

Klasifikasi Penyakit Paru-Paru Menggunakan Algoritma Decision Tree Studi Kasus Pada Data Medis

¹Zaehol Fatah, ²Alviatur Rizqiyah

^{1,2}Sistem Informasi, Universitas Ibrahimy, Situbondo

Email : ¹sukorejobergerak@gmail.com ^{2*}alviaturrizqiyah42@gmail.com

Abstrak

Penyakit paru-paru dan paru-paru adalah salah satu masalah kesehatan yang penting, dan membutuhkan diagnosis yang cepat dan tepat agar bisa membantu dokter dalam mengambil keputusan. Tujuan penelitian ini adalah untuk membuat sistem yang menggunakan Decision Tree untuk mengklasifikasikan penyakit paru-paru, yang merupakan pendekatan yang mudah dipahami dan efisien. Data yang digunakan berasal dari sumber terbuka dan telah diproses terlebih dahulu, yaitu membersihkan data, normalisasi, dan memilih fitur dengan metode Chi-Square. Alasan memilih Decision Tree adalah karena kemampuannya dalam membuat model yang bisa dijelaskan oleh para dokter dan mampu menangani berbagai jenis data. Hasil uji coba menunjukkan bahwa model ini mampu mengenali jenis penyakit paru-paru dengan tingkat akurasi yang baik, serta menampilkan gambar pohon keputusan yang membuat proses diagnosis lebih transparan. Penelitian ini menunjukkan kemungkinan menggunakan Decision Tree sebagai alat bantu dalam sistem bantuan keputusan medis, khususnya dalam klasifikasi penyakit paru-paru, dan membuka peluang untuk dikembangkan lebih lanjut dengan sistem berbasis kecerdasan buatan di bidang kesehatan.

Kata Kunci: *Penyakit Paru-paru, Algoritma Decision Tree, Klasifikasi Medis, Seleksi Fitur, Sistem Pendukung Keputusan.*

Abstract

Lung disease is one of the significant health problems and requires rapid and accurate diagnosis to assist doctors in decision-making. This study aims to develop a system that can classify lung diseases using the Decision Tree method, which is an easily understandable and efficient approach. The data used comes from open sources and has been preprocessed, including data cleaning, normalization, and feature selection using the Chi-Square method. The reason for choosing Decision Tree is its ability to create models that can be explained to doctors and handle various types of data. The experimental results show that this model can accurately recognize types of lung diseases and display a decision tree diagram, making the diagnosis process more transparent. This research demonstrates the potential use of a Decision Tree as an aid in medical decision support systems, particularly in classifying lung diseases, and opens up opportunities for further development with artificial intelligence-based systems in the healthcare field.

Keyword: *Lung Disease, Decision Tree Algorithm, Medical Classification, Feature Selection, Decision Support System.*

PENDAHULUAN

Penyakit paru-paru adalah penyebab utama kematian di berbagai negara, terutama di negara-negara yang masih berkembang. Di sana, akses ke layanan kesehatan dan peralatan diagnosis yang canggih masih terbatas. Menurut World Health Organization (WHO, 2024), jumlah orang yang terkena dan meninggal karena kanker paru-paru terus meningkat setiap

tahunnya. Kondisi ini membutuhkan upaya pengecekan dini dan pengelompokan jenis penyakit yang tepat agar para tenaga medis dapat membuat keputusan yang lebih baik dan pengobatan lebih efektif.

Perkembangan teknologi informasi membawa kemudahan dalam penerapan data mining dan kecerdasan buatan (AI) di bidang medis, terutama untuk membantu diagnosis secara otomatis. Salah satu

algoritma yang sering digunakan adalah Decision Tree, karena mudah diartikan dan memberikan model klasifikasi yang jelas bagi para dokter. Penelitian yang dilakukan oleh Khairudin, Machfud, dan Cahyono (2025) menunjukkan bahwa algoritma C4.5 Decision Tree mampu mengklasifikasikan penyakit kanker paru-paru dengan akurasi sebesar 89%, recall 90%, dan F1-score 88% (Salim dan Sugeng, 2018). Penelitian tersebut memanfaatkan dataset publik, meskipun sumber data tidak dijelaskan secara rinci. Jika penelitian ini menggunakan dataset yang sama, maka hasil yang diperoleh diharapkan lebih baik melalui peningkatan proses preprocessing, pemilihan atribut, dan optimasi parameter algoritma.

