik dilakukan menggunakan nilai Gain Ratio, yaitu pengembangan dari Information Gain yang mempertimbangkan tingkat pembagian data agar lebih optimal.

Pada penelitian ini, algoritma C4.5 diterapkan pada dataset kanker paru-paru yang berisi data pasien dengan berbagai atribut seperti demografi, gaya hidup, dan gejala klinis. Proses implementasi dilakukan menggunakan perangkat lunak RapidMiner. Dataset dimuat ke dalam RapidMiner menggunakan operator *Retrieve*, kemudian atribut LUNG_CANCER ditetapkan sebagai label menggunakan operator *Set Role*. Selanjutnya, algoritma Decision Tree (C4.5) dijalankan untuk membangun model klasifikasi berdasarkan perhitungan Gain Ratio pada setiap atribut.

Hasil dari proses ini berupa model pohon keputusan yang menggambarkan hubungan antar atribut medis dan gaya hidup pasien terhadap status kanker paru-paru. Model tersebut digunakan untuk mengklasifikasikan data pasien ke dalam dua kelas, yaitu YES dan NO, sehingga dapat membantu proses prediksi kanker paru-paru secara sistematis dan terstruktur.

Data yang Dipakai

Data yang digunakan dalam penelitian ini memiliki spesifikasi sebagai berikut:

1. Dataset yang digunakan adalah Lung Cancer Dataset
2. Jumlah data yang digunakan sebanyak 309 data
3. Dataset terdiri dari 16 atribut pasien
4. Atribut target (label) adalah LUNG_CANCER
5. Kelas label terdiri dari dua kategori, yaitu YES dan NO

Tahapan penelitian

Tahapan penelitian yang dilakukan dalam penelitian ini adalah sebagai berikut:

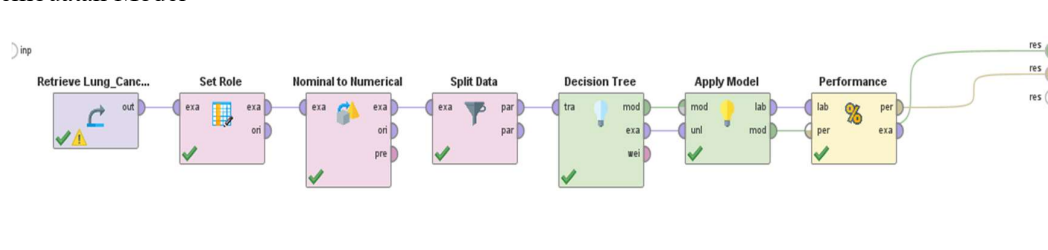
1. Pengumpulan Data

Pada tahap ini dilakukan pengambilan dataset kanker paru-paru yang digunakan sebagai data penelitian. Dataset yang diperoleh merupakan data sekunder yang berisi informasi pasien, mencakup faktor demografi, gaya hidup, serta gejala klinis yang berkaitan dengan kanker paru-paru.

2. Persiapan Data

Tahap persiapan data meliputi proses penyesuaian dan pengolahan data agar siap digunakan dalam proses data mining. Kegiatan pada tahap ini mencakup pemeriksaan struktur data, penentuan atribut yang digunakan, serta penetapan atribut target (label) LUNG_CANCER dengan kelas YES dan NO.

3. Pembuatan Model



Gambar 1. Model Decision Tree C4.5

Pembuatan model klasifikasi dilakukan menggunakan perangkat lunak RapidMiner dengan menerapkan algoritma Decision Tree C4.5. Proses diawali dengan memuat Lung Cancer Dataset menggunakan operator *Retrieve*, kemudian atribut LUNG_CANCER ditetapkan sebagai label melalui operator *Set Role*. Selanjutnya, data bertipe nominal dikonversi menjadi numerik menggunakan operator *Nominal to Numerical* agar dapat diproses oleh algoritma. Dataset kemudian dibagi menjadi data latih dan data uji menggunakan operator *Split Data* untuk keperluan pelatihan dan pengujian model. Pada tahap berikutnya, algoritma Decision Tree C4.5 dijalankan menggunakan kriteria *Gain Ratio* untuk membentuk pohon keputusan yang merepresentasikan pola klasifikasi kanker paru-paru. Model yang dihasilkan selanjutnya diterapkan pada data uji menggunakan operator *Apply Model*, dan performa model dievaluasi menggunakan operator *Performance* untuk mengetahui tingkat akurasi klasifikasi.

