

REVOLUSI DIAGNOSIS: OPTIMASI RANDOM TREE-PSO UNTUK PENYAKIT GINJAL KRONIS

Sartini¹, Sumarna^{2*}, Abdul Hamid³, Ahmad Hafidzul Kahfi⁴, Nicodias Palasara⁵

¹Program Studi Teknologi Informasi, Universitas Bina Sarana Informatika, ²Program Studi Informatika, Universitas Nusa Mandiri, ^{3,5}Program Studi Sistem Informasi, Universitas Bina Sarana Informatika, ⁴Program Studi Informatika, Universitas Bina Sarana Informatika.

^{1,3,4,5}Jln. Kramat Raya No.98, Senen, Jakarta Pusat 10450, ²Jln. Jatiwaringin Raya No. 2, Jakarta Timur 13620

¹sartini.sar@bsi.ac.id, ²sumarna.smn@nusamandiri.ac.id, ³abdul.adh@bsi.ac.id @ bsi.ac.id,
⁴ahmad.azx@bsi.ac.id, ⁵nico.ncp@bsi.ac.id

Abstract

Chronic kidney disease (CKD) is a serious health problem that requires early detection to prevent further complications and improve the quality of life of patients. This study aims to develop a CKD prediction model in an effort to improve the accuracy of early diagnosis of CKD with the dataset used obtained from the UCI Repository. The method used is based on the Random Tree algorithm optimized using Particle Swarm Optimization (PSO) which functions as an optimization method to improve model performance with tree depth and the number of attributes considered at each split, to find the configuration that produces the highest accuracy. The model development process includes the stages of feature selection, model training, and performance evaluation using accuracy, precision, recall, and F1-score metrics. The results showed that the PSO-optimized Random Tree algorithm significantly improved prediction performance compared to the baseline model, with an accuracy of 94.25%. This optimization also reduces model complexity without sacrificing accuracy. These findings indicate the potential for implementing the proposed model to support a more efficient medical decision-making system, especially in early detection of CKD. Further recommendations include the integration of this model into a technology-based system in the clinical environment to reduce the workload of medical personnel.

Keywords : *Chronic Kidney Disease, Early Detection, Feature Selection, Particle Swarm Optimization, Random Tree.*

Abstrak

Penyakit ginjal kronis (PGK) merupakan salah satu masalah kesehatan serius yang memerlukan deteksi dini untuk mencegah komplikasi lebih lanjut dan meningkatkan kualitas hidup pasien. Penelitian ini bertujuan mengembangkan model prediksi PGK dalam upaya meningkatkan akurasi diagnosis dini PGK dengan dataset yang digunakan diperoleh dari UCI Repository. Metode yang dipakai berbasis algoritma Random Tree yang dioptimasi menggunakan Particle Swarm Optimization (PSO) yang berfungsi sebagai metode optimasi untuk meningkatkan kinerja model dengan kedalaman pohon dan jumlah atribut yang dipertimbangkan pada setiap pemisahan, untuk menemukan konfigurasi yang menghasilkan akurasi tertinggi. Proses pengembangan model mencakup tahap seleksi fitur, pelatihan model, dan evaluasi performa menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa algoritma Random Tree yang dioptimasi PSO secara signifikan meningkatkan performa prediksi dibandingkan model baseline, dengan akurasi mencapai 94,25%. Optimalisasi ini juga mengurangi kompleksitas model tanpa mengorbankan akurasi. Temuan ini menunjukkan potensi penerapan model yang diusulkan untuk mendukung sistem pengambilan keputusan medis secara lebih efisien, terutama dalam deteksi dini PGK. Rekomendasi lebih lanjut mencakup integrasi model ini pada sistem berbasis teknologi di lingkungan klinis untuk mengurangi beban kerja tenaga medis.

Kata kunci : *Deteksi Dini, Particle Swarm Optimization, Penyakit Ginjal Kronis, Random Tree, Seleksi Fitur.*

1. PENDAHULUAN

Penyakit Ginjal Kronis disebut juga PGK merupakan salah satu masalah kesehatan global yang semakin meningkat setiap tahunnya [1]. Penyakit ini sering kali tidak terdeteksi pada tahap awal karena gejala yang tidak spesifik [2], sehingga banyak pasien yang baru mendapatkan perawatan ketika penyakit sudah mencapai stadium lanjut [3]. Kondisi ini berkontribusi pada angka morbiditas dan mortalitas yang tinggi serta beban ekonomi yang signifikan bagi pasien dan sistem kesehatan secara keseluruhan [4]. Oleh karena itu, Revolusi diagnosis dengan deteksi dini PGK menjadi aspek penting dalam upaya pencegahan dan pengelolaan yang lebih baik [5].

Revolusi diagnosis merujuk pada perubahan mendasar dalam metode diagnosis tradisional menuju pendekatan yang lebih canggih dan akurat. Penerapan algoritma *Random Tree* yang dioptimasi dengan *Particle Swarm Optimization* (PSO) diharapkan dapat meningkatkan kemampuan deteksi dini PGK secara signifikan. Transformasi ini dianggap revolusioner karena mampu mengatasi keterbatasan metode tradisional dalam hal akurasi, efisiensi, dan ketepatan waktu. Dengan demikian, konsep "Revolusi Diagnosis" berfokus pada pemanfaatan teknologi kecerdasan buatan dengan dampak optimasi untuk mempercepat proses pengambilan keputusan medis secara signifikan.

Dalam beberapa tahun terakhir, pendekatan berbasis kecerdasan buatan telah berkembang pesat dalam dunia medis terutama untuk mendukung diagnosis PGK [6]. Penelitian yang telah dilakukan dengan peningkatan kernel SVM dengan menunjukkan bahwa kombinasi SVM dengan *Particle Swarm Optimization* (PSO) berhasil mencapai akurasi sebesar 99,75% [7]. Penelitian yang menggunakan DT yang dioptimasi oleh PSO, mencapai akurasi 98,50% [8]. Penelitian menggunakan pendekatan NB yang dioptimalkan oleh PSO dan memperoleh akurasi sebesar 95,75% [9]. Selain itu penelitian yang menggabungkan algoritma KNN dengan PSO, yang berhasil meningkatkan akurasi dari 80,00% menjadi 97,75% [10]. Berdasarkan penelitian-penelitian tersebut, terlihat bahwa integrasi algoritma klasifikasi dengan PSO mampu meningkatkan performa model secara signifikan. Namun, penelitian yang menggabungkan *Random Tree* dengan PSO untuk prediksi PGK masih terbatas bahkan belum pernah dilakukan. *Random Tree*, sebagai salah satu algoritma pembelajaran mesin telah terbukti efektif dalam menangani data medis yang kompleks dan beragam [11]. Namun, performa algoritma ini sangat dipengaruhi oleh

pemilihan parameter dan fitur yang digunakan [12]. Untuk mengatasi tantangan ini, algoritma optimasi seperti PSO dapat digunakan untuk meningkatkan akurasi dan efisiensi model prediksi [13]. Inspirasi untuk menggunakan pendekatan ini muncul dari berbagai penelitian yang menunjukkan keberhasilan kombinasi metode optimasi dan pembelajaran mesin dalam bidang kesehatan [14].

