Terbit *online* pada laman : http://teknosi.fti.unand.ac.id/

Jurnal Nasional Teknologi dan Sistem Informasi

| ISSN (Print) 2460-3465 | ISSN (Online) 2476-8812 |

Research Article

Improving Coronary Heart Disease Detection Using K-Means Clustering Techniques

Junaidy Budi Sanger[a], *Liza Wikarsa*[b], *Angelica Theresia Taulu*[c]

[a,b,c]*Program Studi Teknik Informatika, Fakultas Teknik, Universitas Katolik De La Salle Manado, Sulawesi Utara, Indonesia*

INFORMASI ARTIKEL

Sejarah Artikel:
Diterima Redaksi: 28 November 2024
Revisi Akhir: 28 Agustus 2025
Diterbitkan *Online*: 01 September 2025

KEYWORD

Coronary Heart Disease,
Early Disease,
K-Means,
Clustering,
North Sulawesi

KORESPONDENSI

lwikarsa@unikadelasalle.ac.id

A B S T R A C T

The heart is a vital organ in the cardiovascular system, playing a crucial role in blood circulation and supplying the body with oxygen and nutrients. Cardiovascular diseases, particularly coronary heart disease (CHD), are the leading cause of death worldwide. In Indonesia, especially in North Sulawesi, the high prevalence of CHD is indicative of the effects of an unhealthy lifestyle. This study employs the K-Means clustering method to identify the early risk of CHD based on eight common symptoms, including chest pain, nausea, shortness of breath, heartburn, a history of hypertension, obesity, diabetes, and genetics. This innovative approach integrates these early warning signs and categorizes the risk into three groups: low CHD risk (C1), moderate CHD risk (C2), and high CHD risk (C3). The detection results are provided based on responses collected through a questionnaire within an application, aiming to raise awareness of CHD and encourage users to seek further health evaluations and adopt healthier lifestyles.

1. INTRODUCTION

According to data from the World Health Organization (WHO) in 2023, cardiovascular disease is projected to be the leading cause of death globally, with an estimated annual death toll of 17.9 million people. [1]. The Ministry of Health (Kemenkes) reports that, from 2014 to 2019, cardiovascular diseases were the primary cause of death and disability in Indonesia, based on research from the Global Burden of Disease and the Institute for Health Metrics and Evaluation (IHME) [2]. Coronary heart disease (CHD), a major contributor to global mortality, occurs when the coronary arteries, which supply blood to the heart, become blocked by fat buildup [3]. As fat accumulation increases, the arteries narrow, reducing blood flow to the heart [4]. Data from the Basic Health Research (Riskesdas) in 2018 shows that the national prevalence of CHD in Indonesia is 1.5%, with 1.8% in North Sulawesi (SULUT) [5]. It has been reported that 50% of CHD patients are at risk of sudden cardiac death or sudden cardiac arrest [5]. The K-Means clustering method, a common technique in data mining, is used to detect early signs of CHD. Clustering helps identify similar patterns in data, with the first step being data pre-processing. Clustering analysis groups the data and determines a centroid as the central point, which is used for further study, such as characterization, attribute selection, and cluster classification [6]. K-Means, an unsupervised learning method, is widely used for these purposes in research [7]. As an unsupervised learning method, K-means is often used in research for this purpose [8].

The K-Means method has previously been applied in disease-related research. In a study by [6], heart disease risk levels were categorized into three clusters based on age: Cluster 1 (low risk), Cluster 2 (medium risk), and Cluster 3 (high risk). The analysis was conducted on 1,025 data points, with 355 respondents in

https://doi.org/10.25077/TEKNOSI.v11i2.2025.107-117

Cluster 1, 208 in Cluster 2, and 462 in Cluster 3 [9]. Meanwhile, Tedju et al. grouped the level of hypertension in the city of Kupang into 3 clusters, namely standard, pre-hypertension, and stage I hypertension. The results obtained from data from the Kupang City Health Service based on systolic values and diastolic values show that Cluster 1 (average) has a systolic value of 109, 56 and diastolic 70.91, Cluster 2 (pre-hypertension) with a systolic value of 127.55 and diastolic 84.09, and Cluster 3 (stage I hypertension) with a systolic value of 156.26 and diastolic 98.56 [10]. Anggraini et al. also grouped patients' diseases at the Ujung Batu community health center into two clusters based on data from 3,875 cases. In Cluster 1, there were 710 patients with infectious diseases, with 28 instances primarily involving acute upper respiratory infections, mostly from Sukadam Village. In Cluster 2, there were 3,165 patients, with 218 cases predominantly involving acute upper respiratory infections at multiple and unspecified sites, most of whom were from Ngaso Village [11].