4. Pengujian dan Evaluasi

Tahap pengujian model dilakukan pada operator *Apply Model* dengan menerapkan model Decision Tree C4.5 pada data uji, sedangkan evaluasi hasil pengujian dilakukan menggunakan operator *Performance* untuk mengukur akurasi klasifikasi.

Tools yang Digunakan

Tools yang digunakan dalam penelitian ini adalah aplikasi RapidMiner, yang dimanfaatkan untuk melakukan pengolahan data, pembuatan model klasifikasi, serta pengujian dan evaluasi algoritma Decision Tree C4.5.

HASIL DAN PEMBAHASAN

Deskripsi Data

Dataset yang digunakan pada penelitian ini adalah Lung Cancer Dataset yang berjumlah 309 baris data. Dataset ini terdiri dari 16 atribut yang mencakup informasi demografi, gaya hidup, dan gejala klinis pasien. Atribut target (label) adalah LUNG_CANCER dengan dua kategori yaitu “YES” dan “NO”. Struktur dataset ini sejalan dengan penelitian sebelumnya yang memanfaatkan atribut medis dan gaya hidup untuk klasifikasi penyakit paru - paru menggunakan algoritma Decision Tree C4.5 [9]

No	Atribut	Keterangan
1	GENDER	Jenis kelamin pasien (Male/Female)
2	AGE	Usia pasien
3	SMOKING	Riwayat kebiasaan merokok pasien
4	YELLOW_FINGER	Kondisi perubahan warna jari menjadi kuning
5	ANXIETY	Kondisi kecemasan yang dialami pasien
6	PEER_PRESSURE	Pengaruh lingkungan atau tekanan dari rekan sebaya
7	CHRONIC_DISEASE	Riwayat penyakit kronis yang dimiliki pasien
8	FATIGUE	Tingkat kelelahan yang sering dialami pasien
9	ALLERGY	Riwayat alergi yang dimiliki pasien
10	WHEEZING	Gejala bunyi napas mengi saat bernapas
11	ALCOHOL_CONSUMING	Kebiasaan pasien dalam mengonsumsi alkohol
12	COUGHING	Gejala batuk yang dialami pasien
13	SHORTNESS OF BREATH	Gejala sesak napas
14	SWALLOWING DIFFICULTY	Kesulitan pasien dalam menelan
15	CHEST PAIN	Rasa nyeri pada area dada
16	LUNG CANCER	Status diagnosis kanker paru - paru (Label)

