

Analisis Perbandingan Kinerja Naive Bayes Dan C4.5 Untuk Deteksi Penyakit Ginjal

Rayma Yemima Gultom¹, Kurniabudi², Effiyaldi³

Fakultas Ilmu Komputer, Program Studi Magister Sistem Informasi, Universitas Dinamika Bangsa, Jambi, Indonesia

Email: ¹raymayemima@gmail.com, ²kurniabudi@unama.ac.id, ³effiyaldi@unama.ac.id

Email Penulis Korespondensi: raymayemima@gmail.com

Submitted :
24 September
2025

Revision :
11 Maret 2026

Accepted:
12 Maret 2026

Published:
31 Maret 2026

Abstrak— Penyakit Ginjal Kronis (PGK) adalah masalah kesehatan di seluruh dunia dengan peningkatan jumlah penderita yang stabil, termasuk di Indonesia. Identifikasi tepat waktu diabetes melitus gestasional (DMG) sangat penting untuk mengurangi risiko komplikasi dan kematian, namun penilaian manual seringkali membutuhkan waktu dan sumber daya yang cukup besar. Perkembangan teknologi *machine learning* memungkinkan proses diagnosis lebih cepat dan akurat. Penelitian ini bertujuan membandingkan kinerja algoritma *Naive Bayes* dan C4.5 dalam mendeteksi penyakit ginjal menggunakan dataset *Chronic Kidney Disease* (CKD) yang diperoleh dari platform Kaggle. Dataset terdiri dari 400 data pasien dengan 34 atribut medis, seperti tekanan darah, kadar kreatinin, albumin urin, dan parameter laboratorium lainnya. Metode penelitian meliputi pembersihan data, seleksi atribut, transformasi data, pembagian data menjadi 80% data latih dan 20% data uji, serta penerapan kedua algoritma menggunakan perangkat lunak WEKA. Evaluasi kinerja dilakukan dengan matrik akurasi, *precision*, *recall*, F1-score, dan AUC-ROC. Hasil pengujian menunjukkan bahwa pada pembagian data 80:20, algoritma *Naive Bayes* memperoleh akurasi 83,13%, *precision* 84,05%, *recall* 97,68%, F1-score 90,36%, dan AUC-ROC 59,50%. Sementara itu, algoritma C4.5 (J48) mencapai akurasi lebih tinggi yaitu 87,50%, *precision* 85,93%, *recall* 98,45%, F1-score 91,74%, dan AUC-ROC 63,20%.

Kata Kunci: penyakit ginjal kronis, *Naive Bayes*, C4.5, *machine learning*, WEKA, klasifikasi

Abstract— *Chronic kidney disease (CKD) is a global health problem with an increasing number of sufferers, including in Indonesia. Early detection of CKD is crucial to reduce the risk of complications and mortality, but manual examination often requires considerable time and resources. Advances in machine learning technology enable faster and more accurate diagnosis processes. This study aims to compare the performance of the Naive Bayes and C4.5 algorithms in detecting kidney disease using the Chronic Kidney Disease (CKD) dataset obtained from the Kaggle platform. The dataset consists of 400 patient records with 34 medical attributes, such as blood pressure, creatinine level, urine albumin, and other laboratory parameters. The research methods include data cleaning, attribute selection, data transformation, splitting the data into 80% training and 20% testing sets, and applying both algorithms using the WEKA software. Performance evaluation was conducted using accuracy, precision, recall, F1-score, and AUC-ROC metrics. The test results showed that with an 80:20 data split, the Naive Bayes algorithm achieved an accuracy of 83.13%, precision of 84.05%, recall of 97.68%, F1-score of 90.36%, and an AUC-ROC of 59.50%. Meanwhile, the C4.5 (J48) algorithm achieved a higher accuracy of 87.50%, precision of 85.93%, recall of 98.45%, F1-score of 91.74%, and an AUC-ROC of 63.20%.*

Keywords: *chronic kidney disease, Naive Bayes, C4.5, machine learning, WEKA, classification*

