



DETEKSI DINI PENYAKIT KANKER MENGGUNAKAN DECISION TREE

Rania Putri¹, Najwa Maulida Bilqist², Muhammad Audry Fahrezi³, Safitri Juanita^{4*}

^{1,2,3,4}Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur, DKI Jakarta, Indonesia
email koresponden: safitri.juanita@budiluhur.ac.id

(received: 01/02/2026, revised: 10/03/2026, accepted: 29/03/2026)

Abstrak

Kanker merupakan salah satu penyakit tidak menular dengan tingkat mortalitas yang tinggi, yang sebagian besar disebabkan oleh keterlambatan dalam proses diagnosis. Oleh karena itu, diperlukan pendekatan deteksi dini yang efektif, akurat, dan mudah diterapkan untuk membantu mengidentifikasi risiko kanker sejak tahap awal. Penelitian ini bertujuan untuk mengembangkan model deteksi dini penyakit kanker menggunakan algoritma Decision Tree berbasis faktor risiko dan gaya hidup. Dataset yang digunakan berasal dari platform Kaggle, yang memuat atribut umum meliputi usia (*age*), jenis kelamin (*gender*), indeks massa tubuh (BMI), kebiasaan merokok (*smoking*), risiko genetik (*geneticrisk*), aktivitas fisik (*physicalactivity*), konsumsi alkohol (*alcoholintake*), serta riwayat kanker (*cancerhistory*). Label kelas pada dataset bersifat biner, yaitu status diagnosis kanker (ya/tidak), tanpa membedakan jenis kanker secara spesifik. Metodologi penelitian mengacu pada kerangka kerja CRISP-DM. Proses pemodelan dan pengujian dilakukan menggunakan perangkat lunak RapidMiner dengan beberapa skema pembagian data latih dan data uji, yaitu 90:10, 80:20, 70:30, dan 60:40. Evaluasi kinerja model dilakukan menggunakan *confusion matrix* dengan metrik akurasi, precision, recall, dan F1-score. Hasil eksperimen menunjukkan bahwa performa terbaik diperoleh pada skema pembagian data 90:10, dengan nilai akurasi sebesar 92%, *precision* 93.76%, *recall* 83.33%, dan F1-score 88.24%. Hasil ini menunjukkan bahwa algoritma Decision Tree mampu mengklasifikasikan risiko kanker secara efektif berdasarkan faktor risiko umum, serta menghasilkan model yang akurat dan mudah diinterpretasikan. Dengan demikian, penelitian ini memberikan kontribusi dalam pengembangan sistem deteksi dini kanker yang bersifat generik, aplikatif, dan berpotensi diterapkan sebagai sistem pendukung keputusan di bidang kesehatan.

Kata kunci: deteksi dini kanker, decision tree, CRISP-DM, confusion matrix, rapidminer, machine learning.

EARLY DETECTION OF CANCER USING DECISION TREE

Abstract

Cancer is one of the leading non-communicable diseases with a high mortality rate, largely caused by delays in the diagnostic process. Therefore, an effective, accurate, and easily applicable early detection approach is required to identify cancer risk at an early stage. This study aims to develop an early cancer detection model using the Decision Tree algorithm based on risk factors and lifestyle characteristics. The dataset used in this research was obtained from the Kaggle platform and consists of common attributes, including age, gender, body mass index (BMI), smoking habits, genetic risk, physical activity, alcohol intake, and cancer history. The class label in the dataset is binary, representing cancer diagnosis status (yes/no), without distinguishing specific cancer types. The research methodology follows the CRISP-DM framework. Model training and testing were conducted using RapidMiner with several train-test split scenarios, namely 90:10, 80:20, 70:30, and 60:40. Model performance was evaluated using a confusion matrix, with accuracy, precision, recall, and F1-score as the main evaluation metrics. Experimental results indicate that the best performance was achieved using the 90:10 data split, with an accuracy of 92%, precision of 93.76%, recall of 83.33%, and an F1-score of 88.24%. These results demonstrate that the Decision Tree algorithm is effective in classifying cancer risk based on general risk factors while producing an accurate and interpretable model. Thus, this study contributes to the development of a generic and practical early cancer detection system that has the potential to be applied as a decision support system in the healthcare domain.

