

**ANALISIS PERBANDINGAN PERFORMA MODEL K-NEAREST
NEIGHBORS, DAN NAIVE BAYES DENGAN HYPERPARAMETER
OPTIMIZATION DALAM MENDIAGNOSIS PENYAKIT GINJAL
KRONIS (CKD)**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

MUHAMMAD FAZARULLAH

22.11.4761

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS PERBANDINGAN PERFORMA MODEL K-NEAREST
NEIGHBORS, DAN NAIVE BAYES DENGAN *HYPERPARAMETER
OPTIMIZATION* DALAM MENDIAGNOSIS PENYAKIT GINJAL
KRONIS (CKD)**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

MUHAMMAD FAZARULLAH

22.11.4761

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS PERBANDINGAN PERFORMA MODEL K-NEAREST
NEIGHBORS, DAN NAIVE BAYES DENGAN *HYPERPARAMETER
OPTIMIZATION* DALAM MENDIAGNOSIS PENYAKIT GINJAL
KRONIS (CKD)**

yang disusun dan diajukan oleh

Muhammad Fazarullah

22.11.4761

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 25 Februari 2026

Dosen Pembimbing,



Rumini, S.Kom., M.Kom.

NIK. 190302246

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS PERBANDINGAN PERFORMA MODEL K-NEAREST
NEIGHBORS, DAN NAIVE BAYES DENGAN *HYPERPARAMETER
OPTIMIZATION* DALAM MENDIAGNOSIS PENYAKIT GINJAL
KRONIS (CKD)**

yang disusun dan diajukan oleh

Muhammad Fazarullah

22.11.4761

Telah dipertahankan di depan Dewan Penguji
pada tanggal 25 Februari 2026

Susunan Dewan Penguji

Nama Penguji

Rizqi Sukma Kharisma, S.Kom., M.Kom.
NIK. 190302215

Ade Pujiyanto, S.Kom., M.Kom.
NIK. 190302494

Rumini, S.Kom., M.Kom.
NIK. 190302246

Tanda Tangan

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 25 Februari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Muhammad Fazarullah
NIM : 22.11.4761

Menyatakan bahwa Skripsi dengan judul berikut:

ANALISIS PERBANDINGAN PERFORMA MODEL K-NEAREST NEIGHBORS, DAN NAIVE BAYES DENGAN *HYPERPARAMETER OPTIMIZATION* DALAM MENDIAGNOSIS PENYAKIT GINJAL KRONIS (CKD)

Dosen Pembimbing : Rumini, S.Kom., M.Kom.

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, <25 Februari 2026>

Yang Menyatakan,



Muhammad Fazarullah

HALAMAN PERSEMBAHAN

Skripsi ini saya persembahkan kepada kedua orang tua dan keluarga tercinta atas doa, dukungan, serta pengorbanan yang tiada henti dalam mendukung perjalanan pendidikan saya, serta kepada dosen pembimbing dan seluruh pihak yang telah memberikan bimbingan, arahan, dan ilmu sehingga penelitian ini dapat terselesaikan dengan baik. Semoga karya ini dapat menjadi langkah awal dalam memberikan kontribusi di bidang teknologi informasi, khususnya dalam penerapan *machine learning* untuk mendukung proses prediksi dan deteksi dini Penyakit Ginjal Kronis.



KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Shalawat serta salam semoga senantiasa tercurah kepada Nabi Muhammad SAW beserta keluarga dan para sahabatnya. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika Universitas Amikom Yogyakarta.

Dalam penyusunan skripsi ini, penulis memperoleh banyak bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, penulis menyampaikan ucapan terima kasih kepada:

1. Ibu **Rumini, S.Kom., M.Kom.**, selaku dosen pembimbing atas bimbingan, arahan, dan motivasi yang diberikan selama proses penyusunan skripsi.
2. Bapak **Rizqi Sukma Kharisma, S.Kom., M.Kom.**, selaku dosen penguji atas saran dan masukan yang membangun demi penyempurnaan skripsi ini.
3. Bapak **Ade Pujianto, S.Kom., M.Kom.**, selaku dosen penguji atas evaluasi serta masukan yang diberikan.
4. Kedua orang tua dan keluarga penulis atas doa dan dukungan yang tiada henti.
5. Seluruh dosen dan staf Program Studi Informatika Universitas Amikom Yogyakarta serta semua pihak yang telah membantu penyelesaian skripsi ini.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan. Semoga skripsi ini dapat memberikan manfaat bagi pembaca dan perkembangan ilmu pengetahuan.

