Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 1, September 2025
e-ISSN 2685-998X
DOI : 10.30865/json.v7i1.8867
Hal: 33−41

# Heart Disease Classification Based on Medical Record Data Using the Logistic Regression Method

**Syahyana Iswari*, Rozzi Kesuma Dinata, Hafizh Al Kautsar Aidilof**
Departement of Informatics, Universitas Malikussaleh, Lhokseumawe, Indonesia
Email: [1,*]syahyana.210170243@mhs.unimal.ac.id, [2]rozzi@unimal.ac.id, [3]Hafizh@unimal.ac.id
Email Penulis Korespondensi: syahyana.210170243@mhs.unimal.ac.id*
Submitted: **10/07/2025**; Accepted: **18/07/2025**; Published: **08/09/2025**

**Abstract−** Heart disease remains one of the primary causes of mortality globally and poses a significant public health concern, including in Indonesia. Early identification of individuals at risk is essential for lowering death rates and enhancing the success of medical interventions. This research focuses on developing a classification model for heart disease using the Logistic Regression technique, utilizing data extracted from patient medical records. The dataset comprises 100 entries, each containing six key features: age, gender, blood pressure, heart rate, respiratory rate, and chest pain. The model was trained on 80% of the data and evaluated using the remaining 20%. Model performance was assessed using several metrics, including accuracy, precision, recall (sensitivity), F1-score, confusion matrix, and the ROC (Receiver Operating Characteristic) curve. The evaluation results revealed an accuracy of 95%, precision of 100%, recall of 88.89%, F1-score of 94.12%, and an AUC score of 0.99. These outcomes suggest that Logistic Regression is highly effective for classifying heart disease risk and can serve as a valuable tool in early detection systems supported by medical record data.

**Keywords**: Data Mining; Classification; Logistic Regression; Disease.

## INTRODUCTION

The advancement of information technology has become an essential part of human life in the modern era. Today, nearly all human activities generate large volumes of data that continue to grow exponentially. These data originate from various sources, such as digital transactions, social media activities, Internet of Things (IoT) sensors, and electronic medical records. This surge in data volume presents significant opportunities for utilization, but also poses major challenges in terms of data management. To transform data into meaningful and valuable information, it is necessary to employ methods capable of managing and analyzing data effectively and efficiently[1].

A promising approach to address this challenge is data mining, a process that uncovers valuable patterns, associations, and hidden knowledge within large volumes of data by leveraging statistical techniques, artificial intelligence, and machine learning models[2][3]. Classification is one of the key techniques in data mining, used to assign data into predefined categories based on their attributes[4]. This technique enables systems to predict the category of a new entity based on patterns identified in historical data. The application of classification methods has been widely utilized across various sectors, such as finance for credit risk detection, commerce for analyzing customer behavior, and the industrial sector for product quality control[5]. One of the fields that greatly benefits from the application of classification is the healthcare sector. In the medical domain, data classification can assist in the diagnostic process, evaluate risk factors, and support clinical decision-making. The increasing complexity of healthcare data, particularly in digital medical records, has driven the need for intelligent systems that can aid medical professionals in performing more accurate analysis and diagnosis.

One of the diseases with a high mortality rate and significant global concern is heart disease. Although classified as a non-communicable disease, heart disease accounts for an exceptionally high number of deaths[6]. As reported by the World Health Organization (WHO) in 2022, heart disease accounts for nearly 17.9 million deaths each year, making up roughly 31% of all deaths worldwide. In the Asian region, the disease causes around 712,100 fatalities, with Indonesia placing second in Southeast Asia, recording an estimated 371,000 deaths.[7]. A major contributing factor to this high mortality rate is the low level of early detection and the general lack of public awareness regarding the risks associated with unhealthy lifestyles[8]. The symptoms of heart disease often go unnoticed as they tend to appear gradually, ranging from chest pain radiating to the left arm, shortness of breath, extreme fatigue, to fainting. Key contributing factors consist of unhealthy eating patterns, insufficient physical exercise, tobacco use, and elevated levels of emotional stress[9]. Given these circumstances, it is crucial to develop early detection systems based on medical data to identify patients at risk of heart disease before the symptoms become more severe.

