

## Model Hibrida K-Nearest Neighbors Berbasis Genethic Algorithm untuk Prediksi Penyakit Ginjal Kronis

Sinta Rukiastiandari<sup>1\*</sup>, Luthfia Rohimah<sup>2</sup>, Aprillia<sup>3</sup>, Chodidjah<sup>4</sup>, Fara Mutia<sup>5</sup>

<sup>1</sup>Program Studi Teknologi Informasi, Universitas Bina Sarana Informatika

<sup>2</sup>Program Studi Sistem Informasi, Universitas Bina Sarana Informatika

<sup>3,4</sup>Program Studi Manajemen, Universitas Bina Sarana Informatika

<sup>5</sup>Program Studi Administrasi Bisnis, Universitas Bina Sarana Informatika

\*sinta.sru@bsi.ac.id

### Abstrak

Penyakit Ginjal Kronis, yang sering disingkat dengan PGK, merupakan salah satu penyakit serius yang menjadi perhatian utama masyarakat dan dunia medis. Penyakit ini dapat menyebabkan berbagai komplikasi serius jika tidak ditangani dengan tepat dan dini. Oleh karena itu, prediksi yang akurat terhadap PGK sangatlah penting untuk mendukung intervensi dini yang dapat memperlambat progresi penyakit, mencegah komplikasi lebih lanjut, dan meningkatkan peluang kesembuhan pasien. Penelitian ini bertujuan untuk meningkatkan akurasi prediksi PGK dengan mengembangkan model hibrida yang menggabungkan algoritma K-Nearest Neighbors (KNN) dengan optimasi menggunakan Genetic Algorithm (GA). Dalam pendekatan ini, algoritma KNN digunakan untuk membangun model prediksi, sementara GA berperan sebagai alat optimasi yang meningkatkan kinerja model. Efektivitas model yang telah dioptimasi dievaluasi menggunakan metrik utama seperti akurasi, presisi, recall, dan area under the curve (AUC). Hasil menunjukkan adanya peningkatan kinerja yang signifikan, dengan akurasi meningkat sebesar 17,75%, presisi naik 23,84%, dan recall meningkat 5,34%. Penelitian ini memberikan kontribusi penting bagi pengembangan teknologi data mining untuk aplikasi klinis dan membuka peluang untuk perbaikan lebih lanjut di masa depan dalam meningkatkan akurasi prediksi penyakit kronis seperti PGK.

**Kata kunci :** Genethic Algorithm, K-Nearest Neighbours, Model Hibrida, Penyakit Ginjal Kronis

### Abstract

*Chronic Kidney Disease, which is often abbreviated as PGK, is a serious disease that is of major concern to society and the medical world. This disease can cause various serious complications if not treated properly and early. Therefore, accurate prediction of CKD is very important to support early intervention that can slow disease progression, prevent further complications, and increase the patient's chances of recovery. This research aims to increase the accuracy of PGK predictions by developing a hybrid model that combines the K-Nearest Neighbors (KNN) algorithm with optimization using the Genetic Algorithm (GA). In this approach, the KNN algorithm is used to build a prediction model, while GA acts as an optimization tool that improves model performance. The effectiveness of the optimized model is evaluated using key metrics such as accuracy, precision, recall, and area under the curve (AUC). The results show a significant increase in performance, with accuracy increasing by 17.75%, precision increasing by 23.84%, and recall increasing by 5.34%. This research makes an important contribution to the development of data mining technology for clinical applications and opens up opportunities for further improvements in the future in increasing the prediction accuracy of chronic diseases such as CKD.*

**Keywords :** *Chronic Kidney Disease, Genethic Algorithm, Hybrid Model, K-Nearest Neighbours.*

### 1. Pendahuluan

Penyakit merupakan salah satu penyebab utama morbiditas dan mortalitas di seluruh dunia [1].

Dengan meningkatnya populasi dan perubahan gaya hidup, berbagai penyakit baru muncul sementara penyakit lama tetap menjadi tantangan

besar<sup>[2]</sup>. Di antara penyakit-penyakit tersebut, penyakit yang mempengaruhi organ vital seperti ginjal sering kali mendapatkan perhatian khusus karena dampaknya yang signifikan terhadap kualitas hidup pasien<sup>[3]</sup>. Penyakit ginjal kronis (PGK) adalah masalah kesehatan global yang serius karena komplikasi berat jika tidak ditangani<sup>[4][5]</sup>. PGK ditandai dengan penurunan fungsi ginjal secara progresif, sehingga mengganggu peran ginjal dalam menyaring limbah, mengatur cairan tubuh, dan menjaga keseimbangan elektrolit<sup>[6]</sup>. Ketidakmampuan ginjal ini dapat menyebabkan berbagai gangguan kesehatan, seperti hipertensi, anemia, dan masalah tulang<sup>[7]</sup>. Penyakit ini sering kali berkembang secara perlahan dan sering tidak menunjukkan gejala hingga mencapai tahap yang lebih lanjut, yang dapat berujung pada gagal ginjal stadium akhir dan memerlukan dialisis atau transplantasi ginjal<sup>[8]</sup>.

