



Implementation of Naïve Bayes Classifier and K-Nearest Neighbor Algorithms for Chronic Kidney Disease Classification

Implementasi Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor untuk Klasifikasi Penyakit Ginjal Kronik

Vina Wulandari^{1*}, Windy Junita Sari², Zhevin Alfian³,
Legito⁴, Teguh Arifianto⁵

^{1,2,3}Program Studi Sistem Informasi, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia
⁴Prodi Sistem Informasi, Fakultas Sains dan Teknik, Universitas Tjut Nyak Dhien Medan, Indonesia
⁵Jurusan Teknologi Elektro Perkeretaapian, Politeknik Perkeretaapian Indonesia Madiun, Indonesia

Email: ¹12150321941@students.uin-suska.ac.id, ²12150324759@students.uin-suska.ac.id,
³12150311690@students.uin-suska.ac.id, ⁴legitostt@gmail.com, ⁵teguh@ppi.ac.id

Received Nov 27th 2023; Revised Mar 10th 2024; Accepted Sept 19th 2024
Corresponding Author: Vina Wulandari

Abstract

The kidneys are vital organs that play a crucial role in the body. They are responsible for maintaining the body's metabolic balance by eliminating toxins and metabolic waste in the form of urine. Chronic kidney disease is a condition in which the kidneys experience a long-term decline in function. The prevalence of chronic kidney disease patients in Indonesia is significant. Therefore, the classification of chronic kidney disease is conducted using the Naïve Bayes Classifier (NBC) and K-nearest Neighbor (KNN) algorithms, which have good accuracy. Based on the research results, the classification of chronic kidney disease using the NBC algorithm has an accuracy of 94.25%, an average recall value of 94.23%, precision of 98.40%, and an AUC of 0.961. On the other hand, the classification using the K-NN algorithm has an accuracy of 77.79%, recall of 95.06%, precision of 80.20%, and an AUC of 0.627. Both results indicate that the classification using the NBC algorithm outperforms the K-NN algorithm.

Keywords: Chronic Kidney Disease, Classification, Data Mining, K-NN, NBC

Abstrak

Ginjal adalah salah satu organ vital yang memiliki peranan sangat penting dalam tubuh dan memiliki fungsi untuk menjaga keseimbangan metabolisme tubuh dengan mengeluarkan racun dari dalam tubuh dan limbah metabolisme dalam bentuk urine. Penyakit ginjal kronik ialah kondisi di mana ginjal mengalami penurunan fungsi yang berlangsung dalam jangka waktu yang lama. Jumlah nilai prelevansi penderita PGK di Indonesia yang terbilang besar. Maka dari itu dilakukan klasifikasi Penyakit ginjal kronik dengan algoritma *Naïve Bayes Classifier (NBC)* dan *K-nearest Neighbor (K-NN)* yang mempunyai nilai akurasi yang baik. Berdasarkan Hasil penelitian yang diperoleh klasifikasi PGK menggunakan algoritma NBC memiliki akurasi sebesar 94,25%, rata-rata nilai recall 94,23%, presisi 98,40% dan AUC 0,961, Sedangkan klasifikasi menggunakan algoritma K-NN memiliki akurasi sebesar 77,79%, recall 95,06%, presisi 80,20% dan AUC sebesar 0,627. Dari kedua hasil menunjukkan bahwa klasifikasi menggunakan algoritma NBC lebih baik dibanding menggunakan algoritma K-NN.

Kata Kunci: Data Mining, Klasifikasi, K-NN, NBC, Penyakit Ginjal Kronik

1. PENDAHULUAN

Dari sekian banyak orang diantaranya tidak mengetahui apa fungsi dari ginjal mereka dan dimana letaknya [1]. Padahal ginjal merupakan salah satu organ penting yang memiliki peran besar dalam menjaga keseimbangan metabolisme tubuh dengan cara mengeluarkan racun dan limbah metabolisme melalui urine. Oleh karena itu, sangatlah penting bagi kita untuk menyadari betapa pentingnya memiliki ginjal yang sehat. Penyakit ginjal merupakan kondisi yang mengancam nyawa karena seringkali sulit dideteksi oleh penderitanya

sendiri karena kurangnya gejala yang khusus. Namun, ketika kondisi ginjal semakin memburuk, dapat membahayakan nyawa penderita [2].

Penyakit ginjal kronik (PGK) ialah keadaan yang dialami pada ginjal karena tidak lagi berfungsi secara normal dalam rentan waktu yang cukup lama. Kondisi ini dapat mengakibatkan kerusakan ginjal secara permanen. [3]. Sehingga harus dilakukannya cuci darah (hemodialisis) secara rutin atau bahkan dapat dilakukan tindakan transplantasi ginjal untuk menggantikan ginjal yang rusak dengan ginjal yang baru [2]. Menurut data *World Health Organization* (WHO) PGK masuk dalam 10 besar kategori penyakit yang menyebabkan kematian di dunia [4]. Kementerian Kesehatan Republik Indonesia pada tahun 2013 menjelaskan bahwa sebanyak 499.800 masyarakat Indonesia atau sama dengan 2 dari 1000 masyarakat Indonesia menderita PGK [5]. Karena tingginya jumlah penderita penyakit ginjal kronik (PGK), diperlukan metode klasifikasi melalui data mining untuk memprediksi PGK berdasarkan tanda-tanda gejala yang mengarah pada indikasi penyakit tersebut.