Logistic Regression is among the most commonly applied classification techniques in the medical domain. It is a statistical method used to estimate the likelihood of a binary outcome such as the presence or absence of a disease based on multiple independent variables. Logistic Regression is capable of processing both small- and large-scale datasets and provides clearly interpretable results, making it a relevant method for analyzing patient medical records[10]. Several previous studies have demonstrated that Logistic Regression performs well in analyzing medical data. For example, a study by [11] successfully classified anemia cases in adolescent girls with an accuracy of 79.23%, a precision of 80.18%, and a recall value of 96.39%. Meanwhile, a study by [12]

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 1, September 2025
e-ISSN 2685-998X
DOI : 10.30865/json.v7i1.8867


proved that Logistic Regression was capable of detecting heart disease risk with an accuracy rate of 88% and an Area Under the Curve (AUC) of 0.95, indicating excellent classification performance. Another study by [13] showed the effectiveness of this method in classifying diabetes cases, achieving a maximum accuracy of 81%. Although widely applied, each study involves different data characteristics and contextual applications[14]. Although Logistic Regression has been widely applied in various studies, many of them utilize complex medical attributes that may not align with the context of healthcare systems in Indonesia. To date, there is a lack of research specifically focusing on assessing the performance of Logistic Regression in classifying heart disease risk using basic clinical features extracted from medical records, such as age, gender, blood pressure, heart rate, respiratory rate, and chest pain. This study is intended to address that gap.

The primary objective of this research is to assess the performance of the Logistic Regression method in identifying the risk of heart disease based on six clinical features. This study emphasizes not only classification accuracy but also the model's potential as a decision-support tool for early diagnosis within healthcare systems. The findings are expected to contribute to the development of more systematic and data-driven approaches for the early detection and prevention of heart disease.

## 2. RESEARCH METHODOLOGY

The research is organized into a sequence of systematic steps designed to effectively reach the intended goals. A comprehensive overview of the research process is illustrated in Figure 1.

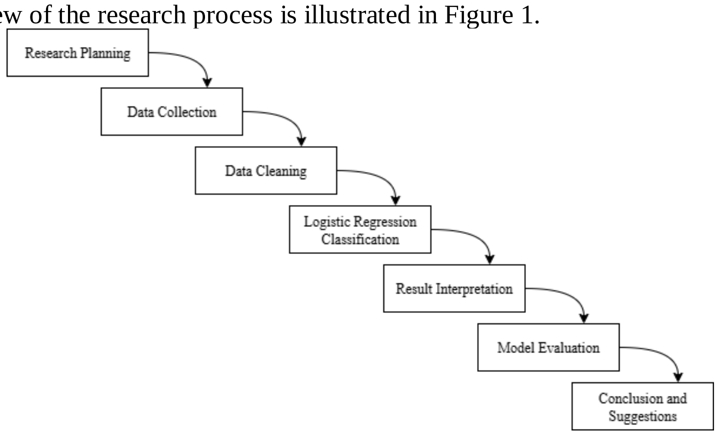


**Figure 1.** Research Stage

1. Research Planning
   The study commences with a well-organized planning phase aimed at defining the research direction, goals, and methodologies to be employed. During this stage, the researcher pinpoints the core issue, which involves developing a classification model to identify the likelihood of heart disease using patient medical records. Logistic Regression was chosen as the algorithm due to its effectiveness in handling binary classification tasks, such as differentiating between individuals who are at risk and those who are not.
2. Data Collection
   This study employed a dataset obtained from existing published research. The dataset includes 100 records, each with six features: gender, age, blood pressure, heart rate (HR), respiratory rate (RR), and chest pain. It also contains a single target variable indicating the classification result for coronary heart disease.

