

Klasifikasi Dini Penyakit Alzheimer Menggunakan Algoritma ID3

Diana Gabriela¹, Imelda Sri Dumayanti², Harlen Gilbert Simanullang³^{1,2,3}Fakultas Ilmu Komputer, Universitas Methodist Indonesia

Info Artikel

Histori Artikel:

Received, Juli, 20, 2025
Revised, Agust, 15, 2025
Accepted, Sept, 10, 2025

Keywords:

Alzheimer, Deteksi dini, ID3,
Pohon Keputusan, Klasifikasi.

ABSTRAK

Alzheimer merupakan salah satu jenis demensia paling umum yang berdampak pada kemampuan kognitif, ingatan, dan perilaku penderitanya. Deteksi sejak dini menjadi krusial agar penanganan dapat dilakukan lebih cepat dan progresivitas penyakit dapat diperlambat. Penelitian ini mengimplementasikan algoritma *Iterative Dichotomiser 3* (ID3) sebagai metode klasifikasi untuk mendeteksi *Alzheimer*. Eksperimen dilakukan melalui dua pendekatan, yaitu (1) perhitungan manual dengan 160 data latih dan 40 data uji, serta (2) implementasi otomatis berbasis *Python* dengan pustaka *Scikit-Learn* menggunakan pembagian data 80% latih dan 20% uji. Hasil pengujian manual menunjukkan akurasi 65%, *precision* 60%, dan *recall* 67% dengan atribut *Age* sebagai akar pohon keputusan. Sedangkan implementasi *Python* menghasilkan akurasi 65%, *precision* untuk kelas “Terdeteksi” sebesar 64%, *recall* 70%, dan *f1-score* 0,67, dengan akar pohon yang sama. Secara keseluruhan, algoritma ID3 terbukti mampu membentuk model klasifikasi yang mudah dipahami dan divisualisasikan, serta memiliki potensi untuk mendukung sistem pendukung keputusan pada deteksi dini *Alzheimer*, khususnya di layanan kesehatan primer.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Penulis Koresponden:

Diana Gabriela,
Fakultas Ilmu Komputer,
Universitas Methodist Indonesia, Medan,
Jl. Hang Tuah No.8, Medan - Sumatera Utara.
Email: purbadiana587@gmail.com

