Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan
Universitas Muhammadiyah Muara Bungo
Volume 3, No 1, 2026
e-ISSN 3089-3763
Hal:60-66

PERBANDINGAN ALGORITMA NAIVE BAYES DAN K-NEAREST NEIGHBOR UNTUK KLASIFIKASI RISIKO PENYAKIT GINJAL KRONIS PADA DATASET CHRONIC KIDNEY DISEASE

Muhammad Al Madany[1], Ahmad Homaidi[2]
Universitas Ibrahimy
muhammadalmadany73@gmail.com

ABSTRAK

Penyakit Ginjal Kronis (CKD) merupakan penyakit tidak menular dengan prevalensi meningkat yang mengakibatkan penurunan fungsi ginjal permanen, sehingga deteksi dini sangat krusial untuk diagnosis serta pengambilan keputusan medis. Penelitian ini bertujuan membandingkan performa algoritma Naïve Bayes dan K-Nearest Neighbor (K-NN) dalam mengklasifikasikan risiko CKD. Metodologi penelitian mengadopsi kerangka kerja CRISP-DM, mulai dari "business understanding" hingga "deployment". Penelitian ini memanfaatkan dataset berjumlah 796 data pasien dengan 11 atribut (10 variabel independen dan 1 variabel dependen). Tahap "preprocessing" mencakup pembersihan data, transformasi data kategorikal ke numerik, normalisasi, serta pembagian data "training" dan "testing". Evaluasi model diukur menggunakan "confusion matrix", akurasi, "precision", "recall", dan "F1-score". Hasil pengujian membuktikan bahwa algoritma Naïve Bayes menunjukkan performa unggul dengan akurasi 99,38%, "precision" 98,77%, "recall" 100%, serta "F1-score" 99,38%. Sebaliknya, algoritma K-NN mencatatkan akurasi 98,75%, "precision" 97,56%, "recall" 100%, dan "F1-score" 98,77%. Dengan tingkat kesalahan klasifikasi lebih rendah, Naïve Bayes terbukti lebih optimal dan konsisten. Hasil riset ini diharapkan menjadi referensi berharga bagi pengembangan sistem pendukung keputusan deteksi dini CKD.

Kata Kunci: Naive Bayes, K-Nearest Neighbor, Chronic Kidney Disease, Klasifikasi, Data Mining

ABSTRACT

Chronic Kidney Disease (CKD) is a non-communicable disease with an increasing prevalence that results in permanent loss of kidney function; therefore, early detection is crucial for diagnosis and medical decision-making. This study aims to compare the performance of Naïve Bayes and K-Nearest Neighbor (K-NN) algorithms in classifying CKD risk The research methodology employs the CRISP-DM framework, covering stages from business understanding to deployment. The study utilizes a dataset containing 796 patient records with 11 attributes (10 independent variables and 1 dependent variable). The preprocessing stage involves data cleaning, converting categorical data to numerical format, normalization, and splitting the data into training and testing sets. Model evaluation is conducted using a confusion matrix, accuracy, precision, recall, and F1-score. Test results indicate that the Naïve Bayes algorithm demonstrated superior performance, achieving an accuracy of 99.38%, precision of 98.77%, recall of 100%, and an F1-score of 99.38%. In contrast, the K-NN algorithm recorded an accuracy of 98.75%, precision of 97.56%, recall of 100%, and an F1-score of 98.77%. With a lower classification error rate, Naïve Bayes proved to be more optimal and consistent. These findings are expected to serve as a valuable reference for the development of decision support systems for the early detection of CKD.

Keywords: *Naïve Bayes, K-Nearest Neighbor, Chronic Kidney Disease, classification, data mining*

Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan
Universitas Muhammadiyah Muara Bungo
Volume 3, No 1, 2026
e-ISSN 3089-3763
Hal:60-66

Pendahuluan

Seiring dengan berkembangnya teknologi, berbagai penelitian telah memanfaatkan metode machine learning untuk membantu proses prediksi dan klasifikasi penyakit ginjal kronis. Berdasarkan penelitian terdahulu, prediksi penyakit gagal ginjal kronis telah dilakukan menggunakan metode Naïve Bayes dan K-Nearest Neighbor dengan hasil akurasi yang cukup tinggi. Namun, penelitian tersebut masih memiliki keterbatasan, antara lain penggunaan algoritma yang terbatas, evaluasi model yang hanya berfokus pada akurasi, serta penggunaan satu dataset tanpa pengujian generalisasi. Selain itu, belum dilakukan analisis feature selection dan optimasi parameter model. Oleh karena itu, penelitian ini diusulkan untuk mengatasi keterbatasan tersebut dengan menerapkan perbandingan algoritma Naïve Bayes dan K-Nearest Neighbor dalam mengklasifikasikan risiko penyakit ginjal kronis serta melakukan evaluasi performa model yang lebih komprehensif.