**Table 1.** Heart Disease Dataset Attributes

| No | Attribute | Data Type | Description |
|---|---|---|---|
| 1 | Gender | Categorical (0/1) | Indicates the gender of the patient (0 = female, 1 = male). |
| 2 | Age | Numerical (years) | The age of the patient, which is a major risk factor for heart disease. |
| 3 | Blood Pressure | Numerical (mmHg) | The patient's resting blood |

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 1, September 2025
e-ISSN 2685-998X
DOI : 10.30865/json.v7i1.8867


| | | | |
|---|---|---|---|
| | | | pressure. High blood pressure can increase heart risk. |
| 4 | HR | Numerical (bpm) | Heart Rate (beats per minute). Used to assess heart function. |
| 5 | RR | Numerical (breaths/min) | Respiratory Rate, or the number of breaths per minute, which may indicate respiratory or cardiac issues. |
| 6 | Chest Pain | Categorical (0/1) | Indicates whether the patient experiences chest pain (0 = no, 1 = yes). |
| 7 | Classification | Categorical (0-n) | Classification label or target indicating the patient's level of heart disease. |

3. Data Cleaning
Prior to model development, the medical record dataset was preprocessed to maintain data reliability and consistency. Any duplicate records were eliminated to avoid introducing bias during training. Incomplete data were addressed through imputation, where numerical values were replaced using the mean and categorical values using the mode. Categorical fields like gender and chest pain were transformed into numerical representations via label encoding. Furthermore, Min-Max normalization was applied to scale all numerical variables within a 0 to 1 range. This step ensured that each attribute contributed equally in the classification process using Logistic Regression.
4. Logistic Regression Classification
The subsequent step involves performing classification using the Logistic Regression technique. This method is a type of predictive modeling, bearing similarities to the widely known Ordinary Least Squares (OLS) Regression.[15] [16]. In contrast to linear regression, which estimates continuous outcomes, logistic regression is employed for classification tasks—specifically to assess the likelihood that an event falls into a particular category[17]. During this stage, the processed data is divided into two segments: one for training and the other for testing, using an 80:20 split ratio. The training portion is used to build the classification model, while the testing portion is used to evaluate how well the model performs. In Logistic Regression, the relationship between the input (independent) variables and the output (dependent) variable is modeled using a logit function. This function is essentially a linear equation that is transformed through a sigmoid function to produce a probability score, which indicates the likelihood of a particular classification outcome.[18].
According to [x], Logistic Regression is a binary classification algorithm used to estimate the probability that a given input belongs to one of two possible classes. The steps involved in this method are described as follows:
   a. Dataset Initialization
   The Logistic Regression model operates on a structured dataset that contains independent variables (features) and a dependent variable (target), where the target is binary (e.g., 0 or 1).
   b. Model Equation
   The model first computes a linear combination of the input features:
   $$z = \beta 0 + \beta 1X1 + \beta 2X2 + \cdots + \beta nXn \quad (1)$$
   This linear output is then transformed into a probability using the sigmoid function:
   $$\mathrm{P(Y = 1|X)} = \frac{1}{1+e^{-(z)}} = \frac{1}{1+e^{-(\beta 0+\sum_{i=1}^{n}\beta ixi)}} \quad (2)$$
   c. Coefficient Estimation

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 1, September 2025
e-ISSN 2685-998X
DOI : 10.30865/json.v7i1.8867


The model parameters (coefficients $\beta$) are estimated using Maximum Likelihood Estimation (MLE). This process identifies the coefficient values that maximize the likelihood of correctly predicting the observed outcomes in the training data.

d. Classification Decision Rule
Once the probability  is calculated, a threshold is applied to assign the class label. A commonly used threshold is 0.5, defined by:

$$\hat{Y}=\begin{cases} 1, if\ P(Y=1)|X \geq 0.5 \\ 0, if\ P(Y=\ 0)|X \leq 0.5 \end{cases} \quad (3)$$

The model generates a logit score, which is a linear combination of the input variables, and transforms it into a probability ranging from 0 to 1 through the sigmoid function. This probability represents the chance that a patient may be diagnosed with heart disease typically classified as positive if it exceeds a predefined threshold, such as 0.5[19].