1. PENDAHULUAN

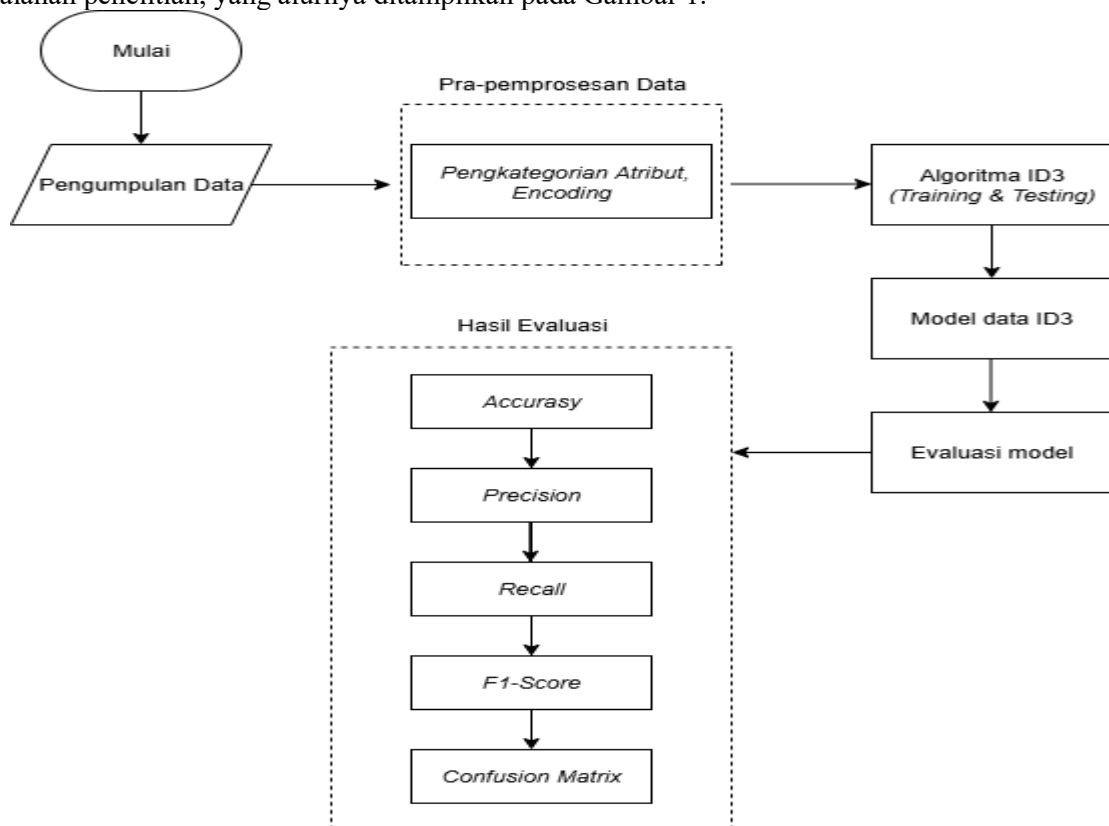
Penyakit *Alzheimer* merupakan gangguan *neurodegeneratif* yang menyerang otak dan menyebabkan penurunan fungsi kognitif, hilangnya memori, serta perubahan perilaku. Penyakit ini umumnya terdiagnosis pada individu berusia di atas 65 tahun, meskipun tidak menutup kemungkinan terjadi pada usia yang lebih muda^[1]. Berdasarkan laporan *World Health Organization*, *Alzheimer* menjadi penyebab 60–70% kasus demensia di dunia, dan jumlah penderita diperkirakan terus meningkat seiring bertambahnya usia harapan hidup. Studi epidemiologi terbaru juga menegaskan bahwa faktor risiko seperti genetik, gaya hidup, serta penyakit penyerta (diabetes, hipertensi, dan kardiovaskular) berkontribusi signifikan terhadap perkembangan penyakit ini. Kemajuan di bidang *machine learning* menawarkan solusi dalam mendukung proses diagnosis berbasis data. Salah satu algoritma yang banyak digunakan untuk klasifikasi adalah Decision Tree, yang membentuk model berupa pohon keputusan dengan node akar, node pengujian, dan node daun sebagai representasi kelas^[2]. Pendekatan ini erat kaitannya dengan kecerdasan buatan (*Artificial Intelligence/AI*), di mana *machine learning* merupakan salah satu cabang utamanya. AI dapat diartikan sebagai kecerdasan yang disematkan pada sistem buatan agar mampu memproses data, mempelajari pola, serta menyesuaikan perilaku dalam menyelesaikan suatu permasalahan^[3]. Kajian ini dilakukan dengan menggunakan dataset *Alzheimer's Disease* dari *Kaggle* yang terdiri atas 200 data numerik dengan 15 atribut terkait faktor risiko kesehatan, seperti usia, indeks massa tubuh, riwayat merokok, pola konsumsi alkohol, aktivitas fisik, kualitas tidur, riwayat keluarga, serta kondisi medis penyerta. Dataset dibagi menjadi 80% data latih dan 20% data uji untuk membangun serta mengevaluasi model. Algoritma

ID3 (*Iterative Dichotomiser 3*) dipilih karena menghasilkan struktur klasifikasi yang interpretatif dan berpotensi. Studi ini berfokus pada penerapan metode tersebut untuk menyusun struktur pohon keputusan yang optimal, serta mengevaluasi kinerjanya dalam mengklasifikasikan data. Batas pembahasan ditetapkan pada pemanfaatan ID3 dengan Python sebagai alat pengolahan data. Model yang dikembangkan diharapkan dapat menghasilkan model klasifikasi yang sederhana dan mudah dipahami, Serta dapat dimanfaatkan baik oleh tenaga medis maupun masyarakat umum, menjadi dasar pengembangan aplikasi deteksi dini Alzheimer, serta memberikan kontribusi bagi penelitian selanjutnya dibidang klasifikasi berbasis pohon keputusan.

