CLASSIFICATION OF HEART DISEASE USING THE K-NEAREST NEIGHBOR ALGORITHM AND LOGISTIC REGRESSION

I Kadek Agga Sugitha[1*]; Agung Triayudi[2]; Endah Tri Esti Handayani[3]

Fakultas Teknologi Komunikasi dan Informatika[1, 2, 3], Informatika[1]
Universitas Nasional, Jakarta, Indonesia[1, 2, 3]
https://www.unas.ac.id/[1, 2, 3]
ikadekaggasugitha.2020@student.unas.ac.id[1*], agungtriayudi@civitas.unas.ac.id[2],
endahtriesti@civitas.unas.ac.id[3]
(*) Corresponding Author

Abstract— *Heart disease is a major cause of death in the world, including in Indonesia, with increasing rates and death rates that carry a huge burden on health and society. Lack of awareness of early signs contributes significantly to this challenge. This study aims to prevent heart disease through early diagnosis using K-Nearest Neighbor (K-NN) and Logistic Regression algorithms. The database, obtained from Kaggle.com, includes 15 clinical units for cardiac diagnosis. The test shows that the K-NN method with k = 3 achieves the highest performance on the experimental data (30%), with 90% precision, 93% precision, 87% recall, and 90% f1 - score. In comparison, Logistic Regression and sigmoid achieved 86% precision, 83% precision, 90% recall, and 86% f1-score on the same experimental data. These results show that K-Nearest Neighbor is better than Logistic Regression as a classification algorithm for heart disease database. Applying these findings to the web-based Streamlit system is expected to improve the efficiency and timeliness of heart disease screening.*

Keywords: *classification, heart disease, k-nearest neighbors, logistic regression, streamlit.*

Abstrak— *Penyakit jantung merupakan penyebab kematian utama di dunia, termasuk di Indonesia, dengan angka dan tingkat kematian yang terus meningkat sehingga menjadi beban besar bagi kesehatan dan masyarakat. Kurangnya kesadaran terhadap tanda-tanda awal memberikan kontribusi yang signifikan terhadap tantangan ini. Penelitian ini bertujuan untuk mencegah penyakit jantung melalui diagnosis dini menggunakan algoritma K-Nearest Neighbor (K-NN) dan Regresi Logistik. Basis data yang diperoleh dari Kaggle.com mencakup 15 unit klinis untuk diagnosis jantung. Pengujian menunjukkan bahwa metode K-NN dengan k = 3 mencapai kinerja tertinggi pada data eksperimen (30%), dengan presisi 90%, presisi 93%, recall 87%, dan f1-score 90%. Sebagai perbandingan, Regresi Logistik dan sigmoid mencapai presisi 86%, presisi 83%, recall 90%, dan f1-score 86% pada data eksperimen yang sama. Hasil ini menunjukkan bahwa K-Nearest Neighbor lebih baik daripada Regresi Logistik sebagai algoritma klasifikasi untuk basis data penyakit jantung. Penerapan temuan ini pada sistem Streamlit berbasis web diharapkan dapat meningkatkan efisiensi dan ketepatan waktu pemeriksaan penyakit jantung.*

Kata Kunci*: klasifikasi, penyakit jantung, k-nearest neighbors, regresi logistik, streamlit.*

INTRODUCTION

Cardiovascular disease, or heart disease, includes all conditions arising from impaired heart function, often due to plaque buildup in coronary arteries, which obstructs blood flow and increases the risk of heart attacks and other complications (Li et al., 2020). It is the leading cause of death globally, with a 2021 American Heart Association (AHA) report stating that cardiovascular diseases, including coronary heart disease (CHD), heart failure (HF), and stroke, are the main causes of global mortality. Despite medical advances reducing some risk factors like smoking and hypertension, the prevalence of other risks such as obesity and diabetes continues to rise (Virani et al., 2021).