Yogyakarta, 25, Februari 2026

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH.....	xiii
INTISARI	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA	7
2.1 Studi Literatur	7
2.2 Dasar Teori.....	15

2.2.1 Penyakit Ginjal Kronis (<i>Chronic Kidney Disease / CKD</i>)	15
2.2.2 Machine Learning	15
2.2.3 Klasifikasi	16
2.2.4 Algoritma <i>K-Nearest Neighbors</i> (KNN).....	16
2.2.5 Algoritma <i>Naive Bayes</i>	17
2.2.6 <i>Hyperparameter</i> dan <i>Hyperparameter Optimization</i>	19
2.2.7 Data Preprocessing.....	19
2.2.8 Evaluasi Kinerja Model Klasifikasi.....	20
BAB III METODE PENELITIAN	22
3.1 Alur Penelitian.....	22
3.1.1 Pengumpulan Dataset.....	23
3.1.2 <i>Preprocessing Data</i>	24
3.1.3 <i>Splitting Data</i>	26
3.1.4 Pemodelan <i>machine learning</i>	26
3.1.5 <i>Hyperparameter Optimization (Grid Search)</i>	42
3.1.6 Evaluasi Peforma	43
3.2 Alat dan Bahan	44
3.2.1 Dataset Penelitian.....	44
3.2.2 <i>Tools and Library</i>	48
3.2.3 <i>Hardware</i>	48
3.2.4 <i>Software</i>	49
BAB IV HASIL DAN PEMBAHASAN	50
4.1 Dataset.....	50
4.2 <i>Preprocessing Data</i>	51
4.2.1 <i>Missing Value</i>	52

4.2.2	Label Encoding	56
4.2.3	Normalisasi Data	59
4.2.4	Feature Correlation Matrix	60
4.3	Split Data	61
4.4	Hyperparameter Optimization dan Permodelan	62
4.4.1	K-Nearest Neighbor sebelum Hyperparameter Optimization	62
4.4.2	K-Nearest Neighbor setelah Hyperparameter Optimization	63
4.4.3	Naïve Bayes sebelum Hyperparameter Optimization	64
4.4.4	Naïve Bayes setelah Hyperparameter Optimization	65
4.5	Evaluasi Performa	66
4.5.1	K-Nearest Neighbor sebelum Hyperparameter Optimization	66
4.5.2	K-Nearest Neighbor setelah Hyperparameter Optimization	69
4.5.3	Naïve Bayes sebelum Hyperparameter Optimization	72
4.5.4	Naïve Bayes setelah Hyperparameter Optimization	74
4.5.5	Hasil Perbandingan kedua Algoritma	78
4.5.6	Hasil Deployment Prediksi CKD	80
BAB V PENUTUP		83
5.1	Kesimpulan	83
5.2	Saran	84
REFERENSI		85

