

# Identifikasi Penyakit Ginjal Kronis Menggunakan Algoritma K-Nearest Neighbour (k-NN)

Allan Krisnanta Wijaya<sup>1</sup>, Rahimatus Syifa<sup>2</sup>, Ilyasa Nanda Rahmadiano<sup>3</sup>, Rinci Kembang Hapsari<sup>4\*</sup>

<sup>1,2,3,4</sup> Teknik Informatika, Institut Teknologi Adhi Tama Surabaya

<sup>1</sup>[alallan.wjy450@gmail.com](mailto:alallan.wjy450@gmail.com)

<sup>2</sup>[sifaspt317@gmail.com](mailto:sifaspt317@gmail.com)

<sup>3</sup>[ilyasananda16@gmail.com](mailto:ilyasananda16@gmail.com)

<sup>4</sup> Teknik Informatika, Institut Teknologi Adhi Tama Surabaya

\*Corresponding author email: [rincikembang@itats.ac.id](mailto:rincikembang@itats.ac.id)

**Abstrak**— Indonesia sekarang dihadapkan pada berbagai persoalan terutama kesehatan yang lazim seperti di negara maju maupun di negara berkembang lainnya yaitu penyakit-penyakit kronis. Hal ini akibat proses degeneratif dan perubahan gaya hidup. Penyakit tersebut, antara lain seperti gagal ginjal kronis, hipertensi, diabetes mellitus, penyakit jantung koroner, stroke, dan sebagainya. Ginjal kronis adalah keadaan dimana penurunan progresif atau fungsi ginjal menurun secara bertahap dalam beberapa bulan atau tahun akibat kerusakan jaringan ginjal. Tujuan dari penelitian ini adalah mengidentifikasi penyakit ginjal kronis. Perkembangan data mining telah banyak digunakan dalam bidang Kesehatan. Dalam penelitian ini menggunakan konsep klasifikasi dalam data mining. Algoritma klasifikasi yang digunakan untuk proses identifikasi penyakit ginjal kronis dalam penelitian ini adalah algoritma k-Nearest Neighbor (k-NN). Proses algoritma k-NN berdasarkan jarak terdekat antara data latih dan data uji. Klasifikasi k-NN sangat tergantung pada value k yang merupakan penentu jarak terdekat. Kinerja algoritma k-NN sangat bergantung juga pada kualitas data yang digunakan. Dataset penelitian ini merupakan data sekunder yang diunduh dari Kaggle.com. Dataset dengan judul Chronic Kidney Disease. Performa sistem diuji dengan menggunakan konsep 5-fold validasi, Dimana data latih dibagi menjadi 5 subset yang sama ukurannya. Selama proses validasi, satu fold digunakan sebagai data validasi, sementara sisanya digunakan sebagai data latih. Dari pengujian sebanyak 5 kali yang telah dilakukan, perhitungan metode k-NN ini didapatkan hasil terbaik yakni Akurasi 63,33%, Spesifisitas 66,11%, Sensitivitas 60,89%.

**Kata Kunci**— Penyakit Ginjal Kronis, k-Nearest Neighbor, Klasifikasi

## I. PENDAHULUAN

Indonesia sekarang dihadapkan berbagai persoalan terutama kesehatan yang biasa dihadapi di negara maju maupun di negara berkembang lainnya yaitu penyakit-penyakit kronis akibat perubahan gaya hidup dan proses degeneratif, seperti gagal ginjal kronis, diabetes mellitus, hipertensi, stroke, penyakit jantung koroner, dan lain - lain. Data Riskesdas tahun 2018 menunjukkan bahwa penyakit tidak menular pada tahun 2018 dan penyakit gagal ginjal kronis merupakan salah satu penyakit yang pevelensinya meningkat setiap tahun dari penyakit tidak menular. Kementerian Kesehatan pun memaparkan hal yang sama bahwa penyakit gagal ginjal kronis

ini setiap tahunnya mengalami peningkatan dan memerlukan penanganan yang serius [1].

