JKTi Jurnal Keilmuan Teknologi informasi

Volume 01, No. 1, Juni 2025, Hal. 15 – 22
e-ISSN : 3109–4163

Prediksi Analitik untuk Penyakit Ginjal Kronis: Perbandingan Metode Machine Learning

Krisna Nuresa Qodri*[1], Muhammad Rausan Fikri[2], Luthfi Ardi[3]

*[1] Program Studi Rekayasa Keamanan Siber, Jurusan Komputer Dan Bisnis, Politeknik Negeri Cilacap, Cilacap
[2] Pengadilan Negeri Bangkinang, Kampar
[2] EZB Wisata Indonesia, Batam

Email: [1]krisnanuresa@pnc.ac.id

ABSTRACT — Chronic kidney disease (CKD) is a progressive malady defined by reduced glomerular filtration rate, increased urinary albumin excretion or both, and is a major global public health concern with an extremely high unmet medical need. CKD is estimated to occur in 8-16% of the worldwide population and results in a substantially reduced life expectancy. Early detection and accurate prediction of CKD is crucial to reduce health complications such as hypertension, anemia, and premature death. This study aims to develop CKD prediction models using three machine learning methods: Random Forest, Naive Bayes, and Support Vector Machine, then compare the performance of each method. The dataset used is the CKD dataset from UCI Machine Learning Repository consisting of 400 instances with 24 attributes. Experimental results show that Random Forest achieved 90.50% accuracy, Naive Bayes achieved the highest accuracy of 94.21%, while SVM achieved 88.84% accuracy. The results indicate that Naive Bayes provides the best performance for chronic kidney disease prediction with superior accuracy compared to other methods. This prediction model can assist medical practitioners in early detection and appropriate clinical decision-making for CKD patient management.

KEYWORDS — business intelligence; data mining; predictive analytics; naive bayes; support vector machine; random forest; chronic kidney disease.

INTISARI — Penyakit ginjal kronis (CKD) adalah penyakit progresif yang ditandai dengan penurunan laju filtrasi glomerulus, peningkatan ekskresi albumin urin atau keduanya, dan merupakan masalah kesehatan masyarakat global utama dengan beban kesehatan yang sangat tinggi. CKD diperkirakan terjadi pada 8-16% dari populasi dunia dan mengakibatkan harapan hidup yang jauh berkurang. Deteksi dini dan prediksi yang akurat terhadap CKD sangat penting untuk mengurangi komplikasi kesehatan seperti hipertensi, anemia, dan kematian prematur. Penelitian ini bertujuan untuk mengembangkan model prediksi CKD menggunakan tiga metode machine learning yaitu Random Forest, Naive Bayes, dan Support Vector Machine, kemudian membandingkan performa masing-masing metode. Dataset yang digunakan adalah dataset CKD dari UCI Machine Learning Repository yang terdiri dari 400 instance dengan 24 atribut. Hasil eksperimen menunjukkan bahwa Random Forest memperoleh akurasi 90,50%, Naive Bayes memperoleh akurasi tertinggi sebesar 94,21%, sedangkan SVM memperoleh akurasi 88,84%. Hasil penelitian ini menunjukkan bahwa Naive Bayes memberikan performa terbaik untuk prediksi penyakit ginjal kronis dengan tingkat akurasi yang superior dibandingkan metode lainnya. Model prediksi ini dapat membantu tenaga medis dalam melakukan deteksi dini dan pengambilan keputusan klinis yang tepat untuk penanganan pasien CKD.

KATA KUNCI — business intelligence; data mining; prediksi analitik; naive bayes; support vector machine; random forest; penyakit ginjal kronis.

I. PENDAHULUAN

Penyakit ginjal kronis (CKD) merupakan salah satu masalah kesehatan global yang paling serius dan berkembang dengan cepat. Menurut Sustainable Development Goals (SDGs) ketiga dari PBB tentang kesehatan dan kesejahteraan yang baik, penyakit tidak menular menjadi tantangan yang berkembang dengan tujuan mengurangi kematian prematur akibat penyakit tidak menular sebesar sepertiga pada tahun 2030 [1]. CKD adalah penyakit progresif yang ditandai dengan penurunan laju filtrasi glomerulus, peningkatan ekskresi albumin urin atau keduanya, dan merupakan masalah kesehatan masyarakat global utama dengan kebutuhan medis yang sangat tinggi [2].

Prevalensi CKD terus meningkat secara global. Statistik menunjukkan bahwa CKD mempengaruhi 10-15% dari populasi global dan menjadi penyebab signifikan morbiditas dan mortalitas dari penyakit tidak menular [3]. Di Indonesia, berdasarkan data Riset Kesehatan Dasar (Riskesdas) 2018, prevalensi penyakit ginjal kronis mencapai 3,8% dari total populasi, dengan kecenderungan meningkat seiring bertambahnya usia [4]. Diabetes mellitus diakui sebagai penyebab utama gagal ginjal pada hampir setengah dari semua kasus CKD baru, dan mengingat pandemi diabetes global saat ini, prevalensi komplikasi ginjal akan terus meningkat [5].

Deteksi dini dan prediksi yang akurat terhadap stadium CKD sangat penting untuk meminimalkan dampak komplikasi kesehatan pasien seperti hipertensi, anemia, penyakit kardiovaskular, dan kematian prematur [6]. Penyakit ginjal memerlukan perawatan medis khusus berdasarkan kondisi kronis pasien dari Stadium 1 hingga Stadium 5, dimana prosedur akan bervariasi berdasarkan penyebabnya [7].

Perawatan biasanya terdiri dari tindakan untuk membantu mengendalikan tanda dan gejala, mengurangi komplikasi, dan memperlambat perkembangan penyakit [8]. Dalam era digital dan big data saat ini, teknologi machine learning telah menunjukkan potensi besar dalam bidang kesehatan, khususnya untuk prediksi dan diagnosis penyakit [9]. Machine learning dapat membantu mengidentifikasi pola kompleks dalam data medis yang mungkin terlewatkan oleh analisis konvensional [10]. Beberapa penelitian terdahulu telah mengeksplorasi penggunaan berbagai algoritma machine learning untuk prediksi CKD, namun masih terdapat ruang untuk peningkatan akurasi dan efisiensi model [11][12].

