Proceeding of International Conference on Advanced Technologies, Innovations and Science
Vol 1, No 1, June 2026, Page 510-518
ISSN XXXX-XXXX (Media Online)
https://journals.adaresearch.or.id/icatis

Classification Analysis of Chronic Kidney Disease Using the Naive Bayes Algorithm to Support Early Detection

Dian Nita*, Isma Aulia Muharni, Yuanda Stefanny

Information Systems Study Program, Sekolah Tinggi Ilmu Komputer Tunas Bangsa, Pematangsiantar, Indonesia
Jl Kartini, Proklamasi, Siantar Barat, 21143, Pematang Siantar, North Sumatera, Indonesia
Email: [1,*]diannitaa6@gmail.com, [2]ismaauliiiaaa17@gmail.com, [3]yuandasatu1@gmail.com
Correspondence Author Email: diannitaa6@gmail.com
Submitted: **01/06/2026**; Accepted: **12/06/2026**; Published: **30/06/2026**

Abstract−Chronic Kidney Disease (CKD) is one of the non-communicable diseases with an increasing prevalence and has become a serious problem in the healthcare sector due to delayed diagnosis and lack of early detection. This study aims to analyze the performance of the Naive Bayes algorithm in classifying chronic kidney disease to support early detection systems. The research used a Chronic Kidney Disease dataset consisting of 400 patient records with 26 medical attributes, including blood pressure, albumin, hemoglobin, blood glucose, and other health indicators. Data processing was carried out using Python with several stages, including data cleaning, missing value checking, categorical data encoding using LabelEncoder, and data normalization using StandardScaler. The dataset was divided into training data and testing data with a ratio of 80:20, resulting in 320 training data and 80 testing data. The classification process was performed using the Naive Bayes algorithm and evaluated using accuracy, precision, recall, f1-score, and confusion matrix. The experimental results showed that the proposed model achieved an accuracy of 97.5%, indicating that the Naive Bayes algorithm is capable of classifying chronic kidney disease effectively. The findings of this study demonstrate that machine learning techniques can support early detection systems for chronic kidney disease and assist healthcare services in improving diagnostic accuracy.

Keywords: Chronic Kidney Disease; Naive Bayes; Machine Learning; Classification; Early Detection

1. INTRODUCTION

Chronic kidney disease (CKD) is a noncommunicable disease that is a major public health concern due to its steadily rising prevalence each year. This condition results from a gradual decline in kidney function over a long period of time, causing the kidneys to be unable to filter blood effectively(Indrianti et al., 2024; Qin et al., 2020; Roy et al., 2021; Wantoro et al., 2026). This condition causes metabolic waste products to accumulate in the body and can lead to various serious complications, such as hypertension, anemia, kidney failure, cardiovascular disorders, and even death (Prasetyo et al., 2024; Ramadhan, Wahyu, et al., 2025; Suwari & Hidayati, 2025; Zuhri et al., 2025). According to the World Health Organization, chronic kidney disease is one of the causes of rising global mortality rates due to delayed diagnosis and low public awareness of the early symptoms of kidney disease (Organization, 2026)(Wang, 2020).

Advances in information technology and artificial intelligence are having a major impact on the healthcare sector, particularly in the processing of medical data. One application of these technologies is the use of machine learning techniques in disease classification. Machine learning enables computer systems to learn patterns from historical data to generate predictions or automated decisions (Adittia Fathah, 2025; Apip & Kurniawan, 2026; Made et al., 2026; Rahayu et al., 2024). In the field of healthcare, machine learning-based classification methods are considered capable of improving diagnostic efficiency, accelerating the analysis of patient data, and helping medical professionals make more accurate decisions (Prasetyo et al., 2024)(Ratnasari et al., 2025). One of the classification algorithms widely used in health research is the Naive Bayes algorithm. This algorithm is based on Bayes' theorem and assumes independence among attributes, resulting in a simple computational process that still delivers good classification performance (Apip & Kurniawan, 2026)(Made et al., 2026).

This study uses a chronic kidney disease dataset consisting of 400 patient records with 26 attributes. Based on the results of data processing using Python, the dataset contains numerical and categorical attributes such as age, blood pressure, specific gravity, albumin, random blood glucose, hemoglobin, hypertension, diabetes mellitus, and other medical attributes. The preprocessing stages were carried out through data cleaning, removal of the ID attribute, handling of missing values, encoding of categorical data using LabelEncoder, and data normalization using StandardScaler. After the preprocessing was complete, the dataset was divided into a training set of 320 records and a testing set of 80 records. A classification model was then built using the Naive Bayes algorithm, yielding an accuracy of 97.5% with a confusion matrix of [[50,0,1],[1,0,0],[0,0,28]]. These results indicate that the Naive Bayes algorithm demonstrates excellent classification performance in detecting chronic kidney disease.

Several previous studies have examined the application of machine learning algorithms for the classification of chronic kidney disease. A study by (Indrianti et al., 2024) used machine learning methods to diagnose chronic kidney disease and found that classification algorithms can improve diagnostic accuracy compared to manual methods. However, that study did not perform a complete data preprocessing step, particularly regarding data normalization. Another study conducted by (Adittia Fathah, 2025) compared several classification algorithms, such as Random Forest, Support Vector Machine, and Naive Bayes, for the detection of chronic kidney disease. The results of the study showed that Random Forest has a high accuracy rate but requires greater computational complexity than Naive Bayes. In addition, research by (Simbolon et al., 2025) applied a classification algorithm to patients with diabetes and chronic kidney disease and

demonstrated that machine learning techniques can improve the process of early disease detection. However, the study did not make extensive use of model evaluation visualization using confusion matrices.