1. PENDAHULUAN

Penyakit Ginjal Kronis (PGK) telah menjadi masalah kesehatan yang signifikan dan meningkat di Indonesia dan secara global. *World Health Organization* (WHO) melaporkan bahwa sekitar 850 juta orang saat ini menderita berbagai gangguan ginjal. Di Indonesia, data tahun 2023 dari Kementerian Kesehatan menunjukkan peningkatan prevalensi penyakit ginjal kronis, yang tetap menjadi penyebab utama kematian. [1]. Hal yang mengkhawatirkan adalah banyak orang tidak menyadari bahwa mereka mengalami penyakit ini karena gejalanya seringkali tidak terlihat pada tahap awal, dan biasanya baru diketahui saat kondisi sudah parah. Untuk mendiagnosis penyakit ginjal kronis, dokter harus memeriksa berbagai faktor seperti tekanan darah pasien, kadar glukosa darah, protein dalam darah (albumin), fungsi ginjal (kreatinin), serta berbagai tes laboratorium lainnya. Dengan kemajuan teknologi, saat ini kita dapat memanfaatkan teknik data mining dan *machine learning* untuk membantu dokter dalam mendiagnosis penyakit ini dengan lebih cepat dan akurat [2].

Hasil regresi logistik berganda menunjukkan bahwa peluang terkena batu ginjal 1,17 kali lebih tinggi pada pasien diabetes, 1,43 kali lebih tinggi pada pasien hipertensi, 2,21 kali lebih tinggi pada pasien penyakit hati berlemak, dan 1,35 kali lebih tinggi pada pasien yang kelebihan berat badan. Tekanan darah, kadar kreatinin, dan albumin urin adalah faktor lain yang terkait dengan batu ginjal. Data Riset Kesehatan Dasar (Riskesdas) tahun 2013 menunjukkan bahwa prevalensi penduduk Indonesia yang menderita Gagal Ginjal adalah 0,2% atau 2 per 1000 penduduk, dan prevalensi Batu Ginjal adalah 0,6% atau 6 per 1000 penduduk, atau 1.499.400 orang Indonesia menderita Batu Ginjal (Riskesdas, 2013). Oleh karena itu, diperlukan pengetahuan mendalam tentang faktor-

faktor yang memicu pembentukan batu ginjal. Dengan mengidentifikasi faktor penyebab batu ginjal, langkah-langkah pencegahan dapat diambil dengan mudah.

Dengan kemajuan teknologi di bidang medis, *machine learning* dapat digunakan untuk memprediksi penyakit ginjal. Algoritma *machine learning* bersifat konstruktif dalam membuat prediksi yang akurat dan memberikan analisis yang akurat. Menurut [3], implementasi *machine learning* dalam bidang kesehatan terbukti mampu meningkatkan efisiensi diagnosis serta membantu pengambilan keputusan klinis berbasis data. Selain itu, studi oleh [4] menekankan bahwa model prediktif berbasis *machine learning* dapat digunakan untuk mendeteksi potensi penyakit sejak dini, sehingga dapat menekan angka kematian dan biaya pengobatan. Hal ini semakin memperkuat urgensi untuk mengembangkan sistem diagnosis cerdas, terutama untuk penyakit-penyakit kronis seperti penyakit ginjal.

Berdasarkan hasil studi bahwa algoritma *machine learning* sudah banyak dilakukan oleh penelitian sebelumnya. Salah satunya dilakukan oleh [5], penelitian tersebut membahas mengenai klasifikasi penyakit *Pneumonia* dengan menggunakan metode *K-Nearest Neighbor* (KNN) dengan hasil penelitian yaitu memperoleh nilai akurasi 90% dengan menggunakan nilai $K=3$. Selanjutnya penelitian yang dilakukan oleh [6] yang menggunakan algoritma *Naïve Bayes* dalam klasifikasi penyakit diabetes mellitus, dengan hasil klasifikasi menunjukkan akurasi sebesar 96.20%. Penelitian yang dilakukan oleh [7] yang membahas mengenai klasifikasi penderita penyakit diabetes menggunakan algoritma *Decision Tree C4.5*. Dengan hasil penelitian yaitu menghasilkan akurasi yang cukup besar yaitu 97,12 %. Kemudian penelitian yang dilakukan oleh [8] yang mencapai akurasi 85,2% untuk deteksi dini diabetes.