In Indonesia, heart disease poses a significant health burden, with a notable increase in prevalence and mortality rates from 1990 to 2019, with stroke and ischemic heart disease being the primary causes of cardiovascular-related deaths (Muharram et al., 2024). Lack of access to information and media correlates with delayed early heart disease examination (Bianto, Kusrini, & Sudarmawan, 2020). Contributing factors include lack of awareness of risk factors, nonspecific early symptoms, limited healthcare access, and high examination costs. Machine learning offers potential solutions by analyzing large datasets and identifying complex patterns, predicting heart disease risk, recognizing early symptom patterns, and improving healthcare access through telemedicine (Prasetyo & Laksana, 2022).

Early detection of heart disease has many benefits, including increased recovery chances, prevention of complications, and cost savings (Yadav, 2020). This study underscores the urgency of early diagnosis in heart disease classification using K-Nearest Neighbor (K-NN) and Logistic Regression algorithms, which allows quicker disease management and helps avoid life-threatening complications (Ullah et al., 2022). Therefore, better access to information and knowledge about heart disease is crucial in prevention, identification, and effective management efforts (Zuama, Rahmatullah, & Yuliani, 2022).

Research over the last five years has identified gaps, with no studies discussing the combined use of K-NN and Logistic Regression algorithms or comparing their performance in a web-based diagnosis system using Streamlit and incorporating new attributes (Smoke)(Kadhim & Radhi, 2023)(Hartono et al., 2023)(Radhika & Thomas George, 2021)(Kavitha, Gnaneswar, Dinesh, Sai, & Suraj, 2021)(Almustafa, 2020). This research aims to detect heart disease by identifying the best-performing algorithm between K-NN and Logistic Regression and implementing the results in a web-based diagnostic system.

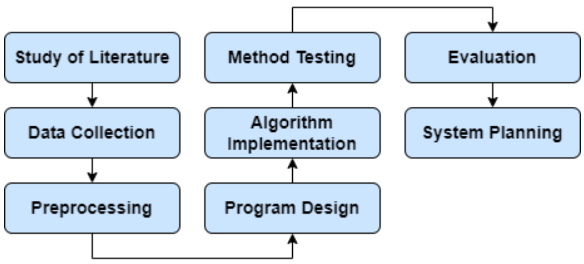
The use of K-Nearest Neighbor (K-NN) was chosen because of its ability to intuitively identify potential heart disease by comparing individual symptoms and clinical attributes with existing patient data, which is highly relevant in a medical context. Logistic Regression was chosen because it provides an in-depth statistical analysis of heart disease risk factors, providing a probabilistic interpretation that helps understand the relationship between risk factors and the likelihood of disease occurrence. The combination of these two methods results in a comprehensive approach, where K-NN focuses on similarity-based classification, while Logistic Regression provides deeper statistical insights, increasing the accuracy and depth of analysis in heart disease research (Sumara, Ari, & Indarti, 2022).

Machine learning plays a vital role in health risk prediction by analyzing multiple factors to support early detection and prevention of diseases. Early detection through technologies such as telemedicine accelerates identification and intervention, ultimately improving health outcomes and reducing costs.

In this context, Logistic Regression is a predictive model used to estimate the likelihood of a data item belonging to a specific class (Ibrahim & Abdulazeez, 2021). It works with both numeric and categorical data, utilizing the sigmoid function to calculate probabilities for binary outcomes, such as 0 or 1 (Alnuaimi & Albaldawi, 2024). As a widely used binary classification algorithm, it is essential in health risk prediction for mapping predictive values to probabilities between 0 and 1 (G et al., 2022). This study explores the use of K-Nearest Neighbor and Logistic Regression algorithms in the prevention, early detection, and management of heart disease, offering insights to help health professionals and researchers reduce its societal impact.

MATERIALS AND METHODS

This study involves several main stages to detect and classify heart disease using the K-Nearest Neighbor (KNN) and Logistic Regression algorithms. These stages are as follows in Figure 1:



Source: (Ariawan, Triayudi, & Sholihati, 2020)
Figure 1. Research Stage Diagram

Study of Literature

Literature studies are conducted to identify discussions related to the topic, providing references for this study on the K-NN algorithm and logistic regression in classification.

Data Collection

This study uses data from Kaggle.com, namely the UCI Heart Disease dataset from the Cleveland Heart Disease Database published by Bayomars in 2019 (www.kaggle.com/code/bayomars12/starter-heart-disease-dataset-408f5662-7). This dataset includes 1026 patients with 14 clinical attributes. The author added 1 attribute (Smoke) that has the potential to affect heart disease, bringing the total to 15 attributes. This dataset is available in Microsoft Excel format (XLSX). This stage also includes understanding the structure and characteristics of the data that has been taken.