2. METODE PENELITIAN

2.1. Desain Penelitian

Desain Penelitian dibuat dalam menunjukkan tahapan yang menuntun peneliti dalam menapaki tahapan studinya secara sistematis. Kerangka ini memuat prosedur untuk merumuskan solusi atas permasalahan penelitian, yang alurnya ditampilkan pada Gambar 1.



Gambar 1. Kerangka Penelitian

2.2. Teknik Pengumpulan Data

Tahap pengumpulan data pada penelitian ini bertujuan memperoleh informasi kesehatan yang akan dianalisis lebih lanjut. Dataset diperoleh dari Kaggle dan melalui proses pra-pemrosesan sebelum dianalisis dengan metode yang telah ditentukan.

2.3. Pra-Pemrosesan

Pra-Pemrosesan adalah tahap awal yang bertujuan menyiapkan dataset agar akurat, konsisten, dan bisa diproses dengan pemodelan machine learning. Adapun tahapan pra-pemrosesan sebagai berikut:

1. Pengkategorian atribut merupakan proses mengubah nilai numerik atau biner menjadi bentuk kategorikal agar dapat diproses oleh algoritma ID3 yang berbasis aturan.

2. Proses encoding dilakukan untuk mengonversi atribut kategorikal menjadi representasi numerik, sedangkan penanganan terhadap data yang hilang diterapkan agar dataset siap digunakan pada tahap pengembangan model klasifikasi^[4].

2.4. Algoritma Iterative Dichotomiser 3 (ID3)

Algoritma ID3 termasuk salah satu metode decision Tree yang banyak digunakan karena mampu membangun model klasifikasi dengan aturan yang sederhana dan mudah dipahami^[5]. Pada *decision tree*, struktur model digambarkan menyerupai pohon, di mana setiap simpul merepresentasikan atribut yang diuji, sedangkan simpul daun menunjukkan hasil klasifikasi atau kelas tertentu. Atribut yang memiliki pengaruh paling besar terhadap pembentukan kelas umumnya ditempatkan pada simpul akar. Setelah nilai entropy diperoleh, langkah selanjutnya adalah menghitung information gain dari setiap atribut untuk menentukan pemisahan data terbaik^[6]. Rumus persamaan entropy bisa dilihat pada berikut:

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i$$

Keterangan:

$Entropy(S)$ = Kumpulan data atau fenomena yang diamati dalam studi

n = Jumlah total kelas dalam *dataset*

p_i = Probabilitas kemunculan data pada kelas i dalam node

Setelah mendapatkan nilai *entropy*, tentukan nilai *information gain* dengan rumus berikut:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \left| \frac{S_i}{S} \right| * Entropy(S_i)$$

Keterangan:

$Gain(S, A)$ = Pengurangan *entropy* setelah *dataset S* dibagi oleh atribut A

n = Jumlah *subset* yang dihasilkan oleh atribut A

$| S_i |$ = Jumlah data dalam *subset S_i*

$| S |$ = Jumlah total data dalam *dataset S*

2.5. Split data Training dan Testing

Penelitian ini terdiri dari 200 data. Dataset dibagi ke dalam dua subset, yaitu data latih dan data uji, dengan proporsi 80% dialokasikan sebagai data latih dan sisanya digunakan sebagai data uji 20%.