Data mining merupakan proses penemuan korelasi, pola, dan tren baru yang bermanfaat melalui ekstraksi informasi dari repositori data yang besar. Tujuan utama dari data mining adalah untuk dapat menemukan, menggali, dan mengekstraksi informasi yang bernilai penting dari dataset yang ada [6]. Data mining dalam ilmu kedokteran digunakan untuk dapat memprediksi penyakit dan membantu dalam mengambil keputusan di bidang klinis. Salah satu contoh penggunaannya adalah dalam analisis PGK. [3]. Untuk mengelompokkan PGK dapat digunakan algoritma NBC dan K-NN.

Algoritma Naïve Bayes (NBC) adalah algoritma klasifikasi sederhana, efisien, dan kinerja yang baik dalam set data, NBC adalah algoritma yang paling populer [7], [8],serta cocok digunakan pada dataset besar dan dapat mengatasi data yang kosong (*missing value*), serta mampu mengelola atribut yang beragam dan gangguan dalam data [9]. Sedangkan Algoritma K-NN adalah algoritma yang memiliki kinerja komputasi yang tinggi dan dapat digunakan untuk menghitung jarak terpendek antara atribut data. Ini adalah algoritma yang mudah digunakan untuk data yang besar dan memiliki banyak karakteristik yang baik, baik dari segi kinerja dan akurasi [10]. Dengan cara memilih nilai K sampel pelatihan yang paling terdekat dengan contoh uji, selain itu dengan menentukan jarak antara data baru dan semua data yang ada didalam data latih. Metriks jarak yang paling umum digunakan adalah menggunakan *euclidean* [11], [12]

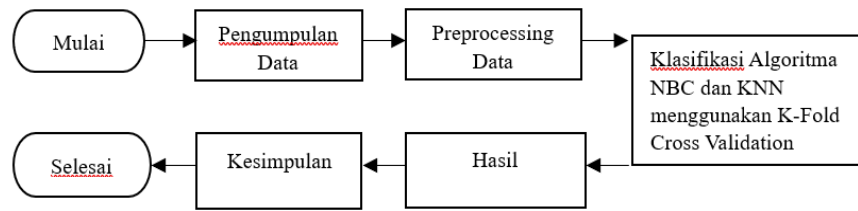
Pada penelitian tahun 2022 yang dilakukan oleh Qurotul A'yuniyah dan rekan-rekannya, Algoritma NBC diterapkan untuk klasifikasi Penyakit Ginjal Kronik (PGK). Dalam penelitian tersebut, data diklasifikasikan menggunakan Algoritma NBC yang diimplementasikan menggunakan *tools* RapidMiner. Hasil penelitian menunjukkan tingkat akurasi sebesar 96,43%, dengan rata-rata *recall* sebesar 93,18%, rata-rata presisi sebesar 93,02%, dan AUC sebesar 93,2%. Dari hasil ini, disimpulkan bahwa penggunaan Algoritma NBC sangat baik dalam mengklasifikasikan data penderita PGK kronik [4]. Kemudian Joseph Teguh Santoso dkk tahun 2021 melakukan penelitian mengenai perbandingan klasifikasi data mining c4.5 dan algoritma naïve bayes dari dataset EDM, dengan dilakukannya pengujian dengan 10-fold cross validation dan uji beda T-Test yang memiliki pengaruh besar terhadap nilai AUC pada metode Naïve Bayes dibandingkan C4.5 dengan rata-rata nilai akurasi naïve bayes sebesar 73,41% dan area *under-curve* sebesar 0,664 [13].

Pada penelitian sebelumnya yang dilakukan oleh Muhammad Naja Maskuri dan rekan-rekannya pada tahun 2022, Algoritma K-Nearest Neighbor (K-NN) diterapkan untuk klasifikasi dan prediksi penyakit stroke. Eksperimen ini menggunakan data latih sebanyak 80 data dan data uji sebanyak 20 data. Hasilnya menunjukkan akurasi sebesar 95%, yang mengindikasikan bahwa algoritma K-NN memberikan prediksi yang akurat dalam 95% kasus. Nilai k yang digunakan dalam perhitungan ini adalah 9, yang berarti algoritma K-NN mempertimbangkan 9 tetangga terdekat untuk membuat prediksi [14]. Lalu, pada tahun 2021, dilakukan sebuah penelitian oleh Umikulsum Indah Lestari dan rekan-rekan mengenai penerapan Metode K-Nearest Neighbor untuk Sistem Pendukung Keputusan dalam Identifikasi Diabetes Melitus. Dalam penelitian ini, implementasi dan pengujian dilakukan dengan menggunakan nilai $k=23$. Hasil pengujian menunjukkan tingkat akurasi sebesar 0,96 atau 96%. Tingkat akurasi tersebut merupakan angka yang cukup tinggi, yang menunjukkan keberhasilan penelitian ini dalam menerapkan metode K-NN sebagai sistem pendukung keputusan dalam mengidentifikasi diabetes melitus secara dini [15].