Ginjal kronis adalah keadaan dimana penurunan progresif atau fungsi ginjal menurun secara bertahap dalam beberapa bulan atau tahun akibat kerusakan jaringan ginjal. Pada awal penyakit ini, penderita sering tidak merasakan gejala apa-apa. Untuk membantu mendiagnosa penyakit ginjal pada pasien baru kronis, bila dilakukan pemeriksaan darah dengan indikasi penyakit ini yang lebih parah [2].

Kemajuan ilmu pengetahuan dan teknologi membermanfaat dalam kehidupan manusia. Salah satu pengetahuan yang dapat diimplementasikan yaitu Data mining. Dalam Data Mining dapat digunakan untuk melakukan diagnosa penyakit.

Teknik Data Mining yang bisa dilakukan adalah Teknik klasifikasi. Klasifikasi adalah salah satu metode analisis data yang dapat digunakan dalam menyelesaikan problem dalam suatu pengelompokan data. Dari penelitian klasterisasi menggunakan algoritma PCA-FCM oleh Dian Pertiwi, Lilik Linawati, Ferdy Samuel Rondonuwu (2021), menghasilkan tingkat akurasi dari algoritma PCA-FCM sebesar  $\geq 50\%$  [5][6]. Algoritma klasifikasi lain yang biasa digunakan yaitu algoritma k-NN (K-Nearest Neighbour). Algoritma k-NN merupakan algoritma klasifikasi yang mempunyai beberapa kelebihan, Dimana penerapannya yang sederhana tetapi efektif dalam banyak kasus. Algoritma nearest neighbor merupakan algoritma yang menggunakan pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama yang berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada yang memiliki kesamaan [3].

Algoritma k-NN telah diimplementasikan dalam mengembangkan sistem prediksi kanker paru-paru, yang menjawab kebutuhan penting akan deteksi dini penyakit kanker paru-paru. Dalam penelitian ini dataset terdiri dari 10.000 catatan pasien yang mencakup demografi, riwayat kesehatan, dan fitur radiografi, kami melakukan prapemrosesan dan normalisasi sebelum melatih, memvalidasi, dan menguji model KNN. Hasil pengujian secara efektif menunjukkan potensi dalam mencapai tingkat akurasi sebesar 95,0% dalam memprediksi kanker paru-paru, sebagaimana dibuktikan dengan pra-pemrosesan yang ketat dan optimalisasi kumpulan data yang terdiri dari 10.000 catatan pasien [4].

Dalam penelitian lain, k-NN digunakan untuk memprediksi jenis penyakit, termasuk penyakit COVID-19, dari gambar

rontgen dada. Hasil penelitian menunjukkan nilai akurasi berada pada rentang antara 1 hingga 0,965, dengan konstanta akurasi validasi sebesar 0,95 [5].

Algoritma k-NN mempunyai beberapa kelebihan antara lain tangguh terhadap data uji yang noisy serta apabila data latihnya besar menjadi efektif. Sedangkan kekurangan dari algoritma k-NN yaitu lamanya proses prediksi dengan klasifikasi karena menggunakan seluruh dataset untuk diproses. Hal utama dalam melakukan prediksi adalah cross-validation yaitu untuk menemukan poin terbaik atau nilai optimum dari hasil analisis, dan juga teknik komposisi yang menentukan berapa banyak data pengujian dan pelatihan yang dibutuhkan. Adapun dalam proses klasifikasi digunakan confusion matrix untuk mendapatkan akurasi klasifikasi dari total kelas aktif dan tidak aktif dari dataset untuk kedua algoritma yang digunakan [6].

Adapun tujuan dari penelitian ini untuk mengetahui yang terkena penyakit ginjal kronis dengan mengimplementasikan metode KNN yang mana pengklasifikasian data yang tidak diketahui berdasarkan data uji dan latih yang diketahui. Hal ini dapat dilakukan dengan menggunakan prosedur matematis untuk mengevaluasi nilai dari kriteria yang berbeda.