2.6. Evaluasi Model

Diproses ini model dilakukan mengukur kualitas serta performa algoritma dengan memanfaatkan data uji yang terpisah dari data latih. Pada penelitian ini, pengukuran dilakukan menggunakan Confusion Matrix, yaitu sebuah tabel yang menampilkan distribusi hasil prediksi dibandingkan dengan label sebenarnya. Melalui *Confusion Matrix*, dapat dihitung beberapa metrik, antara lain *accuracy*, *precision*, *recall*, dan *f1-score*.

Berikut dibawah ini rumus-rumus^[7]:

$$Accuracy = \frac{TP + FN}{TP + FP + TN + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1\ Score = \frac{Precision + Recall}{Precision + Recall}$$

3. HASIL DAN PEMBAHASAN

3.1. Implementasi Algoritma ID3 Berbasis Python

pada penelitian ini dilakukan menggunakan bahasa pemrograman *Python* dengan dukungan pustaka *Scikit-Learn* yang dipilih karena menyediakan berbagai metode *machine learning*, termasuk Decision Tree, serta fitur pendukung seperti *train-test split*, *encoding* data, evaluasi performa, dan visualisasi model. Selain itu, penelitian ini juga memanfaatkan pustaka lain, yaitu *Pandas* untuk membaca dan mengolah dataset, *NumPy* untuk perhitungan numerik, *Matplotlib* untuk menampilkan hasil visualisasi berupa pohon keputusan dan *confusion matrix*.

3.2. Pengumpulan data

Dataset yang digunakan bersifat sekunder dan diperoleh dari platform *Kaggle*. Data ini terdiri atas 200 sampel dengan 15 atribut yang mencerminkan berbagai faktor risiko, Informasi tersebut kemudian digunakan sebagai dasar dalam membangun dan mengevaluasi model klasifikasi berbasis algoritma ID3. Contoh data yang digunakan dalam penelitian ini dapat dilihat pada Gambar 2.

	Age	BMI	Smoking	AlcoholConsumption	PhysicalActivity	DietQuality	SleepQuality	FamilyHistoryAlzheimers
0	67	28.099989	0	3.510408	4.411354	1.401290	7.663635	0
1	76	31.388537	0	13.953106	9.538425	8.934142	8.930221	1
2	75	19.716992	1	8.817248	1.541340	6.281219	8.319863	1
3	74	21.887089	1	8.276900	6.324575	5.653982	4.379219	1
4	87	30.629049	1	6.245747	9.111274	8.117240	7.088391	0
...	...	...	...	...	...	...	...	...
195	66	32.013806	1	9.308706	4.352402	5.432374	9.624312	1
196	61	39.121757	0	1.561126	4.049964	6.555306	7.535540	0
197	75	17.857903	0	18.767261	1.360667	2.904662	8.555256	0
198	77	15.476479	0	4.594670	9.886002	8.120025	5.769464	0
199	90	35.132825	1	10.416973	4.120716	6.096417	7.275428	1

Gambar 2 Dataset penelitian

3.2.1 Pem-prosesan data

1. Atribut dalam dataset memiliki tipe numerik sehingga perlu dikonversi menjadi kategori. Hal ini dilakukan karena algoritma ID3 hanya dapat mengolah data bersifat kategorikal. Proses transformasi dilakukan dengan membagi nilai numerik ke dalam interval tertentu, misalnya usia dibagi ke dalam kategori usia awal, madya, dan lanjut, sedangkan beberapa atribut bersifat biner dengan dua nilai, yakni 0 untuk Tidak dan 1 untuk Iya. Hasil pengkategorian atribut ditampilkan pada Gambar 3.