Berdasarkan hasil pembahasan yang telah dijelaskan, penelitian sebelumnya hanya menggunakan satu jenis algoritma untuk klasifikasi, Pada penelitian ini menggunakan algoritma NBC dan K-NN untuk mengklasifikasikan data penyakit Ginjal Kronik (PGK). Metode ini memanfaatkan kelebihan masing-masing algoritma untuk meningkatkan akurasi dan ketepatan klasifikasi pada data PGK. Diharapkan bahwa penggunaan kedua algoritma ini akan menghasilkan hasil yang lebih baik dalam pengenalan dan klasifikasi PGK. Sehubungan dengan hal tersebut, penelitian ini diberi judul Implementasi Algoritma Naive Bayes Classifier (NBC) dan K-Nearest Neighbor (K-NN) untuk klasifikasi penyakit Ginjal Kronik.

2. METODOLOGI PENELITIAN

Tahap penelitian yang dilakukan melewati beberapa proses, berikut alur penelitian yang akan dilakukan dan dapat dilihat pada Gambar 1.



Gambar 1. Metodologi penelitian

Metodologi penelitian yang digunakan terlihat pada Gambar 1. Penelitian dimulai dengan mengumpulkan data yang diperlukan terlebih dahulu, kemudian tahap Preprocessing data untuk menghapus atribut yang tidak berguna dan record data yang mengalami noise, selanjutnya dilakukan proses klasifikasi menggunakan algoritma NBC dan K-NN dengan Teknik pembagian data latih dan uji menggunakan K-Fold Cross Validation dan langkah terakhir didapatkannya hasil klasifikasi sehingga dapat ditarik kesimpulan dari penelitian yang dilaksanakan.

2.1 Pengumpulan Data

Tahap pengumpulan data yang dibutuhkan bersumber dari situs web *kaggle.com* dengan nama dataset penyakit ginjal kronik yang merupakan dataset tahun 2021, terdiri dari 25 atribut, dan 400 data sampel. Data set ini digunakan sebagai acuan utama untuk dapat dilakukannya analisis dalam penelitian. Kemudian atribut yang digunakan tertera pada Tabel 1.

Tabel 1. Atribut yang digunakan

No	Fitur	Deskripsi
1	Usia	Usia subjek data pada saat pengisian data
2	Blood Pressure	Ukuran tekanan darah dalam tubuh manusia yang digunakan jantung untuk memompa darah keseluruh tubuh
3	Specific Gravity	Perbandingan massa suatu zat dengan massa zat lainnya
4	Albumin	Protein yang membentuk sebagian besar dari plasma darah
5	Sugar	Kadar gula yang ada dalam tubuh
6	Red Blood Cells	Jumlah Kandungan sel darah merah dalam tubuh
7	Pus Cell	leukosit atau sel darah putih untuk melawan infeksi dalam tubuh
8	Pus Cell Clumps	Pembentukan kelompok sel darah putih yang disebabkan infeksi pada ginjal
9	Bacteria	Mikroorganisme bersel dalam tubuh
10	General Dental service	Kesehatan gigi dan mulut
11	Ureum	Zat sisa berasal dari pemecahan asam amino dan protein di dalam hati
12	Serum Creatinine	Sampah dari hasil metabolisme otot yang mengalir dalam sirkulasi darah
13	Sodium	garam atau elektrolit yang mengatur jumlah air dalam tubuh
14	Potassium	Mineral dalam tubuh
15	Hemoglobin	Protein didalam sel darah merah untuk memberi warna pada sel darah
16	Mean Corpuscular Volume	Perhitungan untuk rata-rata ukuran dari sel darah merah
17	White Blood Cell Count	Jumlah keseluruhan sel darah putih dalam tubuh
18	Red Blood Cell Count	Jumlah keseluruhan sel darah merah dalam tubuh
19	Hypertension	Kondisi tekanan darah pada dinding arteri yang tinggi
20	Diabetes Mellitus	Jumlah glukosa atau kadar gula darah yang tinggi
21	Coronary Artery Disease	Penyempitan dan sumbatan pembuluh darah pada jantung
22	Appetite	Nafsu makan subjek data
23	Pedal Edema	Penumpukan cairan di bagian betis dan kaki bagian bawah
24	Anemia	Tubuh kekurangan sel darah merah yang sehat
25	Classification	Identifikasi data subjek

2.2 Preprocessing Data

Setelah data terkumpul, langkah berikutnya adalah melakukan *preprocessing* data. Data yang memiliki missing value, duplikat data dan data yang mengalami noise akan dibersihkan terlebih dahulu [16]. Kemudian setelah dilakukan *data cleaning* jumlah data yang digunakan adalah 156 sampel dari 400 data yang memiliki missing value, dengan jumlah pembagian data disetiap kelasnya yaitu NOT CKD sebanyak 114 dan CKD sebanyak 42 data, kemudian data yang sudah di *preprocessing* terlebih dahulu akan di proses menggunakan algoritma NBC dan K-NN.