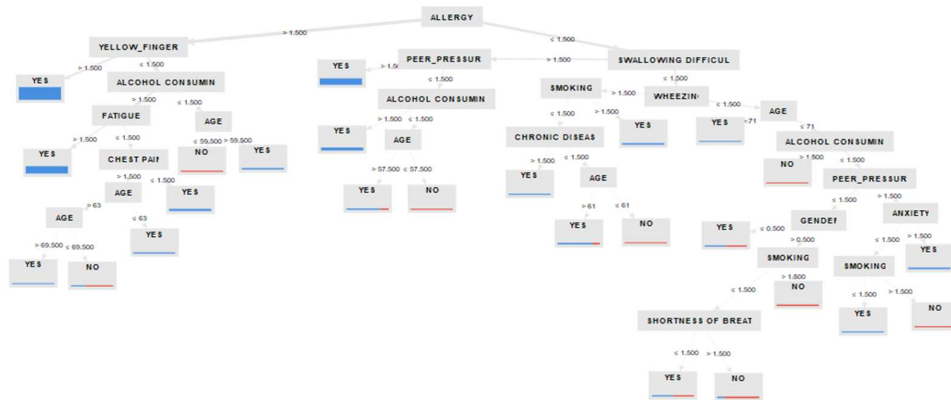
Implementasi Model C4.5

Proses klasifikasi dilakukan menggunakan perangkat lunak RapidMiner dengan mengimplementasikan algoritma Decision Tree C4.5 menggunakan kriteria Gain Ratio. Tahapan proses meliputi *Retrieve Data*, *Set Role* (menetapkan LUNG_CANCER sebagai label), dan *Apply Model* untuk pengujian. Pendekatan ini sejalan dengan metode yang digunakan dalam penelitian [10] yang menyatakan bahwa algoritma C4.5 efektif digunakan dalam proses klasifikasi kanker paru - paru berbasis data medis

Hasil Analisis Pohon Keputusan (Decision Tree)

Algoritma C4.5 menghasilkan sebuah pohon keputusan yang memetakan pola hubungan antar variabel medis secara hierarkis. Berdasarkan visualisasi pohon keputusan yang dihasilkan, ditemukan bahwa atribut ALLERGY menjadi node akar, yang menunjukkan bahwa variabel tersebut memiliki pengaruh paling besar dalam proses klasifikasi kanker paru-paru. Hasil ini sejalan dengan penelitian (Fatah & Wahid, 2025) yang menyatakan bahwa atribut medis tertentu dapat menjadi faktor dominan dalam klasifikasi penyakit paru - paru menggunakan algoritma C4.5

Setelah atribut ALLERGY, pohon keputusan bercabang ke beberapa atribut lain seperti YELLOW_FINGER, PEER_PRESSURE, dan SWALLOWING_DIFFICULTY. Selanjutnya, proses klasifikasi dipengaruhi oleh faktor gaya hidup seperti SMOKING dan ALCOHOL_CONSUMING, serta gejala klinis seperti WHEEZING, FATIGUE, dan CHEST_PAIN hingga menghasilkan keputusan akhir berupa status kanker paru-paru



Gambar 2. Model Pohon Keputusan (Decision Tree) Algoritma C4.5

Berdasarkan pohon keputusan yang terbentuk, dapat diambil beberapa aturan keputusan, di antaranya:

1. IF ALLERGY = Yes AND YELLOW_FINGER = Yes THEN LUNG_CANCER = Yes
2. IF ALLERGY = No AND SWALLOWING_DIFFICULTY = No THEN LUNG_CANCER = No
3. IF ALLERGY = Yes AND PEER_PRESSURE = Yes AND ALCOHOL_CONSUMING = Yes THEN LUNG_CANCER = Yes
4. IF ALLERGY = Yes AND YELLOW_FINGER = No AND SMOKING = Yes THEN LUNG_CANCER = Yes
5. IF ALLERGY = No AND PEER_PRESSURE = No AND CHEST_PAIN = No THEN LUNG_CANCER = No

Aturan - aturan ini menunjukkan bahwa kombinasi antara faktor kesehatan dan gaya hidup berperan penting dalam menentukan klasifikasi kanker paru -paru, sebagaimana juga ditemukan pada penelitian sebelumnya yang menggunakan algoritma C4.5 bidang medis [10]

Perlu dicatat bahwa tidak semua atribut pada dataset, seperti COUGHING, muncul pada pohon keputusan akhir. Hal ini disebabkan oleh nilai Gain Ratio yang lebih rendah dibandingkan atribut lainnya, sehingga atribut tersebut tidak dipilih sebagai node pemisah utama dalam proses pembentukan pohon keputusan menggunakan algoritma C4.5

Evaluasi Performa Model

Evaluasi performa model dilakukan menggunakan operator *Performance* pada perangkat lunak RapidMiner. Berdasarkan hasil pengujian, algoritma Decision Tree C4.5 menghasilkan akurasi sebesar 96,76% dengan classification error sebesar 3,24% yang menunjukkan bahwa model memiliki tingkat kesalahan klasifikasi yang rendah dan mampu melakukan prediksi dengan baik.