## II. LANDASAN TEOR

### A. Normalisasi Data

Data yang terdapat dalam dataset masih banyak yang tidak lengkap, sehingga pra-pemrosesan data diperlukan. Pengolahan data dilakukan dalam penelitian ini dengan cara memasukkan data menggunakan teknik mengganti nilai yang hilang dengan nilai menggunakan perhitungan rata-rata, median, dan modus. Teknik normalisasi data dalam penelitian ini digunakan teknik min-max score untuk menormalkan data pada skala yang sama [1].

Normalisasi Min-Max adalah metode normalisasi yang dilakukan dengan mentransformasi linier terhadap data asli sehingga dihasilkan perbandingan nilai yang seimbang antara data sebelum dan sesudah proses [2]. Perhitungan min-max score menggunakan persamaan(1) [3][6][7][8]:

$$X_{\text{baru}} = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (1)$$

dengan,

$X_{\text{baru}}$  = Hasil dari normalisasi

$X$  = Nilai yang lama

$\max(X)$  = Nilai tertinggi pada dataset

$\min(X)$  = Nilai terendah pada dataset

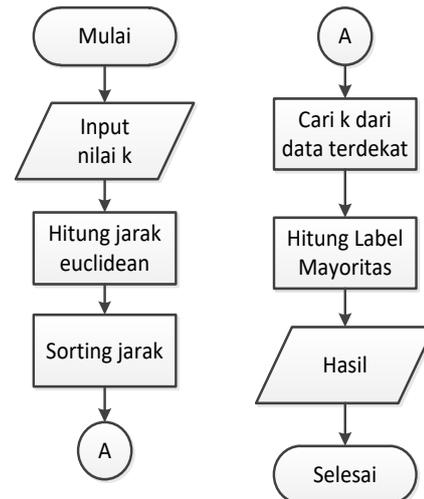
### B. k-Nearest Neighbor

Pada Tahun 1951 Fix dan Hodges memperkenalkan metode non-parametrik, yaitu metode k-NN. Dalam klasifikasi dan identifikasi model, metode ini telah banyak digunakan. K-NN bekerja berdasarkan jarak terpendek antara data pelatihan dan data uji[9]. Klasifikasi k-NN sangat tergantung pada nilai k yang merupakan penentu jarak terpendek [10],[11].

Kualitas data yang digunakan sangat memengaruhi kinerja algoritma ini. KNN memiliki keunggulan seperti kemudahan implementasi dan pemeliharaan terhadap

kumpulan data besar [1]. Langkah-langkah proses klasifikasi dengan menggunakan k-NN yaitu [12],[13]:

1. Menentukan value dari parameter k (jumlah tetangga paling dekat)
2. Menghitung kuadrat jarak Euclidian (*query instance* setiap objek terhadap data sampel yang diberikan)
3. Mengurutkan objek-objek tersebut kedalam klaster yang mempunyai jarak Euclidian paling kecil
4. Mengumpulkan kategori y (kelas tetangga terdekat)
5. Dengan menggunakan kategori y yang paling mayoritas, maka bisa diprediksi value query instance yang telah dihitung



Gbr. 1 Alur Algoritma k-NN

Alur proses dalam algoritma k-NN, ditunjukkan pada Gbr 1. Rumus euclidean distance pada persamaan berikut [14],[15], [16]:

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \quad (2)$$

dengan,

$x_{1i}$  = Data Training ke-i

$x_{2i}$  = Data Uji/Testing ke-i

$i$  = Variabel Data

$d_i$  = Jarak Data ke-i

$p$  = Dimensi Data

### C. Cross Validation

Cross-validation adalah cara untuk menemukan poin terbaik atau nilai optimum dari hasil analisis, dan juga teknik komposisi yang menentukan berapa banyak data pengujian dan pelatihan yang dibutuhkan [17]. *K-fold cross-validation* merupakan cross-validation yang sesuai untuk mengevaluasi performa proses metode algoritma dengan membagi sampel data secara random dan kelompok data maksimum bernilai k kali. Sehingga salah satu dari grup k-fold ini digunakan sebagai data uji sedangkan group lainnya akan digunakan sebagai data latih [18], [19].