	Age	BMI	AlcoholConsumption	PhysicalActivity	DietQuality	SleepQuality	SystolicBP	DiastolicBP	CholesterolTotal
Lanjut Usia Awal		Obesitas	Rendah	Sedang	Buruk	Sedang	Hipertensi Tahap 1	Hipertensi Tahap 2	Normal
Lanjut Usia Madya		Obesitas	Sedang	Sangat Tinggi	Sangat Baik	Sangat Baik	Normal	Prehipertensi	Tinggi
Lanjut Usia Madya		Normal	Sedang	Rendah	Baik	Baik	Normal	Hipertensi Tahap 2	Normal
Lanjut Usia Awal		Normal	Sedang	Tinggi	Sedang	Buruk	Normal	Normal	Tinggi
Lanjut Usia Akhir		Obesitas	Rendah	Sangat Tinggi	Sangat Baik	Sedang	Hipertensi Tahap 1	Normal	Normal
...	...	...	...	...	...	...	...	...	...
Lanjut Usia Awal		Obesitas	Sedang	Sedang	Sedang	Sangat Baik	Normal	Hipertensi Tahap 1	Batas Tinggi
Lanjut Usia		Obesitas	Tidak Mengonsumsi	Sedang	Baik	Sedang	Prehipertensi	Hipertensi Tahap 2	Tinggi

Gambar 3 Pengkategorian atribut

2. Setelah dilakukan proses pengkategorian, setiap atribut yang telah berbentuk kategori kemudian diubah kembali ke dalam representasi numerik melalui proses encoding. Tahap ini diperlukan agar data dapat diolah secara komputasional oleh algoritma ID3. Terdapat dua jenis encoding:

a. *One-hot encoding*

One-Hot Encoding bekerja dengan cara membuat kolom biner baru untuk setiap kategori unik dalam suatu atribut. Setiap baris data akan diberi nilai 1 pada kolom yang sesuai dengan kategorinya, dan 0 pada kolom lainnya. dapat dilihat pada gambar 3

b. *Ordinal encoding*

Setiap kategori diberi angka yang mencerminkan tingkatan atau skala, nilai lebih tinggi menunjukkan posisi atau status lebih besar. Misalnya, pada atribut Kategori BMI, label seperti Kurus Berat, Normal, dan Obesitas diubah menjadi nilai 1, 3, dan 4 sesuai dengan urutan tingkatannya. Hasil dari *ordinal encoding* dapat dilihat pada gambar 4

Age	BMI	Smoking	AlcoholConsumption	PhysicalActivity	DietQuality	SleepQuality	FamilyHistoryAlzheimers	Cardiovascu
1	4	1	0	3	1	3	0	
2	4	1	1	2	2	2	1	
2	3	2	1	0	0	0	1	
1	3	2	1	4	4	1	1	
0	4	2	0	2	2	3	0	
1	3	1	3	4	2	3	1	
1	3	1	1	4	4	1	0	
2	1	1	2	1	0	3	1	
1	2	1	3	3	3	1	0	
2	4	1	2	2	3	0	0	

Gambar 4 *Encoding*

3.2.2 Split data 80:20

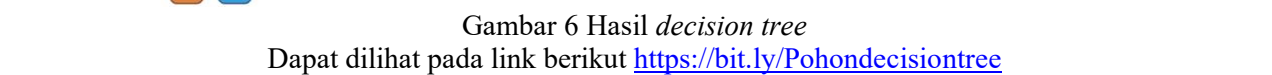
Setelah melalui tahap pra-pemrosesan dan encoding, dataset kemudian dibagi menjadi dua subset, yakni data latih dan data uji dengan proporsi 80%:20%. Hasil pembagian data tersebut ditunjukkan pada Gambar 5.