Penelitian yang dilakukan oleh Marzuq, Wicaksono, dan Setiawan (2023) menggunakan kombinasi algoritma Decision Tree dan Random Forest untuk meningkatkan akurasi klasifikasi hingga 90,4%, serta menekankan pentingnya visualisasi model dalam memahami hasil diagnosis klinis.

Sementara itu, penelitian internasional oleh Krishnaiah, Narsimha, dan Chandra (2013) membuktikan bahwa algoritma Decision Tree efektif dalam menganalisis gejala pasien untuk mendukung diagnosis awal kanker paru-paru.

Selanjutnya, Krishna et al. (2024) menggabungkan Decision Tree dengan CNN VGG16 pada citra CT-scan, dan berhasil mencapai akurasi hingga 99,67%, menunjukkan bahwa kombinasi antara pendekatan simbolik dan visual dapat secara signifikan meningkatkan kinerja klasifikasi. Penelitian-penelitian tersebut sebagian besar menggunakan dataset publik, baik berupa data klinis maupun citra medis. Apabila penelitian ini menggunakan dataset yang sama, maka hasil yang diperoleh diharapkan lebih baik melalui peningkatan pada proses preprocessing, optimasi parameter model, dan pemilihan fitur yang lebih relevan.

Di Indonesia, penelitian oleh (Deigo Anugrah Pratama, Ibnu Rizal Mutaqin dan

Kevin Rafael Manuela, 2023) menggunakan Decision Tree dengan RapidMiner untuk menentukan risiko kanker paru-paru berdasarkan gejala pasien, dan mendapatkan hasil akurasi yang tinggi. Penelitian oleh (Larassati, Zaidiah dan Afrizal, 2022) menunjukkan bahwa Decision Tree bisa digunakan untuk mendeteksi penyakit pernapasan dengan dataset lokal Indonesia, sampai mendapatkan akurasi 88,5%. Penelitian lain oleh (Aulia dan Adelia, 2025) mengaplikasikan algoritma Decision Tree untuk mendeteksi pneumonia pada anak-anak dengan data rekam medis, dan menunjukkan bahwa metode ini bisa mempercepat proses diagnosis secara signifikan.

Selain itu, studi dari (Dinova dan Purnama, no date) menunjukkan bahwa penerapan AI berbasis Decision Tree bisa membantu dokter menilai tingkat keparahan penyakit paru secara lebih objektif. Penelitian dari (Jatnika Fahmi Idris, Rafid Ramadhani dan Muhammad Malik Mutoffar, 2024) juga mendukung hal ini dengan membuat sistem bantuan keputusan diagnosis TBC berbasis Decision Tree dengan akurasi mencapai 91%.

Berdasarkan berbagai penelitian yang dilakukan, dapat disimpulkan bahwa algoritma Decision Tree memiliki keunggulan dalam hal kemudahan dalam diinterpretasi, kecepatan dalam mengolah data, dan tingkat ketepatan hasil yang tinggi. Karena itu, metode ini sangat cocok untuk digunakan dalam proses klasifikasi penyakit paru-paru sebagai bagian dari sistem bantuan pengambilan keputusan medis yang berbasis teknologi kecerdasan buatan.

METODE

Metode penelitian ini menggunakan pendekatan kuantitatif untuk membuat model klasifikasi penyakit paru-paru berdasarkan algoritma decision tree. Penelitian ini terdiri dari lima tahap utama, yaitu: mengumpulkan data, memproses

data sebelum digunakan, memilih fitur yang relevan, membangun model, dan mengevaluasi kinerja model tersebut.