Keywords: early cancer detection, decision tree, CRISP-DM, confusion matrix, RapidMiner, machine learning.

1. Pendahuluan

Kanker merupakan salah satu penyakit tidak menular yang menjadi penyebab utama kematian di dunia dan terus menunjukkan tren peningkatan setiap tahunnya [1], [2]. Tingginya angka morbiditas dan mortalitas akibat kanker sebagian besar disebabkan oleh keterlambatan diagnosis, sehingga pasien baru mendapatkan penanganan pada stadium lanjut [3]. Oleh karena itu, deteksi dini kanker menjadi langkah krusial untuk meningkatkan peluang keberhasilan terapi, menekan angka kematian, serta mengoptimalkan perencanaan intervensi medis sejak tahap awal.

Perkembangan teknologi informasi, khususnya di bidang machine learning, membuka peluang baru dalam mendukung deteksi dini penyakit kanker secara lebih cepat dan objektif [4]. Salah satu algoritma yang banyak digunakan dalam bidang kesehatan adalah Decision Tree, karena kemampuannya menghasilkan model klasifikasi yang mudah diinterpretasikan dan sesuai untuk mendukung pengambilan keputusan medis [5]. Berbagai penelitian sebelumnya telah menunjukkan bahwa Decision Tree mampu memberikan performa yang baik dalam mendeteksi berbagai jenis kanker, seperti kanker paru-paru [6], payudara [7], tulang [8], dan tiroid [9]. Namun, sebagian besar penelitian tersebut masih berfokus pada satu jenis kanker tertentu serta menggunakan data klinis atau laboratorium yang kompleks, yang umumnya hanya tersedia di fasilitas kesehatan dan sulit diterapkan untuk skrining awal berbasis populasi.

Di sisi lain, ketersediaan dataset publik yang memuat faktor risiko umum, seperti usia, jenis kelamin, indeks massa tubuh, kebiasaan merokok, aktivitas fisik, konsumsi alkohol, risiko genetik, dan riwayat kanker, membuka peluang untuk mengembangkan model deteksi dini kanker yang lebih generik dan mudah diakses. Meskipun demikian, pemanfaatan dataset berbasis faktor risiko dengan label kanker bersifat biner (ya/tidak) masih relatif terbatas dalam penelitian sebelumnya, khususnya yang mengkaji efektivitas algoritma Decision Tree sebagai alat skrining awal kanker secara umum.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengembangkan dan mengevaluasi model deteksi dini penyakit kanker menggunakan algoritma Decision Tree dengan memanfaatkan dataset publik dari Kaggle yang berbasis faktor risiko dan gaya hidup. Penelitian ini difokuskan pada klasifikasi status kanker dan riwayat kanker secara biner, tanpa membedakan jenis kanker secara spesifik. Proses pemodelan dan evaluasi dilakukan menggunakan perangkat lunak RapidMiner, dengan tujuan memperoleh model yang tidak hanya memiliki tingkat akurasi yang tinggi, tetapi juga mudah direplikasi dan diinterpretasikan. Diharapkan hasil penelitian ini dapat memberikan kontribusi dalam pengembangan sistem skrining awal kanker yang bersifat praktis, fleksibel, dan berpotensi diterapkan secara luas sebagai pendukung pengambilan keputusan awal di bidang kesehatan.

2. Tinjauan Literatur

Berbagai penelitian terdahulu menunjukkan bahwa algoritma Decision Tree (DT) merupakan metode yang efektif dalam mendukung deteksi dini penyakit kanker, terutama karena kemampuannya menghasilkan model klasifikasi yang transparan dan mudah diinterpretasikan. Pada kasus kanker paru-paru, studi perbandingan menunjukkan bahwa Decision Tree mampu mencapai akurasi sebesar 64,85% pada dataset dengan tujuh atribut, serta mengidentifikasi bahwa faktor merokok dan sesak napas merupakan fitur yang paling berpengaruh dalam proses klasifikasi [10].