2.3 Klasifikasi

Klasifikasi adalah proses menempatkan dokumen atau teks secara otomatis dalam kategori berdasarkan isi teksnya. Klasifikasi membantu mengorganisir dan mengelompokkan jumlah besar dokumen teks yang ada. Dengan menggunakan algoritma klasifikasi, cara ini dapat dilakukan dengan mudah dan cepat. Dasar dari algoritma klasifikasi adalah seleksi fitur, yang bertujuan untuk menemukan kombinasi atribut yang mungkin dalam data. Tujuannya adalah untuk menemukan subset terbaik dari fitur-fitur yang ada untuk digunakan dalam prediksi. Dengan pemilihan fitur ini dapat membantu meningkatkan performa klasifikasi teks, baik dari sisi kecepatan maupun efektivitas dalam pembelajaran [17].

2.4 Naïve Bayes Classifier (NBC)

Naive Bayes Classifier (NBC) adalah metode klasifikasi yang bersifat terawasi yang didasarkan pada teorema Bayes yang berasal dari probabilitas bersyarat [18], Pengklasifikasi Naive Bayes sangat digunakan dan telah berkembang ke dalam lima kategori tingkat lanjut. Ini mencakup pengembangan struktur untuk mempertimbangkan ketergantungan atribut, perluasan struktur, seleksi contoh, pemberian bobot pada contoh, pemilihan atribut, dan pemberian bobot pada atribut untuk meningkatkan daya diskriminasi [19]. Metode yang digunakan adalah menghitung himpunan kemungkinan dengan menjumlahkan kombinasi dan frekuensi dari nilai record yang digunakan. NBC memiliki beberapa varian klasik yaitu *Multinomial*, *Bernoulli*, dan *Gaussian*. Selain itu juga dikembangkan varian klasik yang salah satunya ialah komplemen naive bayes yaitu dengan pengembangan dan adaptasi multinomial. Dalam penerapan algoritma NBC rumus yang diterapkan terlihat pada persamaan 1 [20]:

$$P(A|B) = P(A)/P(B) \times P(B|A) \quad (1)$$

Keterangan:

- A : Data sampel yang tidak diketahui label kelasnya.
- B : Kelas hasil pengelompokkan
- $P(A|B)$: Peluang kemungkinan terjadinya A jika B diketahui
- $P(B|A)$: Peluang kemungkinan terjadinya B jika A diketahui
- $P(A)$: Peluang kemungkinan terjadinya A
- $P(B)$: Peluang kemungkinan terjadinya B

2.5 K-Nearest Neighbor (K-NN)

K-Nearest Neighbor (K-NN) adalah salah satu metode pengelompokkan dalam data mining, dimana K-NN mengelompokkan sekumpulan data berdasarkan data pembelajaran yang telah diklasifikasikan atau diberi label. K-NN masuk kedalam kelompok *supervised learning* yaitu hasil *query* yang baru dikelompokkan instans berdasarkan mayoritas kedekatan dengan kategori yang ada di K-NN. Berikut Langkah yang digunakan untuk melakukan klasifikasi dengan algoritma K-NN [21]:

1. Menentukan parameter k, dengan minimal nilai K adalah 1 dan jumlah maksimalnya dari data latih yang ada.
2. Melakukan Perhitungan jarak antara data latih dan data uji. Perhitungan yang paling sering digunakan untuk menghitung jarak pada perhitungan algoritma K-NN adalah Euclidean. Untuk menghitung Jarak dengan rumus pada persamaan 2.

$$\text{Euclidean} = \sqrt{\sum_i^n = 1 (p_i - q_i)^2} \quad (2)$$

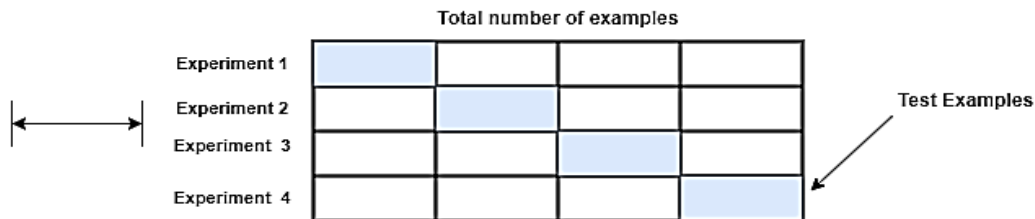
Keterangan:

- p_i : data pelatihan
 - q_i : pengujian data
 - i : variabel data
 - n : ukuran data
3. Selanjutnya mengurutkan jarak yang telah terbentuk dimulai dari yang terbesar ke yang terkecil
 4. Tentukanlah jarak yang paling dekat dengan barisan K
 5. Cocokkan dengan kelas yang simbang
 6. Menemukan besaran kelas dari tetangga terdekat dan menugaskan pada kelas yang tersebut sebagai data kelas yang akan dievaluasi dan diklasifikasi.