Berdasarkan penelitian yang telah dilakukan dengan menggunakan metode *machine learning* menunjukkan bahwa, *machine learning* dapat digunakan untuk mendeteksi suatu penyakit yang dapat membantu memudahkan dalam pencegahan. Dari penelitian yang telah dilakukan didapatkan hasil akurasi tertinggi untuk metode *Naïve Bayes* hasil akurasi sebesar 96% dan metode *Decision Tree* sebesar 97%, namun algoritma *Naïve Bayes* dan algoritma *C4.5* memiliki perbedaan karakteristik dan pendekatan. Dimana *Naïve Bayes* merupakan algoritma berbasis probabilistik yang berkerja secara efisien pada data dengan asumsi independensi antar atribut, sedangkan untuk algoritma *C4.5* membangun model pohon keputusan berdasarkan nilai *information gain*. Oleh karena itu, perlu dilakukan komparasi antara kedua metode tersebut. Tujuannya adalah untuk mengetahui metode yang paling optimal dalam hal akurasi, presisi dan efisiensi, serta memberikan kontribusi terhadap pengembangan sistem pendukung keputusan medis berbasis data yang lebih akurat.

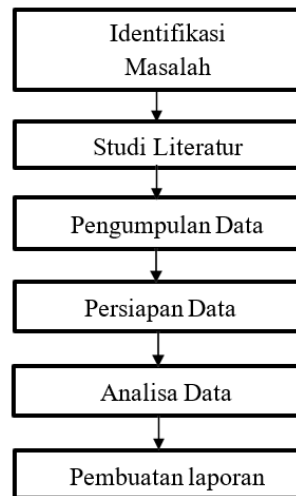
Beberapa studi komparatif telah dilakukan untuk membandingkan kedua algoritma ini. Penelitian oleh [9] penelitian yang menggunakan data set dengan akurasi 72,50% untuk algoritma *C4.5* sedangkan untuk *Naïve Bayes* mendapatkan nilai akurasi sebesar 68,57%. Namun hasil berbeda ditemukan oleh [10] yang menunjukkan bahwa nilai akurasi *C4.5* sebesar 74% lebih besar dibandingkan model *Naïve Bayes* dengan akurasi 67%. Sedangkan penelitian yang dilakukan oleh [11] dengan hasil nilai akurasi *C4.5* sebesar 95% dan *Naïve Bayes* sebesar 87%. Kemudian penelitian yang dilakukan oleh [12] yang membahas mengenai prediksi penyakit jantung menghasilkan nilai akurasi *C4.5* sebesar 83% dan untuk *Naïve Bayes* sebesar 85%.

Dari perbandingan yang telah dilakukan bahwa peneliti hanya membandingkan nilai akurasinya pada kedua metode. Menurut [13] dalam tahapan evaluasi dan penyempurnaan model *machine learning* perlu dilakukan pengujian model, pengujian model tersebut meliputi pengujian akurasi, *Precision*, *recall*, *F1-Score*, *MSE* atau *AUC-ROC*. Sehingga pada penelitian ini akan melakukan komparasi metode *Naïve Bayes* dan *C4.5* untuk melihat perbandingan tingkat akurasi, *Precision*, *recall*, *F1-Score*, *MSE* atau *AUC-ROC*.