Gambar 1. Tahapan Penelitian

Pengumpulan Data

Dalam penelitian ini, data yang digunakan berasal dari dataset publik yang tersedia di platform Kaggle, dengan alamat: <https://www.kaggle.com/datasets/andot03b/src/dataset-predic-terkena-penyakit-paruparu>. Pemilihan dataset dari Kaggle dilakukan karena platform tersebut menyediakan data yang terbuka, terverifikasi, dan banyak digunakan dalam penelitian akademik maupun pengembangan model pembelajaran mesin. Selain itu, dataset Kaggle umumnya memiliki format yang terstruktur, lengkap, serta mudah diakses, sehingga memudahkan proses pra-pemrosesan data (data preprocessing) dan pembangunan model klasifikasi. Penggunaan dataset publik juga memungkinkan replikasi hasil penelitian dan perbandingan kinerja model dengan studi-studi sebelumnya yang menggunakan data serupa.

Dataset ini dibuat untuk membantu proses prediksi gangguan paru-paru dengan mempertimbangkan berbagai atribut sosial dan klinis dari pasien, seperti usia, jenis kelamin, kebiasaan merokok, kegiatan olahraga, serta riwayat penyakit lainnya. Dataset ini terdiri dari lebih dari 300 data pasien, dengan setiap data mewakili satu individu. Program Microsoft Excel 2021 digunakan untuk mengubah format data.csv menjadi bentuk yang siap diproses, kemudian dilanjutkan dengan pengolahan menggunakan bahasa pemrograman Python. Pada proses tersebut, beberapa pustaka seperti pandas, scikit-learn, dan matplotlib digunakan untuk melakukan analisis serta pembuatan model.

Tabel 1. Deskripsi Atribut Dataset Pasien

Nama Atribut	Jenis Data	Deskripsi Singkat
Usia	Numerik	Umur pasien dalam tahun
Jenis_Kelamin	Kategorikal	Laki-laki / Perempuan
Merokok	Biner	Ya / Tidak (status perokok aktif)
Bekerja	Biner	Ya / Tidak (status pekerjaan)
Rumah_Tangga	Biner	Ya / Tidak (status sebagai anggota rumah tangga aktif)
Aktivitas_Begadang	Biner	Ya / Tidak (kebiasaan tidur larut malam)
Aktivitas_Olahraga	Biner	Ya / Tidak (kebiasaan berolahraga rutin)
Asuransi	Biner	Ya / Tidak (kepemilikan asuransi kesehatan)
Penyakit_Bawaan	Kategorikal	Riwayat komorbid seperti diabetes, hipertensi, dll.
Hasil	Label Kelas	Diagnosis akhir (Normal, TB, PPOK, dll.)

Praproses Data

Sebelum model akan dilatih, terlebih dahulu dilakukan proses klasifikasi dan pengelolaan data melalui tahap praproses. Tahap ini bertujuan untuk memastikan data memiliki kualitas baik dan siap digunakan dalam analisis. Berikut adalah tahapan praproses yang dilakukan:

a. Pembersihan Data

Dalam tahap ini, dilakukan identifikasi dan perbaikan masalah pada data. Contohnya, mengatasi data yang hilang, data yang tidak akurat, serta menghapus data yang duplikat. Dalam penelitian ini, data yang bermasalah tidak digunakan. Tujuan dari tahap ini adalah memastikan bahwa data yang digunakan memiliki kualitas tinggi dan dapat diandalkan (Illah, Jauharis Sapu dan Damaliana, 2024).

b. Transformasi Data

Transformasi data adalah tahap penting untuk mengubah bentuk data agar sesuai dengan metode data mining. Setelah data diberi label, data berbentuk kategori dimodifikasi menjadi bentuk numerik. Proses ini menggunakan Label Encoding, yaitu metode yang mengubah data berbentuk teks menjadi bentuk angka.