Pada konteks kanker payudara, sejumlah penelitian melaporkan performa Decision Tree yang bervariasi bergantung pada dataset dan algoritma turunan yang digunakan. Salah satu penelitian menunjukkan bahwa algoritma C5.0 lebih unggul dibandingkan ID3 dengan akurasi sebesar 80% menggunakan data laboratorium berbasis dataset publik dari UCI [7]. Penelitian lain yang mengimplementasikan algoritma ID3 dan CART bahkan melaporkan akurasi hingga 98,25% [4]. Selain itu, penelitian lain pada dataset kanker payudara dengan sembilan atribut dan 500 data menghasilkan akurasi sebesar 93,67% menggunakan algoritma Decision Tree [11].

Pendekatan Decision Tree juga diterapkan pada kasus tumor tulang, dengan mengadopsi kerangka kerja Knowledge Discovery in Databases (KDD), dan berhasil mencapai akurasi sebesar 88,5%, yang menunjukkan efektivitas DT dalam mengolah data klinis yang kompleks [8]. Sementara itu, pada kasus kanker tiroid, eksplorasi algoritma C4.5 digunakan untuk memprediksi tingkat kelangsungan hidup pasien berdasarkan respons klinis, dengan akurasi mencapai 97% [9]. Beberapa penelitian lain secara

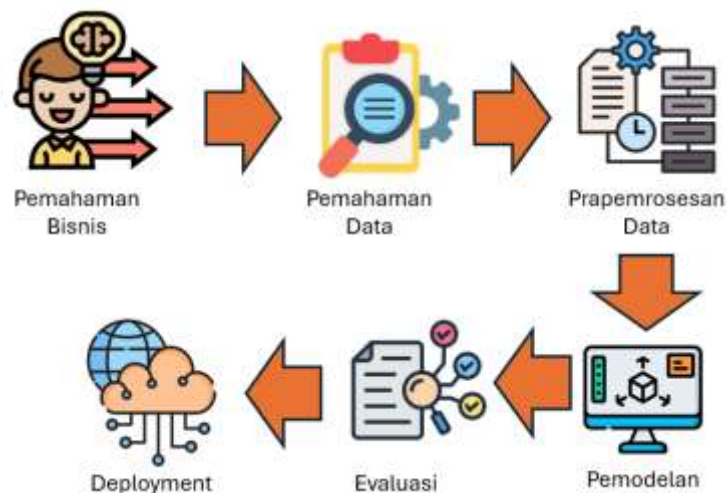
konsisten menerapkan algoritma yang sama untuk mengidentifikasi risiko kanker paru-paru, dengan menegaskan merokok sebagai prediktor utama, di mana salah satu studi melaporkan akurasi sebesar 93,8% [5], [6]. Di luar konteks kanker, algoritma C4.5 juga telah digunakan untuk klasifikasi kondisi pasien hepatitis dengan akurasi sebesar 80,65%, yang semakin memperkuat fleksibilitas algoritma ini dalam bidang kesehatan [12].

Lebih lanjut, berdasarkan hasil berbagai penelitian [4]–[6], [11], [12], implementasi algoritma Decision Tree menggunakan perangkat lunak RapidMiner terbukti mampu menyederhanakan proses pemodelan dan evaluasi kinerja tanpa memerlukan pemrograman yang kompleks, sekaligus menghasilkan performa klasifikasi yang kompetitif pada dataset medis. Secara kolektif, literatur tersebut menegaskan bahwa Decision Tree memiliki keunggulan utama dalam menghasilkan model yang akurat, transparan, dan mudah dipahami, meskipun efektivitasnya sangat dipengaruhi oleh kualitas dan relevansi fitur input yang digunakan, baik berupa data laboratorium maupun gejala klinis.

Berbeda dari penelitian-penelitian sebelumnya yang umumnya berfokus pada satu jenis kanker dan menggunakan fitur klinis atau laboratorium spesifik, penelitian ini menawarkan pendekatan yang lebih generalis dan aplikatif dengan memanfaatkan dataset publik berbasis faktor risiko umum serta label kanker bersifat biner. Pendekatan ini memungkinkan penerapan Decision Tree sebagai alat skrining awal kanker yang lebih fleksibel, mudah direplikasi, dan berpotensi diterapkan pada skala populasi luas, sehingga memberikan kontribusi kebaruan dan relevansi praktis yang kuat untuk dipublikasikan dalam jurnal ilmiah.

3. Metode Penelitian

Pada Gambar 1 terdapat beberapa proses yang merupakan tahapan penelitian.