Model regresi K-NN bekerja dengan mengukur jarak antara data observasi baru dan seluruh data observasi yang terdapat dalam kumpulan data pelatihan. Biasanya, metrik jarak yang sering digunakan adalah metrik jarak *euclidean*. Perhitungan metrik ini mengikuti rumus yang telah ditunjukkan pada rumus 2 [22].

2.6 K-Fold Cross Validation

K-Fold Cross Validation merupakan suatu cara untuk mengevaluasi kinerja model pelatihan yang dibuat metode ini, untuk dapat membagi data latih dan data uji sebanyak k bagian data yang merupakan metode statistik silang yang dapat digunakan untuk memvalidasi kinerja model atau algoritma klasifikasi serta berfungsi agar tidak terjadi tumpang tindih pada data uji. Pada metode ini, dataset yang akan digunakan dibagi terlebih dahulu menjadi K partisi secara acak. Setiap partisi akan digunakan sebagai data *testing* satu kali, sementara partisi yang tersisa digunakan sebagai data *training*. Proses ini diulang sebanyak K kali, di mana setiap kali percobaan menggunakan partisi yang berbeda sebagai data *testing* [23].



Gambar 2. Pengilustrasian K-Fold Cross Validation

Sebuah Pengilustrasian yang terdapat pada gambar 2 dapat dijelaskan bahwa validasi silang 4 kali lipat digunakan dalam percobaan. Rumus tersebut digunakan untuk menghitung nilai akurasi yang dihasilkan oleh K-Fold Cross Validation.

2.7 Supervised dan Unsupervised Learning

Supervised learning adalah teknik dalam *machine learning* dimana model atau mesin belajar dari dataset yang suda diberi label atau kelas yang sesuai. Sedangkan *unsupervised learning* adalah teknik dimana model belajar dari dataset yang tidak memiliki label atau kelas yang ditentukan sebelumnya [24]. Kelebihan dari *supervised learning* adalah memverifikasi kelas data berdasarkan sampel pelatihan dan akurasi klasifikasi. Sedangkan pada *unsupervised learning* kelebihanannya adalah kesalahan operasi kelas yang diminimalkan dan unik diperlakukan sebagai entitas terpisah [19].

3. HASIL DAN PEMBAHASAN

Pada tahap ini dilakukan analisis dan pengujian klasifikasi pada PGK menggunakan algoritma NBC dan K-NN, dimana setiap algoritma akan di uji menaunggunakan K-Fold Cross Validation. berikut proses implementasi algoritma NBC dan K-NN.

3.1 Implementasi Algoritma Naïve Bayes Classifier (NBC)

Dalam penelitian ini, sebanyak 156 data yang telah dibersihkan (*data cleaning*) diklasifikasikan menggunakan algoritma NBC dan percobaan menggunakan K-Fold Cross Validation dengan menggunakan operator yang tersedia dalam *tools* RapidMiner. Kemudian untuk mendapatkan hasil klasifikasi dari algoritma NBC, data akan diproses dan diuji kedalam K-Fold Cross Validation, serta melakukan evaluasi untuk mendapatkan nilai akurasi, presisi, recall dan AUC. Evaluasi ini bertujuan untuk menghitung dan menguji performa penerapan algoritma NBC dalam klasifikasi penyakit ginjal kronik, dan menentukan model terbaik. Hasil pengujian menggunakan metode 10-Fold Cross Validation pada algoritma NBC dapat dilihat pada tabel 2, 3 dan 4.

Tabel 2. Hasil Akurasi Algoritma NBC

Accuracy:94.25%+/-4.63%(micro average:94.23%)			
	True yes	True no	Class precision
Pred.yes	32	7	82.05%
Pred.no	2	115	98.29%
Class recall	94.12%	94.26%	

Tabel 3. Hasil Presisi Algoritma NBC

Precision:98.23%+/-5.60%(micro average:94.26%) (positive class:no)			
	True yes	True no	Class precision
Pred.yes	32	7	82.05%
Pred.no	2	115	98.29%
Class recall	94.12%	94.26%	

Tabel 4. Hasil Recall Algoritma NBC

Precision:98.23% +/-5.60% (micro average:94.26%) (positive class:no)			
	True yes	True no	Class precision
Pred.yes	32	7	82.05%
Pred.no	2	115	98.29%
Class recall	94.12%	94.26%	

Berdasarkan hasil pengolahan data menggunakan 10-Fold Cross Validation pada algoritma NBC, yang direpresentasikan dalam Tabel 2, 3, dan 4, didapatkan rata-rata nilai akurasi sebesar 94,25% dengan standar deviasi sebesar 4,63%. Sementara itu untuk nilai presisi yang dihasilkan yaitu sebesar 98,40%, dengan standar deviasi sebesar 3,38%. Dan nilai recall sebesar 94,23% dengan standar deviasi sebesar 5.60%.

Untuk hasil AUC dari klasifikasi menggunakan algoritma NBC dengan K-Fold Cross Validation didapatkan hasil nilai sebesar 0,961. Dimana nilai tersebut masuk kedalam kategori baik. Hasil AUC algoritma NBC dapat dilihat pada gambar 3.