Another study was conducted by (Ramadhan, Wahyu, et al., 2025) which compared the J48 Decision Tree, CART, and Naive Bayes algorithms for classifying chronic kidney disease. The study showed that the Naive Bayes algorithm has a faster training process than the other algorithms, but did not implement data normalization using StandardScaler. Research by (Widiati et al., 2024) It also implemented the Naive Bayes and K-Nearest Neighbor algorithms for classifying chronic kidney disease and achieved fairly good classification performance (Riany et al., 2023; Riany & Testiana, 2023). However, the study has not yet conducted a comprehensive data preprocessing phase and has not provided a detailed explanation of the model evaluation results using a confusion matrix and classification report.

Based on previous studies, it is evident that most studies focus solely on accuracy results without paying attention to optimal data preprocessing steps. In addition, some studies still have limitations in terms of model evaluation visualization and in-depth analysis of classification performance (Ani et al., 2025; Fauzi & Santoso, 2025; Sari & Supriatna, 2023). Therefore, this study differs from previous research in that it implements a comprehensive set of preprocessing steps, ranging from data cleaning, handling missing values, and encoding categorical data to data normalization using StandardScaler, and model evaluation using classification reports and confusion matrices (Setiawan et al., 2025). With this approach, the resulting classification model is more optimal and easier to analyze.

The main challenge in diagnosing chronic kidney disease is delayed detection due to early symptoms that patients often fail to recognize. Many patients only become aware of their condition when their kidney function has already suffered significant damage, making treatment more difficult and increasing healthcare costs. This situation underscores the importance of a diagnostic support system capable of facilitating early detection automatically and accurately. The solution proposed in this study is the application of a machine learning-based Naive Bayes algorithm to classify chronic kidney disease based on patients' medical data. This system is expected to assist medical professionals in analyzing health data more quickly and efficiently.

The urgency of this research lies in the importance of developing an artificial intelligence-based disease classification system that is highly accurate yet simple to implement. As the volume of digital medical data increases, the use of classification algorithms has become crucial for aiding decision-making processes in the healthcare field. This study also contributes to the development of data mining applications in the medical field, particularly in the classification of chronic kidney disease using the Naive Bayes algorithm. Furthermore, the results of this study are expected to serve as a reference for future research in the development of machine learning-based medical decision support systems.

The state-of-the-art aspect of this study lies in the integration of comprehensive data preprocessing with the Naive Bayes algorithm to improve the classification performance for chronic kidney disease. This study not only achieves an accuracy rate of 97.5% but also presents model evaluations using classification reports and confusion matrices, allowing for a more in-depth analysis of the classification results. With this approach, this study is expected to make a tangible contribution to the development of an accurate, efficient, and easily implementable artificial intelligence-based early detection system for chronic kidney disease within digital healthcare services..

2. RESEARCH METHODS

2.1. Dataset

The data processing steps in this study were carried out systematically to produce an optimal classification model for chronic kidney disease using the Naive Bayes algorithm. The dataset used is the Chronic Kidney Disease dataset, which consists of 400 patient records with 26 medical attributes. This dataset contains numerical and categorical attributes used as parameters for chronic kidney disease classification. The data processing was performed using the Python programming language with the assistance of the Pandas, NumPy, Scikit-Learn, and Matplotlib libraries.

The initial phase of the study involved reading the dataset using the Pandas library. After the dataset was successfully imported, a preliminary analysis was conducted to determine the data structure, data types, number of attributes, and condition of the dataset prior to preprocessing. Based on the analysis results, it was found that the dataset consists of 400 rows of data with 26 attributes, including numerical and categorical attributes such as age, blood pressure, albumin, random blood glucose, hemoglobin, and other medical attributes.

Table 1. Description of Attributes in The Chronic Kidney Disease Dataset

No	Attribute Name	Description
1	age	Patient's Age
2	bp	Blood Pressure
3	sg	Specific gravity
4	al	Albumin
5	su	Sugar
6	rbc	Sel darah merah
7	pc	Pus cell
8	pcc	Pus cell clumps
9	ba	Bacteria

No	Attribute Name	Description
10	bgr	Blood glucose random
11	bu	Blood urea
12	sc	Serum creatinine
13	sod	Sodium
14	pot	Potassium
15	hemo	Hemoglobin
16	pcv	Packed cell volume
17	wc	White blood cell count
18	rc	Red blood cell count
19	htn	Hypertension
20	dm	Diabetes mellitus
21	cad	Coronary artery disease
22	appet	Appetite
23	pe	Pedal edema
24	ane	Anemia
25	classification	CKD classification results

The next step is the data cleaning process to improve the quality of the dataset before the classification process begins. At this stage, the "id" attribute is removed because it has no bearing on the classification of chronic kidney disease. Next, the dataset is checked for missing values to ensure there are no empty entries. The results of this check indicate that the dataset contains no missing values, so the data can be used directly in the next step. Once the data cleaning process is complete, the next step is to encode categorical data using LabelEncoder. This process aims to convert text-based data such as "yes," "no," "good," and "poor" into numerical values so that they can be processed by machine learning algorithms. Next, data normalization is performed using StandardScaler to standardize the data scale, thereby optimizing the classification process and reducing the range differences between attributes.