#### D. Evaluasi

Proses evaluasi digunakan untuk mengukur model klasifikasi dari dataset yang telah dihitung. Evaluasi system digunakan Confusion Matrix. Dimana Confusion Matrix memberikan informasi tentang hasil klasifikasi dari dataset yang dapat diprediksi [12]

TABEL I  
UKURAN FONT UNTUK MAKALAH

| Classification | Predicted Class |            |          |
|----------------|-----------------|------------|----------|
|                |                 | Class= Yes | Class=No |
| Observed Class | Class = Yes     | TP         | FN       |
|                | Class = No      | FP         | TN       |

Rumus untuk menghitung nilai akurasi, spesifisitas, dan sensitivitas adalah sebagai berikut [10], [20]

$$\text{Akurasi} = \frac{TP+TN}{TP+FN+FP+TN} \quad (3)$$

$$\text{Spesifisitas} = \frac{TP}{TP+FN} \quad (4)$$

$$\text{Sensitivitas} = \frac{TN}{FP+TN} \quad (5)$$

dengan,

True Positif (TP) = Jumlah data penyakit ginjal kronis yang terdeteksi penyakit ginjal kronis

False Negatif (FN) = Jumlah data penyakit ginjal kronis yang terdeteksi bukan penyakit ginjal kronis

False Positif (FP) = Jumlah data penyakit bukan ginjal kronis yang terdeteksi penyakit ginjal kronis

True Negatif (TN) = Jumlah data penyakit bukan ginjal kronis yang terdeteksi penyakit bukan ginjal kronis

### III. METODE PENELITIAN

Dalam metodologi penelitian menunjukkan alur penelitian yang kami lakukan. Dimana alur penelitian tersebut ditunjukkan pada Gbr 2.

#### A. Dataset

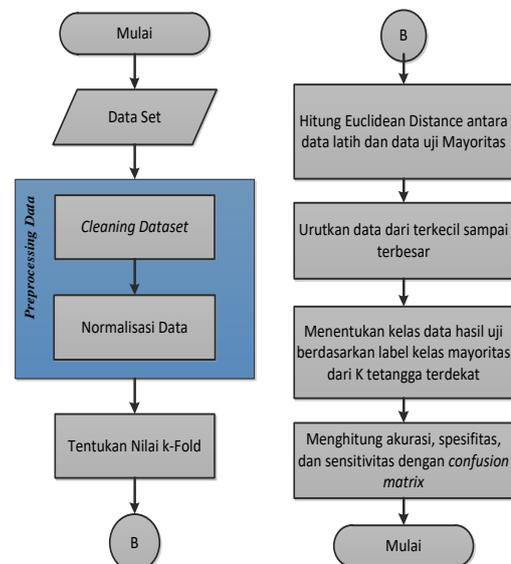
Dalam penelitian ini menggunakan Dataset, yang merupakan data publik yang diambil dari situs Kaggle. Dataset berjudul Chronic Kidney Disease. Penelitian ini menggunakan data set penyakit ginjal kronis ([https://www.kaggle.com/datasets/colearninglounge/chronic-kidney-disease?select=kidney\\_disease\\_train.csv](https://www.kaggle.com/datasets/colearninglounge/chronic-kidney-disease?select=kidney_disease_train.csv)), sejumlah 400 record data dengan 2 label kelas yang ditunjukkan pada atribut target dan total atribut 25 atribut. Label kelas, yaitu label kelas notckd (not chronic kidney disease) dan ckd (chronic kidney disease).

#### B. Preprocessing Data

Dataset yang digunakan masih terdapat missing value (data yang hilang), dimana bentuk informasi atau kelengkapan data kurang, yang disebabkan oleh tidak diberikan informasi atau

kurangnya informasi untuk suatu tentang objek, sulit dicari ataupun tidak adanya informasi tersebut[21].