Gambar 3. Hasil AUC Algoritma NBC

3.2 Implementasi Algoritma K-Nearest Neighbor (K-NN)

Percobaan penelitian selanjutnya dilakukan dengan mengimplementasikan algoritma K-NN dan pengujian menggunakan K-Fold Cross Validation menggunakan tools RapidMiner, dengan 156 data yang telah dibersihkan (*data cleaning*). Selanjutnya untuk mendapatkan hasil klasifikasi dari algoritma K-NN, data PGK akan di proses dan diuji kedalam K-fold Cross Validation. Proses ini memiliki tujuan untuk menemukan nilai akurasi, presisi dan recall dari algoritma K-NN. Data tersebut diuji menggunakan K-Fold Cross Validation sebanyak 10 kali percobaan. Percobaan tersebut dilakukan dengan tujuan untuk menghitung performa dari algoritma K-NN dalam mengklasifikasikan data PGK. Hasil pengujian menggunakan 10-Fold Cross Validation pada algoritma K-NN dapat dilihat pada tabel 5, 6, dan 7.

Tabel 5. Hasil Akurasi Algoritma K-NN

Accuracy:77.79% +/-9.60% (micro average:77.56%)			
	True yes	True no	Class precision
Pred.yes	5	6	45.45%
Pred.no	29	116	80.00%
Class recall	14.71%	95.08%	

Tabel 6. Hasil Presisi Algoritma K-NN

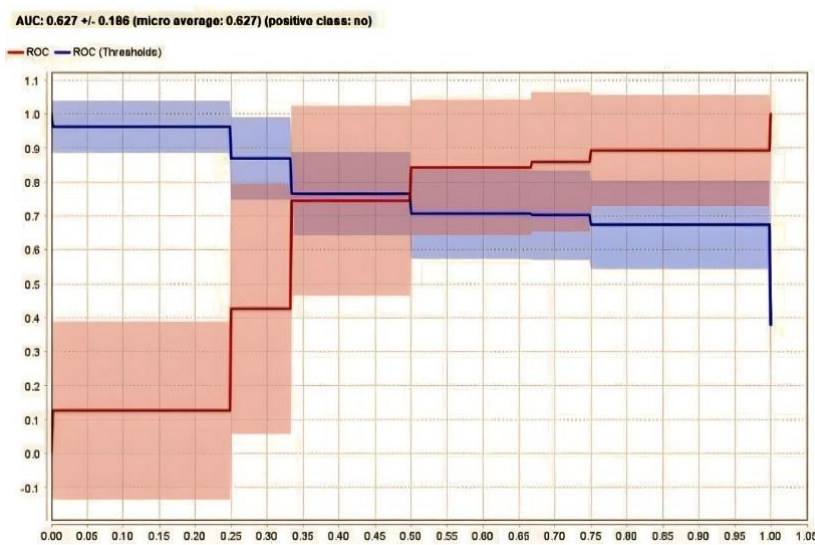
Accuracy:80.20% +/-6.98% (micro average:80.00%) (positive class:no)			
	True yes	True no	Class precision
Pred.yes	5	6	45.45%
Pred.no	29	116	80.00%
Class recall	14.71%	95.08%	

Tabel 7. Hasil Recall Algoritma K-NN

Accuracy:95.06%+/-5.79%(micro average:95.08%) (positive class:no)			
	True yes	True no	Class precision
Pred.yes	5	6	45.45%
Pred.no	29	116	80.00%
Class recall	14.71%	95.08%	

Berdasarkan hasil pengolahan data menggunakan 10-Fold Cross Validation pada tabel 5, 6, dan 7. Diperoleh hasil klasifikasi dengan nilai akurasi sebesar 77,79% dengan standar deviasi sebesar 9,60%. Kemudian nilai presisi yang didapat sebesar 80,20% dengan standar deviasi sebesar 6,98%. Dan nilai *recall* sebesar 95,06% dengan standar deviasi sebesar 5,79%.

Kemudian untuk hasil AUC yang didapatkan dari pengklasifikasian algoritma K-NN dengan K-Fold Cross Validation yaitu sebesar 0,627. Hasil AUC dari algoritma K-NN dapat dilihat pada gambar 4.

**Gambar 4.** Hasil AUC Algoritma K-NN

3.3 Analisis Klasifikasi

Tabel 8. Perbandingan Klasifikasi Algoritma NBC dan K-NN

Algoritma	Akurasi	Presisi	Recall	AUC
Naive Bayes Classifier (NBC)	94,25%	98,40%	94,23%	0,961
K-Nearest Neighbor (K-NN)	77,79%	80,20%	95,06%	0,627

Hasil pengujian model klasifikasi menunjukkan tingkat akurasi dari algoritma NBC sebesar 94,25%, sedangkan tingkat akurasi dari algoritma K-NN sebesar 77,79%. Untuk tingkat presisi, algoritma NBC mendapatkan nilai sebesar 98,40%, sedangkan algoritma K-NN sebesar 80,20%. Kemudian untuk nilai recall dari algoritma NBC sebesar 94,23% dan algoritma K-NN sebesar 95,06%. Untuk hasil AUC, algoritma NBC mendapat nilai sebesar 0,961, sedangkan algoritma K-NN sebesar 0,627. Dari hasil ini dapat disimpulkan bahwa algoritma NBC memiliki performa akurasi, presisi dan AUC lebih baik dari algoritma K-NN dalam mengklasifikasi data penyakit ginjal kronik. Sedangkan untuk performa recall, algoritma K-NN lebih baik dari algoritma NBC.