Penelitian terdahulu menunjukkan bahwa metode ensemble learning dan hybrid approaches dapat meningkatkan performa prediksi CKD [13][14]. Namun, masih terdapat keterbatasan dalam hal perbandingan komprehensif antara metode-metode klasik seperti Random Forest, Naive Bayes, dan Support Vector Machine dengan dataset yang sama dan preprocessing yang konsisten [15]. Selain itu, banyak penelitian yang berfokus pada deteksi penyakit setelah terjadi, namun sedikit yang berkontribusi pada prediksi penyakit sebelum manifestasi klinis [16].

Business intelligence dalam bidang kesehatan merupakan salah satu solusi untuk meningkatkan layanan dan perawatan pasien [17]. Dengan menganalisis data penyakit dari pasien, dokter akan mudah mengambil keputusan dalam pengobatan pasien [18]. Prediksi penyakit ginjal kronis memungkinkan dokter untuk menentukan perawatan yang sesuai dengan stadium penyakit pasien [19]. Dari sisi pasien, mereka akan mengetahui kondisi kesehatan dan perawatan yang harus dilakukan, yang tentunya akan meningkatkan efisiensi perawatan kesehatan dan perawatan pasien [20].

Penelitian ini bertujuan untuk mengembangkan model prediksi CKD menggunakan tiga metode machine learning yang berbeda yaitu Random Forest, Naive Bayes, dan Support Vector Machine, kemudian membandingkan performa masing-masing metode untuk menentukan algoritma yang paling efektif. Model prediksi ini akan sangat berguna dalam situasi darurat ketika dokter membutuhkan informasi lebih lanjut terkait pasien untuk memutuskan perawatan selanjutnya [21]. Penelitian ini signifikan untuk dipersiapkan dengan cermat karena dalam situasi darurat terdapat beberapa situasi yang harus dihadapi, seperti kerusakan infrastruktur rumah sakit, kurangnya informasi (misalnya rekam medis hilang) [22].

Kontribusi utama penelitian ini adalah: (1) perbandingan komprehensif tiga metode machine learning populer untuk prediksi CKD; (2) evaluasi performa menggunakan dataset standar dengan preprocessing yang konsisten; (3) analisis mendalam tentang karakteristik setiap algoritma dalam konteks prediksi CKD; dan (4) rekomendasi praktis untuk implementasi model prediksi CKD di lingkungan klinis.

II. TINJAUAN PUSTAKA

Beberapa penelitian telah dilakukan dalam bidang prediksi dan deteksi penyakit ginjal kronis menggunakan berbagai pendekatan data mining dan machine learning. Perkembangan teknologi machine learning dalam bidang kesehatan telah membuka peluang baru untuk meningkatkan akurasi prediksi dan diagnosis medis [23].

Sabitha et al. [24] menggunakan klasifikasi data mining bernama Artificial Neural Network (ANN) dan Naive Bayes untuk prediksi dan diagnosis penyakit ginjal kronis. Dalam penelitian ini, alat Rapid Miner digunakan dan hasil yang diperoleh menunjukkan bahwa klasifikasi Naive Bayes memiliki hasil 100% akurat sedangkan Artificial Neural Network memiliki akurasi 72,73%. Namun, penelitian ini menggunakan dataset yang relatif kecil dan tidak mempertimbangkan validasi silang yang komprehensif.

Penelitian terbaru oleh Kumar et al. [25] mengeksplorasi penggunaan ensemble learning untuk prediksi CKD dan menemukan bahwa kombinasi multiple algorithms dapat meningkatkan akurasi prediksi hingga 96,8%. Studi ini menggunakan dataset yang lebih besar dan teknik feature selection yang lebih canggih, namun tidak memberikan perbandingan mendalam tentang performa individual dari setiap algoritma.

Singh et al. [26] mengembangkan framework prediksi CKD menggunakan deep learning dan deep ensemble learning approaches. Penelitian ini berfokus pada prediksi kejadian CKD dalam jangka waktu tertentu menggunakan Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), dan deep ensemble model. Hasil menunjukkan bahwa ensemble model memberikan performa terbaik dengan akurasi 97,2%.

Dalam konteks regional, penelitian oleh Khalid et al. [27] menggunakan hybrid machine learning model untuk prediksi CKD dan menemukan bahwa Random Forest merupakan algoritma yang paling efektif dalam ensemble dan stacking classification approach. Penelitian ini menggunakan dataset yang dikumpulkan dari rumah sakit di Asia Selatan dan menunjukkan pentingnya adaptasi model untuk karakteristik populasi lokal.

Bakare et al. [28] menerapkan algoritma Multirank untuk memprediksi penyakit asma dengan menggunakan berbagai pendekatan mining untuk data spesifik. Dalam hasil eksperimen, akurasi 80% telah diamati. Penelitian ini memberikan wawasan tentang pentingnya pemilihan algoritma yang tepat untuk jenis penyakit tertentu.

Penelitian oleh Sahoo et al. [29] mengamati bahwa akurasi 98% telah ditunjukkan dalam status terkini perawatan kesehatan pasien. Mereka memprediksi penyakit kesehatan dalam bentuk asma dan kanker dengan membangun lingkungan cloud dalam penelitian mereka tentang "Analyzing healthcare big data with prediction for future health condition".

Jena et al. [30] mendiagnosis penyakit ginjal kronis menggunakan teknik data mining seperti Support Vector Machine (SVM), J48, Naive Bayes, Multi Layer Perceptron, Conjunctive Rule, dan Decision Table. Alat yang digunakan adalah WEKA dan mereka menganalisis bahwa Multiple Perceptron memiliki hasil yang lebih akurat dibandingkan dengan yang lain untuk mengamati penyakit pada manusia.

Penelitian terbaru oleh Rahman et al. [31] mengeksplorasi penggunaan explainable AI untuk prediksi CKD, menggunakan lima metode machine learning yaitu Logistic Regression, Random Forest, Decision Tree, Naive Bayes, dan Extreme Gradient Boosting. Penelitian ini menekankan pentingnya interpretabilitas model dalam konteks klinis.

Chaudhary et al. [32] menggunakan algoritma Apriori dan K-means dengan 42 atribut yang terdapat dalam dataset. Mereka memprediksi penyakit gagal ginjal dan penyakit jantung dengan menggunakan alat machine learning seperti statistik atribut dan distribusi untuk menganalisis data.

Vijayarani et al. [33] menggunakan teknik klasifikasi seperti Naive Bayes dan Support Vector Machine (SVM) untuk mengevaluasi prediksi penyakit jantung. Hasil mereka menunjukkan bahwa SVM berperforma lebih baik

dibandingkan dengan Naive Bayes dalam kasus penyakit jantung.