Nilai Kappa sebesar 0,844 menunjukkan tingkat kesesuaian yang sangat baik antara hasil prediksi model dan data aktual, sehingga model dapat dikatakan memiliki reliabilitas yang tinggi. Selain itu, nilai weighted mean recall sebesar 89,86% menunjukkan bahwa model mampu mendeteksi sebagian besar kasus kanker paru-paru dengan baik, sedangkan nilai weighted mean precision sebesar 94,93% menunjukkan bahwa sebagian besar hasil prediksi positif yang dihasilkan oleh model adalah benar.

Secara keseluruhan, hasil evaluasi ini menunjukkan bahwa algoritma Decision Tree C4.5 memiliki performa klasifikasi yang baik dan stabil dalam menganalisis data kanker paru-paru. Hasil penelitian ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa algoritma C4.5 efektif digunakan dalam klasifikasi penyakit kanker paru-paru berbasis data medis [10]

accuracy: 96.76%

	true YES	true NO	class precision
pred. YES	214	6	97.27%
pred. NO	2	25	92.59%
class recall	99.07%	80.65%	

Gambar 3. Nilai Akurasi Model C4.5

classification_error: 3.24%

	true YES	true NO	class precision
pred. YES	214	6	97.27%
pred. NO	2	25	92.59%
class recall	99.07%	80.65%	

Gambar 4. Nilai Classification Error Model C4.5

kappa: 0.844

	true YES	true NO	class precision
pred. YES	214	6	97.27%
pred. NO	2	25	92.59%
class recall	99.07%	80.65%	

Gambar 5. Nilai Kappa Model Decision Tree C4.5

weighted_mean_recall: 89.86%, weights: 1, 1

	true YES	true NO	class precision
pred. YES	214	6	97.27%
pred. NO	2	25	92.59%
class recall	99.07%	80.65%	

Gambar 6. Nilai Weighted Mean Recall Model Decision Tree C4.5

weighted_mean_precision: 94.93%, weights: 1, 1

	true YES	true NO	class precision
pred. YES	214	6	97.27%
pred. NO	2	25	92.59%
class recall	99.07%	80.65%	

Gambar 7. Weighted Mean Precision Model Decision Tree C4.5

SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan metode data mining dengan algoritma Decision Tree C4.5 mampu memberikan performa yang baik dalam mengklasifikasikan penyakit kanker paru-paru. Proses penelitian dilakukan melalui beberapa tahapan, mulai dari input data, eksplorasi data (Exploratory Data Analysis), penanganan missing value, pengecekan data duplikat, analisis korelasi antar fitur, pemisahan fitur dan label, pembagian data latih dan data uji, hingga implementasi dan evaluasi model klasifikasi.

Hasil evaluasi model menggunakan confusion matrix menunjukkan bahwa algoritma Decision Tree C4.5 menghasilkan nilai akurasi sebesar 89%, dengan recall 90%, precision 86%, dan F1-score 88%, yang menandakan bahwa model memiliki kemampuan yang baik dalam mengklasifikasikan data kanker paru-paru. Hasil ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa algoritma C4.5 efektif dalam menganalisis data pasien dan mendukung proses pengambilan keputusan medis secara cepat dan akurat, khususnya dalam memperkirakan risiko kanker paru-paru berdasarkan gejala dan faktor gaya hidup pasien.

Selain itu, penggunaan algoritma C4.5 terbukti mampu memanfaatkan atribut demografi, kebiasaan hidup, dan gejala klinis pasien secara signifikan dalam membangun model prediksi kanker paru-paru. Model yang dihasilkan bersifat mudah dipahami karena berbentuk pohon keputusan, sehingga berpotensi digunakan sebagai alat bantu dalam proses deteksi dini dan evaluasi awal kondisi pasien sebelum dilakukan pemeriksaan medis lanjutan seperti X-Ray, CT Scan Thorax, maupun pemeriksaan lainnya.