Pada tahap preprocessing data, dilakukan dua proses, yaitu : 1) *Cleaning* data, pada dataset penelitian ini masih terdapat *missing value* dan *dirty data*. Sehingga diperlukan *cleaning* data untuk membersihkan data-data tersebut; 2) *Normalisasi*, proses ini diperlukan untuk melakukan standarisasi data. Dalam proses normalisasi ini dilakukan transformasi atau mengubah data asli, dengan menskalakan data semua atribut dalam rentang 0 – 1. Dimana normalisasi data digunakan metode Min-Max.



Gbr. 2 Alur Tahapan Penelitian

### IV. HASIL DAN PEMBAHASAN

Dalam penelitian ini, isi dataset terdiri dari 400 record data dengan 25 atribut. Dataset tersebut dalam kondisi missing value dan dirty, sehingga dilakukan preprocessing, dengan melakukan cleaning dataset dan Normalisasi. Sehingga dataset yang digunakan sebagai pelatihan dan pengujian dalam penelitian ini sejumlah 90 record. Dimana terdiri dari 90 data (45 kelas chronic kidney disease dan 45 kelas not-chronic kidney disease), dengan 10 atribut. Atribut tersebut, yaitu : age, bp, bgr, bu, sc, sod, pot, hemo, wc, dan rc

TABEL II  
FOLD 1 JARAK DATA UJI TERHADAP DATA LATIH

| Data Latih | Data Uji |       |     |       |
|------------|----------|-------|-----|-------|
|            | Ke-1     | Ke-2  | ... | Ke-18 |
| 1          | 1.187    | 1.578 | ... | 1.442 |
| 2          | 0.638    | 0.438 | ... | 0.244 |
| 3          | 0.608    | 1.052 | ... | 0.928 |
| ...        | ...      | ...   | ... | ...   |
| ...        | ...      | ...   | ... | ...   |
| 71         | 0.603    | 0.988 | ... | 0.880 |
| 72         | 0.644    | 1.063 | ... | 1.020 |

Karena keterbatasan jumlah records dalam dataset, pada proses pengujian kami melakukan dengan *k-fold cross validation*. Dengan pembagian *cross validation* sebanyak *k-fold* = 5.

Sehingga dataset terbagi menjadi 80% sebagai data dan 20% sebagai data uji.

Dalam 1 fold terdiri 18 data uji (9 records chronic kidney disease dan 9 records not-chronic kidney disease) dan 72 data latih (36 records chronic kidney disease dan 36 records not-chronic kidney disease). Setiap iterasi, data uji dihitung jaraknya menggunakan Euclidean Distance, menggunakan rumus (2) terhadap data latih.

TABEL III  
FOLD 2 JARAK DATA UJI TERHADAP DATA LATIH

| Data Latih | Data Uji |       |     |       |
|------------|----------|-------|-----|-------|
|            | Ke-1     | Ke-2  | ... | Ke-18 |
| 1          | 1.402    | 0.522 | ... | 0.993 |
| 2          | 1.080    | 0.757 | ... | 0.892 |
| 3          | 1.314    | 0.522 | ... | 0.654 |
| ...        | ...      | ...   | ... | ...   |
| ...        | ...      | ...   | ... | ...   |
| 71         | 0.745    | 1.861 | ... | 1.711 |
| 72         | 1.442    | 0.244 | ... | 1.001 |

Hasil perhitungan jarak data uji dan data latih pada fold 1 disampaikan pada Tabel II, fold 2 disampaikan pada Tabel III, fold 3 ditunjukkan pada Tabel IV, fold 4 ditunjukkan pada Tabel V, fold 5 ditunjukkan pada Tabel VI.