5. Result Interpretation
The classification results indicate that each clinical attribute such as age, blood pressure, chest pain, and heart rate plays a role in determining whether an individual falls into the high-risk category for coronary heart disease. The Logistic Regression algorithm generates a probability value for each patient, reflecting the likelihood of that individual having heart disease.

6. Model Evaluation
To determine the performance and reliability of the proposed model, several evaluation metrics are applied. These include accuracy, precision, recall, and F1-score, along with a confusion matrix that provides insights into the model’s correct and incorrect predictions.
The classification performance based on these metrics is outlined below:
1. Accuracy reflects the proportion of total predictions that were correct:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

2. Precision represents how often the model correctly identifies positive cases out of all predicted positives:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

3. Recall (also referred to as sensitivity) measures the model’s ability to detect actual positive cases:

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

4. F1-Score provides a balance between precision and recall, calculated as the harmonic mean of the two:

$$\text{F1-score} = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (7)$$

7. Conclusion and Recommendations
The final stage of this study is the formulation of conclusions and recommendations. The conclusion includes a summary of the classification results, particularly the effectiveness of the Logistic Regression method in identifying patients at risk of heart disease based on medical data. In addition, important recommendations are provided for future research to enhance and refine similar studies.

# 3. RESULT AND DISCUSSION

## 3.1 Dataset

The dataset utilized in this research contains patient information associated with heart disease, derived from medical records gathered between 2021 and 2023, totaling 100 individual entries. A detailed breakdown of the dataset is provided in the table below.

**Table 2.** Heart Disease Dataset Sample

| No | Gender | Age | Blood Pressure | HR (bpm) | RR (breaths/min) | ... | Classification | Chest Pain |
|---|---|---|---|---|---|---|---|---|
| 1 | Male | 45 | 160/100 | 96 | 22 | ... | Coronary Heart Disease | Present |
| 2 | Male | 45 | 130/80 | 102 | 20 | ... | Coronary Heart Disease | Present |

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 1, September 2025
e-ISSN 2685-998X
DOI : 10.30865/json.v7i1.8867


| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 3 | Male | 45 | 150/90 | 98 | 18 | ... | Coronary Heart Disease | Present |
| … | … | … | … | … | … | … | … | … |
| 100 | Female | 78 | 160/100 | 92 | 20 | ... | Arrhythmia | Present |

### 3.2 Normalization Data

The data values were transformed using the Min-Max normalization technique to standardize the scale across all attributes, ensuring that distance calculations within the algorithm are not disproportionately influenced by any particular feature.

**Table 3.** Normalization Data

| No | Gender | Age | Blood Pressure | HR (bpm) | RR (breaths/min) | Chest Pain | Classification |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 1 | 0.545 | 0.333333 | 1 | 1 |
| 2 | 1 | 0 | 0.571428571 | 0.682 | 0.2 | 1 | 1 |
| 3 | 1 | 0 | 0.857142857 | 0.591 | 0.066667 | 1 | 1 |
| … | … | … | … | … | … | … | … |
| 100 | 0 | 1 | 1 | 0.455 | 0.2 | 1 | 0 |

### 3.3 Training Data Results

Based on the probability (likelihood) values generated by the logistic regression model, if the value is ≤ 0.5, the individual is classified as not having coronary heart disease. However, if the value is ≥ 0.5, the individual is classified as having coronary heart disease.

**Table 4.** Training Data Results

| Data | Score | Probability | Likelihood | LogLikehood | DiseaseClassification |
|---|---|---|---|---|---|
| 1 | -0.38 | 0.4043 | 0.5956 | -0.518 | 0 |
| 2 | 0.552 | 0.634769 | 0.365 | -1.007 | 1 |
| 3 | 0.817 | 0.6936 | 0.693 | -0.365 | 1 |
| … | … | … | … | … | … |
| 80 | -0.03 | 0.49248 | 0.5075 | -0.6782 | 0 |

### 3.4 Testing Data Calculation Results

To obtain the results for the testing data, which consists of 20 patient records, it is necessary to determine the logit values, probabilities, and log-likelihood values. The following are the results of the testing data calculations.