➡ Jumlah Data Training: 160
Jumlah Data Testing : 40

Gambar 5 Split data

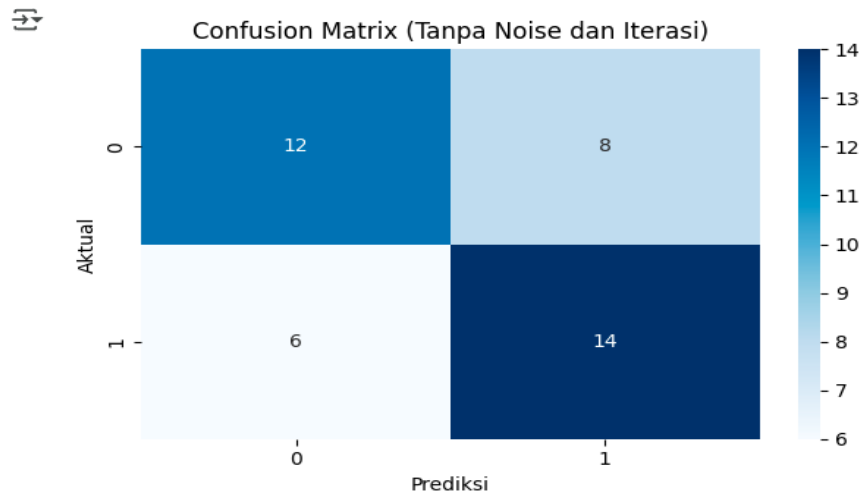
3.2.3 Pelatihan algoritma ID3

Selanjutnya dilakukan pelatihan model menggunakan algoritma ID3 (*Iterative Dichotomiser 3*). Algoritma ini membangun pohon keputusan berdasarkan konsep *entropy* dan *information gain*, di mana atribut dengan nilai gain tertinggi dipilih sebagai akar pohon atau node utama. Dari hasil perhitungan, atribut *Age* memperoleh nilai *information gain* terbesar sehingga ditempatkan sebagai root node, yang menjadi titik awal pemisahan data untuk klasifikasi *Alzheimer*. Hasil *Decision Tree* dapat dilihat pada Gambar 6.



3.2.4 Evaluasi Model

Untuk memudahkan pemahaman terhadap hasil prediksi, distribusi data ditampilkan dalam bentuk diagram heatmap confusion matrix. Matriks ini digunakan untuk menunjukkan perbandingan antara hasil prediksi model dengan kelas sebenarnya. Visualisasi hasil tersebut dapat dilihat pada Gambar 7..



Gambar 7 Confusion matrix

Berdasarkan dari gambar diatas, terdapat 12 data positif yang berhasil diprediksi dengan benar (*true positive*) dan 14 data negatif yang juga diklasifikasikan dengan tepat (*true negative*). Namun, masih terdapat kesalahan prediksi, yaitu 8 data positif yang salah diklasifikasikan sebagai negatif (*false negative*) serta 6 data negatif yang salah diklasifikasikan sebagai positif (*false positive*).

Evaluasi Model (Tanpa Noise dan Iterasi)					
Akurasi	:	0.6500			
Presisi	:	0.6515			
Recall	:	0.6500			
F1-Score	:	0.6491			
Classification Report:					
		precision	recall	f1-score	support
	0	0.67	0.60	0.63	20
	1	0.64	0.70	0.67	20
	accuracy			0.65	40
	macro avg	0.65	0.65	0.65	40
	weighted avg	0.65	0.65	0.65	40

Gambar 8 Hasil evaluasi

Berdasarkan hasil evaluasi model ID3, diperoleh nilai precision dan recall yang berbeda untuk masing-masing kelas. Pada kelas "Terdeteksi", model menunjukkan performa yang lebih baik dibandingkan kelas "Tidak Terdeteksi". Precision pada kelas "Terdeteksi" menunjukkan bahwa sebagian besar prediksi terhadap kasus positif sudah tepat, sementara recall-nya mencerminkan bahwa model mampu mengenali sebagian besar kasus yang seharusnya terdeteksi. Sebaliknya, pada kelas "Tidak Terdeteksi", precision dan recall masih rendah, yang berarti model kurang akurat dalam memprediksi dan mengenali kasus yang sebenarnya tidak terdeteksi. Hal ini menunjukkan bahwa meskipun model cukup mampu mengenali kelas positif, masih diperlukan perbaikan untuk meningkatkan akurasi prediksi pada kelas negatif agar performa model menjadi lebih seimbang.