Gambar 1. Tahapan Penelitian Deteksi Dini Penyakit Kanker menggunakan Decision Tree

Tahapan pada penelitian menggunakan model CRISP-DM [13], adapun penjelasan tiap tahapan pada Gambar 1 adalah sebagai berikut:

3.1. Pemahaman Bisnis

Kanker merupakan salah satu penyakit tidak menular dengan tingkat mortalitas tertinggi di dunia. Di Indonesia, prevalensi kanker berdasarkan *Riset Kesehatan Dasar* (Riskesdas) tahun 2018 tercatat sebesar 1,79% [1]. Tingginya angka tersebut menunjukkan urgensi pengembangan metode deteksi dini kanker yang cepat dan efektif, sehingga proses identifikasi dapat dilakukan lebih awal dan penanganan pasien dapat diberikan secara tepat waktu.

3.2. Pemahaman Data

Sumber dataset bersumber dari Dataset ini kami mengambil dari website : <https://www.kaggle.com/datasets/rabieelkharoua/cancer-prediction-dataset> [14]. Total Dataset : 500



data. Kolom Dataset : *Age*, *Gender*, *BMI*, *Smoking*, *GenetikRisk*, *PhysicalActivity*, *AlcoholIntake*, *CancerHistory*, *Diagnosis*. Contoh dataset terdapat pada Tabel 1. Penjelasan dari setiap kolom sebagai berikut:

1. *Age* : Nilai integer yang mewakili usia pasien, berkisar antara 20 hingga 80.
2. *Gender* : Nilai biner yang mewakili jenis kelamin, di mana 0 menunjukkan Pria dan 1 menunjukkan Wanita.
3. *BMI* : Nilai kontinu yang mewakili Indeks Massa Tubuh, berkisar antara 15 hingga 40.
4. *Smoking* : Nilai biner yang menunjukkan status merokok, di mana 0 berarti Tidak dan 1 berarti Ya.
5. *GeneticRisk* : Nilai kategoris yang mewakili tingkat risiko genetik untuk kanker, dengan 0 menunjukkan Rendah, 1 menunjukkan Sedang, dan 2 menunjukkan Tinggi.
6. *PhysicalActivity* : Nilai kontinu yang menunjukkan jumlah jam per minggu yang dihabiskan untuk aktivitas fisik, berkisar dari 0 hingga 10.
7. *AlcoholIntake* : Nilai kontinu yang menunjukkan jumlah unit alkohol yang dikonsumsi per minggu, berkisar dari 0 hingga 5.
8. *CancerHistory* : Nilai biner yang menunjukkan apakah pasien memiliki riwayat kanker pribadi, di mana 0 berarti Tidak dan 1 berarti Ya.
9. *Diagnosis* : Nilai biner yang menunjukkan status diagnosis kanker, di mana 0 menunjukkan Tidak Ada Kanker dan 1 menunjukkan Kanker.

Tabel 1. Atribut Penelitian

Nama Atribut	Kode Atribut
<i>Age</i>	A
<i>Gender</i>	B
<i>BMI</i>	C
<i>Smoking</i>	D
<i>GeneticRisk</i>	E
<i>PhysicalActivity</i>	F
<i>AlcoholIntake</i>	G
<i>CancerHistory</i>	H
<i>Diagnosis</i>	I

Tabel 2. Contoh Dataset Penelitian

No.	A	B	C	D	E	F	G	H	I
1	45	1	22	1	2	2	1	0	1
2	38	0	25	0	1	4	0	0	0
3	60	1	30	1	2	1	2	1	1
4	50	0	28	0	1	5	0	0	0
5	35	1	24	0	0	3	0	0	0
...	...	...	...	...	...	...	...	...	...
499	65	1	29	1	2	0	2	1	1
500	40	0	27	0	0	4	0	0	0

3.3. Prapemrosesan Data

Pada tahap ini melakukan pemilihan atribut yang akan dipakai pada penelitian yaitu *Age*, *Gender*, *BMI*, *Smoking*, *GeneticRisk*, *PhysicalActivity*, *AlcoholIntake*, *CancerHistory*. Selain pemilihan atribut, dilakukan juga beberapa proses prapemrosesan data sebagai berikut:

- Pengecekan dan Penghapusan Duplikasi Data, untuk memastikan tidak ada data yang tercatat lebih dari satu kali.
- Penanganan Data Kosong (*Missing Values*), yaitu dengan melakukan imputasi menggunakan nilai rata-rata (*mean*) untuk atribut numerik dan modus (*mode*) untuk atribut kategorikal.
- Normalisasi Data Numerik menggunakan metode *Min-Max Scaling* agar semua nilai atribut numerik berada dalam rentang 0 hingga 1, sehingga mempermudah proses pemodelan.