Preprocessing

This stage involves cleaning and preparing the data to be ready for use in machine learning models. Activities performed include handling missing data, data normalization, data transformation, and extraction of relevant features. The main goal is to ensure that the dataset taken from Kaggle can be used effectively for training heart disease classification models.

Program Design

This program is built using Python and VS Code, which is used to write and share programs online. In this program, a machine learning model for heart disease classification is built using the K-Nearest Neighbors (K-NN) and Logistic Regression algorithms. The goal of this program is to learn patterns in the available data and generate metrics such as accuracy, precision, recall, and f1-score. These metrics will help evaluate the performance of the model in classifying data.

Algorithm Implementation

The selected algorithm is implemented and optimized to achieve the best performance through the process of training the model using preprocessed data and validation using test data.

In K-Nearest Neighbors, the configuration is done by selecting the optimal k value and using the Euclidean Distance measurement technique. Choosing the right k value is very important to achieve a balance between bias and variance. Meanwhile, Logistic Regression is optimized by selecting the sigmoid activation function, which helps in modeling the probability of events.

Method Testing

The testing method of this research involves manual calculation and evaluation of classification performance using the K-Nearest Neighbors (K-NN) and Logistic Regression algorithms to detect heart disease.

K-NN determines a new instance class based on the majority of classes from k-nearest neighbors, while Logistic Regression models the probability of the target class using the sigmoid function. Both methods offer different perspectives in data analysis, contributing to improving the accuracy of heart disease prediction.

Evaluation

The evaluation process is carried out by testing the scores obtained from the K-Nearest Neighbors (K-NN) and Logistic Regression algorithms using the method:

A. Cross-Validation

An essential technique for evaluating machine learning models, offering accurate and reliable performance measurements. It works by dividing the dataset into multiple folds, allowing the model to be trained on one part while being tested on another. This approach provides a stable estimate of performance, helping to identify potential bias or variance.

Each fold yields an accuracy value, which is averaged to calculate mean accuracy, while the standard deviation of these accuracy values assesses the model's consistency. A low standard deviation indicates stable performance, whereas a high deviation may signal potential instability influenced by specific data points.

B. Confusion Matrix

This provides detailed insights into classification performance by displaying actual versus predicted classes. In a classification context, the diagonal values from the top left to the bottom right represent correct predictions, while off-diagonal numbers indicate errors. Estimating a classification model's accuracy is essential for predicting future performance and selecting the best classifier among available models.

Confusion matrix can be calculated as follows:

$$\text{Accuracy} = \frac{TP + TN}{(TP + TN + FP + FN)} \tag{1}$$

Accuracy measures the extent to which a classification model is able to correctly classify all samples, both positive and negative.

$$\text{Precision} = \frac{TP}{(TP+FP)} \tag{2}$$

Precision measures the extent to which a classification model is able to correctly identify positive samples. Recall to identify all true positive samples.

$$\text{Recall} = \frac{TP}{(TP+FN)} \tag{3}$$

$$F1 - \text{Score} = 2 \times \frac{(\text{Presisi} \times \text{Recall})}{(\text{Presisi} + \text{Recall})} \tag{4}$$

F1-Score is the harmonic mean of precision and recall. It gives equal weight to both.

C. ROC-AUC

A key metric in machine learning and statistical analysis for evaluating the performance of classification models, especially in binary classification. It measures a model's ability to distinguish between two classes by calculating the area under the ROC curve, which plots the false positive rate against the true positive rate across various thresholds. An AUC value close to 1 signifies strong performance in class differentiation, while a value near 0.5 indicates performance no better than

PILAR Nusa Mandiri: Journal of Computing and Information System
Vol. 20, No. 2 September 2024 | DOI: 10.33480/pilar.v20i2.5742

random guessing. A higher AUC value demonstrates a model's enhanced classification capability.