Prediksi penyakit ginjal kronis memainkan peran penting dalam pencegahan dan manajemen penyakit ini. Dengan prediksi yang akurat, intervensi medis dapat dilakukan lebih awal yang dapat memperlambat progresi penyakit dan meningkatkan kualitas hidup pasien<sup>[9]</sup>. Metode prediksi tradisional sering kali bergantung pada evaluasi klinis dan pemeriksaan laboratorium, tetapi dengan perkembangan teknologi metode berbasis data seperti data mining semakin digunakan untuk mengidentifikasi pola dan faktor

risiko yang dapat membantu dalam diagnosis dini dan perawatan yang lebih efektif<sup>[10]</sup>. Data mining adalah proses ekstraksi informasi berharga dari kumpulan data besar. Dalam konteks medis, data mining dapat digunakan untuk menganalisis data pasien termasuk hasil tes laboratorium, riwayat medis, dan faktor risiko lainnya untuk mengidentifikasi pola yang dapat mengindikasikan kemungkinan perkembangan penyakit.

Dari penelitian yang sudah dilakukan dengan berbagai macam metode, cara kerja dan aplikasi yang berbeda, tetapi beberapa studi menunjukkan keterbatasan dalam hal akurasi, kemampuan generalisasi, dan efisiensi model. Dengan letak gap penelitian yang didapat sebagian besar metode yang ada belum mengintegrasikan algoritma prediksi berbasis jarak, seperti K-Nearest Neighbors (KNN), dengan teknik optimasi seperti Genetic Algorithm (GA) untuk meningkatkan kinerjanya. Maka kebaruan penelitian ini yaitu menggunakan metode K-Nearest Neighbor (K-NN) yang di optimasi dengan Genethic Algorithm (GA) yang sebelumnya tidak pernah dilakukan oleh peneliti lain. Kelebihan KNN adalah kemampuannya untuk melakukan klasifikasi berbasis jarak dengan cara yang sederhana dan intuitif, namun model ini sangat bergantung pada parameter seperti jumlah tetangga (k) dan metrik jarak yang digunakan, yang sering kali ditentukan secara manual dan

suboptimal. Di sisi lain, GA merupakan metode optimasi yang terinspirasi dari teori evolusi, mampu mengeksplorasi ruang pencarian secara efisien untuk menemukan solusi optimal.

Penelitian ini bertujuan untuk mengatasi keterbatasan metode prediksi sebelumnya dengan mengembangkan model hibrida yang menggabungkan KNN dan GA. Model ini diharapkan mampu memberikan prediksi yang lebih akurat dibandingkan dengan pendekatan sebelumnya, serta mendukung diagnosis dini dan intervensi yang lebih efektif untuk memperlambat progresi penyakit dan meningkatkan kualitas hidup pasien. Tujuan spesifik penelitian ini adalah mengintegrasikan algoritma KNN untuk klasifikasi berbasis jarak dengan GA sebagai alat optimasi parameter, guna meningkatkan kinerja prediksi dan menghasilkan model yang lebih adaptif.

Penelitian ini diharapkan memberikan kontribusi signifikan dalam mengatasi keterbatasan metode prediksi tradisional. Dengan meningkatkan performa model prediksi melalui integrasi KNN dan GA, penelitian ini juga bertujuan menyediakan alat diagnosis dini yang lebih efektif, serta memberikan wawasan baru tentang aplikasi teknik data mining untuk mendukung pengambilan keputusan klinis di masa depan