Penyakit Ginjal Kronis (PGK) merupakan salah satu penyakit tidak menular yang menjadi beban penyakit di Indonesia dan termasuk dalam kondisi double burden of disease. Prevalensi PGK pada penduduk Indonesia usia ≥15 tahun mengalami peningkatan, dari 2 per 1000 penduduk pada tahun 2013 menjadi 3,8 per 1000 penduduk pada tahun 2018. Penyakit ginjal kronis bersifat progresif dan ditandai dengan penurunan fungsi ginjal yang berlangsung secara bertahap dan permanen, sehingga pada stadium lanjut dapat berkembang menjadi gagal ginjal. faktor risiko yang berperan penting dalam peningkatan kejadian PGK di Indonesia adalah hipertensi dan diabetes melitus, yang terbukti secara signifikan meningkatkan risiko terjadinya penyakit ginjal kronis pada masyarakat.

Deteksi penyakit ginjal kronis menjadi sangat penting untuk membantu proses penanganan dan pengambilan keputusan medis. Dalam praktiknya, proses identifikasi dan klasifikasi penyakit ginjal memerlukan ketelitian tinggi serta analisis terhadap data medis pasien. Oleh karena itu, pemanfaatan metode komputasi seperti data mining menjadi solusi alternatif yang banyak digunakan dalam bidang kesehatan, khususnya untuk prediksi dan klasifikasi penyakit. Data mining menawarkan metodologi dan solusi teknis untuk mengolah data medis serta membangun model prediksi penyakit ginjal dengan menggunakan algoritma klasifikasi. Beberapa penelitian menunjukkan bahwa penerapan algoritma klasifikasi mampu memberikan hasil prediksi yang baik dan membantu proses identifikasi penyakit ginjal kronis secara berbasis data.

Dua Algoritma yang banyak digunakan dalam penelitian data kesehatan adalah Naïve Bayes dan K-Nearest Neighbor (K-NN). Algoritma yang banyak digunakan dalam penelitian data kesehatan salah satunya adalah Naïve Bayes. Naïve Bayes merupakan metode klasifikasi yang bekerja berdasarkan pendekatan probabilistik dengan asumsi independensi antar atribut. Algoritma ini dikenal memiliki proses komputasi yang sederhana dan cepat, sehingga efisien dalam pengolahan data medis dengan jumlah atribut yang cukup banyak. Beberapa penelitian menunjukkan bahwa algoritma Naïve Bayes mampu memberikan hasil klasifikasi yang stabil dan tingkat akurasi yang baik dalam prediksi penyakit, sehingga sering digunakan dalam penelitian klasifikasi penyakit berbasis data medis.

Di sisi lain, K-Nearest Neighbor (K-NN) merupakan algoritma klasifikasi berbasis kedekatan jarak yang bekerja dengan cara menentukan kelas suatu data baru berdasarkan kemiripannya dengan data latih yang paling dekat. Algoritma K-NN melakukan proses klasifikasi dengan mencari sejumlah K data terdekat dalam data training yang memiliki karakteristik paling mirip dengan data uji. Dalam bidang kesehatan, K-NN sering digunakan untuk membantu proses klasifikasi penyakit karena mampu memanfaatkan kedekatan kasus pasien baru dengan kasus pasien sebelumnya. Hasil penelitian

Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan
Universitas Muhammadiyah Muara Bungo
Volume 3, No 1, 2026
e-ISSN 3089-3763
Hal:60-66

menunjukkan bahwa penerapan algoritma K-NN pada data medis mampu menghasilkan tingkat akurasi yang baik, terutama ketika pemilihan nilai K dilakukan secara tepat dan perhitungan jarak diterapkan dengan benar.