Dengan cara ini, data teks dapat diproses lebih lanjut dalam pembuatan model prediksi (Gori, Sunyoto dan Al Fatta, 2024).

c. Pengodean Label Target

Variabel hasil, yaitu atribut yang menjadi target, dikodekan dalam bentuk biner. Misalnya, "Ya" diberi nilai 1 (menunjukkan adanya penyakit), sedangkan "Tidak" diberi nilai 0 (tidak ada penyakit). Jika ada lebih dari satu jenis penyakit, maka dilakukan pengodean multikelas.

d. Validasi Tipe Data dan Skala

Tipe data setiap atribut dicek agar sesuai dengan kebutuhan algoritma klasifikasi. Jika algoritma yang digunakan sensitif terhadap skala data, seperti KNN atau SVM, maka data numerik seperti usia dinormalisasi menggunakan metode Min-Max Scaling atau Standardization.

Semua tahapan tersebut dilakukan sebelum proses pelatihan model menggunakan pustaka scikit-learn untuk memastikan bahwa data yang digunakan sudah bersih, terstruktur, dan dapat diproses secara efektif.

Pemilihan Fitur

Seleksi fitur adalah cara memilih beberapa variabel atau fitur yang paling penting dan berpengaruh dalam memperkirakan atau menjelaskan suatu hal yang diamati. Tujuannya adalah menyederhanakan model, memudahkan pemahaman terhadap hasil, serta menghindari masalah overfitting (Illah, Jauharis Sapu dan Damaliana, 2024).

Untuk menentukan atribut yang paling relevan dalam klasifikasi penyakit paru-paru, dilakukan seleksi fitur dengan metode Information Gain. Metode ini dipilih karena mampu mengukur seberapa besar kontribusi suatu fitur dalam mengurangi tingkat ketidakpastian (entropi) terhadap kelas yang menjadi label. Teknik ini cocok digunakan pada data yang terdiri dari fitur kategorikal atau biner, seperti Merokok, Jenis_Kelamin, dan Aktivitas_Olahraga.

a. Rumus Information Gain

Perhitungan Information Gain (IG) didasarkan pada perubahan entropi sebelum dan setelah data dibagi berdasarkan suatu atribut. Rumus Information Gain adalah sebagai berikut:

$$Gain(C, A) = Entropy(C) - \sum_{v \in Values(A)} \frac{|C_v|}{|C|} \times Entropy(C_v)$$

b. Tahapan Implementasi

Implementasi perhitungan Information Gain dilakukan menggunakan bahasa pemrograman Python dengan pustaka pandas dan scikit-learn. Berikut tahapan proses seleksi fitur:

1. Melakukan label encoding pada atribut kategorikal agar bisa diproses oleh algoritma klasifikasi.
2. Menghitung entropi awal dataset berdasarkan distribusi kelas pada atribut Hasil.
3. Setelah data dibagi, menghitung entropi berdasarkan nilai setiap atribut.
4. Menghitung nilai Information Gain masing-masing atribut dengan mengurangi entropi awal dan entropi setelah split.
5. Mengurutkan atribut berdasarkan nilai Information Gain terbesar.

c. Hasil Perhitungan Information Gain

Information Gain adalah teknik yang digunakan untuk menyaring fitur yang relevan dan mengurangi jumlah dimensi fitur dalam data yang akan dianalisis (Putri et al., 2025). Berikut adalah hasil perhitungan nilai Information Gain untuk setiap fitur dalam dataset:

Tabel 2. Hasil Seleksi Fitur Berdasarkan Informasi Gain

No	Atribut	Entropi Setelah Split	Informa tion Gain
1	Merokok	0.5177	0.4810
2	Jenis_Kelamin	0.7790	0.2196
3	Bekerja	0.8577	0.1409
4	Penyakit_Bawaan	0.8723	0.1263
5	Aktivitas_Olahra ga	0.8835	0.1152
6	Asuransi	0.9761	0.0225
7	Rumah_Tangga	0.9900	0.0087
8	Usia	0.9962	0.0024
9	Aktivitas_Begada ng	0.9976	0.0011

a. Analisis dan Interpretasi

Dari hasil tabel di atas, atribut Merokok adalah fitur yang paling berpengaruh dalam memengaruhi hasil klasifikasi penyakit paru-paru dengan nilai Information Gain sebesar 0.4810. Atribut lain seperti Jenis_Kelamin, Penyakit_Bawaan, dan Aktivitas_Olahraga juga memberikan pengaruh yang cukup besar terhadap prediksi kelas. Atribut dengan nilai Information Gain rendah, seperti Aktivitas_Begadang, Usia, dan Rumah_Tangga, mungkin bisa dihilangkan agar model lebih sederhana dan lebih efisien dalam proses klasifikasi.

b. Kesesuaian dengan Literatur

Hasil penelitian ini sesuai dengan penelitian yang dilakukan oleh Rilo Pambudi, Abdul Rahman Harahap, Farhan Dwitama Saputra, dan Muhamad Jusub (Pambudi et al., 2024), yang menggunakan algoritma decision tree berbasis C4.5 dalam klasifikasi penyakit paru-paru. Dalam penelitian tersebut, atribut Merokok dan Jenis_Kelamin juga ditemukan sebagai fitur utama dalam memengaruhi hasil klasifikasi.

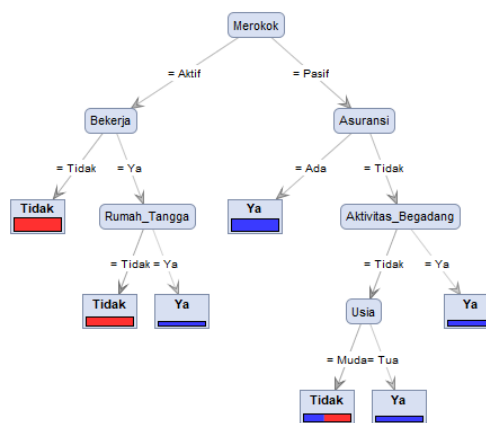
Pembangunan Model

Model klasifikasi penyakit paru-paru dibuat menggunakan algoritma Decision Tree di RapidMiner Studio. Dataset yang sudah melalui tahap praproses dimasukkan ke dalam proses analisis, dengan atribut Hasil ditetapkan sebagai label kelas.

Langkah-langkah pembuatan model meliputi:

1. Mengimpor data menggunakan operator Read Excel.
2. Menetapkan label menggunakan Set Role.
3. Melakukan konversi atribut kategorikal dengan Nominal to Binomial.
4. Menerapkan algoritma Decision Tree dengan kriteria pemisahan Information Gain dan pengaturan maksimum kedalaman pohon.
5. Menggunakan Apply Model and Performance (Classification) untuk mengevaluasi performa model.

Setelah model dilatih, RapidMiner menghasilkan visualisasi pohon keputusan yang menunjukkan alur klasifikasi berdasarkan atribut paling informatif. Pohon tersebut dapat dilihat di bawah ini:



Gambar 2. Visualisasi Pohon Keputusan

Evaluasi Model

Evaluasi model bertujuan untuk melihat seberapa baik akurasi dan keandalan algoritma Decision Tree dalam mengklasifikasikan data penyakit paru-paru. Evaluasi dilakukan di RapidMiner menggunakan operator Performance (Classification) yang menghasilkan beberapa metrik utama seperti akurasi, precision, recall, dan F1-score. Dari hasil evaluasi diperoleh nilai sebagai berikut:

1. Akurasi mencapai 89,47%, artinya model mampu mengklasifikasikan sebagian besar data dengan benar.
2. Precision mencapai 90%, menunjukkan bahwa mayoritas prediksi positif oleh model benar.
3. Recall mencapai 85%, artinya model mampu mendeteksi sebagian besar kasus penyakit.
4. Skor F1 sebesar 87,4% menunjukkan keseimbangan antara tingkat ketepatan prediksi dan kemampuan mendeteksi kasus nyata.