System Planning

The program starts by importing libraries such as Pandas and loading the heart disease dataset from Kaggle into VS Code using pd.read_excel. The data is then processed by cleaning up empty values and separating the data into training and testing sets. After the features (X) and targets (y) are selected, a model is built using the K-Nearest Neighbors (K-NN) and Logistic Regression algorithms, followed by prediction using the testing data.

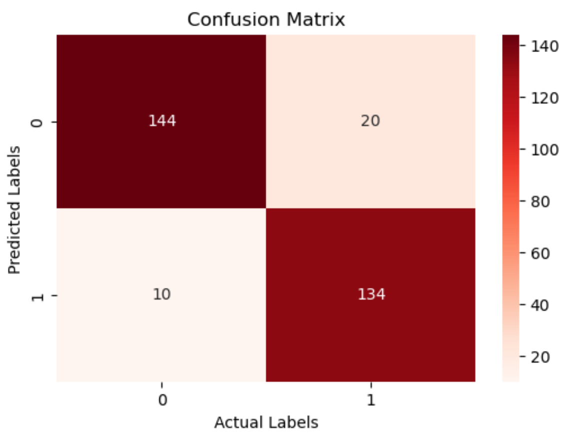
Model evaluation is done through a confusion matrix to measure accuracy, precision, recall, and f1-score, as well as cross-validation scores and ROC-AUC values. The performance of both models is then compared to determine the most accurate algorithm. After saving the program, the output produces a heart disease diagnosis system that is displayed in the Streamlit framework for web access.

RESULTS AND DISCUSSION

The research results used a dataset divided into 70% training data and 30% testing data. Of the 1024 data, it was divided into 717 training data and 308 testing data.

Implementation of K-Nearest Neighbor

Figure 2 shows the confusion matrix resulting from implementation of the K-NN algorithm in Python for the classification process was carried out with a value of K=3, resulting in testing with a number of true positives (TP) of 134, false positives (FP) of 10, false negatives (FN) of 20, and true negatives (TN) of 144.



Source: (Research Result, 2024)
Figure 2. Confusion Matrix KNN Algorithm

Classification Report Algoritma KNN:

	precision	recall	f1-score	support
0	0.88	0.94	0.91	154
1	0.93	0.87	0.90	154
accuracy			0.90	308
macro avg	0.90	0.90	0.90	308
weighted avg	0.90	0.90	0.90	308

Source: (Research Result, 2024)
Figure 3. K-NN Evaluation Results

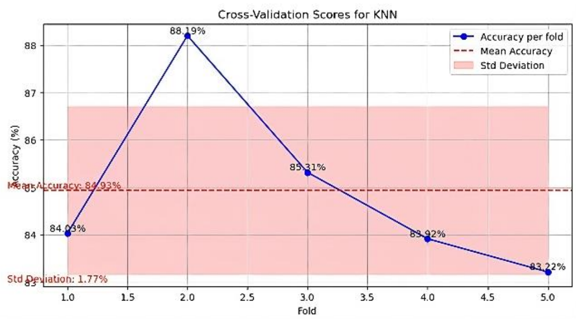
Based on the results of the Classification Report in Figure 3, the accuracy, precision, recall, and F1-Score are as below:

$$\text{Accuracy} = \frac{134 + 144}{134 + 144 + 20 + 10}$$
$$= 0.90259$$
$$= 90\%$$

$$\text{Precision} = \frac{134}{134 + 10}$$
$$= 0.93057$$
$$= 93\%$$

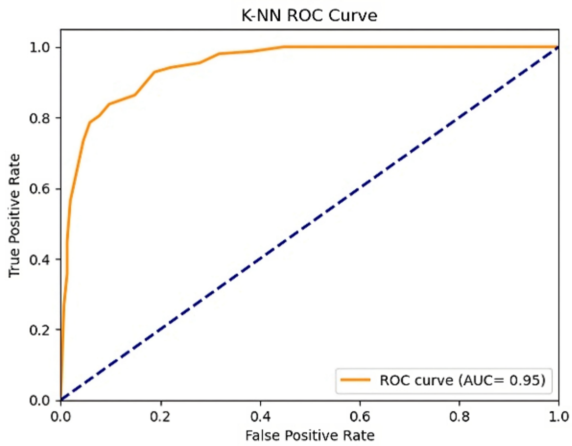
$$\text{Recall} = \frac{134}{134 + 20}$$
$$= 0.87012$$
$$= 87\%$$

$$\text{F1} - \text{Score} = 2 \text{ x } \frac{0.87012 \times 0.93057}{0.87012 + 0.93057}$$
$$= 0.89928$$
$$= 90\%$$