Meskipun demikian, penelitian ini masih memiliki keterbatasan, terutama pada jumlah dan variasi data yang digunakan. Oleh karena itu, penelitian selanjutnya disarankan untuk melakukan tuning hyperparameter, menambah jumlah data pasien, serta mengintegrasikan algoritma C4.5 dengan metode lain atau teknologi di bidang kesehatan guna meningkatkan performa dan keandalan model dalam mendukung praktik klinis sehari-hari.

UCAPAN TERIMA KASIH

Peneliti menyampaikan ucapan terima kasih kepada pihak yang sudah berkontribusi dalam pelaksanaan penelitian dan penyusunan artikel ini.

REFERENSI

- Z. Fatah, “Jurnal Ilmiah Multidisiplin Nusantara Implementasi Metode Decision Tree Dalam Prediksi Kanker Paru Paru Dengan Rapidminer Jurnal Ilmiah Multidisiplin Nusantara,” vol. 2, no. November, pp. 176–184, 2024.
- B. Andriksa, C. Permana, and M. Djamaluddin, “Penerapan Python Dalam Data Mining Untuk Prediksi Kanker Paru Organisasi Kesehatan Dunia (WHO) menyatakan bahwa kanker merupakan kelompok penyakit yang berasal dari hampir seluruh organ tubuh dimana sel-sel yang terdapat pada organ tubuh tersebut tumb,” vol. 6, no. 2, 2023.
- L. Kurniawati, D. Priyanto, N. Sulistianingsih, and M. Syahrir, “Perbandingan Metode Berbasis Decision Tree untuk Mendeteksi Penyakit Paru Comparison of Decision Tree-Based Methods in Lung Disease Detection,” vol. 7, no. 1, pp. 51–62, 2025, doi: 10.30812/bite.v7i1.4909.
- Z. F. Aviatus Sholiha, “Klasifikasi Penyakit Paru-Paru Menggunakan Data Mining Decision Tree,” vol. 4, no. April, pp. 46–51, 2025.
- D. A. Pratama, I. R. Mutaqin, and K. R. Manuela, “Analisis Terjadinya Kanker Paru-Paru Pada Pasien Menggunakan Decision Tree : Penerapan Algoritma C4 . 5 Dan RapidMiner Untuk Menentukan Risiko Kanker Pada Gejala Pasien Deigo Anugrah Pratama Program Studi Sistem Informasi FTI Universitas Bina Sarana Infor,” vol. 2, no. 4, 2023.
- V. Kumar, D. Gupta, S. Juneja, and S. Kumari, “Multi-model machine learning framework for lung cancer risk prediction : A comparative analysis of nine classifiers with hybrid and ensemble approaches using behavioral and hematological parameters,” *SLAS Technol.*, vol. 33, no. June, p. 100314, 2025, doi: 10.1016/j.slast.2025.100314.
- C. A. Maulana, R. P. Pratama, and B. O. Lubis, “Penentuan Sparepart Genset Paling Sering Digunakan Pada Operator Indosat Ooredoo Hutchison Dengan,” vol. 5, no. 1, pp. 300–309, 2024.
- J. Duque, “ScienceDirect ScienceDirect ScienceDirect Data Mining for Knowledge Management Data Mining for Knowledge Duque * Management,” *Procedia Comput. Sci.*, vol. 239, no. 2022, pp. 257–264, 2024, doi: 10.1016/j.procs.2024.06.170.
- M. F. Wahid, “Penerapan Data Mining Menggunakan Algoritma C4 . 5 untuk Klasifikasi Penyakit Paru-Paru Universitas Ibrahimy,” vol. 3, no. 5, pp. 646–653, 2025.
- K. Jainudin and A. Abdullah, “Klasifikasi Penyakit Kanker Paru-Paru Menggunakan Metode,” vol. 8, no. 3, pp. 232–240, 2025.