DAFTAR TABEL

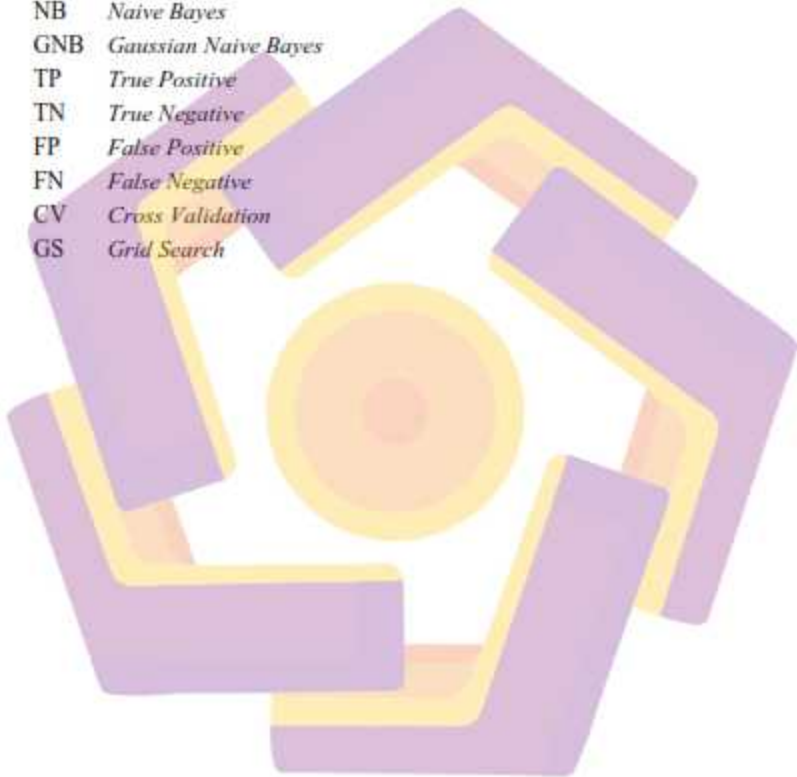
Tabel 2.1 Keaslian Penelitian	14
Tabel 3.1 Informasi Dataset	23
Tabel 3.2 Data Training	27
Tabel 3.3 Data Testing	28
Tabel 3.4 Hasil Jarak <i>Euclidean</i>	31
Tabel 3.5 Analisis Tetangga Terdekat dan Class	31
Tabel 3.6 Data Training	34
Tabel 3.7 Data Testing	34
Tabel 3.8 Nilai Rata-Rata Setiap Fitur pada Masing-Masing Kelas	35
Tabel 3.9 Nilai Varians Setiap Fitur pada Masing-Masing Kelas	36
Tabel 3.10 Nilai <i>Gaussian Probability Density Function</i>	37
Tabel 3.11 Hasil Perhitungan <i>Likelihood</i> , <i>Posterior Probability</i> , dan Keputusan Klasifikasi	41
Tabel 3.12 Deskripsi Fitur dan Tipe Data	47
Tabel 4.1 Dataset	50
Tabel 4.2 Sebelum dan Sesudah Penanganan <i>Missing Value</i>	55
Tabel 4.3 Hasil Label Encoding pada Fitur Kategorikal	59
Tabel 4.4 Komposisi Pembagian Data	61
Tabel 4.5 Parameter <i>Grid Search</i> Algoritma <i>K-Nearest Neighbor</i>	63
Tabel 4.6 Parameter <i>Grid Search</i> Algoritma <i>Naive Bayes</i>	65
Tabel 4.7 Metrik Evaluasi Klasifikasi Model KNN	68
Tabel 4.8 Parameter Optimal KNN Hasil <i>Grid Search</i>	70
Tabel 4.9 Metrik Evaluasi Model KNN Setelah <i>Grid Search</i>	72
Tabel 4.10 Metrik Evaluasi Model <i>Naive Bayes</i> (Default)	74
Tabel 4.11 Parameter Optimal <i>Naive Bayes</i> Hasil <i>Grid Search</i>	75
Tabel 4.12 Metrik Evaluasi Model <i>Naive Bayes</i> setelah <i>Grid Search</i>	77
Tabel 4.13 Perbandingan Performa Algoritma <i>K-Nearest Neighbor</i> dan <i>Naive Bayes</i>	79

DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	22
Gambar 4.1 Jumlah <i>Missing Value</i> pada Setiap Fitur	52
Gambar 4.2 Penanganan <i>Missing Value</i> pada Dataset	53
Gambar 4.3 Hasil Penanganan <i>Missing Value</i> pada Dataset	54
Gambar 4.4 Dataset Sebelum Proses Label Encoding	56
Gambar 4.5 Proses Konversi Data Kategorikal ke Numerik	57
Gambar 4.6 Dataset Setelah Proses Label Encoding	58
Gambar 4.7 Heatmap Korelasi Antar Fitur	60
Gambar 4.8 <i>Confusion Matrix</i> Model <i>K-Nearest Neighbor</i> Parameter	67
Gambar 4.9 <i>Confusion Matrix</i> KNN (<i>Grid Search</i>)	71
Gambar 4.10 <i>Confusion Matrix</i> Naive Bayes (Default)	73
Gambar 4.11 <i>Confusion Matrix</i> Naive Bayes (<i>Grid Search</i>)	76
Gambar 4.12 Perbandingan <i>Recall</i> Model KNN dan Naive Bayes	78
Gambar 4.13 Hasil Deploy Model untuk Prediksi NotCKD	81
Gambar 4.14 Hasil Deploy Model untuk Prediksi CKD	82