**Table 5.** Testing Data Calculation Results

| Actual | Score | Probability | Likelihood | LogLikehood | Predicted |
|---|---|---|---|---|---|
| 0 | 0.17 | 0.5443986 | 0.45560 | -0.7861 | 1 |

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 1, September 2025
e-ISSN 2685-998X
DOI : 10.30865/json.v7i1.8867


| | | | | | |
|---|---|---|---|---|---|
| 0 | 0.296 | 0.5736634 | 0.426336 | -0.8525 | 1 |
| 0 | -0.4565 | 0.3878148 | 0.61218 | -0.490 | 0 |
| … | … | … | … | … | … |
| 1 | 0.613 | 0.6486 | 0.64867 | -0.43281 | 1 |

The Logistic Regression model produces output in the form of logit values, which are then converted into probabilities using the sigmoid function. If the probability value exceeds the threshold of 0.5, the patient is classified as “at risk”; otherwise, the patient is classified as “not at risk.” For example, one test data point with a logit value of 0.17 produces a probability of 0.5444, and is therefore classified as “at risk,” although the actual label is “not at risk.” This indicates a case of misclassification. Conversely, for data with a negative logit value, such as -0.4565, the model produces a probability below 0.5 and correctly classifies the patient as “not at risk.” Evaluation results across all test data show that most of the model’s predictions match the actual labels. The likelihood and log-likelihood values are used to assess how well the model makes decisions for each test record. Overall, these values indicate that the model has a high degree of confidence in the majority of its predictions.

### 3.5 Model Evaluation

```
Classification Report:
                             precision    recall  f1-score   support

Tidak Sakit Jantung Koroner       0.92      1.00      0.96        11
    Penyakit Jantung Koroner       1.00      0.89      0.94         9

                    accuracy                           0.95        20
                   macro avg       0.96      0.94      0.95        20
                weighted avg       0.95      0.95      0.95        20
```

**Figure 2.** Classification Model Performance Testing Results

Based on the classification report, the classification model demonstrates excellent performance in distinguishing between patients with and without coronary heart disease. For the "No Coronary Heart Disease" category, the model achieved a precision of 0.92, a perfect recall of 1.00, and an F1-score of 0.96. Meanwhile, for the "Coronary Heart Disease" category, the model achieved a precision of 1.00, a recall of 0.89, and an F1-score of 0.94. Overall, the model reached an accuracy of 95%, with average values of precision, recall, and F1-score at 0.96, 0.94, and 0.95, respectively. These results indicate that the Logistic Regression method performs accurately and consistently, making it suitable for early detection of heart disease using patient medical records..

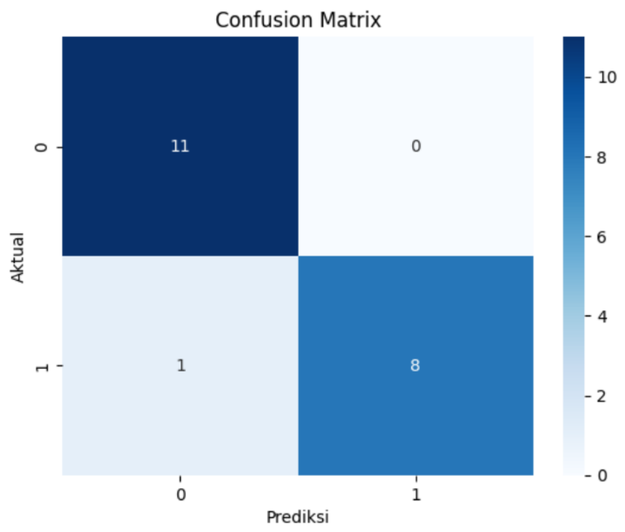


**Figure 3.** Confusion Matrix

The analysis of the confusion matrix reveals that the Logistic Regression model performs exceptionally well in classifying heart disease cases. Among the 20 test data samples, the model accurately identified 11 individuals without heart disease (True Negatives) and correctly classified 8 patients as having heart disease (True Positives).