Table 2. Data Prepocessing Steps

Stages	Description
Data Cleaning	Removing unnecessary attributes
Missing Value Checking	Ensure that the dataset does not contain any missing values
Encoding Data	Converting categorical data to numerical data
Normalisasi Data	Scale the data using StandardScaler
Data Splitting	Split the dataset into training and testing data

The next step is to split the dataset into training and testing data using the `train_test_split` method. The dataset is split in an 80:20 ratio, with 80% of the data used for training and 20% used for testing. The training data is used to train the Naive Bayes model, while the testing data is used to evaluate the performance of the chronic kidney disease classification model.

Table 3. Splitting The Training and Testing Datasets

Data Types	Data Volume	Percentage
Data Training	320	80%
Data Testing	80	20%
Total Dataset	400	100%

Once the dataset has been split, the next step is to train the model using the Naive Bayes algorithm. This algorithm operates based on Bayes' theorem by calculating the probability of each attribute belonging to a specific class and selecting the highest probability as the classification result. After the model has been trained using the training data, a prediction process is performed on the test data to assess the model's ability to accurately classify chronic kidney disease.

2.2. Research Framework

This study employs a quantitative research method with a machine learning-based classification approach to analyze chronic kidney disease using the Naive Bayes algorithm. The study was conducted using the Chronic Kidney Disease (CKD) dataset, which consists of 400 patient records with 26 medical attributes such as age, blood pressure, blood sugar levels, albumin, hemoglobin, diabetes mellitus, and other health attributes (M. Fahmi & Fadlillah, 2023) . Data processing was performed using the Python programming language with the assistance of the Pandas, NumPy, Scikit-Learn, and Matplotlib libraries. The objective of this study is to determine the capability of the Naive Bayes algorithm in classifying chronic kidney disease to support the early detection process with a high level of accuracy.

The research framework begins with the collection of a chronic kidney disease dataset, data preprocessing, the division of data into training and testing sets, model training using the Naive Bayes algorithm, and model evaluation using accuracy, precision, recall, F1-score, and a confusion matrix. The preprocessing stage is conducted to improve data quality before the classification process begins. These steps include removing the ID attribute, checking for missing values,

Proceeding of International Conference on Advanced Technologies, Innovations and Science
Vol 1, No 1, June 2026, Page 510-518
ISSN XXXX-XXXX (Media Online)
https://journals.adaresearch.or.id/icatis

encoding categorical data using LabelEncoder, and normalizing data using StandardScaler. Based on the results of data processing using Python, 320 training data points and 80 testing data points were obtained.

The research hypothesis states that the Naive Bayes algorithm is capable of classifying chronic kidney disease with a high level of accuracy, making it suitable for use in supporting the early detection of chronic kidney disease(H. Fahmi & Sutisna, 2024; Pulungan et al., 2024). The independent variables in this study consist of patient medical attributes such as blood pressure, albumin levels, random blood glucose levels, hemoglobin (hemo), and other health attributes. Meanwhile, the dependent variable in this study is the classification result for chronic kidney disease (classification). The research framework used in this study is shown in Figure 1.

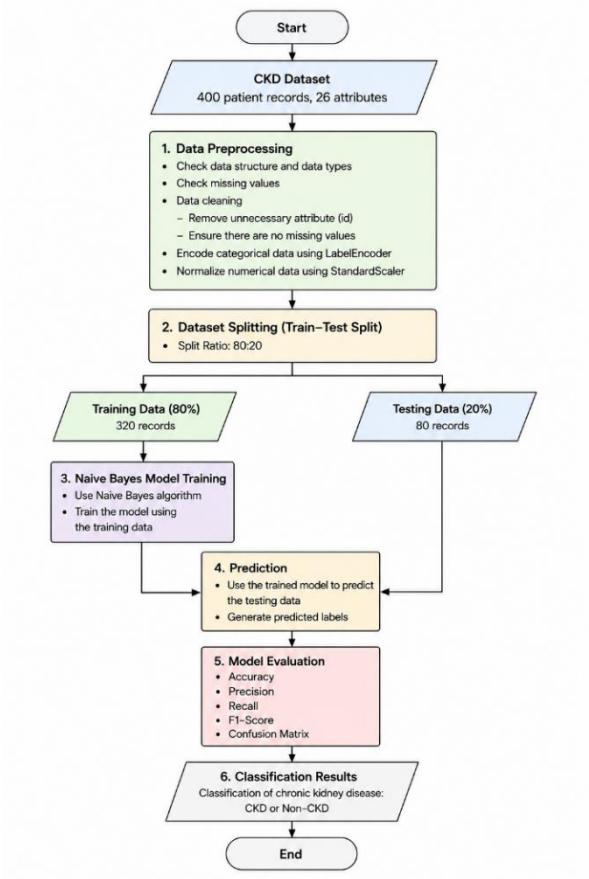

Figure 1. Research Framework for the Classification of Chronic Kidney Disease Using the Naive Bayes Algorithm

As shown in Figure 1, the study began with the collection of a chronic kidney disease dataset consisting of 400 patient records and 26 medical attributes. The dataset was then imported using Python, and an initial analysis was conducted to determine the dataset structure, data types, and to check for missing values. The next step was data cleaning, which involved removing unnecessary attributes and ensuring data quality remained high. Following this, categorical data was encoded and data was normalized using StandardScaler. The dataset was then split into training and testing sets before classification was performed using the Naive Bayes algorithm (Ani et al., 2025; Sari & Supriatna, 2023). The trained model is then used to make predictions on the test data and evaluated using accuracy, precision, recall, F1-score, and a confusion matrix. The evaluation results are visualized in the form of class distribution graphs and a confusion matrix, yielding an accuracy of 97.5%, indicating that the Naive Bayes algorithm is capable of classifying chronic kidney disease with excellent performance.