Source: (Research Result, 2024)
Figure 4. K-NN Cross-Validation Graph

The cross-validation results for the K-Nearest Neighbors (K-NN) algorithm with five tests (folds) can be seen in Figure 4. The accuracy scores of each fold vary from around 0.83 to 0.88, with an average accuracy of 0.849 and a standard deviation of 0.0176, which illustrates the variation in results between folds. This indicates that the K-NN model has quite good and stable performance on the tested data.



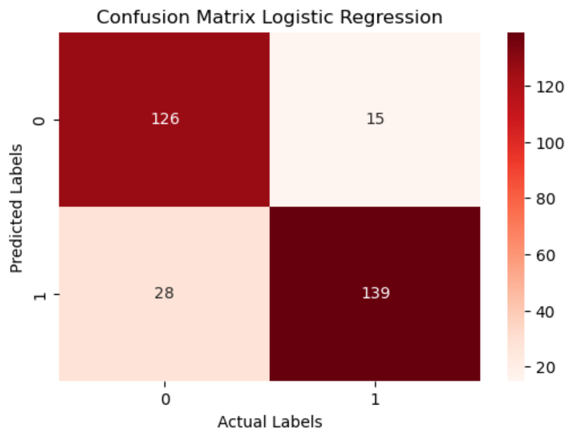
Source: (Research Result, 2024)
Figure 5. K-NN ROC Curve

The explanation of the K-NN model on the ROC curve can be shown in Figure 5 which shows high TPR and low FPR at most thresholds, indicating the model's ability to detect most positive cases while minimizing false positives. With an AUC value of 95% (0.94731), the overall performance of the K-NN model is very good.

Combining the actual label data and the predicted data, the results show that 278 data were predicted correctly and 30 data were predicted incorrectly. This shows that the K-NN classification model is able to make accurate predictions on most of the data, although there are still some errors.

Implementation of Logistic Regression

The implementation of the Logistic Regression algorithm using Python for classification produces a Confusion Matrix shown in Figure 6 with a true positive (TP) value of 139, false positive (FP) of 28, false negative (FN) of 15, and true negative (TN) of 126.



Source: (Research Result, 2024)
Figure 6. Confusion Matrix Logistic Regression

Description of the Classification Report Results in Figure 7 shows accuracy, precision, recall, and F1-Score as below:

$$\text{Accuracy} = \frac{139 + 126}{139 + 126 + 28 + 15}$$
$$= 0.85714$$
$$= 86\%$$

$$\text{Precision} = \frac{139}{139 + 28}$$
$$= 0.83233$$
$$= 83\%$$

$$\text{Recall} = \frac{139}{139 + 15}$$
$$= 0.90259$$
$$= 90\%$$

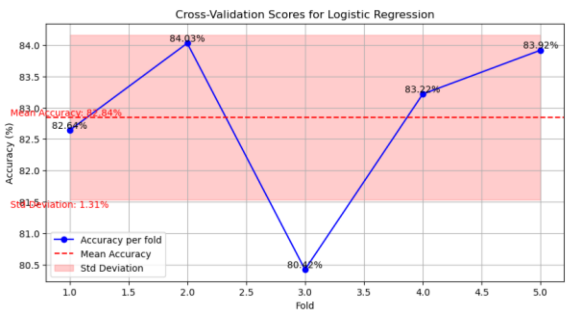
$$\text{F1} - \text{Score} = 2 \times \frac{0.83233 \times 0.90259}{0.83233 + 0.90259}$$
$$= 0.8660$$
$$= 87\%$$

```
Classification Report Algoritma Logistic Regression:
              precision    recall  f1-score   support

           0       0.89      0.82      0.85       154
           1       0.83      0.90      0.87       154

    accuracy                           0.86       308
   macro avg       0.86      0.86      0.86       308
weighted avg       0.86      0.86      0.86       308
```