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 1, September 2025
e-ISSN 2685-998X
DOI : 10.30865/json.v7i1.8867


Notably, there were no incorrect classifications of healthy individuals as sick (False Positives = 0), although one case of heart disease was missed (False Negative = 1). The evaluation metrics reflect the model's strong performance, achieving an accuracy of 95%, a perfect precision score of 100%, a recall of 88.89%, and an F1-score of 94.12%. These results confirm the model’s ability to produce accurate and balanced predictions. Therefore, Logistic Regression is considered a suitable approach for early detection of heart disease using data from patient medical records, offering valuable support for faster and more accurate clinical decision-making processes.

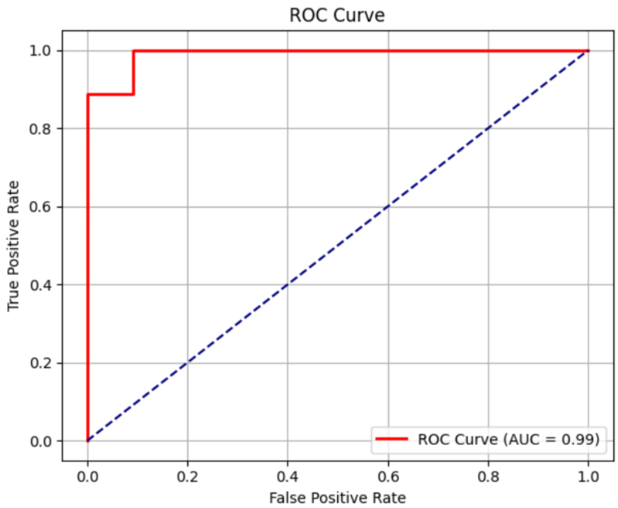


**Figure 4.** ROC and AUC

The figure above displays the Receiver Operating Characteristic (ROC) Curve, which is used to evaluate the performance of a classification model—in this case, the Logistic Regression model applied to detect the risk of heart disease. The ROC curve illustrates the relationship between the True Positive Rate (TPR), or sensitivity, and the False Positive Rate (FPR) across various classification thresholds. In the graph, the red line represents the ROC curve of the model, while the blue dashed line serves as a baseline, representing the performance of a random classifier. The farther the curve is from the baseline and the closer it moves toward the top-left corner, the better the model's performance. From the graph, it is evident that the curve rises steeply and approaches the top-left corner, indicating that the model can effectively distinguish between patients with and without heart disease. This is further supported by the Area Under the Curve (AUC) value of 0.99, which is very close to the perfect score of 1.0. Such a high AUC value demonstrates that the model has excellent classification accuracy and reliability, with minimal error rates. In other words, the Logistic Regression model used in this study is capable of correctly identifying the majority of patients with near-perfect discriminative ability. These results align with earlier outcomes obtained from the confusion matrix and the classification report, both of which further support the model’s strong effectiveness in detecting heart disease at an early stage using data derived from patient medical records.

### 3.6 Discussion

This study introduces a model for classifying the risk of heart disease by utilizing patient medical records through the Logistic Regression approach. The dataset comprises 100 entries gathered between the years 2021 and 2023, with each record featuring six essential attributes: gender, age, blood pressure, heart rate, respiratory rate, and chest pain. For model development, the data was split into training and testing sets using an 80:20 ratio. The model’s classification results confirm its effectiveness in identifying individuals at potential risk and those not at risk of heart disease. The prediction process involved calculating logit values based on the input variables and converting them into probabilities using the sigmoid function, applying a classification threshold of 0.5. The model achieved strong evaluation results, including 95% accuracy, 100% precision, 88.89% recall, and an F1-score of 94.12%. The confusion matrix reported 11 true negative cases, 8 true positives, 1 false negative, and no false positives. Additionally, the model attained an Area Under the Curve (AUC) value of 0.99 on the Receiver Operating Characteristic (ROC) curve, reinforcing its excellent ability to distinguish between high-risk and low-risk patients. In summary, the application of Logistic Regression has proven to be both dependable and accurate for heart disease classification using medical record data, making it a promising tool for early detection systems that aim to support clinical decision-making with data-driven insights.