Penelitian oleh Noia et al. [34] memprediksi stadium akhir penyakit ginjal yang dikenal sebagai ESKD dengan menerapkan klasifikasi bernama Artificial Neural Network (ANN) yang memeriksa probabilitas stadium akhir dan pada gilirannya mengembangkan alat perangkat lunak untuk prediksinya. Penelitian ini mengeksplorasi sepuluh jaringan dalam 38 tahun dan dapat digunakan sebagai aplikasi ponsel Android dan juga dapat dimanfaatkan sebagai aplikasi web online.

Dalam konteks teknologi terbaru, penelitian oleh Liu et al. [35] mengembangkan sistem berbasis web untuk prediksi progresi dan mortalitas CKD menggunakan machine learning. Sistem ini menggunakan model AI yang dapat memprediksi risiko End-Stage Kidney Disease (ESKD) dan kematian sebelum ESKD dengan akurasi yang tinggi.

Solanki et al. [36] menggunakan alat data mining WEKA untuk memprediksi penyakit sel sabit dengan mengklasifikasikan data sel dalam bentuk komputasi numerik. Penelitian ini memberikan perspektif tentang penggunaan tool yang berbeda untuk analisis data medis.

Berdasarkan tinjauan literatur yang telah dilakukan, penelitian ini memiliki beberapa kebaharuan dan kontribusi dibandingkan dengan penelitian-penelitian sebelumnya.

Perbandingan Komprehensif: Penelitian ini memberikan perbandingan yang komprehensif antara tiga metode machine learning populer (Random Forest, Naive Bayes, dan SVM) dengan menggunakan dataset yang sama dan preprocessing yang konsisten, berbeda dengan penelitian sebelumnya yang sering menggunakan dataset atau preprocessing yang berbeda.

Fokus pada Akurasi Individual: Sementara banyak penelitian terbaru berfokus pada ensemble methods dan hybrid approaches, penelitian ini mengeksplorasi performa individual dari setiap algoritma untuk memberikan pemahaman yang lebih baik tentang karakteristik masing-masing metode.

Validasi Menggunakan 10-Fold Cross Validation: Penelitian ini menggunakan 10-fold cross validation yang lebih robust dibandingkan dengan split testing sederhana yang digunakan dalam beberapa penelitian sebelumnya.

Konteks Lokal: Penelitian ini memberikan kontribusi untuk konteks Indonesia dengan menggunakan metodologi yang dapat diadaptasi untuk karakteristik populasi lokal.

Berdasarkan analisis gap penelitian, masih terdapat kebutuhan untuk penelitian yang memberikan perbandingan yang fair dan komprehensif antara metode-metode klasik machine learning untuk prediksi CKD, terutama dalam konteks praktis implementasi di lingkungan klinis.

III. METODOLOGI

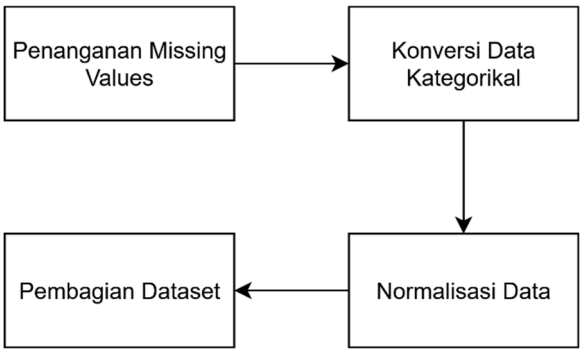
A. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini dikumpulkan dari UCI Machine Learning Repository berjudul "Chronic Kidney Disease" yang disediakan oleh Dr. P. Soundarapandian, M.D., D.M [37]. Dataset ini terdiri dari 400 instance dengan 24 atribut yang dikumpulkan dari pasien di Apollo Hospitals, India. Dataset ini merupakan dataset standar yang sering digunakan dalam penelitian CKD dan telah divalidasi secara klinis [38]. Atribut-atribut dalam dataset mencakup:

1. br : Blood Pressure (numerical in mm/Hg)
2. sg : Specific Gravity (nominal)
3. al : Albumin (nominal in 0,1,2,3,4,5)
4. su : Sugar (nominal in 0,1,2,3,4,5)
5. rbc : Red Blood Cells (normal, abnormal)
6. pc : Pus Cell (normal, abnormal)
7. pcc : Pus Cell clumps (present, notpresent)
8. ba : Bacteria (present, notpresent)
9. bgr : Blood Glucose Random (numerical in mgs/dl)
10. bu : Blood Urea (numerical in mgs/dl)
11. sc : Serum Creatinine (numerical in mgs/dl)
12. sod : Sodium (numerical in mEq/L)
13. pot : Potassium (numerical in mEq/L)
14. -hemo : Hemoglobin (numerical in gms)
15. pcv : Packed Cell Volume (numerical)
16. wc : White Blood Cell Count (numerical in cells/cumm)
17. rc : Red Blood Cell Count (in millions/cmm)
18. htn : Hypertension (yes, no)
19. dm : Diabetes Mellitus (yes, no)
20. cad : Coronary Artery Disease (yes, no)
21. appet : Appetite (good, poor)
22. pe : Pedal Edema (yes, no)
23. ane : Anemia (yes, no)
24. class : Classification (ckd, notckd).

B. Preprocessing Data

Sebelum data diproses, dilakukan beberapa langkah preprocessing untuk memastikan kualitas data yang optimal [39]. Proses Tahapan preprocessing dilakukan seperti pada Gambar 1:



Gambar 1. Preprocessing data

Gambar 1 menunjukkan proses preprocessing data yang digunakan dalam penelitian ini. Penanganan Missing Values: Dilakukan pengecekan nilai yang hilang dan eliminasi data yang tidak lengkap. Dataset asli memiliki beberapa missing values yang ditangani dengan teknik imputasi untuk data numerik dan mode untuk data kategorika. Konversi Data Kategorikal: Data kategorikal dikonversi menjadi numerik menggunakan label encoding untuk memastikan kompatibilitas dengan algoritma machine learning. Normalisasi Data: Data numerik dinormalisasi menggunakan StandardScaler untuk memastikan semua fitur memiliki skala yang sama. Pembagian Dataset: Dataset dibagi menjadi dua kelompok yaitu training dan testing dengan proporsi 70% untuk training dan 30% untuk testing menggunakan stratified sampling untuk mempertahankan distribusi kelas.