4. KESIMPULAN

Penelitian ini dilakukan menggunakan *tools* RapidMiner terhadap dataset Penyakit Ginjal Kronik yang di peroleh dari web kaggle.com yang dilakukan pengujian menggunakan algoritma NBC dengan K-Fold Cross Validation lalu menghasilkan nilai akurasi sebesar 94,25%, presisi sebesar 98,40%, *recall* sebesar 94,23% dan AUC sebesar 0,961, selanjutnya juga dilakukan penelitian menggunakan algoritma K-NN dengan K-Fold Cross Validation maka didapatkan hasil akurasi sebesar 77,79%, presisi sebesar 80,20%, recall sebesar 95,06% dan AUC sebesar 0,627, maka dapat disimpulkan bahwa hasil klasifikasi menggunakan algoritma NBC lebih baik dibandingkan menggunakan algoritma K-NN, sehingga dapat dilakukan prediksi terhadap pasien mana saja yang terkena PGK dan NOT PGK yang nantinya dapat diangkat sebagai acuan untuk penelitian selanjutnya.

REFERENSI

- [1] R. G. Langham *et al.*, “KIDNEY HEALTH FOR ALL: BRIDGING THE GAP IN KIDNEY HEALTH EDUCATION AND LITERACY,” *Brunei International Medical Journal*, vol. 2022, no. 18, pp. 51–62, 2022, doi: 10.1016/j.nefro.2022.02.003.
- [2] I. Yulianti, R. Amegia Saputra, M. Sukrisno Mardiyanto, and A. Rahmawati, “Optimasi Akurasi Algoritma C4.5 Berbasis Particle Swarm Optimization dengan Teknik Bagging pada Prediksi Penyakit Ginjal Kronis Optimization of C4.5 Algorithm Based On Particle Swarm Optimization with Bagging Technique on Prediction of Chronic Kidney Disease,” 2020. [Online]. Available: <https://archive.ics.uci.edu/ml/>
- [3] A. K. Hermawan and A. Nugroho, “Analisa Data Mining Untuk Prediksi Penyakit Ginjal Kronik Dengan Algoritma Regresi Linier,” *Bulletin of Information Technology (BIT)*, vol. 4, no. 1, pp. 37–48, 2023, doi: 10.47065/bit.v3i1.
- [4] Q. A’yuniyah *et al.*, “Implementasi Algoritma Naïve Bayes Classifier (NBC) untuk Klasifikasi Penyakit Ginjal Kronik,” *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 1, p. 72, Sep. 2022, doi: 10.30865/json.v4i1.4781.
- [5] “IMPLEMENTASI ALGORITMA C4.5 DAN K-MEANS PADA DIAGNOSIS PENYAKIT GINJAL KRONIS.”
- [6] D. A. Manalu and G. Gunadi, “IMPLEMENTASI METODE DATA MINING K-MEANS CLUSTERING TERHADAP DATA PEMBAYARAN TRANSAKSI MENGGUNAKAN BAHASA PEMROGRAMAN PYTHON PADA CV DIGITAL DIMENSI,” *Infotech: Journal of Technology Information*, vol. 8, no. 1, pp. 43–54, Jun. 2022, doi: 10.37365/jti.v8i1.131.
- [7] A. W. Syaputri, E. Irwandi, and M. Mustakim, “Naïve Bayes Algorithm for Classification of Student Major’s Specialization,” *Journal of Intelligent Computing & Health Informatics*, vol. 1, no. 1, p. 17, 2020, doi: 10.26714/jichi.v1i1.5570.
- [8] Mustakim *et al.*, “Data Sharing Technique Modeling for Naive Bayes Classifier for Eligibility Classification of Recipient Students in the Smart Indonesia Program,” *Journal of Physics: Conference Series*, vol. 1424, no. 1, 2019, doi: 10.1088/1742-6596/1424/1/012009.
- [9] T. Arifin and D. Ariesta, “PREDIKSI PENYAKIT GINJAL KRONIS MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER BERBASIS PARTICLE SWARM OPTIMIZATION,” *Jurnal Tekno Insentif*, vol. 13, no. 1, pp. 26–30, Apr. 2019, doi: 10.36787/jti.v13i1.97.
- [10] Mustakim, R. Hastarimasuci, P. Papilo, Zarkasih, Zaitun, and A. Nazir, “Variable Selection to Determine Majors of Student using K-Nearest Neighbor and Naïve Bayes Classifier Algorithm,” *Journal of Physics: Conference Series*, vol. 1363, no. 1, 2019, doi: 10.1088/1742-6596/1363/1/012057.
- [11] D. S. Jodas, L. A. Passos, A. Adeel, and J. P. Papa, “PL-K-NN: A Python-based implementation of a parameterless k-Nearest Neighbors classifier [Formula presented],” *Software Impacts*, vol. 15, Mar. 2023, doi: 10.1016/j.simpa.2022.100459.
- [12] A. Sumayli, “Development of advanced machine learning models for optimization of methyl ester biofuel production from papaya oil: Gaussian process regression (GPR), multilayer perceptron (MLP), and K-nearest neighbor (K-NN) regression models,” *Arabian Journal of Chemistry*, vol. 16, no. 7, Jul. 2023, doi: 10.1016/j.arabjc.2023.104833.
- [13] J. T. Santoso, N. L. W. S. R. Ginantra, M. Arifin, R. Riinawati, D. Sudrajat, and R. Rahim, “Comparison of Classification Data Mining C4.5 and Naïve Bayes Algorithms of EDM Dataset,” *TEM Journal*, vol. 10, no. 4, pp. 1738–1744, 2021, doi: 10.18421/TEM104-34.
- [14] M. N. Maskuri, K. Sukerti, and R. M. Herdian Bhakti, “Penerapan Algoritma K-Nearest Neighbor (K-NN) untuk Memprediksi Penyakit Stroke Stroke Disease Predict Using K-NN Algorithm,” *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, vol. 4, no. 1.
- [15] U. I. Lestari, A. Y. Nadhiroh, and C. Novia, “Penerapan Metode K-Nearest Neighbor Untuk Sistem Pendukung Keputusan Identifikasi Penyakit Diabetes Melitus,” vol. 8, no. 4, 2021, [Online]. Available: <http://jurnal.mdp.ac.id>
- [16] M. N. Sidqi, D. P. Rini, and S. Samsuryadi, “Optimization of Deep Neural Networks with Particle Swarm Optimization Algorithm for Liver Disease Classification,” *Computer Engineering and Applications*, vol. 12, no. 1, 2023.
- [17] J. Elektronik and I. Komputer Udayana, “Klasifikasi Jurnal menggunakan Metode K-NN dengan Mengimplementasikan Perbandingan Seleksi Fitur,” [Online]. Available: <https://scholar.google.com>
- [18] M. Vishwakarma and N. Kesswani, “A new two-phase intrusion detection system with Naïve Bayes machine learning for data classification and elliptic envelop method for anomaly detection,” *Decision Analytics Journal*, vol. 7, Jun. 2023, doi: 10.1016/j.dajour.2023.100233.
- [19] J. Dai, X. hong Hao, Z. Li, P. Li, and X. peng Yan, “Adaptive target and jamming recognition for the pulse doppler radar fuze based on a time-frequency joint feature and an online-updated naive bayesian classifier with minimal risk,” *Defence Technology*, vol. 18, no. 3, pp. 457–466, Mar. 2022, doi: 10.1016/j.dt.2021.02.008.