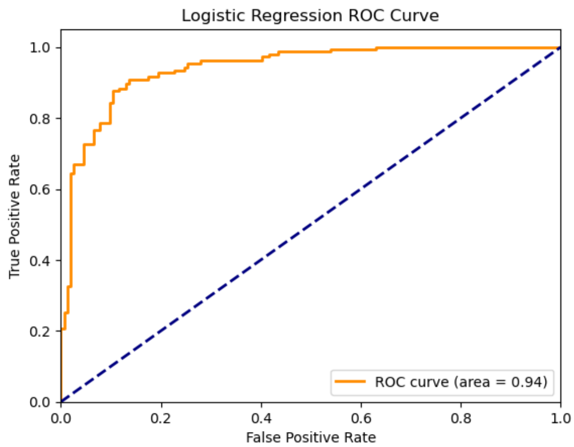
Source: (Research Result, 2024)
Figure 7. Logistic Regression Evaluation Results



Source: (Research Result, 2024)
Figure 8. Cross-Validation Logistic Regression Graph

Evaluation of the Logistic Regression model using cross-validation with 5 folds shows an average accuracy of 0.828, marked by the dashed red line on the plot, and a standard deviation of 0.013, indicating small variations between folds as shown in Figure 8. The accuracy scores for each fold are: Fold 1 has an accuracy of 0.826, Fold 2 increases to 0.840, Fold 3 drops drastically to 0.804 (the lowest), Fold 4 increases again to 0.832, and Fold 5 reaches 0.839.

PILAR Nusa Mandiri: Journal of Computing and Information System
Vol. 20, No. 2 September 2024 | DOI: 10.33480/pilar.v20i2.5742



Source: (Research Result, 2024)
Figure 9. Kurva ROC Logistic Regression

The AUC value ranges between 0 and 1, with higher values indicating a better model ability to distinguish between positive and negative classes. An AUC of 94% means that the model has a very high ability to predict the correct class as shown in Figure 9.

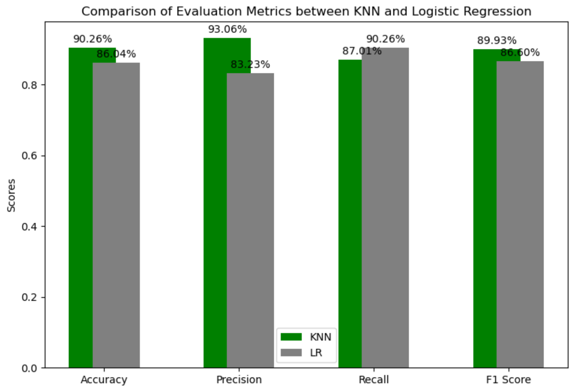
Combining the actual label data and the predicted data from the logistic regression classification, 265 data were predicted correctly and 43 data were predicted incorrectly.

Comparison Results

Figure 10 is a diagram illustrating the comparison of evaluation results between the K-Nearest Neighbor (K-NN) algorithm and Logistic Regression. Based on the performance evaluation of the K-Nearest Neighbor (K-NN) and Logistic Regression methods using accuracy, precision, recall, and f1-score, K-NN with Euclidean distance and a k value of 3 on 30% testing data achieved 90% accuracy, 93% precision, 87% recall, and a 90% f1-score. In contrast, Logistic Regression with a sigmoid function on the same data yielded 86% accuracy, 83% precision, 90% recall, and an 86% f1-score.

K-NN's superiority can be attributed to its nonparametric nature, which allows it to handle complex and nonlinear data without assumptions about data distribution. K-NN excels with irregular decision boundaries by considering nearest neighbors and effectively captures local patterns with small k values. Its instance-based learning adapts well to intricate feature relationships, potentially outperforming the linear approach of Logistic Regression.