- *Encoding Data Kategorikal*, yaitu mengubah data kategorikal menjadi data numerik yang dapat diproses oleh algoritma *Decision Tree*.
- Pemeriksaan *Outlier*, untuk mendeteksi dan menangani nilai ekstrim yang dapat mempengaruhi performa model, baik dengan penghapusan atau penyesuaian data.

3.4. Pemodelan

Pada tahap ini melakukan pemodelan menggunakan algoritma *Decision Tree*, pemodelan ini menggunakan perangkat lunak RapidMiner. Tahapan yang dilakukan pada saat pelatihan dan pengujian menggunakan metode *Split*. Pada penelitian ini menggunakan perbandingan hasil dengan berbagai model latih dan uji, yaitu dengan 90:10, 80:20, 70:30, 60:40.

3.5. Evaluasi

Pada tahap ini menggunakan metode *confusion matrix* menggunakan akurasi, presisi, recall dan F1-Score.

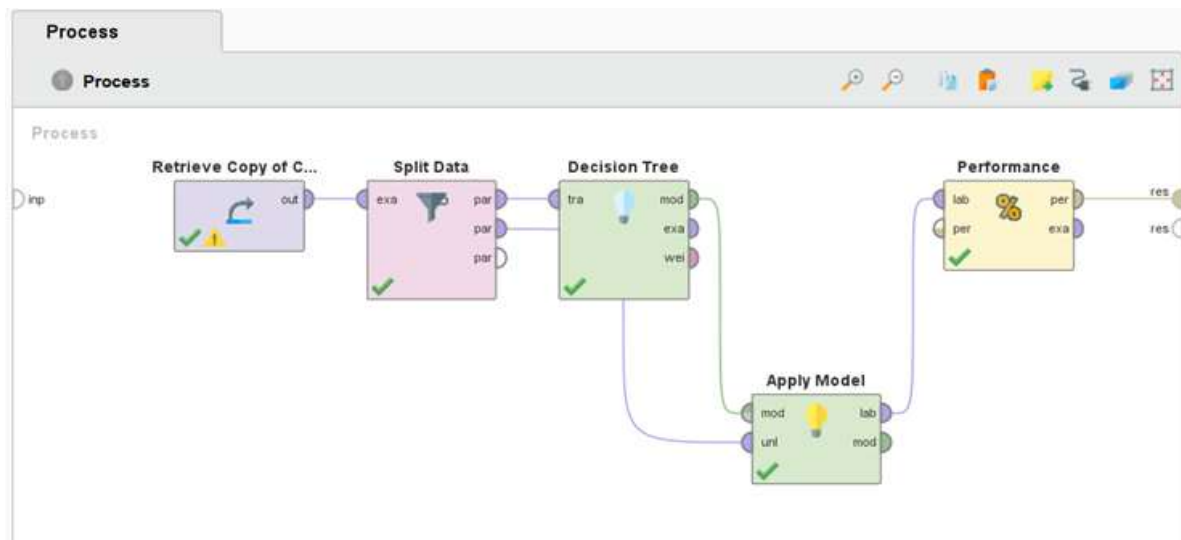
3.6. Deployment

Pada tahap ini, melakukan analisis mendalam dari hasil pemodelan.

4. Hasil dan Pembahasan

4.1. Tahapan Proses Pemodelan Decision Tree Menggunakan Rapid Miner

Pada Gambar 2 menampilkan beberapa tahapan proses eksperimen yang dilakukan dengan mengimplementasikan algoritma *Decision Tree* untuk deteksi dini penyakit kanker menggunakan aplikasi Rapid Miner.