Pada dunia kesehatan kebutuhan nyata akan sistem deteksi dini penyakit ginjal yang cepat, akurat, dan berbasis teknologi. Dengan meningkatnya jumlah penderita penyakit ginjal dan dampaknya yang signifikan terhadap kualitas hidup serta beban pembiayaan kesehatan, maka dibutuhkan pendekatan cerdas dan inovatif berbasis data untuk membantu tenaga medis dalam proses diagnosis. Komparasi algoritma *Naïve Bayes* dan *C4.5* menjadi penting karena masing-masing memiliki keunggulan berbeda yang dapat dimaksimalkan. Penelitian ini tidak hanya memiliki kontribusi akademis dalam memperkaya literatur terkait klasifikasi penyakit, tetapi juga berpotensi memberikan solusi praktis berupa sistem pendukung keputusan klinis yang efisien, terjangkau, dan dapat diimplementasikan di fasilitas layanan kesehatan, khususnya di daerah dengan keterbatasan sumber daya medis dan teknologi.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian



Gambar 1. Tahapan Penelitian

Berdasarkan alur penelitian pada gambar diatas, maka dapat diuraikan menjadi pembahasan pada masing-masing tahapan. alur penelitian yang digunakan sebagai berikut:

- a. Identifikasi Masalah
Identifikasi masalah [4] adalah tahap awal dalam proses penelitian yang bertujuan untuk menemukan, merumuskan, dan menjelaskan permasalahan utama yang akan diteliti secara sistematis dan logis. Pada tahap ini, penulis melakukan identifikasi dan merumuskan permasalahan penelitian, dengan meninjau topik umum dan isu nyata dan juga melakukan analisis dataset kaggle serta meninjau penelitian sebelumnya.
- b. Studi Literatur
Studi literatur dilakukan [3] sebagai tahapan persiapan untuk mencari dasar-dasar teori yang digunakan dalam penelitian, selain itu juga dilakukan pengumpulan data tentang penyakit ginjal kronis. Pada tahap ini dilakukan pencarian landasan-landasan teori yang relevan melalui berbagai artikel, buku, jurnal dan juga internet untuk yang bertujuan melengkapi konsep dan teori dan keilmuan yang baik serta menambah wawasan dan pengetahuan peneliti.
- c. Pengumpulan Data
Pada tahap ini dilakukan [9] pengumpulan data yang akan diproses dengan mencari data yang tersedia, memperoleh data tambahan yang dibutuhkan, mengintegrasikan semua data kedalam dataset, termasuk variabel yang diperlukan dalam proses. Yang kemudian menghasilkan data yang sesuai dengan penelitian. Dalam penelitian ini penulis menggunakan dataset public untuk mendeteksi penyakit ginjal yang diunduh dari laman kaggle.
- d. Persiapan Data
Pada proses persiapan data dilakukan melalui beberapa tahapan, yang akan dijelaskan di bawah ini [14]
 1. Pemilihan Data
Sebelum melakukan penggalian informasi dalam *data mining*, perlu dilakukan seleksi pada kumpulan data operasional. Pada tahap ini, peneliti memilih data untuk memastikan kesesuaian dengan kebutuhan analisis.
 2. Pengolahan Data
Di tahap ini, peneliti menyiapkan data yang diperoleh dari dataset kaggle.
 3. Transformasi Data
Pada tahap ini, peneliti melakukan proses transformasi terhadap data yang telah dipilih, agar data tersebut siap untuk proses *data mining*. Peneliti mengubah atau menggabungkan data ke dalam format yang sesuai untuk analisis lebih lanjut.
 4. Klasifikasi dengan Algoritma C4.5 dan *Naïve Bayes*
Di tahap ini, penulis menganalisis hasil perhitungan menggunakan metode C4.5 dan *Naïve Bayes* serta membandingkannya melalui perangkat lunak Weka.
- e. Analisis Data
Pada Analisis data [17] merupakan tahapan penting dalam suatu proses penelitian yang bertujuan untuk mengolah, menafsirkan, dan menarik kesimpulan dari data yang telah dikumpulkan. Proses analisis data