Secara keseluruhan, model memiliki performa yang baik dan stabil, sehingga dapat digunakan untuk membantu proses klasifikasi awal penyakit paru-paru berdasarkan gaya hidup dan kondisi

kesehatan pasien. Nilai metrik yang seimbang juga menunjukkan bahwa model tidak hanya memperhatikan ketepatan prediksi, tetapi juga kelengkapan dalam mendeteksi kasus yang benar-benar terjadi.

HASIL DAN PEMBAHASAN

Model yang dibuat dengan algoritma Decision Tree menghasilkan struktur pohon keputusan yang mudah diinterpretasikan. Atribut "Merokok" muncul sebagai akar pohon (root node), yang artinya merupakan faktor utama yang memengaruhi hasil klasifikasi. Cabang berikutnya melibatkan atribut seperti "Bekerja", "Asuransi", dan "Aktivitas_Begadang", yang membentuk jalur keputusan berdasarkan kondisi pasien. Struktur pohon ini menunjukkan bagaimana kombinasi faktor gaya hidup dan kondisi sosial dapat digunakan untuk memprediksi kemungkinan seseorang menderita penyakit paru-paru.

Hasil

Hasil evaluasi dilakukan menggunakan data uji di RapidMiner dengan operator Performance (Classification). Hasil metrik evaluasi disajikan dalam gambar berikut:

Tree

```
Merokok = Aktif
| Bekerja = Tidak: Tidak {Ya=0, Tidak=8866}
| Bekerja = Ya
| | Rumah_Tangga = Tidak: Tidak {Ya=0, Tidak=4814}
| | Rumah_Tangga = Ya: Ya {Ya=1530, Tidak=0}
Merokok = Pasif
| Asuransi = Ada: Ya {Ya=7429, Tidak=0}
| Asuransi = Tidak
| | Aktivitas_Begadang = Tidak
| | | Usia = Muda: Tidak {Ya=1602, Tidak=1968}
| | | Usia = Tua: Ya {Ya=1784, Tidak=0}
| | Aktivitas_Begadang = Ya: Ya {Ya=2007, Tidak=0}
```

Gambar 3. Hasil Decision Tree

Hasil menunjukkan bahwa model memiliki akurasi yang bagus, artinya model mampu mengklasifikasikan data secara tepat. Tingkat presisi sebesar 90% berarti sebagian besar hasil prediksi yang menyatakan positif benar-benar positif, sedangkan tingkat recall sebesar 85% menunjukkan model mampu menangkap kebanyakan kasus yang sesungguhnya positif. Nilai F1-Score yang seimbang yaitu

87,4% menunjukkan model memiliki performa yang stabil antara akurasi dan ketelitian.

Hasil ini selaras dengan berbagai penelitian sebelumnya. Misalnya, penelitian tentang prediksi risiko penyakit paru-paru menggunakan algoritma machine learning melaporkan bahwa Decision Tree memberikan performa yang baik dengan akurasi hingga 94,72% dalam mengklasifikasikan penyakit paru-paru menggunakan dataset besar dari Kaggle. Selain itu, penelitian tentang klasifikasi kanker paru-paru dengan Decision Tree menunjukkan bahwa metode ini mampu mencapai tingkat akurasi dan recall yang baik (Aulia dan Adelia, 2025). Penelitian di Indonesia juga menunjukkan bahwa algoritma C4.5, yang merupakan salah satu varian pohon keputusan, mampu mencapai akurasi 94,33% dalam mendeteksi penyakit paru-paru (Kurniawati et al., 2025).