Gambar 2. Tahapan Pemrosesan menggunakan Rapid Miner untuk Model Deteksi Dini Kanker menggunakan *Decision Tree*

4.2. Perbandingan Hasil Model Split pada Model Decision Tree

Berdasarkan proses yang ditampilkan pada Gambar 2, maka pada Tabel 3 menampilkan hasil perbandingan pengujian *Split* dengan beberapa skema latih dan uji. Hasil pada Tabel 2 menunjukkan bahwa performa terbaik dari model deteksi dini kanker menggunakan algoritma *Decision Tree* adalah model *split data* dengan skema latih dan uji 90:10, dengan nilai akurasi 92%. Faktor yang paling mempengaruhi hasil diagnosis adalah: usia, kebiasaan merokok, riwayat kanker pribadi, dan risiko genetik.

Tabel 3. Hasil Perbandingan Model Latih dan Uji menggunakan Model Decision Tree

No	Split Data	Akurasi	Recall	Precision	F1-Score
1.	70:30	87.33%	81.48%	83.02%	82.24%
2.	60:40	83.50%	70.83%	80.95%	75.54%
3.	80:20	90.00%	83.33%	90.91%	86.92%
4.	90:10	92.00%	83.33%	93.75%	88.24%

bold merupakan nilai tertinggi

4.3. Hasil Model Decision Tree dalam bentuk Pohon Keputusan

Gambar 3 menampilkan model pohon keputusan yang dibangun menggunakan dataset publik bersumber di Kaggle yang memiliki delapan atribut, yaitu usia (*age*), jenis kelamin (*gender*), BMI, status merokok (*smoking*), risiko genetik (*genetic risk*), aktivitas fisik (*physical activity*), konsumsi alkohol (*alcohol intake*), dan riwayat kanker (*cancer history*). Model pohon keputusan tersebut mengidentifikasi risiko genetik (*genetic risk*) sebagai simpul akar, yang mengindikasikan peran dominannya dalam proses deteksi dini kanker. Percabangan berikutnya melibatkan riwayat kanker (*cancer history*) dan aktivitas fisik (*physical activity*), yang kemudian diikuti oleh faktor-faktor gaya hidup lainnya seperti konsumsi alkohol (*alcohol intake*) dan status merokok (*smoking*), serta atribut demografis seperti usia (*age*), jenis kelamin (*gender*), dan BMI.

Setiap simpul terminal merepresentasikan hasil klasifikasi deteksi dini kanker yang diperoleh dari kombinasi faktor genetik, klinis, dan gaya hidup. Struktur pohon ini menunjukkan kemampuan model pohon keputusan dalam menangkap pola risiko yang kompleks secara interpretable, serta memungkinkan pemahaman yang transparan mengenai bagaimana kombinasi faktor risiko berkontribusi terhadap keputusan klasifikasi akhir.

**Gambar 3.** Visualisasi aturan pengambilan keputusan untuk Deteksi Dini Kanker menggunakan Decision Tree

Pada Tabel 4 menampilkan aturan (*rule*) yang diekstraksi dari model pohon keputusan menunjukkan bahwa risiko genetik (*genetic risk*) merupakan faktor utama dalam proses deteksi dini kanker. Pada individu dengan risiko genetik rendah, keputusan klasifikasi dipengaruhi oleh faktor gaya hidup seperti aktivitas fisik, konsumsi alkohol, kebiasaan merokok, serta faktor demografis seperti usia, jenis kelamin, dan BMI. Sebaliknya, pada individu dengan risiko genetik tinggi, riwayat kanker dan usia menjadi determinan utama, dengan BMI dan konsumsi alkohol berperan sebagai faktor pemurni keputusan. Representasi rule ini menegaskan kemampuan pohon keputusan dalam menangkap interaksi kompleks antara faktor genetik, klinis, dan gaya hidup secara transparan dan mudah diinterpretasikan.