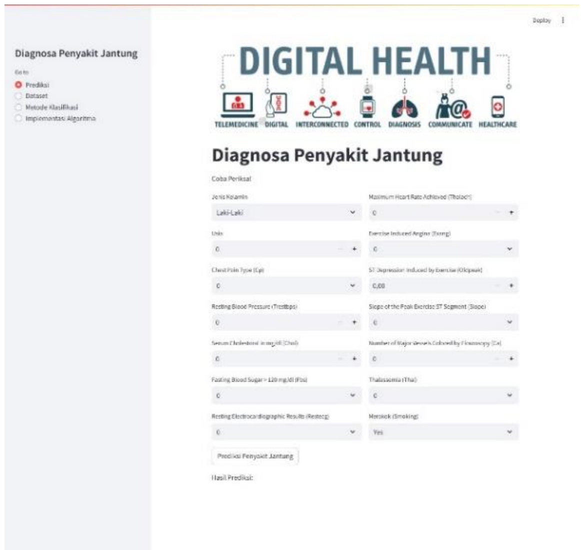
While Logistic Regression shows slightly higher recall, K-NN provides a better balance between precision and recall, as reflected in its higher f1-score, making it more effective at minimizing both false positives and false negatives in this classification task.



Source: (Research Result, 2024)
Figure 10. Comparison of Algorithms Result Graph

The cross-validation results show the advantages of each K-Nearest Neighbors (KNN) and Logistic Regression models. The Logistic Regression model has an average accuracy of 82.84% with a standard deviation of 0.0132, indicating low variation and consistent performance between folds. The highest fold reaches 84.03% (Fold 2), and the lowest 80.42% (Fold 3), indicating good generalization ability without overfitting. Meanwhile, the KNN model has a higher average accuracy of 84.93%, but with a standard deviation of 0.0177, indicating greater variation between folds. The highest fold reaches 88.19% (Fold 2) and the lowest 83.22% (Fold 5), indicating KNN's sensitivity to data variation and potential performance fluctuations depending on the subset of data used.

Interface View



Source: (Research Result, 2024)
Figure 11. Heart Prediction diagnostic display

After comparing the methods, the next step is to create a program output on the Streamlit framework with a display like Figure 11. To run it, enter the values according to the attributes available in the dataset, then click the "Heart Disease Prediction" button. The diagnosis results will be displayed afterwards.

CONCLUSION

This study implements the K-NN and Logistic Regression algorithms to support the diagnosis of heart disease, comparing the classification results of both using accuracy, precision, recall, and f1-score. With K-NN (k value = 3) on 30% testing data, an accuracy of 90%, precision of 93%, recall of 87%, and f1-score of 90% were obtained. Logistic Regression produces an accuracy of 86%, precision of 83%, recall of 90%, and f1-score of 86% on the same testing data. Based on these results, K-NN is proven to be superior to Logistic Regression in classifying heart disease datasets. The implementation of these classification results on the web-based Streamlit framework can help in monitoring heart diagnosis. To improve future research, it would be useful to explore the performance of these models with larger and more diverse datasets, and to experiment with advanced techniques such as feature selection and hyperparameter tuning, which may further improve the accuracy and robustness of diagnostic predictions.

REFERENCE

Almustafa, K. M. (2020). Prediction of heart disease and classifiers' sensitivity analysis. *BMC Bioinformatics*, *21*(1), 1–18. https://doi.org/10.1186/s12859-020-03626-y

Alnuaimi, A. F. A. H., & Albaldawi, T. H. K. (2024). An overview of machine learning classification techniques. *BIO Web of Conferences*, *97*, 1–24. https://doi.org/10.1051/bioconf/20249700133

Ariawan, M. D., Triayudi, A., & Sholihati, I. D. (2020). Perancangan User Interface Design dan User Experience Mobile Responsive Pada Website Perusahaan. *Jurnal Media Informatika Budidarma*, *4*(1), 161. https://doi.org/10.30865/mib.v4i1.1896

Bianto, M. A., Kusrini, K., & Sudarmawan, S. (2020). Perancangan Sistem Klasifikasi Penyakit Jantung Mengunakan Naïve Bayes. *Creative Information Technology Journal*, *6*(1), 75. https://doi.org/10.24076/citec.2019v6i1.231

G, A., Ganesh, B., Ganesh, A., Srinivas, C., Dhanraj, & Mensinkal, K. (2022). Logistic regression technique for prediction of cardiovascular disease. *Global Transitions Proceedings*, *3*(1), 127–130. https://doi.org/10.1016/j.gltp.2022.04.008