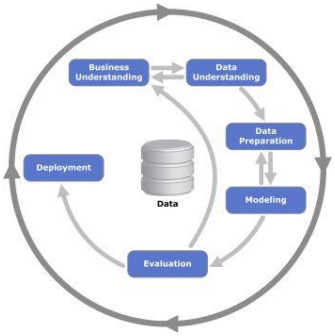
Dataset Chronic Kidney Disease yang tersedia pada repositori seperti Kaggle banyak digunakan dalam penelitian klasifikasi karena memuat data klinis dan laboratorium yang relevan dengan diagnosis penyakit ginjal kronis. Dataset ini dimanfaatkan untuk membangun model prediksi menggunakan metode data mining dan machine learning. Namun, perbedaan karakteristik dan mekanisme kerja setiap algoritma menyebabkan variasi performa meskipun diterapkan pada dataset yang sama. Penelitian komparatif menunjukkan bahwa algoritma Naive Bayes dan K-Nearest Neighbor (K-NN) menghasilkan kinerja yang berbeda dalam hal akurasi dan metrik evaluasi lainnya, yang dipengaruhi oleh parameter serta metode pengolahan data.

Sehubungan dengan kondisi tersebut, diperlukan suatu penelitian yang secara sistematis membandingkan kinerja Algoritma Naive bayes dan K-Nearest Neighbor(K-NN) dalam proses klasifikasi risiko penyakit ginjal kronis. Hasil dari perbandingan ini diharapkan mampu memberikan rekomendasi mengenai algoritma yang paling efektif dan optimal untuk diimplementasikan dalam pengembangan Decision Support System (DSS) guna mendukung deteksi dini Chronic Kidney Disease(CKD).

Materi dan Metode

1. Jenis Penelitian

Penelitian ini menggunakan metode pengembangan sistem CRISP-DM (Cross-industry standard process for data mining), yaitu kerangka kerja standar internasional yang mengatur proses data mining secara sistematis melalui enam taham: business understanding, data understanding, data preparation, modeling, evaluation, dan deployment. Metode ini membantu penelitian berjalan lebih terstruktur dan menghasilkan model klasifikasi risiko penyakit ginjal kronis yang lebih valid serta dapat dipertanggungjawabkan



Gambar 1. Metode CRISP-DM

a. *Busines Understanding* (Pemahaman Bisnis)

Tahapan Business Understanding bertujuan untuk memahami permasalahan dan kebutuhan proyek dari sudut pandang bisnis atau domain penelitian. Pada tahap ini dilakukan identifikasi masalah nyata yang ingin diselesaikan, penentuan tujuan data mining, serta perumusan strategi dan kriteria keberhasilan yang akan dicapai agar hasil analisis relevan dan bermanfaat bagi pengambilan keputusan.

b. *Data Understanding* (Pemahaman Data)

Tahapan Data Understanding berfokus pada proses pengumpulan data awal dan eksplorasi dataset. Pada fase ini dilakukan pemahaman terhadap karakteristik dan kualitas data, termasuk identifikasi nilai hilang (missing values), inkonsistensi, dan anomali data. Tahap ini penting untuk memastikan data yang digunakan layak dan sesuai dengan tujuan penelitian.

c. *Data Preparation* (Persiapan Data)

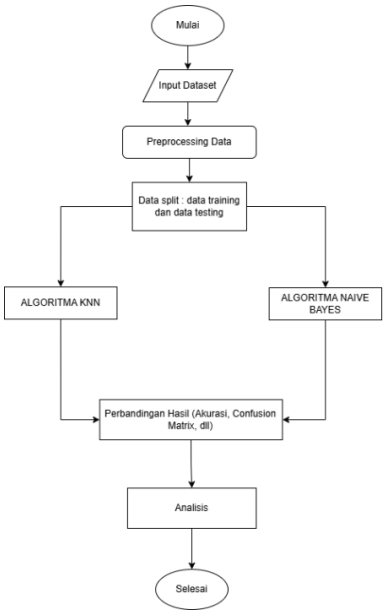
Tahapan Data Preparation merupakan proses pemilihan, pembersihan, konstruksi, dan transformasi data ke dalam format yang sesuai untuk pemodelan. Proses ini mencakup penanganan data hilang, penghapusan duplikasi, normalisasi, serta seleksi atribut. Tahap ini umumnya menjadi fase yang paling memakan waktu dalam proyek data mining karena sangat menentukan kualitas model yang dihasilkan.

Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan
Universitas Muhammadiyah Muara Bungo
Volume 3, No 1, 2026
e-ISSN 3089-3763
Hal:60-66

d. *Modelling* (Pemodelan data)
Tahapan Modeling adalah proses pemilihan dan penerapan teknik pemodelan yang sesuai dengan karakteristik data dan tujuan penelitian. Pada fase ini dilakukan pengaturan parameter, pemilihan algoritma klasifikasi, serta pembangunan model berdasarkan data yang telah dipersiapkan sebelumnya.

e. *Evaluation* (Evaluasi)
Tahapan Evaluation bertujuan untuk memastikan bahwa moedel yang dibangun telah memenuhi tujuan bisnis dan kriteria keberhasilan yang ditetapkan pada tahap awal. Evaluasi dilakukan menggunakan metrik kinerja yang sesuai serta peninjauan menyeluruh terhadap proses dan hasil pemodelan untuk menilai validitas dan keandalan model.

f. *Devlopment* (Penyebaran)
Tahapan *Development* merupakan proses implementasi hasil data mining ke dalam lingkungan atau proses bisnis yang sebenarnya tahapan ini mencakup penerapan model, pemantauan kinerja model setelah implementasi, serta penyusunan dokumentasi dan pelaporan hasil proyek sebagai bahan evaluasi dan pengembangan lanjutan.



Gambar 2. Diagram alur

Gambar tersebut merupakan alur penelitian *(flowchart)* yang digunakan untuk membandingkan performa algoritma K-Nearest Neighbor (KNN) dan Naïve Bayes dalam mengklasifikasikan penyakit Chronic Kidney Disease (CKD). Proses penelitian diawali dengan menginput dataset yang berisi data pasien CKD. Selanjutnya, data menjalani tahap preprocessing, seperti pembersihan data, pengubahan data kategorikal menjadi numerik, dan normalisasi agar data siap digunakan dalam proses klasifikasi. Setelah preprocessing selesai, dataset dibagi menjadi data training dan data testing. Data training digunakan untuk membangun model, sedangkan data testing digunakan untuk menguji kinerja model yang telah dibuat. Kemudian, data yang sama diproses menggunakan dua algoritma yang berbeda, yaitu KNN dan Naïve Bayes, sehingga diperoleh hasil prediksi dari masing-masing metode. Hasil dari kedua algoritma selanjutnya dibandingkan menggunakan beberapa metrik evaluasi, seperti akurasi, confusion matrix, precision, recall, dan F1-score, untuk mengetahui metode yang memiliki performa terbaik. Setelah dilakukan perbandingan, tahap berikutnya adalah analisis hasil, yaitu menginterpretasikan dan mengevaluasi kinerja kedua algoritma berdasarkan nilai yang diperoleh. Proses penelitian diakhiri dengan penarikan kesimpulan mengenai algoritma yang paling optimal dalam mengklasifikasikan penyakit CKD berdasarkan hasil pengujian yang telah dilakukan.

Hasil dan Pembahasan

Dataset yang digunakan terdiri dari 796 data pasien dengan 11 atribut. Sebanyak 10 atribut digunakan sebagai variabel independen, yaitu Age, BMI, Smoking, AlcoholConsumption, PhysicalActivity, BloodPressureSystolic, BloodPressureDiastolic, BloodGlucose, Cholesterol, dan QualityOfLifeScore. Sementara itu, atribut Diagnosis digunakan sebagai variabel dependen atau label klasifikasi. Dataset telah seimbang (balanced), dengan jumlah data pada kelas Diagnosis = 0

Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan
Universitas Muhammadiyah Muara Bungo
Volume 3, No 1, 2026
e-ISSN 3089-3763
Hal:60-66

sebanyak 398 data dan Diagnosis = 1 sebanyak 398 data, sehingga total keseluruhan data berjumlah 796 record. Hal ini membuat proses klasifikasi lebih optimal karena tidak terjadi ketidakseimbangan kelas (class imbalance).