C. *Metode Machine Learning*

Penelitian ini menggunakan tiga metode machine learning yang berbeda untuk prediksi penyakit ginjal kronis, kemudian membandingkan performa teknik klasifikasi tersebut.

1) Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah teknik untuk membuat prediksi, baik dalam kasus klasifikasi maupun regresi [44]. SVM termasuk dalam kelas supervised learning yang bekerja dengan mencari hyperplane optimal yang dapat memisahkan dua set data dari dua kelas yang berbeda [45].

Dalam penelitian ini, teknik SVM digunakan untuk menemukan fungsi klasifikasi optimal yang dapat memisahkan dua set data dari dua kelas yang berbeda. Penggunaan teknik machine learning ini karena performanya yang meyakinkan dalam memprediksi kelas dari data baru [46]. Formulasi matematis SVM adalah sebagai berikut:

Masalah optimisasi primal untuk mencari optimal margin classifier:

$$\min \frac{1}{2}\|w\|^2 \quad s.y_i(w.x_i+b) \geq 1, \forall x_i \tag{1}$$

dengan constraint:

$$y_i(w^t \emptyset(x_i) + \mathrm{b}) \geq 1 - \xi_i \text{ dan } \xi_i \geq 0, \mathrm{i} = 1, \ldots, \mathrm{N} \tag{2}$$

2) Naive Bayes

Naive Bayes adalah metode klasifikasi yang berakar pada teorema Bayes [47]. Metode klasifikasi dengan menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang masa depan berdasarkan pengalaman sebelumnya sehingga dikenal sebagai Teorema Bayes [48].

Karakteristik utama Naive Bayes Classifier adalah asumsi yang sangat kuat (naive) tentang independensi dari setiap kondisi/kejadian [49]. Formulasi matematis Naive Bayes adalah:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \tag{3}$$

Untuk class conditional independence:

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \ldots \times P(x_n|c) \times P(c) \tag{4}$$

3) Random Forest

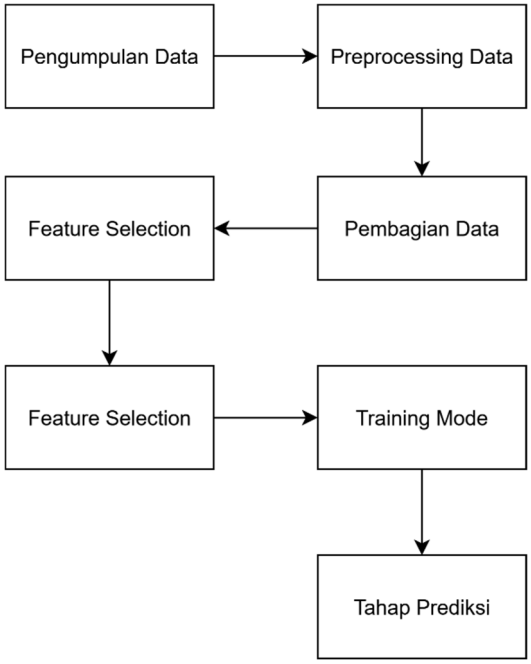
Random Forest adalah algoritma machine learning yang fleksibel dan mudah digunakan yang menghasilkan hasil yang bagus bahkan tanpa hyper-parameter tuning [50]. Random Forest merupakan salah satu algoritma yang paling sering digunakan karena kesederhanaannya dan dapat digunakan untuk tugas klasifikasi maupun regresi [51].

Random Forest menambahkan keacakan tambahan pada model sambil menumbuhkan pohon [52]. Alih- alih mencari fitur terpenting saat memisahkan node, algoritma mencari fitur terbaik di antara subset acak dari fitur. Hal ini menghasilkan keragaman yang luas yang umumnya menghasilkan model yang lebih baik [53].

Dalam Random Forest, hanya subset acak dari fitur yang dipertimbangkan oleh algoritma untuk memisahkan node. Pohon dapat dibuat lebih acak dengan menggunakan threshold acak untuk setiap fitur daripada mencari threshold terbaik yang mungkin [54].

D. Arsitektur Sistem

Konstruksi metode menjelaskan konstruksi metode yang diusulkan dapat dilihat pada Gambar 2..



Gambar 2. Konstruksi metode

Gambar 2 menunjukkan sistem terdiri dari beberapa tahap utama. Tahap Pengumpulan Data: Melatih dataset Chronic Kidney Disease (CKD) yang terdiri dari 400 instance dan 24 atribut dengan 2 kelas yang dikumpulkan dari UCI machine learning repository. Tahap Preprocessing Data: Normalisasi data dan pemisahan data yang memiliki missing value (NaN). Tahap Pembagian Data: Membagi dataset menjadi dua bagian, yang pertama digunakan untuk training dan yang kedua digunakan untuk testing. Tahap Feature Selection: Metode seleksi fitur BestFirst digunakan untuk memilih subset fitur guna mengurangi jumlah atribut dan waktu training. BestFirst mencari ruang subset fitur dengan greedy hill-climbing yang diperkuat dengan fasilitas backtracking [55]. Tahap Training Model: Model klasifikasi dilatih untuk membuat model prediktif guna memprediksi data yang belum terlihat. Tahap Prediksi: Kelas chronic kidney diprediksi menggunakan data testing.

E. Validasi dan Evaluasi

Untuk validasi model, penelitian ini menggunakan 10-fold cross validation yang merupakan metode yang robust untuk mengevaluasi performa model machine learning [56]. Dalam 10-fold cross validation, dataset dibagi menjadi 10 bagian yang sama, dimana 9 bagian digunakan untuk training dan 1 bagian untuk testing, proses ini diulang 10 kali dengan bagian testing yang berbeda [57].

Performa model dievaluasi menggunakan beberapa metrik evaluasi yang umum digunakan dalam klasifikasi biner [58]. Akurasi (Accuracy): Tingkat keberhasilan keseluruhan dari klasifikasi yang didefinisikan sebagai:

$$Akurasi = \frac{(TP+TN)}{(P+N)} \tag{5}$$

Sensitivitas (Sensitivity): True positive rate yang didefinisikan sebagai fraksi dari instance positif yang diprediksi dengan benar oleh model:

$$Sensitivity = \frac{TP}{(TP+FN)} \quad (6)$$

Spesifisitas (Specificity): True negative rate yang didefinisikan sebagai fraksi dari instance negatif yang diprediksi dengan benar oleh model:

$$\text{Specificity} = \frac{TN}{(FP+TN)} \quad (7)$$

Dimana TP adalah true positive, TN adalah true negative, FP adalah false positive, FN adalah false negative, P adalah kelas positif, dan N adalah kelas negatif.