Penelitian ini bertujuan untuk mengembangkan model prediksi PGK berbasis *Random Tree* yang dioptimalkan menggunakan PSO. Secara khusus, penelitian ini akan mengeksplorasi bagaimana algoritma PSO dapat digunakan untuk melakukan seleksi fitur dan optimasi parameter pada *Random Tree* sehingga menghasilkan model dengan akurasi yang lebih tinggi. Hasil yang diharapkan dari penelitian ini adalah model prediksi yang tidak hanya akurat tetapi juga efisien dalam memproses data medis, sehingga dapat diimplementasikan dalam sistem pendukung keputusan di fasilitas kesehatan.

Permasalahan utama yang mendasari penelitian ini dengan rendahnya tingkat deteksi dini PGK akibat kurangnya alat diagnostik yang dapat diandalkan untuk analisis data medis. Selain itu, keragaman data pasien seperti variasi usia, jenis kelamin, dan riwayat medis sering kali menambah kompleksitas dalam proses diagnosis [15]. Berdasarkan pengamatan, model prediksi yang ada saat ini sering kali menghadapi keterbatasan dalam menangani volume data besar dan fitur tidak relevan, yang dapat menurunkan performa keseluruhan model [16]. Dengan demikian, penelitian ini merupakan salah satu dari sedikit studi yang mengoptimasi *Random Tree* dengan PSO dalam domain medis untuk optimasi parameter dan seleksi fitur.

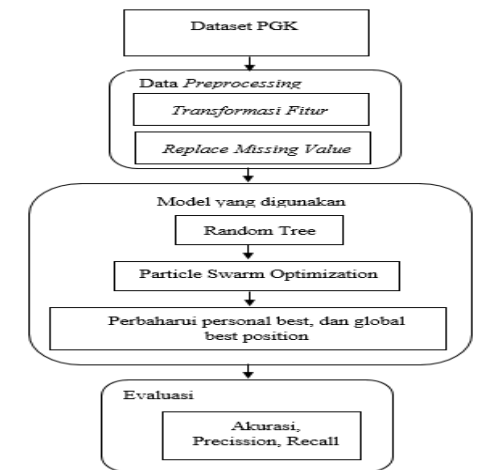
Penelitian ini merupakan salah satu dari sedikit studi yang mengoptimasi *Random Tree* dengan PSO dalam domain medis.

Secara teoritis, *Random Tree* adalah algoritma yang membangun pohon keputusan secara acak untuk meningkatkan keberagaman model dan memperbaiki akurasi prediksi. Meskipun menggunakan pendekatan acak, *Random Tree* dapat mengalami overfitting jika tidak dioptimalkan dengan benar. Di sisi lain, PSO adalah algoritma metaheuristik yang terinspirasi dari perilaku sosial kawanan hewan seperti burung atau ikan, yang efektif dalam menemukan solusi optimal dalam ruang pencarian yang besar. Dengan mengintegrasikan kedua pendekatan ini, penelitian ini diharapkan dapat memberikan kontribusi yang signifikan dalam pengembangan metode prediksi penyakit ginjal kronis yang lebih handal [17].

2. METODOLOGI PENELITIAN

2.1. Skema Alur Penelitian

Penelitian ini menerapkan metode penelitian terapan deskriptif. Proses penelitian dilakukan melalui beberapa tahap yang dirancang dalam bentuk skema alur penelitian. Tahapan-tahapan tersebut dijelaskan secara visual pada Gambar 1 berikut.



Gambar 1. Alur Penelitian

2.2. Pengumpulan Data

Dataset PGK yang digunakan terdiri dari 400 sampel data dengan fitur-fitur medis pasien seperti usia pasien, tekanan darah, urin, sel darah merah, sel darah putih, dan diabetes. Data ini diambil dari sumber terpercaya, yaitu UCI Repository. Distribusi kelas dalam dataset ini seimbang dengan 200 sampel pasien positif PGK dan 200 sampel negatif, memastikan bahwa proses pelatihan model tidak bias terhadap salah satu kelas.

2.3. Data Processing

Tahap ini bertujuan untuk memastikan dataset PGK yang digunakan dalam penelitian ini siap untuk dianalisis oleh model prediksi. Berdasarkan skema alur penelitian pada Gambar 1, langkah utama yang dilakukan dalam preprocessing data adalah transformasi fitur dan mengganti nilai yang hilang (*Replace Missing Value*) [18]. Proses ini melibatkan identifikasi atribut yang memiliki data kosong atau tidak lengkap dan menggantinya dengan metode tertentu, seperti nilai rata-rata [19]. Penggantian nilai yang hilang bertujuan untuk menghindari bias dalam model dan memastikan kelengkapan dataset sebelum tahap analisis selanjutnya[20].

2.4. Model RANDOM TREE dan PSO

Tahap ini merupakan inti dari penelitian, di mana model *Random Tree* digunakan sebagai algoritma klasifikasi. PSO digunakan untuk mengoptimalkan parameter serta performa model. *Random Tree* adalah algoritma pembelajaran mesin yang membangun pohon keputusan tunggal dengan menggunakan prinsip acak untuk meningkatkan keberagaman model dan memperbaiki akurasi prediksi. *Random Tree* bekerja dengan membangun satu pohon keputusan secara acak berdasarkan data pelatihan. Algoritma ini memilih fitur secara acak pada setiap pembagian (*split*) untuk mengurangi variansi dan meningkatkan keberagaman model. Parameter penting dalam *Random Tree*, seperti kedalaman maksimum pohon (*max depth*). Parameter ini berfungsi untuk membatasi seberapa dalam pohon dapat berkembang. Kedalaman yang terlalu besar dapat menyebabkan *overfitting*, di mana model terlalu menyesuaikan diri dengan data latih dan kehilangan kemampuan generalisasi terhadap data baru. Oleh karena itu, menemukan nilai optimal untuk *max_depth* sangat penting agar model dapat mencapai keseimbangan antara bias dan variansi.