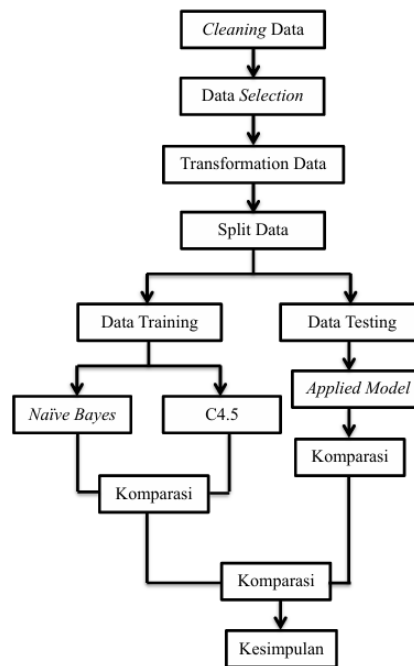
dimulai dari pengumpulan data, dilanjutkan dengan pembersihan data (data cleaning) untuk memastikan keakuratannya, kemudian dilakukan transformasi data agar lebih siap untuk dianalisis.

f. Pembuatan Laporan

Pada tahap ini pembuatan laporan akhir penelitian, dilakukan berdasarkan kerangka kerja yang telah di rancang yang terdiri dari Pendahuluan, Landasan Teori, Metodologi Penelitian, Analisis, Hasil dan Rekomendasi dan Penutup serta lampiran bukti hasil penelitian.

2.2 Alur Eksperimen

Penelitian ini menggunakan dua metode klasifikasi, yaitu *Naïve Bayes* dan *C4.5*. Pada kedua metode tersebut memiliki cara kerja yang berbeda dalam mengklasifikasikan data dan masing-masing memberikan hasil klasifikasi yang spesifik sesuai dengan karakteristik data. Adapun diagram alur experiment terdapat pada gambar 2.



Gambar 2. Diagram Alur Eksperimen

Berdasarkan gambar 2 dapat diuraikan penjelasan sebagai berikut:

a. *Cleaning Data*

Tahapan pertama adalah *cleaning* data yaitu menghapus data-data yang noise serta data yang tidak konsisten atau tidak relevan.

b. *Data Selection*

Tahap dalam proses *data mining* yang bertujuan memilih data yang relevan dan diperlukan dari sekumpulan data yang besar, agar sesuai dengan tujuan analisis atau penelitian.

c. *Transformation Data*

Pada tahap ini, merupakan proses mengubah dan mengonsolidasikan data ke dalam format yang sesuai untuk proses *data mining*.

d. *Split data*

Tahapan ini merupakan tahapan membagi data menjadi dua bagian yaitu *data training* dan *data testing*. Variasi pembagian akan dilakukan dengan 80% data training dan 20% data testing.

e. *Data Training*

1. *Naïve Bayes*

Algoritma ini menggunakan pendekatan probabilistik berdasarkan *Teorema Bayes*, dengan asumsi bahwa semua fitur bersifat independen satu sama lain.

2. *C4.5*

Adalah algoritma pohon keputusan yang mengembangkan ID3. *C4.5* membentuk pohon keputusan berdasarkan informasi gain dan rasio gain, kemudian digunakan untuk klasifikasi data berdasarkan kondisi atribut tertentu.

f. *Komparasi (Training)*

Setelah kedua model selesai dihitung, dilakukan perbandingan awal terhadap model hasil perhitungan.

g. *Data Testing*

Model yang telah dibentuk dari tahap *training* kemudian diterapkan pada *data testing*, yang berfungsi untuk mengevaluasi kinerja model terhadap data yang belum pernah dilihat sebelumnya.

h. *Applied Model*

Hasil dari prediksi model pada data *testing* akan dievaluasi dan dibandingkan berdasarkan metrik berikut :

- a. *Accuracy*: Persentase prediksi yang benar
- b. *Precision*: Tingkat ketepatan prediksi positif
- c. *Recall*: Kemampuan model mendeteksi semua data positif
- d. *F1-Score*: Harmonis antara precision dan recall
- e. *AUC-ROC Curve*: Menunjukkan *trade-off* antara TPR dan FPR

Berikut rumus yang digunakan untuk mencari TPR (*True Positive Rate*), FPR (*False Positive Rate*), Akurasi, dan *F-Measure* pada *confusion matrix* [17]:

$$TPR = \frac{TP}{(TP+FN)} \dots\dots\dots(3.1)$$

$$FPR = \frac{FP}{(FP+TN)} \dots\dots\dots(3.2)$$

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \dots\dots\dots (3.3)$$

$$F\text{-Measure} = \frac{2 * precision + Recall}{Precision + Recall} \dots\dots\dots (3.4)$$

Keterangan :

TP (*True Positive*) adalah dimana nilai sebenarnya positif namun nilai prediksinya menunjukkan positif.