## **2. Tinjauan Pustaka**

### **2.1. Penelitian Terkait**

Beberapa penelitian sebelumnya telah menerapkan berbagai metode untuk prediksi penyakit ginjal kronis menggunakan teknik data mining KNN. Pada tahun 2020, metode KNN diterapkan untuk pertama kalinya dengan melakukan perbandingan terhadap lima metode lain yang sudah ada sebelumnya menggunakan aplikasi WEKA dengan hasil akurasi sebesar 98.50%<sup>[11][12]</sup>. Pada tahun 2021 dilakukan kembali metode KNN dengan mengkomparasi sebanyak empat sampai tujuh metode dengan membagi data dimana 300 patient data digunakan untuk training dan 100 pasien data digunakan untuk testing, sehingga hasil akurasi mendapatkan 71,56<sup>[13]</sup>, 96,10%<sup>[14]</sup>, 98,25%<sup>[15]</sup>, 98.33%<sup>[16]</sup>. Pada tahun 2022 metode K-NN kembali digunakan untuk data yang sama dengan hasil akurasi 94.00%<sup>[17]</sup>. Pada tahun 2024 dengan data penyakit ginjal kronis yang sama dan perlakuan pengolahan data yang sama telah dilakukan dengan beberapa metode SVM, Decision Tree, dan Naïve Bayes yang ditingkatkan dengan teknik optimasi Particle Swarm Optimization (PSO) dengan hasil nilai akurasi diantara 95,75%<sup>[18]</sup>, 98,50%<sup>[19]</sup>, 99.75%<sup>[20]</sup>. Pada tahun yang sama dengan data yang sama dan perlakuan yang sama dilakukan kembali metode K-NN dengan akurasi sebesar 80.00% dengan peningkatan

metode optimasi PSO menghasilkan nilai akurasi sebesar 97.75%<sup>[21]</sup>

## 2.2. Landasan Teori

### 1. Data Mining

Data mining adalah proses pemodelan yang memanfaatkan metode tertentu untuk memperoleh informasi yang berguna bagi ilmu pengetahuan. Selain itu, data mining juga merupakan proses eksplorasi data guna menemukan informasi yang sebelumnya tidak diketahui dari kumpulan data tersebut <sup>[22]</sup>.

### 2. K-NN

K-NN adalah algoritma *machine learning* berbasis instance yang digunakan untuk klasifikasi dan regresi. Algoritma ini bekerja dengan mencari sejumlah tetangga terdekat (K) dari data yang ingin diprediksi, berdasarkan jarak yang dihitung dari data yang sudah ada. K-NN tidak membangun model eksplisit, melainkan menyimpan semua data pelatihan dan melakukan prediksi berdasarkan kemiripan antara data baru dan data tetangganya. Keputusan klasifikasi atau nilai regresi ditentukan oleh mayoritas kelas atau rata-rata nilai dari tetangga terdekat tersebut. Algoritma ini sangat sederhana namun efektif, terutama dalam kasus di mana distribusi data tidak linier <sup>[23]</sup>.

### 3. Genetic Algorithm

Algoritma Genetika (GA) adalah teknik pencarian dan optimasi yang mengambil inspirasi dari

mekanisme seleksi alam yang terdapat dalam teori evolusi biologi. Algoritma ini dirancang untuk mencari solusi optimal atau mendekati optimal dari sebuah masalah dengan meniru konsep evolusi seperti seleksi, rekombinasi (crossover), dan mutasi. Dalam algoritma ini, sekelompok solusi yang disebut kromosom atau individu diolah, di mana setiap individu merepresentasikan sebuah kemungkinan solusi untuk masalah yang dihadapi <sup>[24]</sup>.

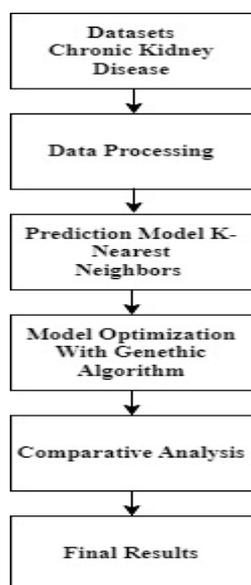
### 4. Area Under Curve (AUC)

AUC adalah ukuran dari luas di bawah kurva Receiver Operating Characteristic (ROC), yang merupakan grafik hasil dari plot antara sensitivitas (True Positive Rate) dan spesifisitas (False Positive Rate). Nilai AUC berkisar antara 0 hingga 1, dan mencerminkan kemampuan model untuk membedakan antara kelas-kelas yang berbeda. Semakin tinggi nilai AUC, semakin baik model dalam memprediksi variabel yang diteliti. Model dengan AUC mendekati 1 menunjukkan performa klasifikasi yang sangat baik, sementara nilai AUC mendekati 0,5 menunjukkan performa yang setara dengan tebak-tebakan acak<sup>[25]</sup>.

## 3. Metode Penelitian

Metodologi penelitian yang dilakukan untuk memprediksi PGK terdiri dari beberapa prosedur yang dituangkan dengan sebuah Gambar 1. Serta jenis penelitian menggunakan kuantitatif yang dikerjakan dengan terstruktur untuk menganalisis

dan mengkaji kualitas hubungan antara bagian atribut-atribut tersebut. Tahap pertama melibatkan pengumpulan data penyakit ginjal kronis yang diperoleh dari UCI *Repository*. Selanjutnya data tersebut dilakukan proses atau yang disebut data processing untuk mendapatkan hasil dari data yang valid untuk digunakan sehingga mendapatkan informasi sesuai yang diinginkan. Tahap berikutnya adalah penerapan algoritma KNN untuk membangun model prediksi berdasarkan data penyakit ginjal kronis. Tahap berikutnya adalah meningkatkan algoritma KNN dengan model optimasi GA untuk membangun model prediksi berdasarkan data PGK. Setelah itu hasil dari pengujian yang nantinya akan di uji coba salah satu rumah sakit swasta.