Hartono, A., Aska Dewi, L., Yuniarti, E., Tahta Hirani Putri, S., Surya Harahap, T., & Hartono, A. (2023). Machine Learning Classification for Detecting Heart Disease with K-NN Algorithm, Decision Tree and Random Forest. *Eksakta : Berkala Ilmiah Bidang MIPA*, *24*(04), 513–522. Retrieved from http://www.eksakta.ppj.unp.ac.id/index.php/eksakta

Ibrahim, I., & Abdulazeez, A. (2021). The Role of Machine Learning Algorithms for Diagnosing Diseases. *Journal of Applied Science and Technology Trends*, *2*(01), 10–19. https://doi.org/10.38094/jastt20179

Kadhim, M. A., & Radhi, A. M. (2023). Heart disease classification using optimized Machine learning algorithms. *Iraqi Journal for Computer Science and Mathematics*, *4*(2), 31–42. https://doi.org/10.52866/ijcsm.2023.02.02.004

Kavitha, M., Gnaneswar, G., Dinesh, R., Sai, Y. R., & Suraj, R. S. (2021). Heart Disease Prediction using Hybrid machine Learning Model. *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, 1329–1333. IEEE. https://doi.org/10.1109/ICICT50816.2021.9358597

Li, J. P., Haq, A. U., Din, S. U., Khan, J., Khan, A., & Saboor, A. (2020). Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare. *IEEE Access*, *8*(Ml), 107562–107582. https://doi.org/10.1109/ACCESS.2020.3001149

Muharram, F. R., Multazam, C. E. C. Z., Mustofa, A., Socha, W., Andrianto, Martini, S., … Yi-Li, C. (2024). The 30 Years of Shifting in The Indonesian Cardiovascular Burden—Analysis of The Global Burden of Disease Study. *Journal of Epidemiology and Global Health*, *14*(1), 193–212. https://doi.org/10.1007/s44197-024-00187-8

Prasetyo, A. B., & Laksana, T. G. (2022). Optimasi Algoritma K-Nearest Neighbors dengan Teknik Cross Validation Dengan Streamlit (Studi Data: Penyakit Diabetes). *Journal of Applied Informatics and Computing (JAIC)*, *6*(2), 194–204. Retrieved from http://jurnal.polibatam.ac.id/index.php/JAIC

Radhika, R., & Thomas George, S. (2021). Heart Disease Classification Using Machine Learning Techniques. *Journal of Physics: Conference Series*, *1937*(1). https://doi.org/10.1088/1742-6596/1937/1/012047

Sumara, R., Ari, N., & Indarti, I. (2022). Identifikasi Faktor Kejadian Penyakit Jantung Koroner

Vol. 20, No. 2 September 2024 | DOI: 10.33480/pilar.v20i2.5742

Terhadap Wanita Usia ≤ 50 Tahun di RSU Haji Surabaya. *Jurnal Manajemen Asuhan Keperawatan*, *6*(2), 53–59. https://doi.org/10.33655/mak.v6i2.134

Ullah, F., Chen, X., Rajab, K., Al Reshan, M. S., Shaikh, A., Hassan, M. A., … Davidekova, M. (2022). An Efficient Machine Learning Model Based on Improved Features Selections for Early and Accurate Heart Disease Predication. *Computational Intelligence and Neuroscience*, *2022*, 1–12. https://doi.org/10.1155/2022/1906466

Virani, S. S., Alonso, A., Aparicio, H. J., Benjamin, E. J., Bittencourt, M. S., Callaway, C. W., … Tsao, C. W. (2021). Heart Disease and Stroke Statistics—2021 Update. *Circulation*, *143*(8), E254–E743. https://doi.org/10.1161/CIR.0000000000000950

Yadav, S. S. (2020). *Application of Machine Learning for the Detection of Heart Disease*. (Icimia), 165–172.

Zuama, R. A., Rahmatullah, S., & Yuliani, Y. (2022). Analisis Performa Algoritma Machine Learning pada Prediksi Penyakit Cerebrovascular Accidents. *Jurnal Media Informatika Budidarma*, *6*(1), 531. https://doi.org/10.30865/mib.v6i1.3488