## 4. CONCLUSION

The results of this study demonstrate that Logistic Regression is a reliable and effective method for predicting heart disease risk based on patient medical records. The model developed in this research utilizes six essential medical attributes—age, gender, blood pressure, heart rate, respiratory rate, and chest pain—to determine whether a patient is at risk or not. The dataset, comprising 100 patient records collected between 2021 and 2023, was divided into training and testing sets using an 80:20 split ratio. The performance evaluation of the model yielded strong results, achieving 95% accuracy, 100% precision, 88.89% recall, and an F1-score of 94.12%. Furthermore, the model achieved an AUC value of 0.99 on the ROC curve, reflecting its excellent capability in distinguishing between patients who are at risk and those who are not. These findings suggest that Logistic Regression is well-suited for use in early detection systems, particularly in digital healthcare applications where speed and accuracy are essential.

## REFERENCES

[1] L. Britanthia, C. Tanujaya, B. Susanto, and A. Saragih, "Perbandingan Metode Regresi Logistik dan Random Forest untuk Klasifikasi Fitur Mode Audio Spotify," *Indonesian Journal of Data and Science (IJODAS)*, vol. 1, no. 3, pp. 68–78, 2020.

[2] Y. Anugerah, D. Khurrota A'yunan, S. Busono, F. Sains, and D. Teknologi, "IMPLEMENTASI DATA MINING DALAM KLASIFIKASI DIAGNOSA KANKER PAYUDARA MENGGUNAKAN ALGORITMA LOGISTIC REGRESSION," *Jurnal TEKINKOM*, vol. 6, no. 2, p. 2023, 2023, doi: 10.37600/tekinkom.v6i2.948.

[3] H. Al Kautsar Aidilof and W. Kurniawan, "KLASIFIKASI BERITA OLAHRAGA PADA PORTAL BERITA ONLINE DENGAN METODE K-NEAREST NEIGHBOUR (KNN) DAN LEVENSHTEIN DISTANCE," 2021. Accessed: Jul. 15, 2025. [Online]. Available: https://doi.org/10.29103/tts.v2i1.3760

[4] R. K. Dinata, H. Akbar, and N. Hasdyna, "Algoritma K-Nearest Neighbor dengan Euclidean Distance dan Manhattan Distance untuk Klasifikasi Transportasi Bus," *ILKOM Jurnal Ilmiah*, vol. 12, no. 2, pp. 104–111, Aug. 2020, doi: 10.33096/ilkom.v12i2.539.104-111.

[5] M. Arhami and M. Nasir, "Data_Mining_Algoritma_dan_Implementasi," *Data Mining - Algoritma dan Implementasi*, 2020.

[6] J. Junifer Pangaribuan and H. Tanjaya, "MENDETEKSI PENYAKIT JANTUNG MENGGUNAKAN MACHINE LEARNING DENGAN ALGORITMA LOGISTIC REGRESSION," 2021.

[7] A. Harisa, P. Wulandari, S. Ningrat, and Y. Yodang, "Pengaruh Terapi Murottal Terhadap Depresi Pada Pasien Congestive Heart Failure Di Pusat Jantung Terpadu RSUP Dr. Wahidin Sudirohusodo," *Dunia Keperawatan: Jurnal Keperawatan dan Kesehatan*, vol. 8, no. 2, p. 269, Aug. 2020, doi: 10.20527/dk.v8i2.8324.