2.3. Algoritma Naive Bayes

The Naive Bayes algorithm is a probability-based classification method that uses Bayes' Theorem to determine the probability that a data point belongs to a specific class based on its attributes (Chotimah & Rozzaqi, 2023; Gultom et al., 2026; Rizal et al., 2023). This algorithm assumes that each attribute is independent, making the probability calculation process simpler and more efficient (Abdussomad et al., 2024; Jannah et al., 2025; Zanis & Ghufron, 2025). The classification process in Naive Bayes is performed by calculating the probability of a hypothesis based on the observed data. The algorithm calculates the probability of each class and selects the class with the highest probability value as the classification result (Chotimah & Rozzaqi, 2023; Faradeya & Subhiyakto, 2025; Rizal et al., 2023).
The Naive Bayes algorithm equation used in this study can be written as follows:

$$P(H|X) = \frac{P(X|H) \times P(H}{P(X)} \quad (1)$$

Description:

a. $P(H|X)$ = the probability of hypothesis H given condition X
b. $P(X|H)$ = the probability of data X occurring given hypothesis H
c. $P(H)$ = the probability of hypothesis H
d. $P(X)$ = the probability of data X

The data normalization equation using StandardScaler can be written as follows:

$$z = \frac{X-\mu}{\sigma} \quad (2)$$

Description:

a. Z = normalized data
b. X = original data values
c. μ = mean of the data
d. σ = standard deviation of the data

This study employed the Naive Bayes algorithm because it is capable of processing various medical attributes—such as blood pressure, blood glucose levels, albumin, and hemoglobin—used in the classification of chronic kidney disease (Pratiwi & Gama, 2026; Ramadhan, Wibowo, et al., 2025; Wantoro, 2026). In addition to having a fast and simple computational process, this algorithm also demonstrates good performance in the classification of chronic kidney disease, making it suitable for supporting early disease detection (Chotimah & Rozzaqi, 2023; Rizal et al., 2023; Wantoro, 2026).

3. RESULTS AND DISCUSSION

3.1. Model Evaluation Techniques

The model evaluation in this study was conducted to assess the performance of the Naive Bayes algorithm in classifying chronic kidney disease. The evaluation process utilized several performance metrics, including accuracy, precision, recall, F1-score, and the confusion matrix. These metrics were used to measure the model's accuracy in making predictions based on the pre-processed test data. Accuracy is used to measure how well a model classifies data overall. The accuracy score represents the ratio of the number of data points correctly predicted to the total number of data points tested. The higher the accuracy score, the better the performance of the resulting classification model (Made et al., 2026).

Precision is used to measure the accuracy of the model's positive predictions. This metric indicates the proportion of data classified as positive that actually belongs to the positive class. Recall, on the other hand, is used to measure the model's ability to identify all positive data in the dataset. Additionally, the F1-score is used to measure the balance between precision and recall, thereby providing a more comprehensive picture of the model's performance.Model evaluation is also performed using a confusion matrix to determine the number of correct and incorrect predictions generated by the classification model. The confusion matrix aids in analyzing model performance by providing information on True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). The accuracy formula can be written as follows:

$$Accuracy = TP + TNTP + TN + FP + FN \quad (3)$$

Description:

a. TP = True Positive
b. TN = True Negative
c. FP = False Positive
d. FN = False Negative

The entire model evaluation process was conducted using the Scikit-Learn library in Python to ensure that the model performance metrics could be analyzed accurately and systematically.

3.2. Data Cleaning

The initial phase of the study involved analyzing a chronic kidney disease dataset using the Python programming language. The dataset consisted of 400 patient records with 26 medical attributes, including both numerical and categorical variables such as age, blood pressure, albumin, random blood glucose, hemoglobin, hypertension, and other health-related attributes. Initial analysis was performed using the head() and info() functions, along with a check for missing values, to assess the dataset's condition prior to preprocessing.

Based on the initial analysis, it was determined that the dataset contained a mix of numerical and categorical attributes, necessitating preprocessing before model training. Additionally, the missing value check revealed that the dataset contained no missing data, allowing it to be used directly in subsequent steps. The preprocessing stages were carried out through data cleaning, removal of the id attribute, encoding of categorical data using LabelEncoder, and data normalization using StandardScaler. The preprocessing process aims to improve the quality of the dataset so that the classification model can work optimally.