PSO adalah algoritma metaheuristik yang diilhami oleh perilaku kawanan hewan, seperti burung atau ikan, untuk mencari solusi optimal [21]. Dalam penelitian ini, PSO digunakan untuk mengoptimalkan parameter *Random Tree* dengan memilih nilai terbaik untuk parameter seperti kedalaman maksimum (*max_depth*) dan jumlah fitur yang dipertimbangkan pada setiap split (*max_features*), untuk meningkatkan akurasi. Selain itu, seleksi fitur dilakukan dengan memilih atribut-atribut yang paling relevan dari dataset PGK untuk meningkatkan efisiensi dan mengurangi kompleksitas model.

2.5. Evaluasi

Tahap evaluasi bertujuan untuk menilai kinerja model prediksi berbasis *Random Tree* yang telah dioptimasi menggunakan PSO. Evaluasi dilakukan untuk memastikan akurasi dan efisiensi model dalam memprediksi PGK. Proses evaluasi dalam penelitian ini mencakup. Pengukuran

kinerja model dengan menggunakan beberapa metrik terutama akurasi dan AUC. Setelah itu Model Random Tree yang dioptimasi dengan PSO dibandingkan dengan random tree tanpa optimasi parameter dan algoritma lain yang sudah dilakukan berdasarkan data yang sama.

3. HASIL DAN PEMBAHASAN

Berdasarkan hasil eksperimen yang dilakukan menggunakan metode Random Tree dan PSO untuk prediksi PGK dengan tahapan pengumpulan data, data processing, model, dan evaluasi.

3.1. Pengumpulan Data

Data penelitian diambil dari dataset Iris yang diunduh melalui UCI Repository. Dataset ini berisi sebanyak 400 data, 24 fitur dan 1 target yang digunakan untuk analisis klasifikasi berdasarkan karakteristik data yang tersedia. Setelah proses pengunduhan, dataset diolah dan disimpan dalam file berformat .csv untuk mempermudah analisis lebih lanjut. Dataset ini ada pada Gambar 2, yang menunjukkan struktur dan elemen data yang digunakan dalam penelitian.

Variabel	Tipe	Deskripsi	Unit
age	Integer		year
bp	Integer	blood pressure	mm/Hg
sg	Categorical	specific gravity	
al	Categorical	albumin	
su	Categorical	sugar	
rbc	Binary	red blood cells	
pc	Binary	pus cell	
pcc	Binary	pus cell clumps	
ba	Binary	bacteria	
bgr	Integer	blood glucose random	mg/dl
bu	Integer	blood urea	mg/dl
sc	Continuous	serum creatinine	mg/dl
sod	Integer	sodium	mEq/L
pot	Continuous	potassium	mEq/L
hemo	Continuous	hemoglobin	gms
pcv	Integer	packed cell volume	
wbcc	Integer	white blood cell count	cells/cmm
rbcc	Continuous	red blood cell count	millions/cmm
htn	Binary	hypertension	
dm	Binary	diabetes mellitus	
cad	Binary	coronary artery disease	
appet	Binary	appetite	
pe	Binary	pedal edema	
ane	Binary	anemia	
class	Binary	ckd or not ckd	

Gambar 2. Dataset CKD

Gambar 2 merupakan struktur dataset yang digunakan dalam penelitian ini. Dataset tersebut terdiri dari beberapa kolom yang masing-masing mewakili atribut atau fitur penting, serta satu kolom target yang menjadi fokus utama klasifikasi. Fitur-fitur dalam dataset telah dipilih dan dirancang dengan mempertimbangan relevansinya terhadap prediksi PGK. Secara spesifik, dataset mencakup atribut yang menggambarkan kondisi kesehatan subjek, seperti:

1. Usia merupakan atribut yang menggambarkan faktor usia sebagai salah satu elemen utama dalam analisis kesehatan. Usia subjek pada

penyakit PGK bervariasi dari 2 hingga 90 tahun.

2. Tekanan darah pasien penyakit PGK diukur dalam satuan mmHg, tekanan darah dianggap sebagai indikator penting dalam mendeteksi kemungkinan hipertensi yang berkontribusi pada PGK.
3. Kreatinin Kapanjangan dari Kadar Kreatinin digunakan untuk menilai fungsi ginjal. Nilai kreatinin yang tinggi dapat mengindikasikan kerusakan ginjal.
4. Hemoglobin: Tingkat hemoglobin memberikan informasi tentang kondisi anemia, yang sering menjadi faktor pendukung dalam perkembangan PGK.
5. Riwayat diabetes ini merupakan data kategori yang menunjukkan status diabetes pasien, yang sering menjadi penyebab utama komplikasi ginjal.
6. Kolom target dalam dataset adalah klasifikasi kondisi pasien, yaitu PGK (Penyakit Ginjal Kronis) atau Tidak PGK, yang merupakan hasil prediksi model. Target ini didasarkan pada parameter kesehatan utama yang ada di dalam dataset.

Dataset ini telah dirancang untuk mendukung analisis klasifikasi yang melibatkan berbagai metode prediktif dengan memberikan gambaran menyeluruh tentang bagaimana elemen data digunakan untuk mencapai tujuan penelitian.

Data ini nantinya akan melewati proses *preprocessing*, seperti normalisasi, imputasi data yang hilang, dan transformasi fitur kategori menjadi bentuk numerik. Pemilihan dataset dari UCI Repository memberikan keuntungan berupa akses ke data yang sudah terstruktur dan digunakan secara luas dalam penelitian ilmiah. Hal ini memastikan bahwa dataset yang digunakan relevan, valid, dan dapat menghasilkan model prediksi yang akurat.

3.2. Data Preprocessing

Sebelum dapat digunakan dalam proses pelatihan model, dataset akan melewati tahapan *preprocessing* untuk memastikan kualitas data yang optimal. Tahapan *preprocessing* dilakukan untuk menangani nilai yang hilang, standar format, serta memastikan kompatibilitas data dengan algoritma yang digunakan. Tahapan

preprocessing dataset PGK meliputi langkah-langkah berikut:

1. Transformasi Fitur Kategori

Dataset PGK mengandung beberapa fitur kategori, seperti rbc, pc, pcc, dan fitur kategori lainnya. Isi dari fitur tersebut normal/abnormal, present/notpresent, good/poor, dan yes/no yang

dapat di ubah menjadi normal/good/present menjadi 0 dan abnormal/poor/notpresent menjadi 1.