Gambar 1. Tahapan Penelitian

### 1. Dataset

Dataset yang digunakan dalam penelitian prediksi akurat penyakit ginjal kronis merupakan dataset

sekunder yang diperoleh dari UCI *Repository* dan berisi data tentang PGK sebanyak 300 data. Dataset ini mencakup berbagai atribut yang relevan dengan kondisi kesehatan pasien, seperti hasil tes laboratorium dan informasi demografis. Data ini merupakan dasar penting untuk analisis lebih lanjut dan pembangunan model prediksi. Dataset ini mencakup berbagai atribut yang relevan untuk analisis termasuk: *Age*, *Blood Pressure (bp)*, *various blood tests*, *Diabetes Mellitus (dm)*, *Coronary Artery Disease (cad)*, *Appetite (appet)*.

### 2. Data Processing.

Tahap Proses pengolahan data penelitian prediksi akurat penyakit ginjal kronis melibatkan pembersihan dan transformasi data mentah menjadi format yang dapat digunakan untuk analisis. Tahapan ini mencakup penanganan data yang hilang dan proses imputasi. Proses imputasi dengan memberikan nilai rata-rata pada masing-masing atribut tersebut. Tujuan dari data processing adalah untuk memastikan bahwa data yang digunakan dalam model adalah akurat, konsisten, dan siap untuk analisis lebih lanjut yang dilakukan secara langsung dengan menggunakan program *Rapid Miner 5.0*.

### 3. Prediction Model K<sub>NN</sub>

Pada tahap ini, algoritma KNN diterapkan untuk membangun model prediksi PGK. KNN adalah metode klasifikasi yang menggunakan kedekatan antara data untuk membuat prediksi. Model ini

dibangun dengan menggunakan dataset yang telah diproses untuk mengidentifikasi pola dan hubungan dalam data yang dapat digunakan untuk memprediksi kondisi pasien.

#### 4. Model Optimization With GA

Selanjutnya, algoritma GA diterapkan untuk mengoptimalkan model KNN. GA digunakan untuk mencari parameter terbaik bagi model KNN dengan menggunakan teknik evolusi [26], seperti seleksi, crossover, dan mutasi. Optimasi ini bertujuan untuk meningkatkan akurasi dan kinerja model prediksi dengan menemukan konfigurasi parameter yang paling efektif.

#### 5. Comparative Analysis

Analisis perbandingan dilakukan untuk mengevaluasi kinerja model KNN yang telah dioptimalkan dengan GA dibandingkan dengan model-model lain yang digunakan dalam penelitian sebelumnya. Proses ini melibatkan perbandingan metrik kinerja seperti akurasi, *presisi*, *recall*, dan AUC untuk menilai efektivitas model dalam memprediksi penyakit ginjal kronis. Nilai akurasi akan menunjukkan seberapa baik model dalam memprediksi secara keseluruhan. Sementara itu, *presisi* dan *recall* akan memberikan gambaran seberapa baik model dalam membedakan spesies dengan tepat.

#### 6. Final Results

Tahap akhir dari penelitian ini melibatkan penyajian hasil akhir dari model prediksi yang telah dikembangkan. Hasil ini mencakup evaluasi

dari model KNN yang telah dioptimalkan dengan GA, serta interpretasi dari metrik kinerja yang diperoleh. Kesimpulan dari penelitian ini akan disajikan untuk menunjukkan sejauh mana model hibrida ini berhasil dalam meningkatkan akurasi prediksi penyakit ginjal kronis dan implikasinya untuk aplikasi klinis.