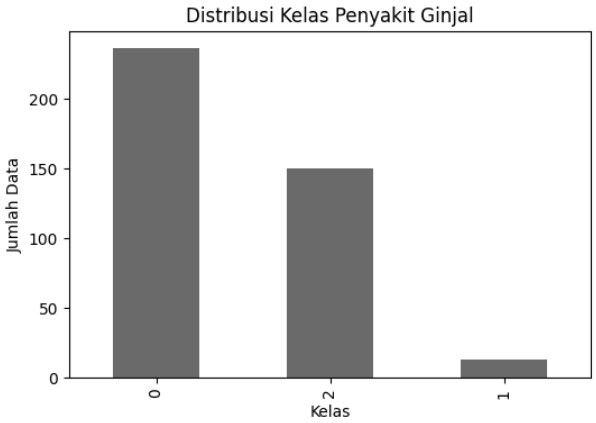

Figure 2. Distribution of Kidney Disease by Severity

Figure 2 shows that the data distribution in the chronic kidney disease dataset is imbalanced. Class 0 has the largest number of data points compared to the other classes, while Class 1 has the fewest. This imbalance in data distribution can affect the performance of the classification model, particularly its ability to predict the minority class. This condition causes the model to more easily recognize the majority class than the minority class. Nevertheless, the Naive Bayes algorithm is still capable of producing excellent classification performance with a high accuracy rate. This indicates that the preprocessing steps performed help improve the model's performance in classifying chronic kidney disease.

After the preprocessing was completed, the dataset was split into training and testing data using the `train_test_split` method with an 80:20 ratio. As a result of this split, there were 320 training data points and 80 testing data points. The training data was used to train the Naive Bayes model, while the testing data was used to evaluate the classification model's performance.

3.3. Results of the Naive Bayes Model Testing

The next step is the model training process using the Naive Bayes algorithm. This algorithm operates based on Bayes' theorem by calculating the probability of each attribute belonging to a specific class and selecting the highest probability as the classification result. The algorithm was implemented using the GaussianNB library from Scikit-Learn in Python.After the model was trained using the training data, a prediction process was performed on the testing data to assess the model's ability to classify chronic kidney disease. Test results indicate that the Naive Bayes algorithm achieves an accuracy of 97.5%. This value demonstrates that the model exhibits excellent classification accuracy in detecting chronic kidney disease.

In addition to accuracy, the model is evaluated using a classification report to assess its performance across each classification class. Test results show that the model has high precision, recall, and F1-score values for most classification classes. However, there is one class with very little data, causing the model to struggle with predictions for that class. This condition is influenced by the imbalanced data distribution in the chronic kidney disease dataset (Simbolon et al., 2025).

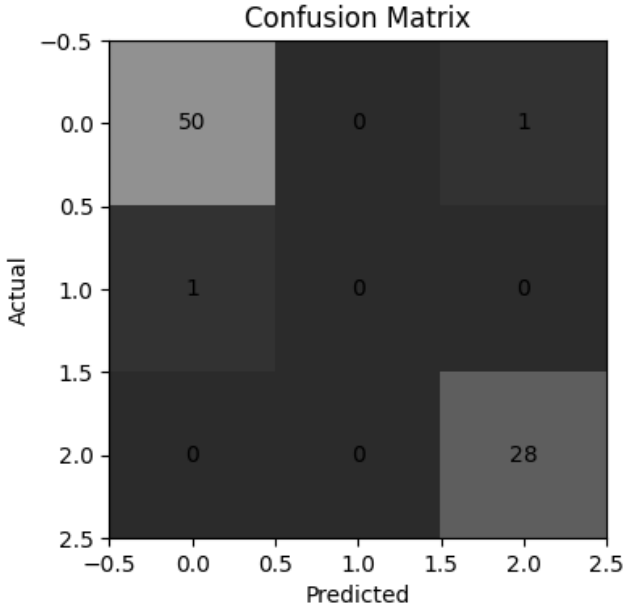

Figure 3. Confusion Matrix Naive Bayes

Figure 3 shows that the Naive Bayes algorithm is highly effective at classifying chronic kidney disease. The model successfully and accurately predicted 50 CKD cases and 28 non-CKD cases, with prediction errors occurring in only a small portion of the data. The confusion matrix results indicate that the model has a high classification accuracy rate, making it well-suited to support the early detection of chronic kidney disease. The confusion matrix results also show that the model is able to recognize patient data patterns quite well, even though the dataset has an imbalanced class distribution. The high number of correct predictions compared to incorrect predictions indicates that the Naive Bayes algorithm is effective for classifying medical data. In addition, the use of preprocessing techniques such as data normalization and data encoding

3.4. Discussion

The results of the study show that the Naive Bayes algorithm is capable of classifying chronic kidney disease with an accuracy rate of 97.5%. These results indicate that the research hypothesis is accepted, namely that the Naive Bayes algorithm can be used as a method for classifying chronic kidney disease with excellent performance. The high accuracy rate is influenced by the optimal data preprocessing process, ranging from data cleaning and encoding of categorical data to data normalization using StandardScaler.

The results of this study are consistent with the research conducted by Roy et al., 2021, which found that the Naive Bayes algorithm performs well in classifying medical data because it performs computations simply while still achieving a high level of accuracy. Research by Qin et al. (2020) also demonstrates that machine learning algorithms can improve the diagnosis of chronic kidney disease compared to manual methods. Additionally, research by Ikhsan et al. (2024) indicates that the Naive Bayes algorithm has a faster training process compared to other classification algorithms.