2. Identifikasi Nilai yang Hilang

Berdasarkan pengamatan awal dataset, terdapat beberapa fitur yang memiliki nilai kosong atau tidak lengkap Pada Gambar 3.

age	bp	sg	al	su	rbc	pc	pcc	ba	bgr	bu	sc	sod	pot	hemo	pcv	wc	rc	htn	dm	cad	appet	pe	ane	class
48	80	1.020	1	0	?	normal	notpresent	notpresent	121	36	1.2	?	?	15.4	44	7800	5.2	yes	yes	no	good	no	no	ckd
7	50	1.020	4	0	?	normal	notpresent	notpresent	?	18	0.8	?	?	11.3	38	6000	?	no	no	no	good	no	no	ckd
62	80	1.010	2	3	normal	normal	notpresent	notpresent	423	53	1.8	?	?	9.6	31	7500	?	no	yes	no	poor	no	yes	ckd
48	70	1.005	4	0	normal	abnormal	present	notpresent	117	56	3.8	111	2.5	11.2	32	6700	3.9	yes	no	no	poor	yes	yes	ckd
51	80	1.010	2	0	normal	normal	notpresent	notpresent	106	26	1.4	?	?	11.6	35	7300	4.6	no	no	no	good	no	no	ckd
60	90	1.015	3	0	?	?	notpresent	notpresent	74	25	1.1	142	3.2	12.2	39	7800	4.4	yes	yes	no	good	yes	no	ckd
68	70	1.010	0	0	?	normal	notpresent	notpresent	100	54	24.0	104	4.0	12.4	36	?	?	no	no	no	good	no	no	ckd
24	?	1.015	2	4	normal	abnormal	notpresent	notpresent	410	31	1.1	?	?	12.4	44	6900	5	no	yes	no	good	yes	no	ckd
52	100	1.015	3	0	normal	abnormal	present	notpresent	138	60	1.9	?	?	10.8	33	9600	4.0	yes	yes	no	good	no	yes	ckd
53	90	1.020	2	0	abnormal	abnormal	present	notpresent	70	107	7.2	114	3.7	9.5	29	12100	3.7	yes	yes	no	poor	no	yes	ckd
50	60	1.010	2	4	?	abnormal	present	notpresent	490	55	4.0	?	?	9.4	28	?	?	yes	yes	no	good	no	yes	ckd
63	70	1.010	3	0	abnormal	abnormal	present	notpresent	380	60	2.7	131	4.2	10.8	32	4500	3.8	yes	yes	no	poor	yes	no	ckd
68	70	1.015	3	1	?	normal	present	notpresent	208	72	2.1	138	5.8	9.7	28	12200	3.4	yes	yes	yes	poor	yes	no	ckd
68	70	?	?	?	?	?	notpresent	notpresent	98	86	4.6	135	3.4	9.8	?	?	?	yes	yes	yes	poor	yes	no	ckd
68	80	1.010	3	2	normal	abnormal	present	present	157	90	4.1	130	6.4	5.6	16	11000	2.6	yes	yes	yes	poor	yes	no	ckd
40	80	1.015	3	0	?	normal	notpresent	notpresent	76	162	9.6	141	4.9	7.6	24	3800	2.8	yes	no	no	good	no	yes	ckd
47	70	1.015	2	0	?	normal	notpresent	notpresent	99	46	2.2	138	4.1	12.6	?	?	?	no	no	no	good	no	no	ckd
47	80	?	?	?	?	?	notpresent	notpresent	114	87	5.2	139	3.7	12.1	?	?	?	yes	no	no	poor	no	no	ckd
60	100	1.025	0	3	?	normal	notpresent	notpresent	263	27	1.3	135	4.3	12.7	37	11400	4.3	yes	yes	yes	good	no	no	ckd
62	60	1.015	1	0	?	abnormal	present	notpresent	100	31	1.6	?	?	10.3	30	5300	3.7	yes	no	yes	good	no	no	ckd

Gambar 3. Dataset CKD Sebelum Processing

Gambar 3 yang ditandai dengan tanda "?" pada dataset. Nilai-nilai yang hilang ini ditemukan diseluruh fitur yang terdapat pada penyakit PGK.

3. Penanganan Nilai Hilang

Untuk menangani nilai hilang, dilakukan metode imputasi data, seperti Gambar 4.

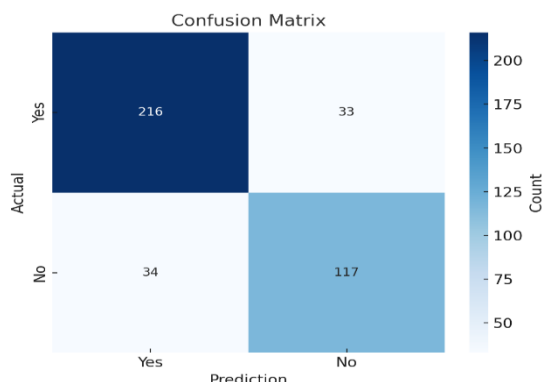
age	bp	sg	al	su	rbc	pc	pcc	ba	bgr	bu	sc	sod	pot	hemo	pcv	wc	rc	htn	dm	cad	appet	pe	ane	class
48	70	1.005	4	0	1	0	1	0	117	56	3.8	111	2.5	11.2	32	6700	3.9	1	0	0	0	1	1	1
53	90	1.020	2	0	0	0	1	0	70	107	7.2	114	3.7	9.5	29	12100	3.7	1	1	0	0	0	1	1
63	70	1.010	3	0	0	0	1	0	380	60	2.7	131	4.2	10.8	32	4500	3.8	1	1	0	0	1	0	1
68	80	1.010	3	2	1	0	1	1	157	90	4.1	130	6.4	5.6	16	11000	2.6	1	1	1	0	1	0	1
61	80	1.015	2	0	0	0	0	0	173	148	3.9	135	5.2	7.7	24	9200	3.2	1	1	1	0	1	1	1
48	80	1.025	4	0	1	0	0	0	95	163	7.7	136	3.8	9.8	32	6900	3.4	1	0	0	1	0	1	1
69	70	1.010	3	4	1	0	0	0	264	87	2.7	130	4.0	12.5	37	9600	4.1	1	1	1	1	1	0	1
73	70	1.005	0	0	1	1	0	0	70	32	0.9	125	4.0	10.0	29	18900	3.5	1	1	0	1	1	0	1
73	80	1.020	2	0	0	0	0	0	253	142	4.6	138	5.8	10.5	33	7200	4.3	1	1	1	1	0	0	1
46	60	1.010	1	0	1	1	0	0	163	92	3.3	141	4.0	9.8	28	14600	3.2	1	1	0	1	0	0	1
56	90	1.015	2	0	0	0	0	0	129	107	6.7	131	4.8	9.1	29	6400	3.4	1	0	0	1	0	0	1
48	80	1.005	4	0	0	0	0	1	133	139	8.5	132	5.5	10.3	36	6200	4	0	1	0	1	1	0	1
59	70	1.010	3	0	1	0	0	0	76	186	15	135	7.6	7.1	22	3800	2.1	1	0	0	0	1	1	1
63	100	1.010	2	2	1	1	0	1	280	35	3.2	143	3.5	13.0	40	9800	4.2	1	0	1	1	0	0	1
56	70	1.015	4	1	0	1	0	0	210	26	1.7	136	3.8	16.1	52	12500	5.6	0	0	0	1	0	0	1
71	70	1.010	3	0	1	0	1	1	219	82	3.6	133	4.4	10.4	33	5600	3.6	1	1	1	1	0	0	1
73	100	1.010	3	2	0	0	1	0	295	90	5.6	140	2.9	9.2	30	7000	3.2	1	1	1	0	0	0	1
71	60	1.015	4	0	1	1	0	0	118	125	5.3	136	4.9	11.4	35	15200	4.3	1	1	0	0	1	0	1
52	90	1.015	4	3	1	0	0	0	224	166	5.6	133	4.7	8.1	23	5000	2.9	1	1	0	1	0	1	1
50	90	1.010	2	0	1	0	1	1	128	208	9.2	134	4.8	8.2	22	16300	2.7	0	0	0	0	1	1	1