FP (*False Positive*) adalah dimana nilai sebenarnya negatif namun nilai prediksinya menunjukkan positif.

TN (*True Negative*) adalah dimana nilai sebenarnya negatif namun nilai prediksinya menunjukkan negatif.

FN (*False Negative*) adalah dimana nilai sebenarnya positif namun nilai prediksinya menunjukkan negatif.

3. HASIL DAN PEMBAHASAN

3.1 Tahapan Eksperimen

Berdasarkan diagram alur eksperimen yang telah dibentuk sebelumnya, penelitian ini terdiri atas beberapa tahapan, yaitu:

3.1.1 *Data Cleaning*

Pada tahap ini peneliti melakukan proses menghilangkan noise dan data yang tidak konsisten atau tidak relevan. Dataset awal berjumlah 450 data kemudian setelah dilakukan *cleaning* data yaitu menghilangkan data yang tidak relevan sehingga data setelah diproses sebanyak 400 data prediksi penyakit ginjal. Contoh data yang tidak relevan dapat dilihat pada tabel 1.

Tabel 1. Contoh Data Yang Tidak Relevan

Age	BMI	AlcoholConsumption	PhysicalActivity	DietQuality	SleepQuality	SystolicBP	DiastolicBP
	31,06941	5,12811186	1,676219708	0,240385559	4,07643366	113	33
34	29,69212	18,60955184	8,377574454	6,503233231	7,652812555	0	67
	37,39482	11,88242933	9,607401096	2,104828363	,392786113	147	
40	31,32968	16,02016486	0,408870975	6,96442209	6,282273622	117	65
43	23,72631	7,944146086	0,780318707	3,097795666	4,021638617	98	66

3.1.2 *Data Selection*

Pada tahap ini data dipilih berdasarkan atribut yang relevan. Tidak semua atribut digunakan, adapun atribut yang tidak relevan yaitu atribut *PatientID*. Karena pada atribut tersebut hanya berisi nomor urut dari data pasien. Sehingga atribut tersebut tidak digunakan untuk tahap penelitian selanjutnya.

3.1.3 *Data Transformation*

Pada tahap ini, merubah data agar data dapat diproses dalam data mining. Dalam tahap ini data pada kelas diagnosis yang awalnya bentuk numeric (1 & 0) diubah kedalam bentuk kata. “yes” dan “no”. Hal ini dilakukan untuk mempermudah pemahaman hasil penelitian. Kelas diagnosis setelah dirubah kedalam bentuk kata dapat dilihat pada tabel 2.

Tabel 2. Transformasi Kelas

No.	Diagnosis	Diagnosis
1	1	yes
2	1	yes
3	0	no
4	0	no
5	1	yes

3.1.4 Split Data

Pada penelitian akan dilakukan proses split data menjadi 2 bagian yaitu data *training* dan data *testing*. Pembagian akan dilakukan dengan 80% data *training* dan 20% data *testing*. Pembagian data *split* seperti pada tabel 3.

Tabel 3. Pembagian Data Split

No.	Training	Testing	Jumlah Data Training	Jumlah Data Testing
1	80%	20%	320	80

3.2 Komparasi Algoritma Naïve Bayes dan C4.5

Setelah melakukan penelitian dalam mengklasifikasikan dataset dengan menggunakan algoritma *naïve bayes* dan C4.5, hasil pengujian berdasarkan *accuracy*, *persicion*, *recall*, F1-Score dan AUC-ROC. Berikut hasil perhitungan evaluasi matrix dengan data training 80% dan data testing 20% adapat dilihat pada tabel 4.