The difference between this study and previous research lies in the more comprehensive data preprocessing stage. In this study, we removed unnecessary attributes, checked for missing values, encoded categorical data, and normalized the data using StandardScaler. Additionally, this study presents visualizations of the classification results using class distribution plots and confusion matrices, allowing for a more in-depth analysis of the model's performance. Although it achieved a high accuracy score, this study still has limitations, particularly regarding the imbalanced distribution of the dataset. The class distribution results show that one class has very few data points compared to the others. This situation makes it difficult for the model to recognize data patterns in the minority class. Therefore, future research could employ data balancing techniques such as SMOTE or other oversampling methods to improve the performance of the classification model. Overall, the results of the study indicate that the Naive Bayes algorithm can be used as a method for classifying chronic kidney disease to support early detection. The high accuracy rate of 97.5% demonstrates that the model is capable of optimally classifying patient data, thereby aiding AI-based decision-making in the healthcare sector.

4. CONCLUSION

Based on the results of the study, it can be concluded that the Naive Bayes algorithm is effective for classifying chronic kidney disease to support early detection. The study was conducted using the Chronic Kidney Disease dataset, which consists of 400 patient records with 26 medical attributes. The research process began with data preprocessing, including data cleaning, removal of unnecessary attributes, encoding of categorical data using LabelEncoder, and data normalization using StandardScaler, followed by model training and testing using the Naive Bayes algorithm. Based on the model testing results, an accuracy of 97.5% was obtained, indicating that the model has an excellent classification accuracy in detecting chronic kidney disease. The confusion matrix results also show that the majority of the data was classified correctly, suggesting that the Naive Bayes algorithm is effective for medical data classification. Furthermore, this study demonstrates that the data preprocessing stage plays a significant role in improving the performance of the classification model. Although it yielded good performance, this study still has limitations, particularly regarding the imbalanced dataset distribution, where one class has a very small number of data points, which affects the model's prediction results. Therefore, future research is expected to address

REFERENCES

Abdussomad, A., Kurniawan, I., & Wibowo, A. (2024). Prediksi Kemungkinan Penyakit Liver menggunakan Algoritma Klasifikasi Naive Bayes. In *Technologia : Jurnal Ilmiah* (Vol. 15, Issue 3, p. 506). Universitas Islam Kalimantan MAB Banjarmasin. https://doi.org/10.31602/tji.v15i3.15288

Adittia Fathah, C. J. (2025). Klasifikasi Gender Menggunakan Data Wajah Dengan Algoritma Gender Classification Using Face Data With The Naïve Bayes Algorithm And K-Nearest Neighbors Algorithm. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIIK*, *12*(1), 99–110. https://doi.org/10.25126/jtiik.2025128724

Ani, B. A. K. P., Siswa, T. A. Y., & Yulianto, F. (2025). Implementasi Algoritma Naive Bayes Untuk Klasifikasi Penyakit Ispa. In *Jurnal Tekno Kompak* (Vol. 20, Issue 1, pp. 236–249). Universitas Teknokrat Indonesia. https://doi.org/10.33365/jtk.v20i1.700

Apip, A., & Kurniawan, A. (2026). Perbandingan Kinerja Algoritma Naive Bayes dan K-Nearest Neighbors untuk Analisis Sentimen Ulasan Aplikasi Lazada Menggunakan Python. In *Jurnal Publikasi Teknik Informatika* (Vol. 5, Issue 1, pp. 218–234). Politeknik Pratama Purwokerto. https://doi.org/10.55606/jupti.v5i1.6437

Chotimah, S. N., & Rozzaqi, A. R. (2023). Klasifikasi Diagnosis Penyakit Ginjal Kronis Dengan Menerapkan Konsep Algoritma Naïve Bayes. In *JIPETIK:Jurnal Ilmiah Penelitian Teknologi Informasi & Komputer* (Vol. 4, Issue 1, pp. 8–15). Universitas PGRI Semarang. https://doi.org/10.26877/jipetik.v4i1.16174

Fahmi, H., & Sutisna. (2024). Implementasi Data Mining Klasifikasi Gejala Penyakit TB Menggunakan Algoritma Naive Bayes pada Studi Kasus Puskesmas Pegangsaan Dua B. In *Jurnal Indonesia : Manajemen Informatika dan Komunikasi* (Vol. 5, Issue 3, pp. 2888–2898). Sekolah Tinggi Manajemen Informatika dan Komputer Indonesia Banda Aceh. https://doi.org/10.35870/jimik.v5i3.970

Fahmi, M., & Fadlillah, F. A. N. (2023). Klasifikasi Postingan Pengguna Facebook Untuk Deteksi Phising Menggunakan Naive Bayes. In *Jurnal Riset Informatika dan Teknologi Informasi* (Vol. 1, Issue 1, pp. 25–29). Jejaring Penelitian dan Pengabdian Masyarakat. https://doi.org/10.58776/jriti.v1i1.53

Faradeya, M. A.-Z., & Subhiyakto, E. R. (2025). Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma Naive Bayes. In *Jurnal Algoritma* (Vol. 22, Issue 1, pp. 115–127). Institut Teknologi Garut. https://doi.org/10.33364/algoritma/v.22-1.2178