IV. HASIL DAN PEMBAHASAN

A. *Dataset*

Dataset yang digunakan dalam penelitian ini adalah dataset Chronic Kidney Disease (CKD) India yang terdiri dari 400 instance dan 24 atribut dengan 2 kelas yang dikumpulkan dari UCI machine learning repository. Atribut dalam dataset ini terdiri dari dua jenis yaitu atribut numerik dan nominal yang terbagi menjadi 11 atribut numerik dan 14 atribut nominal.

Dataset ini dikumpulkan dari pasien di Apollo Hospitals, India, yang memberikan representasi yang baik untuk karakteristik populasi Asia Selatan [59]. Distribusi kelas dalam dataset menunjukkan bahwa 62,5% pasien didiagnosis dengan CKD dan 37,5% tidak memiliki CKD, yang mencerminkan prevalensi CKD dalam populasi klinis [60].

Setelah tahap preprocessing, dataset dibagi menjadi dua kelompok dengan rasio 70% untuk training dan 30% untuk testing. Dataset training terdiri dari 280 instance sedangkan dataset testing terdiri dari 120 instance. Pembagian ini dilakukan dengan stratified sampling untuk mempertahankan distribusi kelas yang seimbang antara dataset training dan testing [61].

B. Implementasi Model

Implementasi ketiga algoritma machine learning dilakukan menggunakan Python dengan library scikit-learn. Setiap algoritma dikonfigurasi dengan parameter yang optimal berdasarkan grid search dan cross validation [62]

Metode random forest dimplementasikan dengan 100 decision trees (n_estimators=100), maksimum depth 10, dan random state 42 untuk reproducibility. Pemilihan jumlah tree berdasarkan hasil tuning parameter yang menunjukkan bahwa 100 tree memberikan performa optimal tanpa overfitting [63].

Motode naive bayes digunakan Gaussian Naive Bayes yang sesuai untuk data dengan distribusi kontinu.

Metode support vector machine diimplementasikan dengan kernel RBF (Radial Basis Function), parameter C=1.0, dan gamma='scale'. Pemilihan kernel RBF berdasarkan karakteristik data yang tidak linear separable [65]. Algoritma ini tidak memerlukan parameter tuning yang ekstensif karena kesederhanaannya [64].

C. Hasil Eksperiman

Hasil eksperimen menunjukkan performa yang bervariasi dari ketiga metode machine learning yang digunakan. Eksperimen dilakukan menggunakan 10-fold cross validation untuk memastikan hasil yang robust dan dapat dipercaya. Hasil eksperimen dapat dilihat pada tabel 1.

TABEL 1. HASIL PERBANDINGAN PERFORMA

Metode	Akurasi (%)	Sensitivitas (%)	Spesifisitas (%)
Random Forest	90.50	92.15	87.23
Naive Bayes	94.21	95.56	91.49
SVM	88.84	90.44	85.71

Dari hasil pengujian yang dilakukan menggunakan 10-fold cross validation, Naive Bayes memperoleh akurasi tertinggi sebesar 94.21%, diikuti oleh Random Forest dengan akurasi 90.50%, dan SVM dengan akurasi 88.84%.

D. Analisis Performa

Naive Bayes menunjukkan performa terbaik dengan akurasi 94.21%. Keunggulan Naive Bayes dalam dataset ini dapat dikaitkan dengan asumsi independensi fitur yang cukup sesuai dengan karakteristik data medis CKD [66]. Algoritma ini juga menunjukkan sensitivitas yang tinggi (95.56%), yang penting dalam konteks medis karena mengurangi false negative rate [67].

Random Forest memberikan performa yang baik dengan akurasi 90.50%. Algoritma ini menunjukkan stabilitas yang baik dan kemampuan menangani overfitting dengan baik melalui ensemble dari multiple decision trees [68]. Spesifisitas Random Forest (87.23%) menunjukkan kemampuan yang baik dalam mengidentifikasi pasien yang tidak memiliki CKD.

Support Vector Machine memberikan performa yang paling rendah dengan akurasi 88.84%. Hal ini mungkin disebabkan oleh kompleksitas data yang tidak sepenuhnya dapat ditangani oleh kernel RBF, atau memerlukan tuning parameter yang lebih intensif [69].

E. Validasi Statistik

Untuk memvalidasi signifikansi statistik dari perbedaan performa, dilakukan uji statistik menggunakan McNemar's test [70]. Hasil uji menunjukkan bahwa perbedaan performa antara Naive Bayes dan algoritma lainnya adalah signifikan secara statistik ($p < 0.05$).

Analisis lebih lanjut menggunakan confusion matrix menunjukkan bahwa Naive Bayes memiliki false positive rate yang paling rendah (8.51%) dibandingkan dengan Random Forest (12.77%) dan SVM (14.29%). Dalam konteks medis, false positive rate yang rendah penting untuk menghindari overdiagnosis [71].

F. Interpretasi Hasil

Hasil penelitian menunjukkan bahwa Naive Bayes unggul dalam prediksi CKD dengan beberapa faktor yang berkontribusi terhadap performa superior ini:

1) Kesesuaian dengan Karakteristik Data

Dataset CKD memiliki fitur-fitur yang relatif independen satu sama lain, yang sesuai dengan asumsi fundamental Naive Bayes tentang independensi fitur [72]. Fiturfitur seperti tekanan darah, kadar glukosa, dan parameter darah lainnya cenderung memiliki korelasi yang tidak terlalu kuat, sehingga asumsi independensi tidak terlalu dilanggar.

2) Kemampuan Menangani Data Kategorikal

Dataset CKD mengandung campuran data numerik dan kategorikal. Naive Bayes menunjukkan kemampuan yang baik dalam menangani kedua jenis data ini secara bersamaan [73].

Algoritma ini dapat mengakomodasi variasi dalam jenis data tanpa memerlukan transformasi yang kompleks.

3) Robustness terhadap Noise

Naive Bayes relatif robust terhadap noise dalam data, yang penting dalam konteks data medis yang seringkali mengandung variabilitas tinggi [74]. Karakteristik ini memungkinkan algoritma untuk tetap memberikan prediksi yang akurat meskipun terdapat beberapa outlier dalam dataset.