TABEL IV  
FOLD 3 JARAK DATA UJI TERHADAP DATA LATIH

| Data Latih | Data Uji |       |     |       |
|------------|----------|-------|-----|-------|
|            | Ke-1     | Ke-2  | ... | Ke-18 |
| 1          | 0,548    | 0,850 | ... | 1.321 |
| 2          | 0,435    | 0,666 | ... | 1.184 |
| 3          | 0,361    | 0,674 | ... | 1.106 |
| ...        | ...      | ...   | ... | ...   |
| ...        | ...      | ...   | ... | ...   |
| 71         | 0,598    | 0,421 | ... | 0.724 |
| 72         | 0,993    | 0,892 | ... | 1.075 |

TABEL V  
FOLD 4 JARAK DATA UJI TERHADAP DATA LATIH

| Data Latih | Data Uji |       |     |       |
|------------|----------|-------|-----|-------|
|            | Ke-1     | Ke-2  | ... | Ke-18 |
| 1          | 0.486    | 0.273 | ... | 0.675 |
| 2          | 0.394    | 0.396 | ... | 0.910 |
| 3          | 0.491    | 0.534 | ... | 0.894 |
| ...        | ...      | ...   | ... | ...   |
| ...        | ...      | ...   | ... | ...   |
| 71         | 0.874    | 0.694 | ... | 0.606 |
| 72         | 1.321    | 1.184 | ... | 1.027 |

Setelah perhitungan pada setiap iterasi fold, dilakukan pengurutan jarak dari yang terdekat/terkecil ke jarak terbesar. Karena dalam implementasi metode k-NN dalam penelitian ini menggunakan k = 10, maka dari data jarak yang sudah diurutkan diambil 10 data jarak terkecil yang digunakan sebagai penentu data uji masuk pada kelas ckd atau non-ckd.

Hasil kelas dilihat dari seberapa banyak kelas chronic kidney disease (ckd) dan not-chronic kidney disease (notckd) dari 10 jarak terkecil, data mayoritas menentukan letak kelasnya. Nilai

confusion matrix didapatkan dari 1-fold sampai 5-old ditunjukkan pada Gambar 3.

| Ford ke-1       |  |         |         | Ford ke-2       |   |     |         |   |   |
|-----------------|--|---------|---------|-----------------|---|-----|---------|---|---|
| Kelas Prediksi  |  |         |         | Kelas Prediksi  |   |     |         |   |   |
| Kelas Observasi |  | ckd     | Not-ckd | Kelas Observasi |   | ckd | Not-ckd |   |   |
|                 |  | ckd     | 8       |                 | 1 |     | ckd     | 6 | 3 |
|                 |  | Not-ckd | 5       |                 | 4 |     | Not-ckd | 4 | 5 |

| Ford ke-3       |  |         |         | Ford ke-4       |   |     |         |   |   |
|-----------------|--|---------|---------|-----------------|---|-----|---------|---|---|
| Kelas Prediksi  |  |         |         | Kelas Prediksi  |   |     |         |   |   |
| Kelas Observasi |  | ckd     | Not-ckd | Kelas Observasi |   | ckd | Not-ckd |   |   |
|                 |  | ckd     | 3       |                 | 6 |     | ckd     | 7 | 2 |
|                 |  | Not-ckd | 5       |                 | 4 |     | Not-ckd | 2 | 7 |

| Ford ke-5       |  |         |         |   |
|-----------------|--|---------|---------|---|
| Kelas Prediksi  |  |         |         |   |
| Kelas Observasi |  | ckd     | Not-ckd |   |
|                 |  | ckd     | 5       | 4 |
|                 |  | Not-ckd | 2       | 7 |

Gbr. 3 Confusion Matrix dengan 5-Fold

Hasil kelas dilihat dari seberapa banyak kelas chronic kidney disease (ckd) dan not-chronic kidney disease (notckd) dari 10 jarak terkecil, data mayoritas menentukan letak kelasnya. Nilai confusion matrix didapatkan dari 1-fold sampai 5-old ditunjukkan pada Gbr 3.