There are several differences between this research and previous research. Using the K-Means method, this research will develop a web-based application to identify early CHD risk levels based on initial symptoms. Based on interviews conducted with the heart specialists in Manado City, it is known that there are eight early symptoms of CHD, namely chest pain, nausea, shortness of breath, heartburn, history of hypertension, obesity, history of diabetes, and genetics. These eight initial symptoms are used as independent variables. Meanwhile, the study's dependent variables consist of three categories of CHD risk levels identified as Cluster 1 (low risk), Cluster 2 (medium risk), and Cluster 3 (high risk) based on previous research findings [9]. The results of detecting the CHD risk level will help individuals follow up by undergoing a more comprehensive health examination. The application was built on a web-based platform, allowing users to access it easily anywhere and anytime.

2. METHOD

The software development method used is Rapid Application Development (RAD), a linear sequential approach that focuses on short development cycles. The component-based construction approach allows rapid progress in developing these models [12]. Several stages in the RAD method involve users, and analysis can be seen in Figure 1 [13].

Figure 1. RAD stages [13]

2.1. Requirements Planning

At this stage, system analysts and users collaborate to define the system's objectives and identify the requirements to achieve these goals. They also analyze all system components required by the users. The output of this phase is data collection and processing, problem identification, target user analysis, definition of requirements specifications, and analysis of the application of the K-Means method.

This research data was collected through a questionnaire distributed using Google Forms. Respondents were parishioners from 16 spiritual regions of the Sacred Heart of Mary Cathedral Parish in Manado. The Cathedral has 1,846 people [14]. Hence, the research sample was calculated using the formula of Isaac and Michael with a significant level of 5%. So, the sample size for this research was 310 respondents.

Table 1. List of Questionnaire

Code	Question
P1	1. How much do you understand about coronary heart disease (CHD)? a. Do not understand b. That's enough to understand c. Very understanding
P2	2. Have you ever had a heart health check-up? a. Do not understand b. That's enough to understand c. Very understanding
P3	3. Do you feel chest pain on the left side? (Pain like burning, sharp stabbing, and may radiate to the jaw or left arm) a. No pain b. It's quite painful c. Very painful
P4	4. Do you feel nauseous when doing daily activities like walking or climbing stairs? a. Not nauseous b. Quite nauseous c. Very nauseous
P5	5. Do you experience shortness of breath when doing daily activities such as walking or climbing stairs? a. Not short of breath b. Just short of breath c. Very short of breath
P6	6. Do you feel heartburn when doing daily activities such as walking or climbing stairs? a. No pain b. It's quite painful

https://doi.org/10.25077/TEKNOSI.v11i2.2025.107-117

		c. Very painful
P7	7.	Do you have a history of hypertension? a. Yes b. No
P8	8.	Are you overweight? (Obesity) a. Yes b. No
P9	9.	Do you have a history of diabetes? a. Yes b. No
P10	10.	Is there a history of coronary heart disease (CHD) in your family? a. Yes b. No

An r-calculated value will be compared to the r-table value to test the validity. The r-table value for N = 300 at the 5% confidence level used is 0.113 [15]. Table 2 below is a comparison table of r-calculation and r-table.

Table 2. Validity Test of r-calculated Values and r-table Values

No.	Code	r-calculate	r-table	Conclusion
1.	P1	0,271	0,113	Valid
2.	P2	0,196	0,113	Valid
3.	P3	0,616	0,113	Valid
4.	P4	0,488	0,113	Valid
5.	P5	0,670	0,113	Valid
6.	P6	0,451	0,113	Valid
7.	P7	0,507	0,113	Valid
8.	P8	0,556	0,113	Valid
9.	P9	0,197	0,113	Valid
10.	P10	0,451	0,113	Valid

Table 2 shows the r-calculated value > r-table value, so it can be concluded that the questions used have been proven valid. As a basis for decision-making to test reliability, a comparison will be carried out with the condition that if the Cronbach's Alpha value is > 0.60, then the question is declared reliable or consistent. The following are the test results of 310 cases that have been carried out.