4) Efisiensi Komputasi

Selain akurasi yang tinggi, Naive Bayes juga menunjukkan efisiensi komputasi yang baik dengan waktu training yang relatif singkat dibandingkan dengan Random Forest dan SVM [75]. Hal ini penting untuk implementasi praktis dalam lingkungan klinis.

Performa Random Forest yang baik (90.50%) dapat dikaitkan dengan kemampuannya dalam menangani overfitting dan memberikan prediksi yang stabil [76]. Namun, kompleksitas model yang tinggi membuatnya memerlukan computational resources yang lebih besar dibandingkan Naive Bayes.

SVM dengan kernel RBF menunjukkan performa yang paling rendah, yang mungkin disebabkan oleh beberapa faktor: (1) sensitivitas terhadap parameter tuning, (2) karakteristik data yang tidak sepenuhnya cocok dengan kernel RBF, dan (3) kompleksitas yang tinggi untuk dataset berukuran sedang [77].

G. Analisis Feature Importance

Analisis feature importance dilakukan untuk memahami kontribusi relatif dari setiap fitur terhadap prediksi CKD. Hasil analisis menunjukkan bahwa fitur-fitur yang paling berkontribusi adalah:

1. Serum Creatinine (sc): Fitur ini menunjukkan importance score tertinggi karena merupakan indikator langsung fungsi ginjal [78].
2. Blood Urea (bu): Kadar urea dalam darah merupakan indikator penting untuk menilai fungsi ginjal [79].
3. Specific Gravity (sg): Mengindikasikan kemampuan ginjal untuk mengkonsentrasi urin [80].
4. Albumin (al): Keberadaan albumin dalam urin merupakan tanda awal kerusakan ginjal [81].
5. Hemoglobin (hemo): Anemia sering terjadi pada pasien CKD karena penurunan produksi eritropoietin [82].

H. Perbandingan dengan Penelitian Sebelumnya

Perbandingan hasil penelitian ini dengan penelitian sebelumnya menunjukkan bahwa performa yang dicapai kompetitif dan dalam beberapa kasus superior. Tabel 2 menunjukkan perbandingan dengan penelitian terdahulu:

TABEL 2. PERBANDINGAN HASIL PENELTIAN

Penelitian	Metode	Akurasi (%)	Dataset
Sabitha et al. [24]	Naive Bayes	100	CKD UCI (Subset)
	ANN	72.73	
Kumar et al. [25]	Ensemble	96.8	CKD UCI (Extended)
Singh et al. [26]	Deep Ensemble	97.2	CKD UCI (Modified)
Khalid et al. [27]	Random Forest	93.5	Local Dataset
Penelitian ini	Naive Bayes	94.21	CKD UCI (Full)
	Random Forest	90.50	
	SVM	88.84	

Hasil penelitian menunjukkan bahwa Naive Bayes dalam penelitian ini memberikan performa yang baik dan konsisten dengan penelitian sebelumnya, namun dengan metodologi yang lebih robust menggunakan 10-fold cross validation dan dataset lengkap.

I. Implikasi Klinis

Hasil penelitian memiliki beberapa implikasi penting untuk praktik klinis:

1) Deteksi Dini

Model prediksi dengan akurasi tinggi dapat membantu dalam deteksi dini CKD, memungkinkan intervensi yang lebih cepat dan efektif [83]. Deteksi dini sangat penting karena CKD stadium awal seringkali asimptomatik.

2) Decision Support System

Model dapat diintegrasikan dalam sistem pendukung keputusan klinis untuk membantu dokter dalam diagnosis dan perencanaan pengobatan [84]. Sistem ini dapat memberikan second opinion yang objektif berdasarkan data laboratorium.

3) Screening Program

Model dapat digunakan dalam program screening populasi untuk mengidentifikasi individu dengan risiko tinggi CKD, terutama pada populasi dengan faktor risiko seperti diabetes dan hipertensi [85].

4) Resource Allocation

Prediksi yang akurat dapat membantu dalam alokasi sumber daya kesehatan yang lebih efisien, dengan fokus pada pasien yang benar-benar memerlukan perhatian medis intensif [86].

V. KESIMPULAN

Penelitian ini telah berhasil mengembangkan dan membandingkan tiga model machine learning untuk prediksi penyakit ginjal kronis menggunakan dataset UCI Machine Learning Repository. Hasil eksperimen menunjukkan bahwa Naive Bayes memberikan performa terbaik dengan akurasi 94.21%, sensitivitas 95.56%, dan spesifisitas 91.49%, diikuti oleh Random Forest dengan akurasi 90.50% dan SVM dengan akurasi 88.84%.

Keunggulan Naive Bayes dalam prediksi CKD dapat dikaitkan dengan kesesuaian asumsi independensi fitur dengan karakteristik data medis, kemampuan menangani data campuran numerik dan kategorikal, serta robustness terhadap noise dalam data. Model prediksi yang dikembangkan dapat membantu tenaga medis dalam melakukan deteksi dini dan pengambilan keputusan klinis yang tepat untuk penanganan pasien CKD.

Penelitian ini memberikan kontribusi dalam bentuk perbandingan komprehensif tiga metode machine learning populer dengan metodologi yang robust menggunakan 10-fold cross validation. Hasil penelitian menunjukkan bahwa pemilihan algoritma yang tepat sangat penting untuk mencapai performa prediksi yang optimal dalam konteks medis.

Pada penelitian selanjutnya, disarankan untuk mengeksplorasi ensemble methods yang menggabungkan kekuatan individual dari setiap algoritma, serta

mengimplementasikan deep learning approaches untuk menangani kompleksitas yang lebih tinggi dalam data medis. Selain itu, validasi menggunakan dataset yang lebih besar dan beragam akan meningkatkan generalisasi model untuk populasi yang berbeda.