Gambar 4. Dataset CKD Setelah Processing

Pada Gambar 4. penanganan nilai yang hilang seperti fitur numerik dengan nilai rata-rata atau nilai tengah digunakan untuk menggantikan nilai yang hilang. Misalnya, tekanan darah (bp) yang kosong akan diisi dengan rata-rata tekanan darah dari seluruh data. Selain itu fitur kategori dengan nilai yang paling sering muncul (*modus*) digunakan untuk menggantikan nilai yang hilang pada fitur kategori. Dataset hasil *preprocessing* inilah yang akan digunakan pada tahap berikutnya, yaitu pelatihan model menggunakan metode Random Tree dan PSO.

3.3. Model Random Tree dan PSO

Pada tahap pertama, *Random Tree* diterapkan sebagai algoritma dasar untuk melakukan klasifikasi pada dataset penyakit ginjal kronis. Random Tree dengan model ini, data dipecah ke dalam dua kelas, yaitu PGK dan Not PGK, berdasarkan kriteria yang ditetapkan dalam pohon keputusan dengan hasil dari confusion matrix pada Gambar 5.

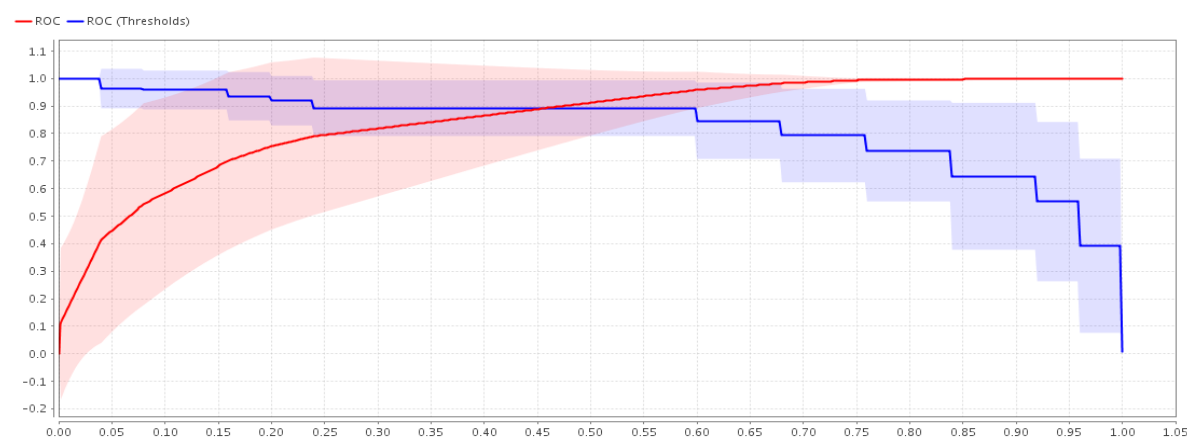


Gambar 5. Confussion Matrix Random Tree

Gambar 5 menunjukkan hasil evaluasi awal dari model Random Tree, yang memberikan 216 kasus di mana pasien benar-benar memiliki PGK (Penyakit Ginjal Kronis) dan model memprediksi dengan sesuai. Terdapat 33 kasus di mana pasien sebenarnya memiliki PGK, tetapi model salah memprediksi mereka tidak memiliki PGK. Selain itu, terdapat 34 kasus di mana pasien sebenarnya

tidak memiliki PGK, tetapi model salah memprediksi mereka memiliki PGK. Akhirnya, terdapat 117 kasus di mana pasien benar-benar tidak memiliki PGK, dan model memprediksi mereka tidak memiliki PGK. Confusion matrix ini menunjukkan bagaimana model Random Tree mampu menangani data PGK dengan tingkat akurasi keseluruhan sebesar 83.25%. Dengan demikian, model ini cukup baik dalam mengidentifikasi pasien dengan PGK.

Hasil evaluasi lebih lanjut didasarkan pada perhitungan *Area Under Curve* (AUC) dan grafik kurva ROC (*Receiver Operating Characteristic*), yang merupakan metrik untuk mengukur kemampuan model dalam membedakan antara kelas positif dan negatif. Model ini dirancang untuk memprediksi secara akurat apakah seorang pasien memiliki PGK (kelas positif) atau tidak memiliki PGK (kelas negatif). Grafik yang ditampilkan bertujuan untuk memberikan wawasan lebih mendalam tentang kinerja prediktif model.



Gambar 6. Kurva ROC Random Forest Data PGK

Pada Gambar 6, ditampilkan kurva ROC, yang menggambarkan hubungan antara tingkat *true positive rate* (TPR) dan *false positive rate* (FPR) pada berbagai *threshold*. Nilai AUC (*Area Under Curve*) sebesar 0.850 ± 0.131 menunjukkan bahwa model memiliki kemampuan klasifikasi yang baik. Kurva merah pada grafik merepresentasikan nilai ROC, sedangkan kurva biru menunjukkan distribusi *threshold*. Daerah bayangan pada grafik mencerminkan tingkat variabilitas kinerja model. Dengan nilai AUC yang cukup tinggi, model Random Tree ini menunjukkan performa yang andal dalam membedakan pasien dengan PGK dan

tanpa PGK. Hal ini mendukung potensi penggunaan model dalam konteks diagnosis klinis.