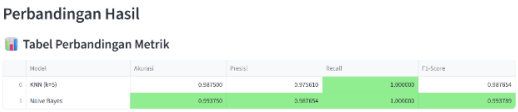
Deskripsi Atribut

Table 1 Atribut

No	Nama Atribut	Tipe Data	Keterangan
1	Age	Integer	Usia pasien (tahun)
2	BMI	Float	Body Mass Index (Indeks Massa Tubuh)
3	Smoking	Integer	Status merokok (0 = Tidak, 1 = Ya)
4	AlcoholConsumption	Integer	Konsumsi alkohol (0 = Tidak, 1 = Ya)
5	PhysicalActivity	Float	Tingkat aktivitas fisik
6	BloodPressureSystolic	Integer	Tekanan darah sistolik (mmHg)
7	BloodPressureDiastolic	Integer	Tekanan darah diastolik (mmHg)
8	BloodGlucose	Integer	Kadar glukosa darah
9	Cholesterol	Integer	Kadar kolesterol
10	QualityOfLifeScore	Float	Skor kualitas hidup pasien
11	Diagnosis	Integer	Label klasifikasi (0 = Tidak CKD, 1 = CKD)

Perbandingan Metrix

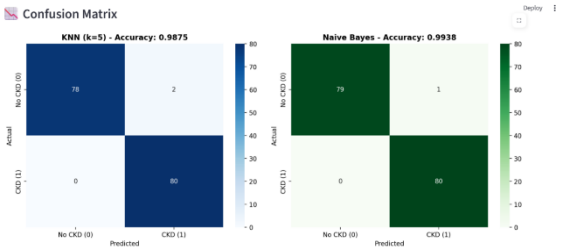
Dalam penelitian ini, algoritma Naive Bayes terbukti lebih unggul dibandingkan K-Nearest Neighbor (KNN) dengan k=5 dalam klasifikasi data Chronic Kidney Disease (CKD). Meskipun keduanya sama-sama memiliki nilai recall 100% yang berarti semua kasus positif berhasil terdeteksi, Naive Bayes menunjukkan performa lebih baik dengan akurasi 99,38%, presisi 98,77%, dan F1-Score 99,38%, sedangkan KNN hanya mencapai akurasi 98,75%, presisi 97,56%, dan F1-Score 98,77%. Perbedaan ini menegaskan bahwa Naive Bayes lebih optimal dan konsisten dalam menghasilkan prediksi, sehingga dapat dianggap sebagai algoritma yang lebih tepat digunakan pada dataset CKD dalam penelitian ini.



Gambar 2. Perbandingan metrix

Confusion Matrix

confusion matrix, kedua algoritma menunjukkan kinerja yang sangat baik dalam mengklasifikasikan data Chronic Kidney Disease (CKD). Pada algoritma KNN (k=5) diperoleh 78 data No CKD (0) yang berhasil diprediksi dengan benar, 80 data CKD (1) yang berhasil diprediksi dengan benar, serta terdapat 2 data No CKD yang salah diklasifikasikan sebagai CKD. Tidak terdapat data CKD yang salah diprediksi sebagai No CKD, sehingga nilai recall mencapai 100%. Sementara itu, algoritma Naive Bayes menghasilkan performa yang lebih baik dengan 79 data No CKD dan 80 data CKD yang diprediksi dengan benar, serta hanya 1 data No CKD yang salah diklasifikasikan sebagai CKD. Sama seperti KNN, tidak ada data CKD yang salah diprediksi sebagai No CKD. Hasil ini menunjukkan bahwa kedua algoritma mampu mendeteksi seluruh kasus CKD dengan sangat baik, namun Naive Bayes memiliki tingkat kesalahan yang lebih rendah dan akurasi yang lebih tinggi (99,38%) dibandingkan KNN (98,75%), sehingga Naive Bayes menjadi algoritma yang paling optimal untuk klasifikasi CKD pada dataset yang digunakan dalam penelitian ini.



Gambar 3. Confusion Matrix

Pada subbab ini dijelaskan teori-teori pendukung yang menjadi landasan dalam penelitian perbandingan algoritma naive bayes dan K- *K-Nearest Neighbor* (KNN) Penelitian

Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan
Universitas Muhammadiyah Muara Bungo
Volume 3, No 1, 2026
e-ISSN 3089-3763
Hal:60-66

ini bertujuan untuk membandingkan kinerja algoritma Naïve Bayes dan K-Nearest Neighbor (KNN) dalam memprediksi penyakit Gagal Ginjal Kronis (Chronic Kidney Disease/CKD) menggunakan dataset CKD dari UCI Repository. Sebelum proses klasifikasi dilakukan, data terlebih dahulu melalui tahap pra-pemrosesan, seperti label encoding untuk mengubah data kategorikal menjadi numerik dan normalisasi untuk menyamakan rentang nilai atribut. Selanjutnya, data dibagi menjadi data latih dan data uji untuk mengevaluasi performa kedua algoritma. Naïve Bayes melakukan prediksi berdasarkan perhitungan probabilitas, sedangkan KNN mengklasifikasikan data berdasarkan kedekatan dengan data tetangga terdekat. Hasil penelitian menunjukkan bahwa KNN menghasilkan akurasi yang lebih tinggi, yaitu 96% pada nilai K = 3, dibandingkan dengan Naïve Bayes yang memperoleh akurasi 90%. Dengan demikian, KNN dinilai lebih efektif dalam mengenali pola pada dataset CKD yang digunakan, meskipun kedua metode sama-sama mampu membantu proses deteksi dini penyakit gagal ginjal kronis. Penelitian ini juga menyarankan penggunaan dataset lain pada penelitian selanjutnya untuk menguji kestabilan dan konsistensi performa kedua algoritma dalam berbagai kondisi data.

Penelitian yang dilakukan oleh Adi Muzakir, Anita Desiani, dan Ali Amran bertujuan untuk membandingkan kinerja algoritma Naïve Bayes dan K-Nearest Neighbor (K-NN) dalam mengklasifikasikan penyakit kanker prostat sebagai upaya deteksi dini penyakit. Penelitian ini memanfaatkan teknik data mining untuk mengolah data pasien dan menentukan metode klasifikasi yang memberikan hasil terbaik. Berdasarkan hasil pengujian, algoritma Naïve Bayes memperoleh akurasi sebesar 80%, presisi rata-rata 71,5%, dan recall 88%, sedangkan algoritma K-NN menghasilkan akurasi sebesar 90%, presisi rata-rata 93%, dan recall 87,5%. Hasil tersebut menunjukkan bahwa K-NN memiliki performa yang lebih baik dibandingkan Naïve Bayes dalam mengklasifikasikan kanker prostat karena mampu mencapai tingkat akurasi dan presisi yang lebih tinggi. Meskipun demikian, Naïve Bayes tetap menunjukkan kinerja yang cukup baik dengan nilai akurasi, presisi, dan recall yang berada di atas 70%, sehingga kedua algoritma dapat digunakan dalam proses klasifikasi penyakit kanker prostat, namun K-NN lebih direkomendasikan berdasarkan hasil penelitian tersebut.

Penelitian ini bertujuan untuk membangun aplikasi klasifikasi penyakit diabetes dengan membandingkan dua algoritma machine learning, yaitu K-Nearest Neighbor (KNN) dan Naïve Bayes. Dataset yang digunakan terdiri dari 50 data pasien yang diperoleh dari Klinik Bidan Saptarum Maslahah Jombang, dengan delapan atribut yang berkaitan dengan kondisi pasien, seperti kadar glukosa, tekanan darah, jumlah kehamilan, insulin, BMI, riwayat diabetes, dan usia. Sebelum proses klasifikasi dilakukan, data terlebih dahulu melalui tahap preprocessing, yaitu normalisasi menggunakan MinMaxScaler dan pembagian data menjadi data latih serta data uji. Hasil pengujian menunjukkan bahwa algoritma KNN dengan nilai K = 3 memperoleh akurasi sebesar 93%, sedangkan Naïve Bayes menghasilkan akurasi sebesar 95%, sehingga Naïve Bayes memiliki performa yang sedikit lebih baik dalam mengklasifikasikan pasien diabetes pada dataset tersebut. Kedua model kemudian diintegrasikan ke dalam aplikasi berbasis Python sehingga dapat digunakan untuk melakukan prediksi secara langsung berdasarkan data yang dimasukkan oleh pengguna. Berdasarkan hasil penelitian, dapat disimpulkan bahwa kedua algoritma efektif dalam mendeteksi penyakit diabetes, namun Naïve Bayes memberikan tingkat akurasi yang lebih tinggi, sehingga lebih unggul pada penelitian ini. Aplikasi yang dikembangkan juga dapat membantu proses deteksi dini diabetes secara cepat dan mendukung pengambilan keputusan dalam penanganan pasien.

Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Naïve Bayes dan K-Nearest Neighbor (K-NN) sama-sama mampu digunakan untuk klasifikasi risiko penyakit ginjal kronis (Chronic Kidney Disease/CKD) dengan performa yang sangat baik. Proses klasifikasi dilakukan menggunakan dataset CKD yang telah melalui tahap preprocessing, normalisasi, dan pembagian data training serta testing menggunakan metode CRISP-DM. Hasil pengujian menunjukkan bahwa algoritma Naïve Bayes memberikan performa yang lebih unggul dibandingkan algoritma K-NN. Naïve Bayes memperoleh nilai akurasi sebesar 99,38%, precision 98,77%, recall 100%, dan F1-score 99,38%, sedangkan K-NN memperoleh akurasi sebesar 98,75%, precision 97,56%, recall 100%, dan F1-score 98,77%. Selain itu, berdasarkan confusion matrix, Naïve Bayes memiliki tingkat kesalahan klasifikasi yang lebih rendah dibandingkan K-NN. Dengan demikian, algoritma Naïve Bayes dinilai lebih optimal dan konsisten dalam mengklasifikasikan risiko penyakit ginjal kronis pada dataset yang digunakan dalam penelitian ini. Hasil penelitian ini diharapkan dapat menjadi referensi dalam pengembangan sistem pendukung keputusan dan membantu proses deteksi dini penyakit ginjal kronis secara lebih cepat dan akurat.

Daftar Pustaka

[1] M. Rani, I. B. P. Pramana, and A. A. G. Oka, "Studi Meta-analisis Hipertensi dan Diabetes Melitus Sebagai Faktor Risiko Penyakit Ginjal Kronis di Indonesia," *E-Jurnal Med. Udayana*, vol. 11, no. 9, p. 49, 2022, doi: 10.24843/mu.2022.v11.i9.p10.

[2] A. Rohman and M. Rochcham, "Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Kelulusan Mahasiswa," *Neo Tek.*, vol. 5, no. 1, pp. 34–40, 2019, doi: 10.37760/neoteknika.v5i1.1379.

[3] S. C. Dewi, C. E. Putra, and A. G. Nugraheni, "Implementasi Metode K-Nearest Neighbors (KNN) dan Naive Bayes untuk Klasifikasi Penyakit Jantung," *Technol. Informatics Insight J.*, vol. 3, no. 2, pp. 76–94, 2024, doi: 10.32639/p5e7b161.

[4] Hasran, "Klasifikasi Penyakit Jantung Menggunakan Metode K-Nearest Neighbor," *Indones. J. Data Sci.*, vol. 1, no. 1, pp. 6–10, 2020, [Online]. Available: http://bit.ly/datasetcardio.

[5] I. Ikko *et al.*, "PENYAKIT GINJAL KRONIS Imaniar Ikko Mulya Rizky , Wika Fitriana Purwaningtyas , Mega Rahmawati : Anlisis Komparatif Algo-ritma Klasifikasi Untuk Prediksi Dini Penyakit Ginjal Kronis PENDAHULUAN Penyakit ginjal kronis (PGK) adalah penyakit kronis yang m," vol. 9, no. 2, pp. 89–95, 2025.

[6] P. C. Ncr *et al.*, "Step-by-step data mining guide," *SPSS inc*, vol. 78, pp. 1–78, 2000, [Online]. Available: http://www.crisp-dm.org/CRISPWP-0800.pdf

[7] N. P. Nugraha, R. Azim, S. Z. Daffa, and P. S. Ningayu, "Perbandingan Akurasi Metode Naïve Bayes dan Metode KNN untuk Memprediksi Gagal Ginjal Kronis," *J. Rekayasa Elektro Sriwij.*, vol. 5, no. 1, pp. 1–10, 2023, doi: 10.36706/jres.v5i1.63.

[8] A. Muzakir, A. Desiani, and A. Amran, "Klasifikasi Penyakit Kanker Prostat Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbor," *Komputika J. Sist. Komput.*, vol. 12, no. 1, pp. 73–79, 2023, doi: 10.34010/komputika.v12i1.9629.

[9] W. D. Prasetya and B. Sujatmiko, "Rancang Bangun Aplikasi dengan Perbandingan Metode K-Nearest Neighbor (KNN) dan Naive Bayes dalam Klasifikasi Penderita Penyakit Diabetes," *J. Informatics Comput. Sci.*, vol. 3,