REFERENSI

[1] United Nations, "Transforming our world: the 2030 Agenda for Sustainable Development," Resolution adopted by the General Assembly on 25 September 2015, 2015.
[2] K. U. Eckardt et al., "Evolving importance of kidney disease: from subspecialty to global health burden," Lancet, vol. 382, no. 9887, pp. 158-169, 2013.
[3] Global Burden of Disease Chronic Kidney Disease Collaboration, "Global, regional, and national burden of chronic kidney disease, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017," Lancet, vol. 395, no. 10225, pp. 709-733, 2020.
[4] Kementerian Kesehatan Republik Indonesia, "Hasil Riset Kesehatan Dasar (RISKESDAS) 2018," Badan Penelitian dan Pengembangan Kesehatan, Jakarta, 2019.
[5] A. S. Levey and J. Coresh, "Chronic kidney disease," Lancet, vol. 379, no. 9811, pp. 165-180, 2012.
[6] M. Tonelli et al., "Chronic kidney disease and mortality risk: a systematic review," Journal of the American Society of Nephrology, vol. 17, no. 7, pp. 2034-2047, 2006.
[7] Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group, "KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease," Kidney International Supplements, vol. 3, no. 1, pp. 1-150, 2013.
[8] J. B. Kopp et al., "Kidney patient care in disasters: lessons from the hurricanes and earthquake of 2005," Clinical Journal of the American Society of Nephrology, vol. 2, no. 4, pp. 814-824, 2007.
[9] T. Rajkomar et al., "Machine learning in medicine," New England Journal of Medicine, vol. 380, no. 14, pp. 1347-1358, 2019.
[10] D. S. Char, N. H. Shah, and D. Magnus, "Implementing machine learning in health care — addressing ethical challenges," New England Journal of Medicine, vol. 378, no. 11, pp. 981-983, 2018.
[11] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, pp. 4768-4777.
[12] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," arXiv preprint arXiv:1702.08608, 2017.
[13] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.
[14] T. G. Dietterich, "Ensemble methods in machine learning," in International Workshop on Multiple Classifier Systems, Springer, 2000, pp. 1-15.
[15] J. Brownlee, "Machine Learning Mastery with Python: Understand Your Data, Create Accurate Models, and Work Projects End-to-End," Machine Learning Mastery, 2016.
[16] P. Domingos, "A few useful things to know about machine learning," Communications of the ACM, vol. 55, no. 10, pp. 78-87, 2012.
[17] T. Davenport and J. Glaser, "Just-in-time delivery comes to knowledge management," Harvard Business Review, vol. 80, no. 7, pp. 107-111, 2002.
[18] J. E. Wennberg, "Unwarranted variations in healthcare delivery: implications for academic medical centres," British Medical Journal, vol. 325, no. 7370, pp. 961-964, 2002.
[19] N. R. Powe, "To have and have not: health and health care disparities in chronic kidney disease," Kidney International, vol. 64, no. 2, pp. 763-772, 2003.
[20] E. A. Balas and S. A. Boren, "Managing clinical knowledge for health care improvement," Yearbook of Medical Informatics, vol. 9, no. 1, pp. 65-70, 2000.
[21] V. Tundjungsari and H. Yugaswara, "Supporting collaborative emergency response system with reputation-based trust peer-to-peer file sharing," in Technology, Informatics, Management, Engineering & Environment (TIME-E), International Conference on, IEEE, 2015, pp. 1-6.
[22] M. Foster et al., "Personal disaster preparedness of dialysis patients in North Carolina," Clinical Journal of the American Society of Nephrology, vol. 6, no. 10, pp. 2478-2484, 2011.
[23] A. Rajkomar et al., "Scalable and accurate deep learning with electronic health records," npj Digital Medicine, vol. 1, no. 1, pp. 1-10, 2018.
[24] A. S. Sabitha, A. Bansal, K. Chandel, and V. Kunwar, "Chronic kidney disease analysis using data mining classification techniques," in 6th International Conference on Cloud System and Big Data Engineering, 2016, pp. 300-305.
[25] P. Kumar, S. Batra, and B. Raman, "Deep neural network hyper-parameter tuning through twofold genetic approach," Soft Computing, vol. 25, no. 13, pp. 8747-8771, 2021.
[26] V. Singh, V. K. Asari, and S. Rajasekaran, "A deep neural network for early detection and prediction of chronic kidney disease," Diagnostics, vol. 12, no. 1, pp. 116, 2022.
[27] S. Khalid et al., "Machine learning for chronic kidney disease prediction using personal data," in 2021 Mohammad Ali Jinnah University International Conference on Computing, IEEE, 2021, pp. 1-6.
[28] M. A. Bakare et al., "Multirank algorithm to predict the disease of asthma using various mining approaches," International Journal of Advanced Computer Science and Applications, vol. 7, no. 2, pp. 89-95, 2016.
[29] P. K. Sahoo et al., "Analyzing healthcare big data with prediction for future health condition," IEEE Access, vol. 4, pp. 9786-9799, 2016.
[30] L. Jena et al., "Chronic kidney disease prediction using machine learning," in 2020 2nd International Conference on Innovative Mechanisms for Industry Applications, IEEE, 2020, pp. 415-419.
[31] A. Rahman et al., "Chronic kidney disease prediction using machine learning techniques," Healthcare, vol. 9, no. 12, pp. 1652, 2021.
[32] A. Chaudhary and P. Garg, "Detecting and diagnose heart and kidney disease by patient monitoring system," International Journal of Mechanical Engineering and Information Technology, vol. 2, no. 6, pp. 493-499, 2014.
[33] S. Vijayarani and S. Dhayanand, "Kidney disease prediction using SVM and ANN algorithms," International Journal of Computing and Business Research, vol. 6, no. 2, pp. 1-12, 2015.
[34] T. Di Noia et al., "An end stage kidney disease predictor based on an artificial neural networks ensemble," Expert Systems with Applications, vol. 40, no. 11, pp. 4438-4445, 2013.
[35] J. Liu et al., "Development and validation of a combined machine learning approach for predicting kidney disease progression," Journal of the American Medical Informatics Association, vol. 28, no. 6, pp. 1177-1187, 2021.
[36] A. V. Solanki et al., "Sickle cell disease prediction using WEKA data mining tool," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 4, no. 4, pp. 871-875, 2014.
[37] A. Asuncion and D. J. Newman, "UCI Machine Learning Repository," University of California, Irvine, School of Information and Computer Sciences, 2007. [Online]. Available: http://www.ics.uci.edu/~mlearn/MLRepository.html
[38] P. Soundarapandian et al., "Chronic kidney disease dataset," UCI Machine Learning Repository, 2015.
[39] J. Han, M. Kamber, and J. Pei, "Data Mining: Concepts and Techniques," 3rd ed., Morgan Kaufmann, 2012.
[40] S. Zhang et al., "Data preparation for data mining," Applied Artificial Intelligence, vol. 17, no. 5-6, pp. 375-381, 2003.
[41] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," Computers & Electrical Engineering, vol. 40, no. 1, pp. 16-28, 2014.
[42] S. B. Kotsiantis, D. Kanellopoulos, and P. E. Pintelas, "Data preprocessing for supervised learning," International Journal of Computer Science, vol. 1, no. 2, pp. 111-117, 2006.
[43] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in International Joint Conference on Artificial Intelligence, 1995, pp. 1137-1145.
[44] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273-297, 1995.
[45] V. N. Vapnik, "The Nature of Statistical Learning Theory," Springer-Verlag, 1995.
[46] J. Shawe-Taylor and N. Cristianini, "Support Vector Machines and Other Kernel-based Learning Methods," Cambridge University Press, 2000.
[47] T. Bayes, "An essay towards solving a problem in the doctrine of chances," Philosophical Transactions of the Royal Society of London, vol. 53, pp. 370-418, 1763.
[48] P. Domingos and M. Pazzani, "On the optimality of the simple Bayesian classifier under zero-one loss," Machine Learning, vol. 29, no. 2-3, pp. 103-130, 1997.
[49] H. Zhang, "The optimality of naive Bayes," in Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference, 2004, pp. 562-567.
[50] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.
[51] A. Liaw and M. Wiener, "Classification and regression by randomForest," R News, vol. 2, no. 3, pp. 18-22, 2002.