Setelah dilakukan model Random Tree maka dilakukan optimasi dengan PSO, dengan menunjukkan hasil perhitungan bobot atribut. Bobot ini menggambarkan tingkat kepentingan atau kontribusi masing-masing atribut terhadap performa model dalam melakukan klasifikasi. atribut dengan bobot lebih tinggi memiliki pengaruh yang lebih besar dalam menentukan hasil prediksi dibandingkan atribut dengan bobot lebih rendah. Hasil ini dapat membantu dalam

proses memilih atribut-atribut yang paling relevan untuk digunakan dalam pelatihan model, sehingga meningkatkan efisiensi dan akurasi. Hasil bobot atribut pada Tabel 1.

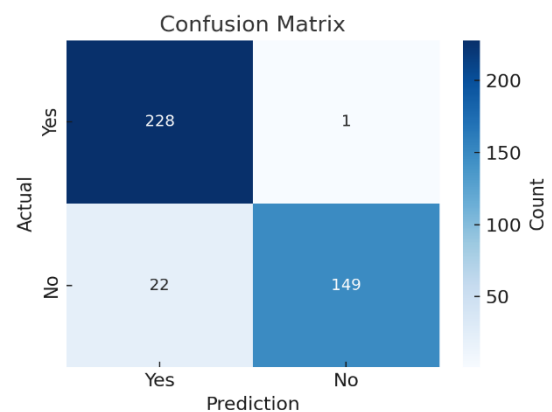
TABEL 1. BOBOT ATRIBUT

Atribut	Bobot
age	1.0
bp	0.9968772649065762
sg	0.9098404924881465
al	1.0
su	0.09866854888185733
rbc	0.37646027486417427
pc	0.8910599238471872
pcc	9.427505609281666E-4
ba	0.8317030231997394
bgr	0.9308835928810184
bu	0.0
sc	0.8682364853381699
sod	0.0
pot	1.0
hemo	0.4950735579041348
pcv	0.07197611799596029
wc	1.0
rc	1.0
htn	1.0
dm	0.0
cad	1.0
appet	0.0
pe	0.32315445109944485
ane	1.0
age	1.0
bp	0.9968772649065762

Berdasarkan Tabel 1 bobot atribut yang ditampilkan, beberapa atribut seperti age, su, dan pc memiliki bobot tertinggi (1.0), menunjukkan bahwa atribut-atribut ini sangat penting untuk model. Sebaliknya, beberapa atribut seperti ba, pot, dan dm memiliki bobot yang rendah atau nol, yang berarti kontribusi mereka terhadap prediksi relatif kecil. Informasi ini dapat digunakan untuk menyederhanakan model dengan menghapus

atribut yang tidak relevan, tanpa mengurangi kinerja prediksi secara signifikan. Analisis bobot ini memberikan wawasan yang berharga untuk meningkatkan interpretabilitas dan efisiensi model prediksi.

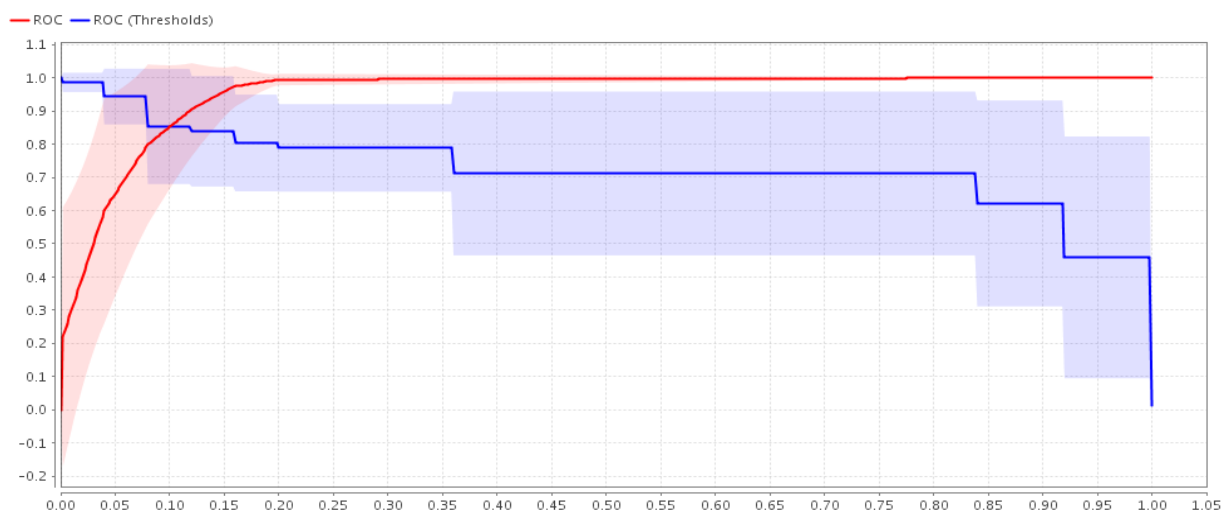
Setelah melakukan analisis terhadap bobot atribut menggunakan metode PSO, evaluasi lebih lanjut dilakukan untuk mengukur akurasi dan kinerja model Random Tree yang dioptimalkan dengan PSO dalam klasifikasi data PGK. Hasil dari pengoptimalan model pada Gambar 7.



Gambar 7. Confussion Matrix PSO

Gambar 7 menunjukkan hasil evaluasi model Random Tree yang dioptimalkan dengan PSO pada data PGK. Model ini memberikan 228 kasus di mana pasien benar-benar memiliki PGK, dan model memprediksi dengan sesuai. Terdapat 1 kasus di mana pasien sebenarnya memiliki PGK, tetapi model salah memprediksi mereka tidak memiliki PGK. Selain itu, terdapat 22 kasus di mana pasien sebenarnya tidak memiliki PGK, tetapi model salah memprediksi mereka memiliki PGK. Akhirnya, terdapat 149 kasus di mana pasien benar-benar tidak memiliki PGK, dan model memprediksi mereka tidak memiliki PGK.

Setelah melakukan evaluasi dengan confusion matrix, analisis lebih lanjut dilakukan untuk mengevaluasi kinerja model Random Tree yang dioptimalkan dengan PSO menggunakan kurva ROC dan nilai AUC.



Gambar 8. Kurva ROC PSO Data PGK

Gambar 8 menunjukkan kurva ROC untuk model Random Tree yang dioptimalkan dengan PSO, dengan nilai AUC sebesar 0.953 ± 0.034 . Nilai ini menunjukkan bahwa model memiliki kemampuan klasifikasi yang sangat baik, hampir mendekati sempurna dalam membedakan antara pasien dengan CKD dan tanpa CKD. Dengan nilai AUC yang tinggi, model ini menunjukkan keandalan yang kuat untuk diterapkan dalam aplikasi klinis, terutama untuk mendukung diagnosis penyakit ginjal kronis.