TABEL VI  
FOLD 5 JARAK DATA UJI TERHADAP DATA LATIH

| Data Latih | Data Uji |       |     |       |
|------------|----------|-------|-----|-------|
|            | Ke-1     | Ke-2  | ... | Ke-18 |
| 1          | 0.650    | 0.391 | ... | 0.644 |
| 2          | 0.499    | 0.326 | ... | 1.063 |
| 3          | 0.337    | 0.517 | ... | 0.968 |
| ...        | ...      | ...   | ... | ...   |
| ...        | ...      | ...   | ... | ...   |
| 71         | 1.217    | 1.192 | ... | 1.225 |
| 72         | 0.675    | 0.910 | ... | 0.878 |

Berdasarkan nilai confusion matrix nilai akurasi, spesefikasi dan sesnsitifitas setiap fold ditunjukkan pada Tabel VII. Dan nilai rata-rata akurasi adalah 63.55%, rata-rata spesifitasi adalah 66,11%, dan rata-rata sensitifitasnya adalah 60,89%.

TABEL VII  
FOLD 3 JARAK DATA UJI TERHADAP DATA LATIH

| Fold | Akurasi | Spesifikasi | Sensitifitas |
|------|---------|-------------|--------------|
| 1    | 66.67   | 88.89       | 44.44        |
| 2    | 66.67   | 75.00       | 60.00        |
| 3    | 38.89   | 33.33       | 44.44        |
| 4    | 77.78   | 77.78       | 77.8         |
| 5    | 66.67   | 55.56       | 77.8         |

## V. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, diperoleh kesimpulan bahwa penelitian implementasi metode k-NN untuk klasifikasi Penyakit Ginjal Kronis mendapatkan hasil akurasi terbaik pada k-fold ke-4, spesifisitas terbaik pada k-fold ke-1, dan sensitivitas terbaik pada k-fold ke-4 dan k-fold ke-5. Sedangkan hasil akurasi terburuk pada k-fold ke-3, spesifisitas terburuk pada k-fold ke-3, dan sensitivitas terburuk pada k-fold ke-1 dan k-fold ke-3. Tingkat akurasi masih tergolong sedang untuk dataset Penyakit Ginjal Kronis.