Fauzi, H., & Santoso, B. (2025). Implementasi Chatbot Untuk Diagnosis Awal Penyakit Saluran Pencernaan Menggunakan Metode Klasifikasi Naive Bayes. In *Jurnal sosial dan sains* (Vol. 5, Issue 12, pp. 829–836). Green Publisher. https://doi.org/10.59188/jurnalsosains.v5i12.32645

Gultom, R. Y., Kurniabudi, & Effiyaldi. (2026). Analisis Perbandingan Kinerja Naive Bayes Dan C4.5 Untuk Deteksi Penyakit Ginjal. In *Jurnal Manajemen Teknologi Dan Sistem Informasi (JMS)* (Vol. 6, Issue 1, pp. 1241–1247). LPPM STIKOM Dinamika Bangsa Jambi. https://doi.org/10.33998/jms.2026.6.1.2603

Indrianti, N. F., Ningsih, A. K., Ilyas, R., Informatika, T., Jenderal, U., Yani, A., Barat, J., & Neighbor, K. (2024). Implementasi Data Mining Untuk Klasifikasi Penyakit Gagal Ginjal Kronis Menggunakan Metode K-Nearest Neighbor. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(2), 2255–2260.

Jannah, M. D., Fauzi, A., Emilia, C., & Hikmayanti, H. (2025). Klasifikasi Penyakit Hipertensi Menggunakan Support Vector Machine dan Naive Bayes. In *Jurnal Algoritma* (Vol. 22, Issue 1, pp. 905–913). Institut Teknologi Garut. https://doi.org/10.33364/algoritma/v.22-1.2316

Made, N., Septhi, O., Wahyudi, A., & Gama, O. (2026). Analisis Performa XGBoost dan Naive Bayes untuk Klasifikasi Dini Penyakit Hipertensi prediksi penyakit . Dalam studi terbaru terkait prediksi penyakit kardiovaskular , (Rayadhani &. *Jurnal Ilmiah Teknik Informatika Dan Komunikasi*, *6*.

Organization, W. H. (2026). *Kidney disease*. World Health Organization.

Prasetyo, M. A., Zyen, A. K., & Kusumodestoni, R. H. (2024). Optimasi Algoritma Naive Bayes Berbasis Kernel Untuk Klasifikasi Penyakit Hati. In *Jurnal Informatika Teknologi dan Sains (Jinteks)* (Vol. 6, Issue 3, pp. 748–756). Jurnal Informatika, Teknologi dan Sains Universitas Teknologi Sumbawa. https://doi.org/10.51401/jinteks.v6i3.4783

Pratiwi, N. M. O. S., & Gama, A. W. O. (2026). Analisis Performa XGBoost dan Naive Bayes untuk Klasifikasi Dini Penyakit Hipertensi. In *Jurnal Ilmiah Teknik Informatika dan Komunikasi* (Vol. 6, Issue 1, pp. 549–563). Politeknik Pratama Purwokerto. https://doi.org/10.55606/juitik.v6i1.2119

Pulungan, M. P., Purnomo, A., Kurniasih, A., Korespondensi, P., Class, I., & Technique, S. M. O. (2024). Penerapan Smote Untuk Mengatasi Imbalance Class Dalam Klasifikasi Kepribadian Mbti Menggunakan Naive Bayes Application Of Smote To Overcome Class Imbalance In The Mbti Personality Classification Using The Naïve Bayes Classifier. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIIK*, *11*(5), 1033–1042. https://doi.org/10.25126/jtiik.2024117989

Qin, J., Chen, L., Liu, Y., Liu, C., Feng, C., & Chen, B. (2020). A machine learning methodology for diagnosing chronic kidney disease. *IEEE Access*, *8*, 20991–21002. https://doi.org/10.1109/ACCESS.2019.2963053

Rahayu, A. H., Sugianto, C. A., & Rohmayani, D. (2024). Analisis Big Data dalam Deteksi Dini Wabah Penyakit Menular untuk Mendukung Sistem Kesehatan Publik masyarakat , ekonomi , dan kehidupan sehari-hari . Penyakit menular seperti COVID-19 , Ebola ,. *Journal of New Trends in Sciences*.

Ramadhan, K. G., Wahyu, G., Wibowo, N., & Maori, N. A. (2025). Komparasi Deteksi Penyakit Ginjal Kronis Menggunakan. *JIRE (Jurnal Informatika & Rekayasa Elektronika)*, *8*(1), 13–21.

Ramadhan, K. G., Wibowo, G. W. N., & Maori, N. A. (2025). Komparasi Deteksi Penyakit Ginjal Kronis Menggunakan Algoritma Support Vector Machine Dan Random Forest. In *Jurnal Informatika dan Rekayasa Elektronik* (Vol. 8, Issue 1, pp. 13–21). STMIK Lombok. https://doi.org/10.36595/jire.v8i1.1288

Ratnasari, R., Feriyanto, D., Pringsewu, U. A., Artikel, I., Learning, D., & Padi, P. D. (2025). Deteksi Penyakit Daun Padi Menggunakan Deep Learning untuk mendukung Produktivitas dan Pertanian Berkelanjutan. *Jurnal Algoritma*, *22*(2), 716–727. https://doi.org/10.33364/algoritma/v.22-2.2900

Riany, A. F., & Testiana, G. (2023). Penerapan Data Mining Untuk Klasifikasi Penyakit. *MDP STUDENT CONFERENCE (MSC)*, 297–305.