Table 3. Case Processing Summary

Case Processing Summary			
		N	%
Cases	Valid	310	100.0
	Excluded[a]	0	.0
	Total	310	100.0

Table 4. Reliability Statistics

Reliability Statistics	
Cronbach's Alpha	N of Items
.630	10

In Table 3, 310 of the tested data are valid. Table 4 shows that the Cronbach's Alpha value is 0.630. Since 0.630 > 0.60, the questions used are declared valid.

Regarding the questionnaire, from the 310 data collected, there were 241 people (77.7%) who did not understand CHD, so it can be seen that there are still many people who do not know about CHD. There are 253 people (81.6%) who have never had a heart health check-up, so it can be seen that there are still many people who do not know that there is a possibility that they are already experiencing the early symptoms of CHD. The most striking symptom of CHD risk in the data collected was that 41 people (around 13.2%) had a history of hypertension. In the questionnaire distributed, there were 193 people (62.3%) who were in the age range > 46 years, 43 people (13.9%) in the age range 36-45 years, 41 people (13.2%) in the age range 15 -25 years, and 33 people (10.6%) in the age range 26-35 years. Overall, people in the region have a low CHD risk level (C1).

2.2. Analysis of the Application of the K-Means Method

This section will analyze the application of the K-Means method to detect and classify the level of CHD risk among individuals in the spiritual area. For this reason, 10 sample data will be used to demonstrate how to calculate using this method. The steps for implementing the K-Means method are as follows [16]:

1. Preparing data sampling
   The following is data on people from the spiritual area and the initial symptoms and causal factors: chest pain, nausea, shortness of breath, heartburn, hypertension, obesity, diabetes, and genetics.
2. Determine the number of clusters. There are 3 clusters distinguished in this research, namely:
   C1 = Low CHD Risk
   C2 = Moderate CHD Risk
   C3 = High CHD Risk.
3. Determine the value of the center point or centroid
   The center point value is determined by calculating the difference between each data point. The centroid is determined based on the total data value, which includes C1 = 8-9, C2 = 10-11, and C3 = 12.

Table 5. Total data value

No.	Total	*Cluster*
1.	12	C3
2.	12	C3
3.	10	C2
4.	10	C2
5.	8	C1
6.	9	C1
7.	11	C2
8.	10	C2
9.	8	C1
10.	10	C2

The centroid values for clusters C1, C2, and C3 are taken from data points 1, 3, and 6 in the 10 data points above. Each cluster has eight columns, each containing the values for a specific symptom. The data selected as C1, C2, and C3 in the sampling were chosen randomly. The following is a data table of centroid values for iteration 1.

Table 6. Centroid Values

C1	**1**	**1**	**1**	**2**	**1**	**1**	**1**	**1**
C2	1	1	1	1	1	2	1	2
C3	1	1	2	1	2	2	1	2

4. Calculate the distance of each piece of data based on the closest distance between the data and the centroid value using the Euclidean Distance formula.

Table 7. Euclidean Distance Calculation: Iteration 1

d(1,1)	**2.24**	**d(6,1)**	**0.00**
d(1,2)	1.41	d(6,2)	1.73
d(1,3)	0.00	d(6,3)	2.24
d(2,1)	2.24	d(7,1)	2.00
d(2,2)	1.41	d(7,2)	1.73
d(2,3)	1.41	d(7,3)	1.73
d(3,1)	1.73	d(8,1)	1.73
d(3,2)	0.00	d(8,2)	1.41
d(3,3)	1.41	d(8,3)	1.41
d(4,1)	1.73	d(9,1)	1.00
d(4,2)	1.41	d(9,2)	1.41
d(4,3)	2.00	d(9,3)	2.00
d(5,1)	1.00	d(10,1)	1.00
d(5,2)	1.41	d(10,2)	2.00
d(5,3)	2.00	d(10,3)	2.45

5. Group the data based on the closest distance to the middle value.

Table 8. Data Grouping Iteration 1

No.	C1	C2	C3	Hasil
1	2.24	1.41	0.00	C3
2	2.24	1.41	1.41	C2
3	1.73	0.00	1.41	C2
4	1.73	1.41	2.00	C2
5	1.00	1.41	2.00	C1
6	0.00	1.73	2.24	C1
7	2.00	1.73	1.73	C2
8	1.73	1.41	1.41	C2
9	1.00	1.41	2.00	C1
10	1.00	2.00	2.45	C1

6. Determine the new centroid/center value.
7. Recalculate by repeating steps 3 and 4.
   Based on the calculations carried out in the previous step, the new centroid value in iteration 2 is as follows:

Table 9. The Values of New Centroid

C1	**1.00**	**1.25**	**1.00**	**1.50**	**1.00**	**1.00**	**1.00**	**1.00**
C2	1.60	1.00	1.00	1.00	1.60	1.60	1.00	1.80
C3	1.00	1.00	2.00	1.00	2.00	2.00	1.00	2.00

Table 10. Calculation of Euclidean Distance, Iteration 2

d(1,1)	**2.08**	**d(6,1)**	**0.56**
d(1,2)	1.31	d(6,2)	1.65
d(1,3)	0.00	d(6,3)	2.24
d(2,1)	2.08	d(7,1)	1.82
d(2,2)	0.72	d(7,2)	0.85
d(2,3)	1.41	d(7,3)	1.73
d(3,1)	1.52	d(8,1)	1.52
d(3,2)	0.96	d(8,2)	0.96
d(3,3)	1.41	d(8,3)	1.41
d(4,1)	1.52	d(9,1)	0.56
d(4,2)	1.15	d(9,2)	1.31
d(4,3)	2.00	d(9,3)	2.00
d(5,1)	0.56	d(10,1)	0.90
d(5,2)	1.31	d(10,2)	1.93
d(5,3)	2.00	d(10,3)	2.45

Table 11. Data Grouping Iteration 2

No	C1	C2	C3	Hasil
1	2.08	1.31	0.00	C3
2	2.08	0.72	1.41	C2
3	1.52	0.96	1.41	C2
4	1.52	1.15	2.00	C2
5	0.56	1.31	2.00	C1
6	0.56	1.65	2.24	C1
7	1.82	0.85	1.73	C2
8	1.52	0.96	1.41	C2
9	0.56	1.31	2.00	C1
10	0.90	1.93	2.45	C1

8. The grouping is declared complete if no data moves to another cluster.
   Based on the calculations, it can be seen that no data has been moved to other clusters, so the calculation is declared complete. Data 1 is included in the high CHD risk level (C3), data 2, 3, 4, 7, 8 are included in the moderate CHD risk level (C2), and data 5, 6, 9, 10 are included in the low CHD risk level (C1).

2.3. RAD Design Workshop

Analysts and programmers collaborate to create the system at this stage, presenting users with prototypes to gather feedback.

https://doi.org/10.25077/TEKNOSI.v11i2.2025.107-117

Developers then refine the system based on user input, analyzing and enhancing modules as needed.

UML modeling consists of Use Case Diagram, Class Diagram, and Activity Diagram. The following is a Use Case Diagram of the Application for Early Detection of Coronary Heart Disease using the K-Means method.

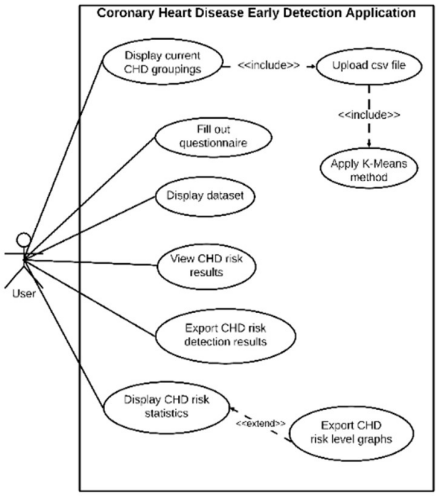

Figure 2. Use Case Diagram

Users can display the latest CHD grouping, upload CSV files, complete questionnaires, view datasets, view CHD risk results, export CHD risk detection results, view CHD risk statistics, and export CHD risk level graphs.

The following is the Class Diagram of the Early Detection Application for Coronary Heart Disease using the K-Means method.