Tabel 4. Aturan (*Rule*) berdasarkan Pohon Keputusan

Aturan	Model Keputusan	Aturan	Model Keputusan
1	IF GeneticRisk $\leq$ 1.500 AND PhysicalActivity $>$ 4.623 THEN Class = 1	10	IF GeneticRisk $\leq$ 1.500 AND AlcoholIntake $>$ 0.154 AND Age $\leq$ 45.500 AND Gender $>$ 0.500 AND Smoking $\leq$ 0.500 THEN Class = 0
2	IF GeneticRisk $\leq$ 1.500	11	IF GeneticRisk $>$ 1.500



Aturan	Model Keputusan	Aturan	Model Keputusan
	AND PhysicalActivity ≤ 4.623 AND BMI ≤ 19.187 THEN Class = 1		AND CancerHistory > 0.500 THEN Class = 1
3	IF GeneticRisk ≤ 1.500 AND PhysicalActivity ≤ 4.623 AND BMI > 19.187 AND Age ≤ 53.500 THEN Class = 1	12	IF GeneticRisk > 1.500 AND CancerHistory ≤ 0.500 AND Age ≤ 50.500 AND AlcoholIntake ≤ 0.825 THEN Class = 0
4	IF GeneticRisk ≤ 1.500 AND PhysicalActivity ≤ 4.623 AND BMI > 19.187 AND Age > 53.500 AND Gender ≤ 0.500 THEN Class = 1	13	IF GeneticRisk > 1.500 AND CancerHistory ≤ 0.500 AND Age ≤ 50.500 AND AlcoholIntake > 0.825 AND Age ≤ 75.500 AND Gender ≤ 0.500 THEN Class = 0
5	IF GeneticRisk ≤ 1.500 AND PhysicalActivity ≤ 4.623 AND BMI > 19.187 AND Age > 53.500 AND Gender > 0.500 THEN Class = 0	14	IF GeneticRisk > 1.500 AND CancerHistory ≤ 0.500 AND Age ≤ 50.500 AND AlcoholIntake > 0.825 AND Age > 75.500 THEN Class = 1
6	IF GeneticRisk ≤ 1.500 AND AlcoholIntake ≤ 0.154 THEN Class = 0	15	IF GeneticRisk > 1.500 AND CancerHistory ≤ 0.500 AND Age > 50.500 AND BMI ≤ 28.817 THEN Class = 0
7	IF GeneticRisk ≤ 1.500 AND AlcoholIntake > 0.154 AND Age > 45.500 THEN Class = 1	16	IF GeneticRisk > 1.500 AND CancerHistory ≤ 0.500 AND Age > 50.500 AND BMI > 28.817 AND Gender > 0.500 THEN Class = 0
8	IF GeneticRisk ≤ 1.500 AND AlcoholIntake > 0.154 AND Age ≤ 45.500 AND Gender ≤ 0.500 THEN Class = 1	17	IF GeneticRisk > 1.500 AND CancerHistory ≤ 0.500 AND Age > 50.500 AND BMI > 28.817 AND Gender ≤ 0.500 AND BMI > 29.029 THEN Class = 1
9	IF GeneticRisk ≤ 1.500 AND AlcoholIntake > 0.154 AND Age ≤ 45.500 AND Gender > 0.500 AND Smoking > 0.500 THEN Class = 1		

5. Kesimpulan

Penelitian ini bertujuan untuk mengembangkan model deteksi dini penyakit kanker berbasis algoritma Decision Tree dengan memanfaatkan faktor risiko umum dan gaya hidup. Berdasarkan hasil pengujian yang telah dilakukan, model Decision Tree menunjukkan performa terbaik pada pembagian data latih dan data uji sebesar 90:10, dengan nilai akurasi sebesar 92%. Hal ini menunjukkan bahwa model mampu memberikan prediksi yang benar terhadap sebagian besar data uji. Nilai recall sebesar 83,33% mengindikasikan kemampuan model yang cukup baik dalam mendeteksi kasus kanker positif, sedangkan nilai precision sebesar 93,75% menunjukkan bahwa prediksi positif yang dihasilkan model

memiliki tingkat ketepatan yang tinggi. Dengan F1-score sebesar 88,24%, model ini memiliki keseimbangan yang baik antara kemampuan deteksi dan ketepatan prediksi.

Meskipun hasil penelitian menunjukkan kinerja yang baik, pengembangan lebih lanjut masih dimungkinkan. Penelitian selanjutnya dapat dilakukan dengan menggunakan dataset yang lebih besar dan lebih beragam, menambahkan fitur klinis yang lebih spesifik, serta membandingkan Decision Tree dengan algoritma machine learning lain guna meningkatkan akurasi dan kemampuan generalisasi model.