4. KESIMPULAN

Penelitian ini membuktikan bahwa algoritma *Iterative Dichotomiser 3* (ID3) dapat diterapkan secara efektif untuk klasifikasi dini penyakit *Alzheimer* melalui tahapan pemodelan yang terstruktur, mulai dari pra-pemrosesan data, pembagian data latih dan uji, pembangunan model pohon keputusan, hingga evaluasi performa. Algoritma ID3 berhasil membentuk pohon keputusan yang memetakan atribut pada

dataset ke dalam dua kelas diagnosis, yaitu Terdeteksi dan Tidak Terdeteksi. Proses perhitungan entropy dan information gain memungkinkan penentuan atribut paling berpengaruh secara sistematis, di mana atribut *Age* menjadi akar pohon keputusan. Berdasarkan hasil evaluasi, model menghasilkan akurasi sebesar 65%, *precision* 0,64 dan *recall* 0,70 untuk kelas Terdeteksi, serta *precision* 0,67 dan *recall* 0,60 untuk kelas Tidak Terdeteksi. Nilai rata-rata *macro* dan *weighted average* untuk *precision*, *recall*, dan *f1-score* berada di angka 0,65, yang menunjukkan performa model cukup stabil dan layak digunakan sebagai dasar pengembangan sistem pendukung keputusan di layanan kesehatan primer. Meski demikian, hasil penelitian ini juga menegaskan perlunya optimasi lanjutan, seperti pemilihan fitur yang lebih relevan dan penanganan ketidakseimbangan data, agar model dapat mencapai tingkat akurasi yang lebih seimbang untuk kedua kelas diagnosis.

REFERENSI

- [1] A. A. Mortara, M. Permatasari, A. Desiani, Y. Andriani, and M. Arhami, "Perbandingan Algoritma C4.5 dan Adaptive Boosting dalam Klasifikasi Penyakit Alzheimer Comparison of C4.5 and Adaptive Boosting Algorithms in Alzheimer's Disease Classification," *J. Teknol. dan Inf.*, vol. 13, 2023, doi: 10.34010/jati.v13i2.
- [2] N. Ranti, M. Iqbal, and K. H. Hanif, "Klasifikasi Penyakit Kanker Payudara Menggunakan Perbandingan Algoritma Machine Learning," vol. 3, no. 1, pp. 1–6, 2022, [Online]. Available: <http://creativecommons.org/licenses/by/4.0/>
- [3] Harlen Gilbert Simanullang, Arina Prima Silalahi, and M. I. Hutapea, *Pendukung Keputusan Seleksi Indeks LQ-45*. 2023. [Online]. Available: www.madzamedia.co.id
- [4] N. Henisaniyya *et al.*, "Sistemasi: Jurnal Sistem Informasi Klasifikasi Class Thyroid menggunakan Algoritma ID3 dan Artificial Neural Network (ANN) Classification of Thyroid Class using ID3 Algorithm and Artificial Neural Network (ANN)," 2025. [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [5] F. Ferdina, N. Satyahadewi, and D. Kusnandar, "Penerapan Algoritma Iterative Dichotomiser 3 (Id3) Dalam Klasifikasi Faktor Risiko Penyakit Diabetes Melitus," *Var. J. Stat. Its Appl.*, vol. 5, no. 2, pp. 139–146, 2023, doi: 10.30598/variancevol5iss2page139-146.
- [6] Miryam Clementine and Arum, "Prediksi Churn Nasabah Bank Menggunakan Klasifikasi Naïve Bayes dan ID3," *J. Process.*, vol. 17, no. 1, pp. 9–18, May 2022, doi: 10.33998/processor.2022.17.1.1170.
- [7] A. Fabian Azmi, A. Voutama, S. Karawang Jl HSRonggo Waluyo, and T. Timur, "PREDIKSI CHURN NASABAH BANK MENGGUNAKAN KLASIFIKASI RANDOM FOREST DAN DECISION TREE DENGAN EVALUASI CONFUSION MATRIX," vol. 13, no. 1, 2024.