3.4. Evaluasi

Pada bagian ini dilakukan evaluasi kinerja model Random Tree dan Random Tree yang dioptimalkan dengan PSO berdasarkan beberapa metrik evaluasi, yaitu accuracy, precision, recall, dan AUC. Dengan membandingkan kinerja kedua model, dapat terlihat dampak optimasi PSO terhadap performa model. Tabel II merangkum hasil evaluasi kinerja kedua model.

TABEL II. HASIL EVALUASI RANDOM TREE DAN PSO

Algoritma	Random Tree	Random Tree + PSO
Accuracy	83,25	94,25
Precision	77,48	87,95
Recall	78,00	99,33
AUC	0,850	0,953

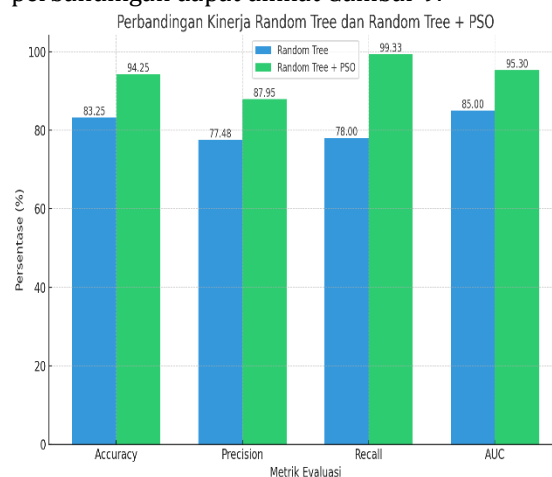
Berdasarkan tabel di atas, dapat disimpulkan bahwa optimasi menggunakan PSO memberikan peningkatan kinerja yang signifikan pada model Random Tree. Accuracy meningkat dari 83,25% menjadi 94,25%, menunjukkan bahwa model dengan PSO lebih andal dalam

menghasilkan prediksi yang benar secara keseluruhan.

Precision meningkat dari 77,48% menjadi 87,95%, yang berarti model dengan PSO lebih sedikit menghasilkan prediksi positif yang salah.

Recall mengalami peningkatan besar dari 78,00% menjadi 99,33%, yang menunjukkan bahwa model dengan PSO hampir selalu berhasil mengidentifikasi pasien dengan PGK (kelas positif).

AUC meningkat dari 0,850 menjadi 0,953, mencerminkan kemampuan model yang lebih baik dalam membedakan antara kelas positif dan negatif pada berbagai threshold. Dengan grafik perbandingan dapat dilihat Gambar 9.



Gambar 9. Grafik Perbandingan Metrik

TABEL III. HASIL EVALUASI KINERJA ALGORITMA DENGAN OPTIMASI PSO

Algoritma	Accuracy Tanpa PSO	Accuracy Optimal PSO
SVM	98,25	99,75

DT	98,25	98,50
NB	95,00	95,75
NN	98,00	99,00
KNN	80,00	97,75
RF	83,25	94,25

Hasil pada TABEL III semakin menegaskan efektivitas pendekatan yang diusulkan dalam penelitian ini, yaitu peningkatan akurasi yang signifikan pada algoritma Random Tree dari 83,25% menjadi 94,25%, menunjukkan bahwa optimasi PSO berhasil meningkatkan performa model dalam mendukung diagnosis penyakit ginjal kronis (PGK). Dibandingkan dengan algoritma lain yang juga dioptimasi PSO (seperti SVM, NN, dan KNN), Random Tree menunjukkan peningkatan akurasi yang sangat kompetitif.

PSO bekerja secara efektif pada Random Tree karena metode ini sangat bergantung pada pemilihan fitur dan pembobotan yang tepat. PSO membantu mengoptimalkan bobot atribut dengan mengeksplorasi ruang solusi secara cepat dan konvergen pada solusi terbaik.

Dibandingkan dengan DT dan NB, peningkatan akurasi pada Random Tree lebih besar karena algoritma pohon keputusan rentan terhadap pemilihan atribut yang kurang optimal tanpa optimasi.

4. Kesimpulan dan Saran

Penelitian menunjukkan bahwa integrasi antara Random Tree dan PSO memberikan peningkatan yang konsisten dan signifikan. Meskipun akurasi Random Tree + PSO (94,25%) masih sedikit lebih rendah dibandingkan SVM + PSO (99,75%), hasil ini menunjukkan bahwa metode Random Tree yang dioptimalkan menggunakan PSO tetap kompetitif dan layak dipertimbangkan, terutama mengingat sifat Random Tree yang efisien dalam menangani dataset yang kompleks.

Peningkatan kinerja yang dicapai melalui optimasi PSO juga mencerminkan kemampuan algoritma ini dalam menyesuaikan parameter model secara efektif, sehingga menghasilkan hasil klasifikasi yang lebih akurat dan dapat diandalkan dalam konteks klinis. Hasil ini mendukung hipotesis penelitian bahwa kombinasi algoritma pembelajaran mesin dengan teknik optimasi dapat meningkatkan kinerja model dalam memprediksi penyakit ginjal kronis. Penelitian ini memberikan kontribusi nyata dalam pengembangan metode prediktif untuk mendukung diagnosis penyakit ginjal kronis, sekaligus menunjukkan potensi besar PSO sebagai alat optimasi yang efektif dalam domain

medis. Dengan hasil yang diperoleh, penelitian ini membuka peluang untuk implementasi lebih lanjut, termasuk dalam sistem pendukung keputusan berbasis klinis yang dapat membantu praktisi kesehatan dalam pengambilan keputusan yang lebih akurat dan dapat mempercepat proses skrining pasien di fasilitas kesehatan, memungkinkan dokter untuk fokus pada kasus-kasus yang lebih kompleks. Dengan kemampuan mendeteksi CKD secara dini, intervensi medis yang tepat dapat segera dilakukan, sehingga memperbaiki prognosis pasien dan mengurangi biaya perawatan jangka panjang.

Saran untuk penelitian masa yang akan datang. Selain PSO, penelitian di masa depan dapat mengeksplorasi teknik optimasi lain seperti *Genetic Algorithm* atau *Differential Evolution* untuk meningkatkan performa model. Selain itu penggunaan dataset yang lebih besar, atau pengujian model dengan teknik ensemble learning.