Daftar Pustaka

- [1] S. K. Hero, "Faktor Risiko Kanker Payudara," *J. Med. Utama*, vol. 03, no. 01, pp. 1533–1538, 2021, [Online]. Available: <https://jurnalmedikahutama.com/index.php/JMH/article/view/310>.
- [2] W. O. Masrida, R. N. H. Haris, and N. A. H. Asapa, "Estimasi Angka Kematian Penyakit Kanker Akibat Rokok di Indonesia Tahun 2020," *J. Penelit. Sains dan Kesehat. Avicenna*, vol. 1, no. 2, pp. 21–26, 2022, doi: 10.69677/avicenna.v1i2.13.
- [3] Y. S. Alfianti, "Kanker Tanpa Benjolan: Memahami Gejala Aneh Yang Kerap Terlupakan," *J. Inov. Glob.*, vol. 3, no. 5, pp. 756–763, 2025.
- [4] I. Setiawan and E. K. Nurnawati, "Penerapan Algoritma Decision Tree Untuk Memprediksi Kanker Payudara menggunakan Data Mining dan Machine Learning," *J. Din. Inform.*, vol. 11, no. 2, pp. 103–112, 2022, [Online]. Available: <https://jdi.upy.ac.id/index.php/jdi/article/view/200>.
- [5] 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," *J. Tek. Mesin, Ind. Elektro dan Inform.*, vol. 2, no. 4, pp. 156–170, 2023, [Online]. Available: <https://ejurnal.politeknikpratama.ac.id/index.php/jtmei/article/view/3004>.
- [6] R. Rofiani, L. Oktaviani, D. Vernand, and T. Hendriawan, "Penerapan Metode Klasifikasi Decision Tree dalam Prediksi Kanker Paru-Paru Menggunakan Algoritma C4.5," *J. TEKNO KOMPAK*, vol. 18, no. 1, pp. 126–139, 2024.
- [7] A. D. F. Mellina, S. Suhartono, and M. A. Yaqin, "Algoritma Decision Tree untuk Prediksi Deteksi Penyakit Kanker Payudara," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 9, no. 1, pp. 70–78, 2024, doi: 10.14421/jiska.2024.9.1.70-78.
- [8] R. D. Destianti, M. D. Gultom, and G. Noviamanti, "Model Klasifikasi Tingkat Keparahan Tumor Tulang Menggunakan Algoritma Decision Tree Berbasis Fitur Klinis," *J. Comput. Sci. Artif. Intell.*, vol. 2, no. 3, pp. 1–9, 2025.
- [9] A. D. Putri, F. Sholekhah, E. Dadynata, L. Efrizoni, R. Rahmadden, and N. Sapina, "Penerapan Algoritma Decision Tree C4.5 untuk Memprediksi Tingkat Kelangsungan Hidup Pasien Kanker Tiroid," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 4, pp. 1485–1495, 2024, [Online]. Available: <https://journal.irpi.or.id/index.php/malcom/article/view/1532>.
- [10] D. C. J. Kartikasari and R. Adawiyah, "Perbandingan Klasifikasi Penyakit Kanker Paru-Paru Menggunakan Decision Tree Dan Random Forest," *Komputika J. Sist. Komput.*, vol. 14, no. 1, pp. 79–85, 2025, doi: 10.34010/kp5h2h96.
- [11] Z. Fatah and R. N. Maulani, "Klasifikasi Data Kanker Payudara Menggunakan Algoritma Decision Tree Berbasis Rapidminer," *Jamastika*, vol. 4, no. 2, pp. 127–132, 2025, [Online]. Available: <https://jurnal.unw.ac.id/index.php/jamastika/article/view/4504/2842>.
- [12] L. Pambudi and S. Sefrika, "Klasifikasi Dampak dan Kondisi Pasien Hepatitis Menggunakan Metode Decision Tree," *J. Artif. Intell. Digit. Bus.*, vol. 4, no. 3, pp. 4769–4775, 2025, doi: 10.31004/riggs.v4i3.2692.
- [13] C. Schröer, F. Kruse, and J. M. Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Model," *Procedia Comput. Sci.*, vol. 181, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.
- [14] R. El Kharoua, "Cancer Prediction Dataset," *Kaggle.com*, 2024. <https://www.kaggle.com/datasets/rabieelkharoua/cancer-prediction-dataset> (accessed Feb. 06, 2025).