DAFTAR LAMBANG DAN SINGKATAN

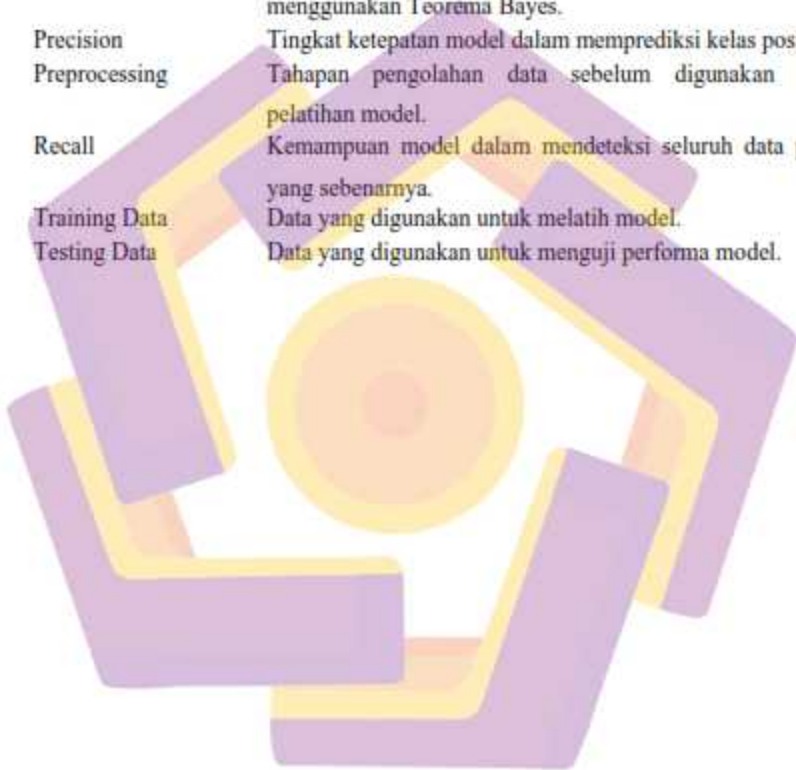
CKD	<i>Chronic Kidney Disease</i>
KNN	<i>K-Nearest Neighbor</i>
NB	<i>Naive Bayes</i>
GNB	<i>Gaussian Naive Bayes</i>
TP	<i>True Positive</i>
TN	<i>True Negative</i>
FP	<i>False Positive</i>
FN	<i>False Negative</i>
CV	<i>Cross Validation</i>
GS	<i>Grid Search</i>



DAFTAR ISTILAH

Accuracy	Tingkat ketepatan model dalam memprediksi seluruh data secara keseluruhan.
Chronic Kidney Disease (CKD)	Penyakit ginjal kronis yang ditandai dengan penurunan fungsi ginjal secara progresif.
Classification Report	Ringkasan hasil evaluasi model yang mencakup nilai accuracy, precision, recall, dan F1-score.
Confusion Matrix	Tabel evaluasi yang menunjukkan jumlah prediksi benar dan salah dalam klasifikasi.
Cross Validation	Teknik evaluasi model dengan membagi data menjadi beberapa bagian untuk menguji kestabilan performa.
Dataset	Sekumpulan data yang digunakan untuk pelatihan dan pengujian model.
Deployment	Proses implementasi model ke dalam sistem aplikasi.
Distance Metric	Metode pengukuran jarak antar data pada algoritma KNN.
Euclidean Distance	Metode pengukuran jarak garis lurus antar dua titik dalam ruang fitur.
F1-Score	Rata-rata harmonis dari precision dan recall.
False Negative	Kondisi ketika data positif diprediksi sebagai negatif oleh model.
Feature (Fitur)	Variabel atau atribut yang digunakan sebagai input dalam model.
Grid Search	Metode pencarian kombinasi parameter terbaik dengan menguji seluruh kemungkinan parameter.
Hyperparameter	Parameter yang ditentukan sebelum proses pelatihan model dimulai.
Hyperparameter Optimization	Proses pencarian kombinasi hyperparameter terbaik untuk meningkatkan performa model.
Klasifikasi	Proses pengelompokan data ke dalam kelas tertentu berdasarkan pola yang dipelajari.
K-Nearest Neighbor (KNN)	Algoritma klasifikasi berbasis jarak yang menentukan kelas berdasarkan mayoritas tetangga terdekat.
Label Encoding	Proses mengubah data kategorikal menjadi bentuk numerik.