Tabel 4. Perbandingan Kinerja Klasifikasi

No.	Split Data	Perbandingan	Algoritma	
			Naïve Bayes	C4.5
1	80% Data Training	Accuracy	83.13%	83.44%
2		Precision	84.05%	83.23%
3		Recall	97.68%	99.61%
4		F1-Score	90.36%	90.69%
5		AUC-ROC	59.50%	85.25%
1	20% Data Testing	Accuracy	90.00%	98.75%
2		Precision	92.06%	100%
3		Recall	95.08%	98.36%
4		F1-Score	93.55%	99.17%
5		AUC-ROC	84.38%	85.25%

4. KESIMPULAN

Berdasarkan hasil analisis yang telah dilakukan, penelitian ini menunjukkan bahwa algoritma Naive Bayes dan C4.5 memiliki karakteristik dan kinerja yang berbeda dalam mendeteksi penyakit ginjal. Algoritma C4.5 cenderung menghasilkan performa yang lebih unggul secara keseluruhan. Dengan akurasi sebesar 83.44% pada data training dan 98.75% pada data testing dan nilai *recall* mencapai 99.61 pada data *training* sedangkan pada data *testing* mencapai 98.36%, C4.5 mampu mengidentifikasi seluruh kasus positif tanpa ada yang terlewat. Sementara itu, algoritma Naive Bayes, menghasilkan nilai akurasi 83.13% pada data testing untuk data training mencapai 90%, tetap menunjukkan kinerja yang baik dalam hal efisiensi. Perbedaan mendasar antara keduanya terletak pada cara kerja: *Naive Bayes* mengandalkan asumsi independensi antar atribut, sedangkan C4.5 memanfaatkan nilai gain dan rasio gain untuk membentuk model klasifikasi berbasis pohon keputusan. Penelitian ini memberikan kontribusi nyata dalam pengembangan model klasifikasi berbasis data mining untuk keperluan medis, khususnya dalam mendeteksi penyakit ginjal. Dengan membandingkan dua algoritma populer, penelitian ini tidak hanya menilai akurasi prediksi, tetapi juga mempertimbangkan aspek interpretabilitas dan kesesuaian algoritma dengan karakteristik data medis. Hasil ini dapat digunakan sebagai landasan dalam perancangan sistem deteksi dini penyakit ginjal yang lebih objektif, terukur. Penelitian ini dapat menjadi dasar bagi pengembangan sistem pendukung keputusan dalam proses medeteksi penyakit ginjal. Dengan implementasi sistem yang berbasis kecerdasan buatan dan analisis data, rumah sakit dapat meningkatkan efisiensi dan efektivitas dalam menyusun strategi dalam menangani penyakit ginjal dengan mendeteksi dini penyakit ginjal.