## 4. Hasil dan Pembahasan

### 4.1. Hasil

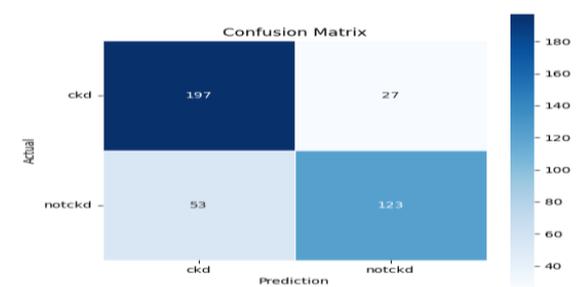
Sumber data yang dipakai adalah data penyakit ginjal kronis, dimana data yang diambil diperoleh dari data publik UCI *repository*. Data tersebut didapatkan dari rumah sakit pada tahun 2015 dengan jumlah data sebanyak 400 data sampel dengan 24 variabel yang mencakup berbagai atribut yang relevan dengan kondisi kesehatan pasien, seperti hasil tes laboratorium, informasi demografis, dan riwayat kesehatan. Beberapa atribut penting dalam dataset ini antara lain umur (*age*), tekanan darah (*bp*), berat jenis urine (*sg*), kadar albumin (*al*), kadar gula (*su*), jumlah sel darah merah (*rbc*), dan kadar hemoglobin (*hemo*). Dataset ini digunakan sebagai dasar untuk membangun dan menguji model prediksi penyakit ginjal kronis yang dikembangkan dalam penelitian ini, seperti yang ditunjukkan pada Tabel 1.

Tabel 1. Contoh Data Akhir Setelah Diolah

age	56.0	58.0	...	25.0	67.0
bp	80.0	110.0	...	60.0	80.0
sg	1017.	1010.	...	1020.	1025.
	0	0	...	0	0

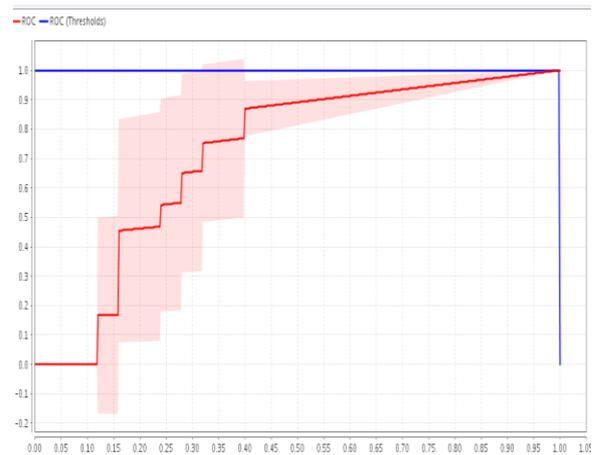
al	1.0	4.0	...	0.0	0.0
su	0.0	0.0	...	0.0	0.0
rbc	1.0	1.0	...	1.0	1.0
pc	1.0	1.0	...	1.0	1.0
pcc	2.0	2.0	...	2.0	2.0
ba	2.0	2.0	...	2.0	2.0
bgr	415.0	251.0	...	119.0	99.0
...	...	...	...	...	...
pe	2.0	2.0	...	2.0	2.0
ane	2.0	2.0	...	2.0	2.0
clas s	ckd	ckd	...	notck d	notck d

Hasil dari penerapan algoritma KNN untuk prediksi PGK ditunjukkan dalam bentuk matriks kebingungan. Matriks kebingungan ini akan memberikan gambaran tentang kinerja model dalam mengklasifikasikan data pasien sebagai PGK atau bukan PGK. Analisis terhadap matriks kebingungan akan dilakukan untuk menilai akurasi, sensitivitas, spesifisitas, serta jenis kesalahan klasifikasi yang terjadi yang akan membantu dalam mengevaluasi efektivitas model KNN dalam deteksi dini penyakit ginjal kronis. Matriks kebingungan seperti yang terlihat pada Gambar 2.



Gambar 2. Confusion Matrix KNN

Dari matriks pada Gambar 2, terlihat bahwa model KNN berhasil memprediksi 197 pasien dengan PGK secara benar (*true positives*) dan 123 pasien yang tidak memiliki PGK secara benar (*true negatives*). Namun model ini juga menghasilkan 53 kesalahan klasifikasi dimana pasien dengan PGK diklasifikasikan sebagai tidak PGK (*false negatives*), dan 27 kesalahan klasifikasi dimana pasien yang tidak memiliki PGK diklasifikasikan sebagai PGK (*false positives*). Dari hasil tersebut dapat menghasilkan Kurva ROC pada gambar 3.



Gambar 3. AUC KNN

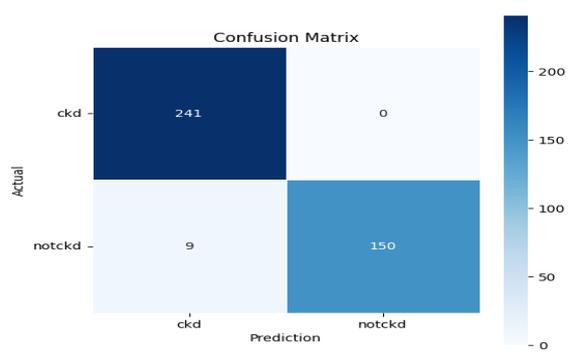
Gambar 3 menunjukkan kurva ROC dari model untuk prediksi PGK. Pada kurva ini, sumbu y mewakili *true positive rate* (sensitivitas), sementara sumbu x mewakili *false positive rate* (1-spesifisitas).