Machine Learning	Metode komputasi yang memungkinkan sistem belajar dari data untuk melakukan prediksi.
Min-Max Scaling	Metode normalisasi yang mengubah nilai fitur ke dalam rentang 0–1.
Missing Value	Data yang hilang atau tidak terisi dalam dataset.
Naive Bayes	Algoritma klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes.
Precision	Tingkat ketepatan model dalam memprediksi kelas positif.
Preprocessing	Tahapan pengolahan data sebelum digunakan dalam pelatihan model.
Recall	Kemampuan model dalam mendeteksi seluruh data positif yang sebenarnya.
Training Data	Data yang digunakan untuk melatih model.
Testing Data	Data yang digunakan untuk menguji performa model.



INTISARI

Penyakit Ginjal Kronis (*Chronic Kidney Disease/CKD*) merupakan penyakit progresif yang dapat menyebabkan penurunan fungsi ginjal secara permanen apabila tidak terdeteksi sejak dini, sehingga diperlukan pendekatan berbasis data untuk membantu proses identifikasi risiko penyakit secara lebih cepat dan akurat. Penelitian ini bertujuan untuk membandingkan kinerja dua algoritma *machine learning*, yaitu *K-Nearest Neighbor* (KNN) dan *Naive Bayes*, dalam mengklasifikasikan kondisi Penyakit Ginjal Kronis pada pasien. Dataset yang digunakan diperoleh dari *UCI Machine Learning Repository* dan melibatkan sejumlah fitur klinis serta hasil pemeriksaan laboratorium yang berkaitan dengan fungsi ginjal. Metode penelitian meliputi tahap *preprocessing* data, pembagian data latih dan data uji, serta penerapan *hyperparameter optimization* menggunakan metode *Grid Search* untuk meningkatkan kinerja model. Evaluasi performa dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*, dengan *recall* dijadikan sebagai metrik utama karena berfokus pada kemampuan model dalam mendeteksi kasus positif CKD. Hasil penelitian menunjukkan bahwa algoritma KNN setelah dilakukan optimasi mengalami peningkatan kinerja yang signifikan dengan nilai *recall* sebesar 98%, *accuracy* sebesar 98,8%, dan *F1-score* sebesar 99%, sedangkan algoritma *Naive Bayes* menunjukkan performa yang stabil dengan nilai *recall* sebesar 96% dan tidak mengalami peningkatan signifikan setelah proses optimasi. Berdasarkan hasil tersebut, algoritma *K-Nearest Neighbor* yang telah dioptimasi direkomendasikan untuk digunakan sebagai sistem pendukung dalam membantu proses prediksi dan deteksi dini Penyakit Ginjal Kronis (CKD) secara komputasional.

Kata kunci: Penyakit Ginjal Kronis, *K-Nearest Neighbor*, *Naive Bayes*, Klasifikasi, *Hyperparameter Optimization*

ABSTRACT

Chronic Kidney Disease (CKD) is a progressive condition that may lead to permanent loss of kidney function if not detected at an early stage, highlighting the need for data-driven approaches to support accurate and timely risk identification. This study aims to compare the performance of two machine learning algorithms, namely K-Nearest Neighbor (KNN) and Naive Bayes, in classifying Chronic Kidney Disease. The dataset used in this study was obtained from the UCI Machine Learning Repository and consists of clinical features and laboratory test results related to kidney function. The research methodology includes data preprocessing, data splitting into training and testing sets, and hyperparameter optimization using the Grid Search method to improve model performance. Model evaluation was conducted using accuracy, precision, recall, and F1-score, with recall selected as the primary metric due to its focus on identifying positive CKD cases. The experimental results indicate that the optimized KNN algorithm achieved a significant performance improvement, reaching a recall of 98%, an accuracy of 98.8%, and an F1-score of 99%, while the Naive Bayes algorithm demonstrated stable performance with a recall of 96% and no significant improvement after optimization. Based on these findings, the optimized K-Nearest Neighbor algorithm is recommended as a decision support system to assist in the computational prediction and early detection of Chronic Kidney Disease.

Keyword: *Chronic Kidney Disease, K-Nearest Neighbor, Naive Bayes, Classification, Hyperparameter Optimization*