[8] S. Muharni and S. Andriyanto, "Sistem Diagnosa Penyakit Jantung Berbasis Case Based Reasoning (CBR)," 2021, Accessed: Feb. 23, 2025. [Online]. Available: https://jurnal.darmajaya.ac.id/index.php/PSND/article/view/2910/1233

[9] M. Program Studi Ilmu Keperawatan, S. Citra Delima Bangka Belitung, J. I. Pinus, K. Pedang, P. Pinang, and K. Bangka Belitung, "DUKUNGAN KELUARGA BERHUBUNGAN DENGAN KUALITAS HIDUP PASIEN GAGAL JANTUNG," 2023. [Online]. Available: http://jurnal.globalhealthsciencegroup.com/index.php/JPPP

[10] K. Klasifikasi Metode Regresi Logistik dan Metode Chaid dengan Pembobotan Sampel Puspa Juwita and P. Hendikawati, "Ketepatan Klasifikasi Metode Regresi Logistik dan Metode Chaid dengan Pembobotan Sampel," 2021. [Online]. Available: http://journal.unnes.ac.id/nju/index.php/JM

[11] R. Ibnas and B. Agung Kurniawan, "Klasifikasi Penderita Anemia Menggunakan Metode Regresi Logistik," 2024. Accessed: Feb. 23, 2025. [Online]. Available: https://journal3.uin-alauddin.ac.id/index.php/msa/article/view/45083

[12] G. Rohma, U. Asyafiiyah, and R. M. Akbar, "METODE LOGISTIC REGRESSION PREDICTION OF PATIENTS INDICATED WITH HEART DISEASE USING LOGISTIC REGRESSION METHOD," vol. 4, no. 1, pp. 19–23, 2024, [Online]. Available: http://ejurnal.unim.ac.id/index.php/submit

[13] Z. Mutaqin, C. Rozikin, and Y. A. Tomo, "KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN ALGORITMA LOGISTIC REGRESSION," 2024. [Online]. Available: https://journalpedia.com/1/index.php/jsti

[14] A. Z. Hafizhah, S. Suakanto, R. Y. Fa'rifah, and E. T. Nuryatno, "Klasifikasi Laju Pernapasan dan Saturasi Oksigen Menggunakan Regresi Logistik," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 2, pp. 448–458, May 2023, doi: 10.29100/jipi.v8i2.3481.

[15] M. Purba, F. Ilmu Komputer, and U. Sjakhyakirti, "Klasifikasi Dataset Teks Pengaduan Masyarakat Terhadap Pemerintah di Sosial Media Menggunakan Logistic Regression Article Info ABSTRAK," *JSAI: Journal Scientific and Applied Informatics*, vol. 7, no. 1, p. 5, 2024, doi: 10.36085.

[16] Amrin, Rudianto, and Sismadi, "JITE (Journal of Informatics and Telecommunication Engineering) Data Mining with Logistic Regression and Support Vector Machine for Hepatitis Disease Diagnosis," *JITE*, vol. 8, no. 2, 2025, doi: 10.31289/jite.v8i2.13218.

[17] M. Fahmuddin, M. K. Aidid, and M. J. Taslim, "IMPLEMENTASI ANALISIS REGRESI LOGISTIK DENGAN METODE MACHINE LEARNING UNTUK MENGKLASIFIKASI BERITA DI INDONESIA," *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, vol. 5, no. 03, pp. 155–162, 2023, doi: 10.35580/variansiunm116.

[18] R. Susetyoko, W. Yuwono, and E. Purwantini, "JIP (Jurnal Informatika Polinema) Model Klasifikasi Pada Seleksi Mahasiswa Baru Penerima KIP Kuliah Menggunakan Regresi Logistik Biner".

[19] N. P. Setiawati, B. A. Nugroho, A. T. Setiawan, and D. Hartanti, "Data Mining Untuk Klasifikasi Diagnosis Tingkat Keparahan Penyakit Diabetes Dengan Algoritma Logistik Regresi," 2023. Accessed: Feb. 23, 2025. [Online]. Available: https://ojs.udb.ac.id/index.php/Senatib/article/view/3254