Kurva ROC yang mendekati sudut kiri atas grafik menunjukkan bahwa model memiliki kinerja yang baik dalam membedakan antara kelas PGK dan not PGK. AUC memberikan ukuran kuantitatif dari kemampuan diskriminatif model. Dalam hal ini,

nilai AUC yang diperoleh adalah 0.500. Nilai AUC sebesar 0.500 menunjukkan bahwa model KNN memiliki kemampuan yang setara dengan penembakan acak dalam membedakan antara kelas CKD dan not CKD.

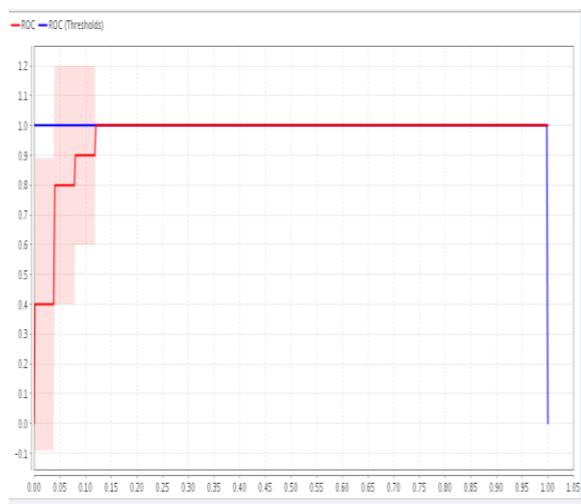
Setelah melakukan penerapan algoritma K-NN, Maka akan dilakukan analisis bobot atribut dengan hasil yang didapat seperti *sg*, *pc*, dan *pcc* menonjol dengan bobot yang lebih tinggi menandakan bahwa mereka memiliki pengaruh signifikan dalam model prediksi PGK. Sementara itu, atribut seperti *age*, *bp*, dan *su* memiliki bobot nol yang menunjukkan bahwa atribut-atribut ini tidak memberikan kontribusi yang berarti dalam model setelah optimasi. Analisis ini penting untuk memastikan bahwa model fokus pada variabel yang benar-benar relevan, sehingga dapat meningkatkan akurasi prediksi dan keandalan model dalam penerapan klinis yang dapat dilihat pada matriks kebingungan.

Matriks ini memberikan gambaran tentang efektivitas model KNN setelah dioptimasi dengan GA dalam memprediksi kasus CKD dan *not* CKD. Hasil dari optimasi ini ditunjukkan melalui matriks kebingungan yang telah terlihat pada Gambar 4.



Gambar 4. *Confusion Matrix* KNN+GA

Gambar 4 menunjukkan hasil dari matriks kebingungan setelah penerapan optimasi dengan GA. Dari gambar tersebut, terlihat bahwa model yang telah dioptimasi berhasil memprediksi 241 pasien dengan PGK secara benar (*true positives*) dan 150 pasien yang tidak memiliki PGK secara benar (*true negatives*). Model ini juga menunjukkan bahwa hanya terdapat 9 kesalahan klasifikasi dimana pasien yang sebenarnya memiliki PGK diklasifikasikan sebagai tidak memiliki PGK (*false negatives*), dan tidak ada kesalahan klasifikasi pada pasien yang tidak memiliki PGK (*false positives*). Kurva ROC yang dihasilkan dari model prediksi KNN –GA untuk penyakit ginjal kronis pada gambar 5.



Gambar 5. AUC KNN + GA

Gambar 5 menunjukkan kurva ROC dari model KNN optimasi GA untuk prediksi PGK. Pada kurva ini, sumbu y mewakili *true positive rate* (sensitivitas), sementara sumbu x mewakili *false positive rate* (1-spesifisitas). Kurva ROC yang mendekati sudut kiri atas grafik menunjukkan bahwa model memiliki kinerja yang baik dalam membedakan antara kelas CKD dan *not* CKD. AUC memberikan ukuran kuantitatif dari kemampuan diskriminatif model. Dalam hal ini, nilai AUC yang diperoleh adalah 0.500. Nilai AUC sebesar 0.500 menunjukkan bahwa model KNN+GA memiliki kemampuan yang setara dengan penebakan acak dalam membedakan antara kelas CKD dan *not* CKD.