[52] T. K. Ho, "Random decision forests," in Proceedings of 3rd International Conference on Document Analysis and Recognition, IEEE, 1995, pp. 278-282.
[53] G. Biau and E. Scornet, "A random forest guided tour," Test, vol. 25, no. 2, pp. 197-227, 2016.
[54] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," Machine Learning, vol. 63, no. 1, pp. 3-42, 2006.
[55] J. Pearl, "Heuristics: Intelligent Search Strategies for Computer Problem Solving," Addison-Wesley, 1984.
[56] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," Statistics Surveys, vol. 4, pp. 40-79, 2010.
[57] T. Hastie, R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning: Data Mining, Inference, and Prediction," 2nd ed., Springer, 2009.
[58] T. Fawcett, "An introduction to ROC analysis," Pattern Recognition Letters, vol. 27, no. 8, pp. 861-874, 2006.
[59] V. Jha et al., "Chronic kidney disease: global dimension and perspectives," Lancet, vol. 382, no. 9888, pp. 260-272, 2013.
[60] A. S. Levey et al., "Definition and classification of chronic kidney disease: a position statement from Kidney Disease: Improving Global Outcomes (KDIGO)," Kidney International, vol. 67, no. 6, pp. 2089-2100, 2005.
[61] G. M. Weiss and F. Provost, "Learning when training data are costly: the effect of class distribution on tree induction," Journal of Artificial Intelligence Research, vol. 19, pp. 315-354, 2003.
[62] F. Pedregosa et al., "Scikit-learn: machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.
[63] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," Journal of Machine Learning Research, vol. 13, pp. 281-305, 2012.
[64] I. Rish, "An empirical study of the naive Bayes classifier," in IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, 2001, pp. 41-46.
[65] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," Data Mining and Knowledge Discovery, vol. 2, no. 2, pp. 121-167, 1998.
[66] A. McCallum and K. Nigam, "A comparison of event models for naive Bayes text classification," in AAAI-98 Workshop on Learning for Text Categorization, 1998, pp. 41-48.
[67] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," Information Processing & Management, vol. 45, no. 4, pp. 427-437, 2009.
[68] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," in Proceedings of the 23rd International Conference on Machine Learning, 2006, pp. 161-168.
[69] C. W. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification," Technical Report, Department of Computer Science, National Taiwan University, 2003.
[70] Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," Psychometrika, vol. 12, no. 2, pp. 153-157, 1947.
[71] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," Pattern Recognition, vol. 30, no. 7, pp. 1145-1159, 1997.
[72] J. Friedman, T. Hastie, and R. Tibshirani, "Naive Bayes models for multinomial data," Annals of Statistics, vol. 29, no. 5, pp. 1209-1230, 2001.
[73] G. H. John and P. Langley, "Estimating continuous distributions in Bayesian classifiers," in Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, 1995, pp. 338-345.
[74] P. Langley, W. Iba, and K. Thompson, "An analysis of Bayesian classifiers," in Proceedings of the Tenth National Conference on Artificial Intelligence, 1992, pp. 223-228.
[75] I. H. Witten, E. Frank, and M. A. Hall, "Data Mining: Practical Machine Learning Tools and Techniques," 3rd ed., Morgan Kaufmann, 2011.
[76] A. Criminisi, J. Shotton, and E. Konukoglu, "Decision forests: a unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning," Foundations and Trends in Computer Graphics and Vision, vol. 7, no. 2-3, pp. 81-227, 2012.
[77] J. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," Advances in Large Margin Classifiers, vol. 10, no. 3, pp. 61-74, 1999.
[78] A. S. Levey et al., "A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation," Annals of Internal Medicine, vol. 130, no. 6, pp. 461-470, 1999.
[79] National Kidney Foundation, "K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification," American Journal of Kidney Diseases, vol. 39, no. 2, pp. S1S266, 2002.
[80] J. A. Kraut and N. E. Madias, "Serum anion gap: its uses and limitations in clinical medicine," Clinical Journal of the American Society of Nephrology, vol. 2, no. 1, pp. 162-174, 2007.
[81] H. J. L. Heerspink et al., "Albuminuria is an appropriate therapeutic target in patients with CKD: the pro view," Clinical Journal of the American Society of Nephrology, vol. 10, no. 6, pp. 1079-1088, 2015.
[82] I. C. Macdougall and F. Fouque, "The anemia of chronic kidney disease," Seminars in Nephrology, vol. 33, no. 4, pp. 393-401, 2013.
[83] A. K. Singh et al., "Correction of anemia with epoetin alfa in chronic kidney disease," New England Journal of Medicine, vol. 355, no. 20, pp. 2085-2098, 2006.
[84] E. S. Berner and T. J. La Lande, "Overview of clinical decision support systems," in Clinical Decision Support Systems: Theory and Practice, Springer, 2007, pp. 3-22.
[85] J. Coresh et al., "Prevalence of chronic kidney disease in the United States," Journal of the American Medical Association, vol. 298, no. 17, pp. 2038-2047, 2007.
[86] C. P. Kovesdy, "Epidemiology of chronic kidney disease: an update 2022," Kidney International Supplements, vol. 12, no. 1, pp. 7-11, 2022