Daftar Pustaka:

- [1] A. Alaiad, H. Najadat, B. Mohsen, and K. Balhaf, "Classification and Association Rule Mining Technique for Predicting Chronic Kidney Disease," *J. Inf. Knowl. Manag.*, vol. 19, no. 1, 2020, doi: 10.1142/S0219649220400158.
- [2] J. Qin, L. Chen, Y. Liu, C. Liu, C. Feng, and B. Chen, "A machine learning methodology for diagnosing chronic kidney disease," *IEEE Access*, vol. 8, pp. 20991–21002, 2020, doi: 10.1109/ACCESS.2019.2963053.
- [3] S. Rezayi, K. Maghooli, and S. Saeedi, "Applying Data Mining Approaches for Chronic Kidney Disease Diagnosis," *Int. J. Intell. Syst. Appl. Eng.*, 2021, doi: DOI: <https://doi.org/10.18201/ijisae.2021473640>.
- [4] R. Pramanik, S. Khare, and M. K. Gourisaria, "Inferring the Occurrence of Chronic Kidney Failure: A Data Mining Solution," 2021, doi: https://doi.org/10.1007/978-981-16-3346-1_59.
- [5] E. M. Senan *et al.*, "Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques," *J. Healthc. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/1004767.
- [6] M. Hosseinzadeh *et al.*, "A diagnostic prediction model for chronic kidney disease in internet of things platform,"

- Multimed. Tools Appl.*, vol. 80, no. 11, pp. 16933–16950, 2021, doi: 10.1007/s11042-020-09049-4.
- [7] G. Wijaya, "Improvement of Kernel SVM to Enhance Accuracy in Chronic Kidney Disease," vol. 9, no. 1, pp. 136–144, 2024, doi: <https://doi.org/10.33395/sinkron.v9i1.13112> e-ISSN.
- [8] W. Widiati, N. Iriadi, I. Ariyati, I. Nawawi, and Sugiono, "Pendekatan Hibrida Decision Tree-Particle Swarm Optimization untuk Deteksi Dini Penyakit Ginjal Kronis," *JASIEK (Jurnal Apl. Sains, Informasi, Elektron. dan Komputer)*, vol. 6, no. 1, pp. 11–21, 2024, doi: 10.26905/jasiek.v6i1.13006.
- [9] H. Nurdin, Suhardjono, A. Wuryanto, D. Yuliandari, and H. Sugiarto, "Naive Bayes and Particle Swarm Optimization in Early Detection Of Chronic Kidney Disease," *J. Tek. Inform.*, vol. 5, no. 3, pp. 703–708, 2024, doi: <https://doi.org/10.52436/1.jutif.2024.5.3.1750>.
- [10] S. Widodo, H. Brawijaya, and S. Samudi, "Building a Predictive Model for Chronic Kidney Disease: Integrating KNN and PSO," *Paradig. -J. Komput. dan Inform.*, vol. 26, no. 1, pp. 58–64, 2024, doi: 10.31294/p.v26i1.3282.
- [11] I. M. D. Priyatama and Ridwansyah, "Klasifikasi Anak Berkebutuhan Khusus Tunagrahita Menggunakan Metode Algoritma C4.5," *Paradigma*, vol. 24, no. 1, pp. 90–95, 2022, doi: <https://doi.org/10.31294/paradigma.v24i1.1087>.
- [12] H. Nurdin, I. Carolina, R. L. Andharsaputri, A. Wuryanto, and Ridwansyah, "Forward Selection as a Feature Selection Method in the SVM Kernel for Student Graduation Data," *Sink. J. dan Penelit. Tek. Inform.*, vol. 8, no. October, pp. 2531–2537, 2024, doi: 10.33395/sinkron.v8i4.14172.
- [13] I. Ariyati, S. Rosyida, K. Ramanda, V. Riyanto, S. Faizah, and Ridwansyah, "Optimization of the Decision Tree Algorithm Used Particle Swarm Optimization in the Selection of Digital Payments," in *Journal of Physics: Conference Series*, 2020, vol. 1641, no. 1, doi: 10.1088/1742-6596/1641/1/012090.
- [14] A. Hamid and Ridwansyah, "Optimizing Heart Failure Detection : A Comparison between Naive Bayes and Particle Swarm Optimization," *Paradigma*, vol. 26, no. 1, pp. 30–36, 2024, doi: <https://doi.org/10.31294/p.v26i1.3284>.
- [15] R. Darmawan, Y. H. Chrisnanto, and G. Abdillah, "KLASIFIKASI DIAGNOSA PENYAKIT TIROID MENGGUNAKAN METODE," *JIRE (Jurnal Inform. Rekayasa Elektron.)*, vol. 7, no. 2, pp. 203–209, 2024.
- [16] C. M. Zakariya et al., "ANALISIS SENTIMEN TERKAIT PROGRAM KARTU PRAKERJA MENGGUNAKAN METODE K-NEAREST NEIGHBORS," *JIRE (Jurnal Inform. Rekayasa Elektron.)*, vol. 7, no. 2, pp. 183–192, 2024.
- [17] S. H. Hassan and A. M. Abdulazeez, "A Review on Utilizing Data Mining Techniques for Chronic Kidney Disease Detection," *Indones. J. Comput. Sci.*, vol. 13, no. 3, pp. 4049–4066, 2024, doi: <https://doi.org/10.33022/ijcs.v13i3.4062>.
- [18] S. Rukiastiandari, Luthfia Rohimah, Aprillia, Chodidjah, and Fara Mutia, "Model Hibrida K-Nearest Neighbors Berbasis Genethic Algorithm untuk Prediksi Penyakit Ginjal Kronis," *Infotek J. Inform. dan Teknol.*, vol. 8, no. 1, pp. 44–55, 2025, doi: 10.29408/jit.v8i1.27918.
- [19] Ridwansyah, M. Iqbal, H. Destiana, Sugiono, and A. Hamid, "Data Mining Berbasis Machine Learning Untuk Analitik Prediktif Dalam Kelulusan," *semantik*, vol. 10, no. 2, pp. 1–10, 2024, doi: <https://doi.org/10.55679/semantik.v10i2.67>.
- [20] R. Ridwansyah, G. Wijaya, and J. J. Purnama, "Hybrid Optimization Method Based on Genetic Algorithm for Graduates Students," *J. Pilar Nusa Mandiri*, vol. 16, no. 1, pp. 53–58, 2020, doi: 10.33480/pilar.v16i1.1180.
- [21] M. Iqbal et al., "Implementation of Particle Swarm Optimization Based Machine Learning Algorithm for Student Performance Prediction," *JITK (Jurnal Ilmu Pengetah. Dan Teknol. Komputer)*, vol. 6, no. 2, pp. 195–204, 2020, doi: 10.33480/jitk.v6i2.1695.IMPLEMENTATION.