Pembahasan

Dari perhitungan Information Gain, atribut "Merokok" memiliki nilai tertinggi yaitu 0,4810, diikuti oleh "Jenis_Kelamin", "Bekerja", dan "Penyakit_Bawaan". Hal ini menunjukkan bahwa faktor-faktor ini memiliki peran penting dalam menentukan jenis penyakit paru-paru. Model Decision Tree tidak hanya menonjol dalam akurasi, tetapi juga mudah dipahami, sehingga cocok digunakan dalam sistem pendukung keputusan medis. Visualisasi pohon keputusan memudahkan tenaga medis memahami logika di balik hasil prediksi, yang menjadi salah satu keunggulan dibandingkan algoritma lain yang sulit diinterpretasikan.

Hasil ini sesuai dengan penelitian sebelumnya oleh (Pambudi et al., 2024), yang menyatakan bahwa faktor merokok dan jenis kelamin memainkan peran penting dalam klasifikasi penyakit paru-paru menggunakan algoritma C4.5. Dengan demikian, model ini dianggap efektif, efisien, dan bisa diterapkan dalam konteks medis.

SIMPULAN (PENUTUP)

Penelitian ini menunjukkan bahwa algoritma Decision Tree mampu mengklasifikasikan penyakit paru-paru secara efektif dengan tingkat akurasi mencapai 89,47%. Nilai precision, recall, dan F1-score juga menunjukkan bahwa model memiliki kinerja yang seimbang. Beberapa atribut seperti Merokok, Jenis_Kelamin, dan Penyakit_Bawaan berpengaruh besar dalam proses klasifikasi, dan hal ini didukung oleh metode seleksi fitur Information Gain. Penggunaan RapidMiner Studio memudahkan proses analisis dan mempermudah interpretasi hasil, sehingga model ini memiliki potensi untuk digunakan sebagai sistem pendukung keputusan medis dalam mendeteksi penyakit paru-paru secara dini. Untuk penelitian selanjutnya, disarankan menggunakan dataset yang lebih besar dan menggabungkan algoritma Decision Tree dengan metode lain agar akurasi dan kemampuan generalisasi model dapat ditingkatkan.

UCAPAN TERIMA KASIH

Penulis ingin mengucapkan terima kasih kepada semua pihak yang telah membantu dalam menyelesaikan penelitian ini. Ucapan terima kasih khusus juga disampaikan kepada dosen pembimbing yang telah memberikan arahan dan saran yang sangat berguna sepanjang proses penelitian. Penulis juga mengucapkan terima kasih kepada tim teknis serta teman-teman yang telah membantu dalam proses pengumpulan data, analisis fitur, dan pembuatan model klasifikasi penyakit paru-paru.

Tidak lupa, penulis menghargai dukungan dari platform penyedia dataset seperti Kaggle, serta perangkat lunak RapidMiner dan Python yang memudahkan pelaksanaan eksperimen secara efisien. Semoga hasil penelitian ini dapat bermanfaat dalam pengembangan sistem pendukung keputusan di bidang kesehatan, terutama dalam mendeteksi penyakit paru-paru sejak dini.