## VI. REFERENSI

- [1] A. Ariani and Samsuryadi, "Klasifikasi Penyakit Ginjal Kronis menggunakan K-Nearest Neighbor," *Pros. Annu. Res. Semin.* 2019, vol. 5, no. 1, pp. 148–151, 2019.
- [2] H. Henderi, "Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer," *IJIS Int. J. Informatics Inf. Syst.*, vol. 4, no. 1, pp. 13–20, 2021, doi: 10.47738/ijis.v4i1.73.
- [3] S. Ribaric and I. Fratric, "Experimental evaluation of matching-score normalization techniques on different multimodal biometric systems," *Proc. Mediterr. Electrotech. Conf. - MELECON*, vol. 2006, pp. 498–501, 2006, doi: 10.1109/melcon.2006.1653147.
- [4] K. Moon and A. Jetawat, "Predicting Lung Cancer with K-Nearest Neighbors (KNN): A Computational Approach," *Indian J. Sci. Technol.*, vol. 17, no. 21, pp. 2199–2206, 2024, doi: 10.17485/ijst/v17i21.1192.
- [5] M. Firdaus, "KNN Machine Learning Architecture for Pneumonia Chest X-Ray Clustering KNN Machine Learning Architecture for Pneumonia Chest X-Ray Clustering," no. June, 2024, doi: 10.26418/telectrical.v2i1.78604.
- [6] S. G. K. Patro and K. K. sahu, "Normalization: A Preprocessing Stage," *Iarjset*, no. March, pp. 20–22, 2015, doi: 10.17148/iarjset.2015.2305.
- [7] A. S. M. Al-rawahnaa, A. Yahya, and B. Al, "Data mining for Education Sector , a proposed concept," *JournalofAppliedDataSciss*, vol. 1, no. 1, pp. 1–10, 2020.
- [8] R. K. Hapsari, A. H. Salim, B. D. Meilani, T. Indriyani, and A. Rachman, "Comparison of the Normalization Method of Data in Classifying Brain Tumors with the k-NN Algorithm," in *The 2nd International Conference on Neural Networks and Machine Learning*, Atlantis Press International BV, 2023, pp. 21–29. doi: 10.2991/978-94-6463-174-6\_3.
- [9] A. Rachmad, R. K. Hapsari, W. Setiawan, T. Indriyani, E. M. S. Rochman, and B. D. Satoto, "Classification of Tobacco Leaf Quality Using Feature Extraction of Gray Level Co-occurrence Matrix (GLCM) and K-Nearest Neighbor (K-NN)," vol. 1. Atlantis Press International BV, 2023. doi: 10.2991/978-94-6463-174-6\_4.
- [10] A. Deviyanto and M. D. R. Wahyudi, "Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 3, no. 1, p. 1, 2018, doi: 10.14421/jiska.2018.31-01.
- [11] A. Salam, J. Zeniarja, and R. S. U. Khasanah, "Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekpress Indonesia)," *Pros. SINTAK*, pp. 480–486, 2018.
- [12] W. Yunus, "Algoritma K-Nearest Neighbor Berbasis Particle Swarm Optimization Untuk Prediksi Penyakit Ginjal Kronik," *J. Tek. Elektro CosPhi*, vol. 2, no. 2, pp. 51–55, 2018.
- [13] L. Farokhah, "Implementasi K-Nearest Neighbor untuk Klasifikasi Bunga Dengan Ekstraksi Fitur Warna RGB," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, p. 1129, 2020, doi: 10.25126/jtiik.2020722608.
- [14] M. Safaat, A. Sahari, and D. Lusiyantri, "Implementasi Metode K-Nearest Neighbor Untuk Mengklasifikasi Jenis Penyakit Katarak," *J. Ilm. Mat. Dan Terap.*, vol. 17, no. 1, pp. 92–99, 2020, doi: 10.22487/2540766x.2020.v17.i1.15184.
- [15] M. Reza Noviansyah, T. Rismawan, D. Marisa Midyanti, J. Sistem Komputer, and F. H. MIPA Universitas Tanjungpura Jl Hadari Nawawi, "Penerapan Data Mining Menggunakan Metode K-Nearest Neighbor Untuk Klasifikasi Indeks Cuaca Kebakaran Berdasarkan Data Aws (Automatic Weather Station) (Studi Kasus: Kabupaten Kubu Raya)," *J. Coding, Sist. Komput. Untan*, vol. 06, no. 2, pp. 48–56, 2018.
- [16] W. W. Pribadi, A. Yunus, and A. Sartika Wiguna, "Perbandingan Metode K-Means Euclidean Distance Dan Manhattan Distance Pada Penentuan Zonasi Covid-19 Di Kabupaten Malang," *J. Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 493–500, 2022.
- [17] R. R. Waliyansyah and C. Fitriyah, "Perbandingan Akurasi Klasifikasi Citra Kayu Jati Menggunakan Metode Naive Bayes dan k-Nearest Neighbor (k-NN)," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 2, p. 157, 2019, doi: 10.26418/jp.v5i2.32473.
- [18] D. Cahyanti, A. Rahmayani, and S. A. Husniar, "Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 39–43, 2020, doi: 10.33096/ijodas.v1i2.13.
- [19] R. K. Hapsari, M. Miswanto, R. Rulaningtyas, H. Suprajitno, and G. H. Seng, "Modified Gray-Level Haralick Texture Features for Early Detection of Diabetes Mellitus and High Cholesterol with Iris Image," *Int. J. Biomed. Imaging*, vol. 2022, 2022, doi: <https://doi.org/10.1155/2022/5336373>.
- [20] R. K. Hapsari, M. Miswanto, R. Rulaningtyas, and H. Suprajitno, "Identification of Diabetes Mellitus and High Cholesterol Based on Iris Image," *J. Hunan Univ. (Natural Sci.)*, vol. 48, no. 10, pp. 151–160, 2021.
- [21] I. W. Gamadarena and I. Waspada, "Implementasi Data Mining untuk Deteksi Penyakit Ginjal Kronis (PGK) menggunakan K-Nearest Neighbor (KNN) dengan Backward Elimination," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 2, p. 417, 2020, doi: 10.25126/jtiik.2020721896.