REFERENCES

- [1] Warite Gire, Eman Sulaiman, and Nirwana, "Hubungan Pengetahuan dan Dukungan Keluarga Dengan Kepatuhan Pasien Gagal Ginjal Kronik Dalam Menjalani Terapi Hemodialisa di RSUD Bahteramas Provinsi Sulawesi Tenggara Tahun 2023," *J. Penelit. Sains dan Kesehat. Avicenna*, vol. 2, no. 3, pp. 60–68, 2023, doi: 10.69677/avicenna.v2i3.62.
- [2] A. A. Qori'ah and Z. Fatah, "Implementasi Prediksi Penyakit Ginjal Kronis dengan Menggunakan Metode Decision Tree," *JUSIFOR J. Sist. Inf. dan Inform.*, vol. 3, no. 2, pp. 180–186, 2024, doi: 10.70609/jusifor.v3i2.5803.
- [3] D. Irawan, H. Oktavianto, M. K. Anam, T. Informatika, and U. M. Jember, "Analisis Penerapan Algoritma Naive Bayes," *JASIE "Jurnal Apl. Sist. Inf. Dan Elektron."*, vol. 1, no. 2, pp. 127–134, 2019.
- [4] K. L. Kohsasih and Z. Situmorang, "Comparative Analysis of C4.5 and Naïve Bayes Algorithms in Predicting Cerebrovascular Disease," *J. Inform.*, vol. 9, no. 1, pp. 13–17, 2022.
- [5] A. A. D. Halim and S. Anraeni, "Analisis Klasifikasi Dataset Citra Penyakit Pneumonia menggunakan Metode K-Nearest Neighbor (KNN)," *Indones. J. Data Sci.*, vol. 2, no. 1, pp. 01–12, 2021, doi: 10.33096/ijodas.v2i1.23.
- [6] A. Ridwan, "Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 15–21, 2020, doi: 10.47970/siskom-kb.v4i1.169.
- [7] F. M. Hana, "Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4 . 5," 2020.
- [8] E. A. Julisawati, "Pemilihan Jurusan SMA Menggunakan Sistem Pendukung Keputusan pada Siswa SMPN 210 Jakarta Menggunakan Metode Profile Matching," vol. 3, no. 3, pp. 397–403, 2023.
- [9] H. Mukaromah, "Perbandingan Kinerja Algoritma Naive Bayes dan," vol. 7, no. 1, pp. 20–24, 2020.
- [10] K. Pratami and Iwan Setiawan Wibisono, "Perbandingan Algoritma C4.5 Dan Naive Bayes Dalam Mendeteksi Hipertensi Di Puskesmas Banyubiru," *J. Mhs. Tek. Inform.*, vol. 2, no. 1, pp. 16–26, 2023, doi: 10.35473/jamastika.v2i1.1937.
- [11] B. A. C. Permana and I. K. D. Patwari, "Komparasi Metode Klasifikasi Data Mining Decision Tree dan Naïve Bayes Untuk Prediksi Penyakit Diabetes," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 63–69, 2021, doi: 10.29408/jit.v4i1.2994.
- [12] Julia Triani, Yovi Pratama, and E. Yanti, "Komparasi Dalam Prediksi Gagal Jantung Dengan Menggunakan Metode C4.5 dan Naïve Bayes," *J. Inform. Dan Rekayasa Komputer(JAKAKOM)*, vol. 3, no. 1, pp. 394–402, 2023, doi: 10.33998/jakakom.2023.3.1.759.
- [13] K. P. Widiatmika, *Pengantar Machine Learning*, vol. 16, no. 2. 2015.
- [14] M. A.-Z. Faradeya and E. R. Subhiyakto, "Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma Naive Bayes," *J. Algoritm.*, vol. 22, no. 1, pp. 115–127, 2025, doi: 10.33364/algoritma/v.22-1.2178.
- [15] J. Ha, M. Kambe, and J. Pe, *Data Mining: Concepts and Techniques*. 2011. doi: 10.1016/C2009-0-61819-5.
- [16] P. A. Sihotang and D. Sitanggang, "Penerapan Metode Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Jantung," *J. Tek. Inf. dan Komput.*, vol. 7, no. 2, p. 899, 2024, doi: 10.37600/tekinkom.v7i2.1535.
- [17] A. A. Permana *et al.*, *Machine Learning*, vol. 45, no. 13. 2023. [Online]. Available: <https://books.google.ca/books?id=EoYBngEACAAJ&dq=mitchell+machine+learning+1997&hl=en&sa=X&ved=0ahUKewiomdqfj8TkAhWGslkKHRCbAtoQ6AEIKjAA>
- [18] S. Beryl Enrico Ritonga and U. Enri, "Perbandingan Algoritme C4.5 Dan Naïve Bayes Dalam Klasifikasi Severe Preeklampsia Menggunakan Hematologi," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 2200–2207, 2024, doi: 10.36040/jati.v8i2.9435.
- [19] R. S. Wahono, *Data Mining Data mining*, vol. 2, no. January 2013. 2023. [Online]. Available: https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part
- [20] M. Ardiansyah, A. Sunyoto, and E. T. Luthfi, "Analisis Perbandingan Akurasi Algoritma Naïve Bayes Dan C4.5 untuk Klasifikasi Diabetes," *Edumatic J. Pendidik. Inform.*, vol. 5, no. 2, pp. 147–156, 2021, doi: 10.29408/edumatic.v5i2.3424.