DAFTAR PUSTAKA

- Willermark, A.P. dan Isind, A. serta Adelia, Q. (2025) 'Prediksi Risiko Penyakit Paru-Paru Menggunakan Algoritma Machine Learning', 3(Juli), hal. 70–79.
- Deigo Anugrah Pratama, Ibnu Rizal Mutaqin dan Kevin Rafael Manuela (2023) 'Analisis Terjadinya Kanker Paru-Paru pada Pasien Menggunakan Decision Tree: Penerapan Algoritma C4.5 dan RapidMiner Untuk Menentukan Risiko Kanker pada Gejala Pasien', *Jurnal Teknik Mesin, Industri, Elektro dan Informatika*, 2(4), hal 156–170. Tersedia di: <https://doi.org/10.55606/jtmei.v2i4.3004>.
- Dinova, M.A. dan Purnama, R.A. (tanpa tanggal) 'Klasifikasi Untuk Mendeteksi Penyakit Paru-Paru dengan Algoritma Decision Tree', hal 1–7.
- Gori, T., Sunyoto, A. dan Al Fatta, H. (2024) 'Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa', *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(1), hal 215–224. Tersedia di: <https://doi.org/10.25126/jtiik.20241118074>.
- Illah, I.Z.A., Jauharis Sapu, W.S. dan Damaliana, A.T. (2024) 'Implementasi Metode Klasifikasi LightGBM dan Analisis Survival dalam Memprediksi Pelanggan Churn', *Jurnal Komtika (Komputasi dan Informatika)*, 8(1), hal 43- 53. Tersedia di: <https://doi.org/10.31603/komtika.v8i1.11194>.
- Jatnika Fahmi Idris, Rafid Ramadhani dan Muhammad Malik Mutoffar (2024) 'Klasifikasi Penyakit Kanker Paru Menggunakan Perbandingan Algoritma Machine Learning', *Jurnal Media Akademik (JMA)*, 2(2). Tersedia di: <https://doi.org/10.62281/v2i2.145>.
- Khairudin, K., Machfud, S. dan Cahyono, Y. (2025) 'Klasifikasi Penyakit Kanker Paru-Paru Menggunakan Metode

- C4.5', *KERNEL: Jurnal Riset Inovasi Bidang Informatika dan Pendidikan Informatika*, 5(2), hal. 83–93. Tersedia di: <https://doi.org/10.31284/j.kernel.2024.v5i2.7315>.
- Krishna, S.U. dan lain-lain (2024) 'Prediksi dan Klasifikasi Kanker Paru Menggunakan Decision Tree dan VGG16 Convolutional Neural Networks', *The Open Biomedical Engineering Journal*, 18(1), hal.1– 14. Tersedia di: <https://doi.org/10.2174/0118741207290271240322061032>.
- Krishnaiah, V., Narsimha, G. dan Chandra, N.S. (2013) 'Diagnosis Sistem Prediksi Kanker Paru Menggunakan Teknik Klasifikasi Data Mining', *International Journal of Computer Science and Information Technologies*, 4(1), hal. 39–45. Tersedia di: <http://www.ijcsit.com/docs/Volume4/Vol4Issue1/ijcsit2013040110.pdf>.
- Kurniawati, L. dan lain-lain (2025) 'Perbandingan Metode Berbasis Decision Tree dalam Deteksi Penyakit Paru-Paru', *Jurnal Bumigora Information Technology (BITe)*, 7(1), hal. 51–62. Tersedia di: <https://doi.org/10.30812/bite.v7i1.4909>.
- Larassati, D., Zaidiah, A. dan Afrizal, S. (2022) 'Sistem Prediksi Penyakit Jantung Koroner Menggunakan Metode Naive Bayes', *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 7(2), hal. 533–546. Tersedia di: <https://doi.org/10.29100/jipi.v7i2.2842>.
- Marzuq, R.D., Wicaksono, S.A. dan Setiawan, N.Y. (2023) 'Prediksi Kanker Paru-Paru Menggunakan Algoritme Random Forest Decision Tree', *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 7(7), hal. 3448–3456.
- Pambudi, R. dan lain-lain (2024) 'Klasifikasi Penyakit Paru-Paru Menggunakan Metode Decision Tree', *OKTAL: Jurnal Ilmu Komputer dan Science*, 3(9), hal. 2397–2398. Tersedia di: <https://journal.mediapublikasi.id/index.php/oktal>.
- Putri, W.M. dan lain-lain (2025) 'KLASIFIKASI PENYAKIT GINJAL KRONIS MENGGUNAKAN Abstraksi Keywords: Pendahuluan Tinjauan Pustaka', 7(1).
- Salim, M.F. dan Sugeng, S. (2018) 'Analisis Rekam Medis Pasien Diabetes Mellitus Melalui Implementasi Teknik Data Mining di RSUP Dr. Sardjito Yogyakarta', *Jurnal Kesehatan Vokasional*, 2(2), halaman 167. Tersedia di: <https://doi.org/10.22146/jkesvo.30331>.