#### 4.2. Pembahasan

Untuk menilai efektivitas model K-NN dan model K-NN yang telah dioptimasi menggunakan GA, maka akan dilakukan analisis perbandingan

dengan model yang telah diterapkan tersebut. Analisis perbandingan ini mencakup evaluasi berbagai metrik kinerja seperti akurasi, presisi, recall, dan AUC. Setiap model dievaluasi berdasarkan kemampuannya dalam mengklasifikasikan data pasien dengan benar sebagai CKD atau *not* CKD. Hasil perbandingan ini memberikan wawasan tentang keunggulan dan kelemahan masing-masing model dalam konteks prediksi penyakit ginjal kronis.

Table 2. Perbandingan Algoritma

Algoritma	KNN	KNN+GA
<i>Accuracy</i>	80.00%	97.75%
<i>Precision</i>	70.82%	94.66%
<i>Recall</i>	82.00%	100.00%
<i>AUC</i>	0.500%	0.500%

Hasil dari analisis perbandingan yang ditampilkan dalam Tabel 2 menunjukkan peningkatan yang signifikan dalam kinerja model KNN setelah dioptimasi menggunakan GA. Akurasi model meningkat dari 80.00% menjadi 97.75%, menunjukkan bahwa optimasi dengan GA secara substansial meningkatkan kemampuan model untuk mengklasifikasikan pasien dengan benar sebagai CKD atau *not* CKD. Selain peningkatan akurasi, presisi model juga mengalami peningkatan dari 70.82% menjadi 94.66%, yang berarti model yang dioptimasi lebih baik dalam mengidentifikasi pasien dengan PGK tanpa memberikan terlalu banyak kesalahan positif. Recall model juga mencapai 100.00% setelah

optimasi, menandakan bahwa model yang dioptimasi berhasil mendeteksi semua kasus PGK dengan benar tanpa ada yang terlewat. Namun, meskipun ada peningkatan yang signifikan dalam akurasi, presisi, dan recall, nilai AUC tetap pada 0.500 untuk kedua model.

Model K-NN tanpa optimasi menunjukkan kelemahan dalam mendeteksi pasien PGK dengan matriks kebingungan menghasilkan 53 false negatives dan 27 false positives. Hal ini menunjukkan bahwa banyak pasien PGK yang tidak terdeteksi (false negatives) atau salah diklasifikasikan sebagai non-PGK (false positives). Kinerja yang kurang optimal ini disebabkan oleh sensitivitas K-NN terhadap parameter seperti jumlah tetangga (k) dan metrik jarak yang digunakan.

Setelah optimasi menggunakan GA, model menunjukkan peningkatan yang signifikan. Analisis bobot atribut menunjukkan bahwa fitur seperti specific gravity (sg), packed cell volume (pc), dan pus cell clumps (pcc) memiliki bobot yang lebih tinggi, menandakan pengaruh besar dalam prediksi. Sebaliknya, atribut seperti age, blood pressure (bp), dan sugar (su) memiliki bobot nol, yang berarti atribut tersebut kurang relevan dalam prediksi. Matriks kebingungan setelah optimasi menunjukkan hanya 9 false negatives dan tidak ada false positives. Hal ini mencerminkan bahwa model yang dioptimasi lebih baik dalam mendeteksi pasien PGK

sekaligus menghindari kesalahan klasifikasi.

Dengan akurasi, sensitivitas, dan keandalan yang lebih baik, model K-NN yang dioptimasi dengan GA dapat diandalkan sebagai alat bantu diagnostik untuk deteksi dini PGK. Penggunaan model ini mendukung intervensi medis yang lebih cepat dan efektif, sehingga dapat memperlambat progresi penyakit serta meningkatkan kualitas hidup pasien

## 5. Kesimpulan

Penelitian ini berhasil mengembangkan model prediksi penyakit ginjal kronis dengan menggunakan algoritma KNN yang dioptimasi dengan GA. Hasil dari penelitian menunjukkan bahwa optimasi menggunakan GA mampu meningkatkan kinerja model KNN secara signifikan, dengan peningkatan akurasi dari 80.00% menjadi 97.75%. Selain itu, presisi dan *recall* juga mengalami peningkatan masing-masing mencapai 94.66% dan 100.00%. Namun, nilai AUC tetap pada 0.500 menunjukkan bahwa kemampuan diskriminatif model untuk membedakan antara PGK dan bukan PGK tidak mengalami peningkatan yang signifikan. Meskipun demikian, hasil penelitian ini menunjukkan bahwa optimasi dengan GA dapat membuat model KNN menjadi lebih andal dan akurat untuk diagnosis dini penyakit ginjal kronis, yang sangat penting untuk aplikasi klinis. Temuan ini memberikan kontribusi signifikan terhadap

pengembangan metode prediksi penyakit ginjal kronis, serta menyediakan dasar yang kuat untuk penelitian lebih lanjut dalam upaya meningkatkan kualitas diagnosis dan manajemen penyakit ginjal kronis.