-
- [20] N. Marito Putry and B. Nurina Sari, “KOMPARASI ALGORITMA K-NN DAN NAÏVE BAYES UNTUK KLASIFIKASI DIAGNOSIS PENYAKIT DIABETES MELITUS,” *Jurnal Sains dan Manajemen*, vol. 10, no. 1, 2022.
- [21] J. Elektronik, I. K. Udayana, A. Made, S. I. Dewi, I. Bagus, and G. Dwidasmara, “Implementation Of The K-Nearest Neighbor (K-NN) Algorithm For Classification Of Obesity Levels”, [Online]. Available: www.kaggle.com
- [22] A. Sumayli, “Development of advanced machine learning models for optimization of methyl ester biofuel production from papaya oil: Gaussian process regression (GPR), multilayer perceptron (MLP), and K-nearest neighbor (K-NN) regression models,” *Arabian Journal of Chemistry*, vol. 16, no. 7, Jul. 2023, doi: 10.1016/j.arabjc.2023.104833.
- [23] N. Puspitasari, R. Rosmasari, F. W. Pratama, and H. Sulastri, “Quality Classification of Palm Oil Varieties Using Naive Bayes Classifier,” *Digital Zone: Jurnal Teknologi Informasi dan Komunikasi*, vol. 13, no. 1, pp. 11–23, May 2022, doi: 10.31849/digitalzone.v13i1.9773.
- [24] E. Retnoningsih and R. Pramudita, “Mengenal Machine Learning Dengan Teknik Supervised dan Unsupervised Learning Menggunakan Python,” *BINA INSANI ICT JOURNAL*, vol. 7, no. 2, pp. 156–165, 2020, [Online]. Available: <https://www.python.org/>
- [25] R. Septiani *et al.*, “Perbandingan Metode Supervised Classification dan Unsupervised Classification terhadap Penutup Lahan di Kabupaten Buleleng,” *Address : Jl. Sesetan No.196 Sesetan*, vol. 16, no. 2, p. 80223, 2019, doi: 10.15294/jg.v16i2.19777.