Riany, A. F., Testiana, G., Informasi, S. S., & Palembang, K. (2023). Penerapan Data Mining untuk Klasifikasi Penyakit Stroke Menggunakan Algoritma Naïve Bayes Sedangkan Provinsi di Indonesia dengan. *JURNAL SAINTEKOM*, *9*(1), 42–54.

Rizal, M., Syahaf, M. Z., Priyambodo, S. R., & Rhamdani, Y. (2023). Optimasi Algoritma Naïve Bayes Menggunakan Forward Selection Untuk Klasifikasi Penyakit Ginjal Kronis. In *Naratif : Jurnal Nasional Riset, Aplikasi dan Teknik Informatika* (Vol. 5, Issue 1, pp. 71–80). Sekolah Tinggi Teknologi Bandung. https://doi.org/10.53580/naratif.v5i1.200

Roy, M. S., Ghosh, R., Goswami, D., & Karthik, R. (2021). Comparative analysis of machine learning methods to detect chronic kidney disease. *Journal of Physics: Conference Series*, *1911*(1), 12005. https://doi.org/10.1088/1742-6596/1911/1/012005

Sari, A. Y., & Supriatna, E. (2023). Penerapan Data Mining Menggunakan Metode Algoritma Naive Bayes Classifier untuk Mendukung Strategi Promosi. In *Jurnal Dimamu* (Vol. 3, Issue 1, pp. 18–28). Universitas Masoem. https://doi.org/10.32627/dimamu.v3i1.837

Setiawan, D., Muhammad, A., Herawati, S., & Dewi, F. (2025). Penerapan Algoritma Klasifikasi untuk Deteksi Dini Penyakit Jantung Koroner Berdasarkan Gejala Klinis koroner berdasarkan data klinis pasien menggunakan algoritma pembelajaran mesin [7]. komputasional berbasis algoritma pembelajaran mesin untuk klasifik. *Teknik: Jurnal Ilmu Teknik Dan Informatika*, *5*(1), 18–26.

Simbolon, M. D., Bukit, D. D., Purba, R. E., Ketaren, F. H., & Prabowo, A. (2025). Perbandingan Algoritma K-Nearest Neighboor dan Naive Bayes Dalam Prediksi Penyakit Ginjal Kronis Pada Lansia. In *Journal of Information System Research (JOSH)* (Vol. 6, Issue 2). Forum Kerjasama Pendidikan Tinggi (FKPT). https://doi.org/10.47065/josh.v6i2.5938

Suwari, D. R., & Hidayati, N. (2025). Klasifikasi Kompetensi Mahasiswa menggunakan Algoritma Naive Bayes. In *Dinamik* (Vol. 30, Issue 2, pp. 281–288). Universitas Stikubank. https://doi.org/10.35315/dinamik.v30i2.10156

Wang, B. (2020). *Kidney disease diagnosis based on machine learning*. https://doi.org/10.1109/CDS49703.2020.00066

Wantoro, A. (2026). Optimasi Klasifikasi Penyakit Ginjal Kronis (Pgk) Menggunakan Kombinasi Metode Seleksi Fitur Dan Algoritma Machine Learning. In *Jurnal Informatika dan Teknik Elektro Terapan* (Vol. 14, Issue 2). Lembaga Penelitian dan Pengabdian kepada Masyarakat Universitas Lampung. https://doi.org/10.23960/jitet.v14i2.9383

Wantoro, A., Pura, L., Putra, A. D., Priandika, A. T., Hendra, V., Isnain, A. R., Informatika, T., Pringsewu, U. A., Dalam, P. P., Sakit, R., Daerah, U., Moeloek, A., Informasi, S., Indonesia, U. T., Indonesia, U. T., Matematika, P., & Indonesia, U. T. (2026). Optimasi Klasifikasi Penyakit Ginjal Kronis (Pgk) Menggunakan Kombinasi Metode Seleksi. *JITET*, *14*(2).

Widiati, W., Iriadi, N., Ariyati, I., Nawawi, I., & Sugiono, S. (2024). Pendekatan Hibrida Decision Tree-Particle Swarm Optimization untuk Deteksi Dini Penyakit Ginjal Kronis. In *JASIEK (Jurnal Aplikasi Sains, Informasi, Elektronika dan Komputer)* (Vol. 6, Issue 1, pp. 11–22). Universitas Merdeka Malang. https://doi.org/10.26905/jasiek.v6i1.13006

Zanis, M. N., & Ghufron, G. (2025). Implementasi Metode Naive Bayes Untuk Klasifikasi Risiko Penyakit Hipertensi. In *Nusantara of Engineering (NOE)* (Vol. 8, Issue 2, pp. 409–416). Universitas Nusantara PGRI Kediri. https://doi.org/10.29407/noe.v8i02.25207

Zuhri, M. R., Kusrini, K., & Ariatmanto, D. (2025). Analisis Perbandingan Algoritma Klasifikasi Untuk Identifikasi Diabetes Dengan Menggunakan Metode Random Forest Dan Naive Bayes. In *Jurnal Informatika Teknologi dan Sains (Jinteks)* (Vol. 7, Issue 1, pp. 11–20). Jurnal Informatika, Teknologi dan Sains Universitas Teknologi Sumbawa. https://doi.org/10.51401/jinteks.v7i1.5146