## 6. Daftar Pustaka

- [1] E. H. A. Rady and A. S. Anwar, "Prediction of kidney disease stages using data mining algorithms," *Informatics Med. Unlocked*, vol. 15, no. December 2019, p. 100178, 2019, doi: 10.1016/j.imu.2019.100178.
- [2] Arif-UI-Islam and S. H. Ripon, "Rule Induction and Prediction of Chronic Kidney Disease Using Boosting Classifiers, Ant-Miner and J48 Decision Tree," *2nd Int. Conf. Electr. Comput. Commun. Eng. ECCE 2019*, pp. 1–6, 2019, doi: 10.1109/ECACE.2019.8679388.
- [3] S. B. Akben, "Early Stage Chronic Kidney Disease Diagnosis by Applying Data Mining Methods to Urinalysis, Blood Analysis and Disease History," *Irbm*, vol. 39, no. 5, pp. 353–358, 2019, doi: 10.1016/j.irbm.2018.09.004.
- [4] S. Revathy, B. Bharathi, P. Jeyanthi, and M. Ramesh, "Chronic kidney disease prediction using machine learning models," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 1, pp. 6364–6367, 2019, doi: 10.35940/ijeat.A2213.109119.
- [5] R. A. S. P. S, K. V Rangarao, and A. Saranya, "Efficient datamining model for prediction of chronic kidney disease using wrapper methods," *Int. J. Informatics Commun. Technol.*, vol. 8, no. 2, p. 63, 2019, doi: 10.11591/ijict.v8i2.pp63-70.
- [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] E. Purwaningsih, "Improving the Performance of Support Vector Machine With Forward Selection for Prediction of Chronic Kidney Disease," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 8, no. 1, pp. 18–24, 2022, doi: 10.33480/jitk.v8i1.3327.
- [8] 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>.
- [9] Fathurrahman, Yupi Kuspani Putra, and M. A. J. Hidayat, "Penerapan Algoritma Naïve Bayes Untuk Analisis Pengaruh Faktor Pendidikan Terhadap Peningkatan Kesehatan Masyarakat," *Infotek J. Inform. dan Teknol.*, vol. 7, no. 2, 2024.
- [10] B. A. C. Permana, Muhammad Sadali, and R. Ahmad, "Penerapan Model Decision Tree Menggunakan Python Untuk Prediksi Faktor Dominan Penyebab Penyakit Stroke," *Infotek J. Inform. dan Teknol.*, vol. 7, no. 1, 2024.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] M. Mamun, A. Farjana, M. Al Mamun, M. S. Ahammed, and M. M. Rahman, "Heart failure survival prediction using machine

- learning algorithm: am I safe from heart failure?," 2022, doi: 10.1109/AlloT54504.2022.9817303.
- [15] P. Chittora *et al.*, "Prediction of Chronic Kidney Disease - A Machine Learning Perspective," *IEEE Access*, vol. 9, pp. 17312–17334, 2021, doi: 10.1109/ACCESS.2021.3053763.
- [16] R. A. M. Elghalid, A. Alwirshiffani, A. A. I. Mohamed, F. H. A. Aldeeb, and A. Andiasha, "Comparison of Some Machine Learning Algorithms for Predicting Heart Failure," 2022, doi: 10.1109/ICEMIS56295.2022.9914325.
- [17] I. Saha, M. K. Gourisaria, and G. M. Harshvardhan, "Classification System for Prediction of Chronic Kidney Disease Using Data Mining Techniques," *Lect. Notes Networks Syst.*, vol. 318, no. May 2017, pp. 429–443, 2022, doi: 10.1007/978-981-16-5689-7\_38.
- [18] 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>.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] 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> Paradigma,.
- [23] V. Riyanto, A. Hamid, and Ridwansyah, "Prediction of Student Graduation Time Using the Best Algorithm," *Indones. J. Artif. Intell. Data Min.*, vol. 2, no. 2, pp. 1–9, 2019, doi: <http://dx.doi.org/10.24014/ijaidm.v2i1.6424>
- [24] Ridwansyah and E. Purwaningsih, "Mengukur Tingkat Error Ketahanan Beton dengan Metode Klasifikasi Neural Network dan Support Vector Machine," *J. Tek. Inform.*, vol. 5, no. 1, 2019, doi: <https://doi.org/10.51998/jti.v5i1.295>.
- [25] Ridwansyah and E. Purwaningsih, "Particle Swarm Optimization Untuk Meningkatkan Akurasi Prediksi Pemasaran Bank," *J. PILAR Nusa Mandiri*, vol. 14, no. 1, pp. 83–88, 2018.
- [26] 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