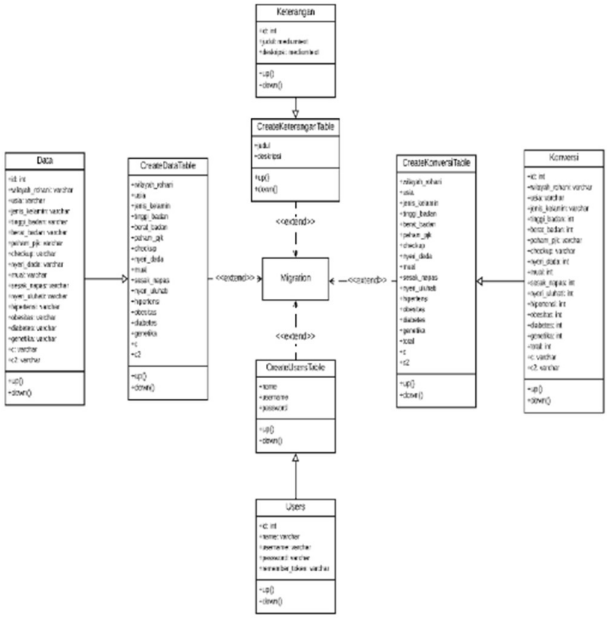

Figure 3. Class Diagram Migration

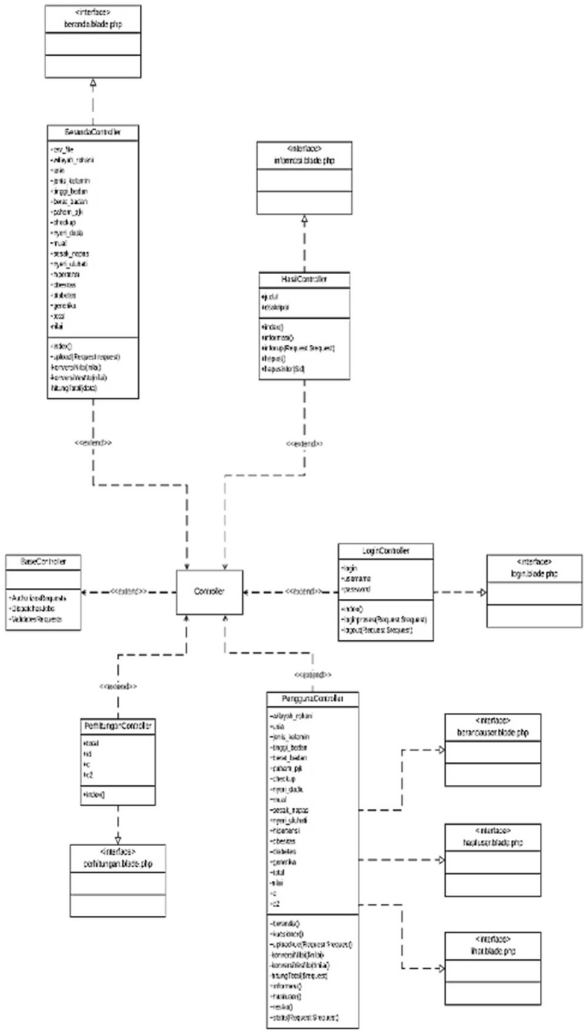

Figure 4. Class Diagram Controller

The following is the Activity Diagram of the Early Detection Application for Coronary Heart Disease using the K-Means method. In Figure 5, an activity depicts the admin in the system who enters the application, upload parishioner data, carry out the calculation process, check the system calculation results, and check the suitability of the results at each iteration.

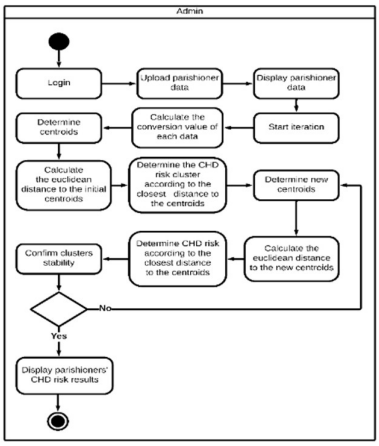

Figure 5. Activity Diagram for Implementing the K-Means Method

Figure 6 depicts the activity of the user who enters the application and be able to see the CHD risk data table for parishioners.

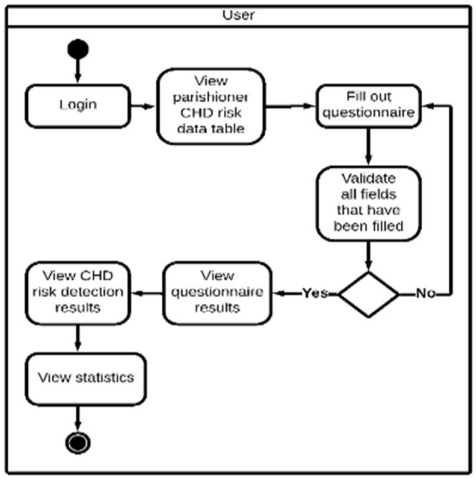

Figure 6. User Activity Diagram

Fill out the questionnaire, verify that all fields have been completed, view the questionnaire results, review the CHD risk detection results, and access the statistics.

In Figure 7, an entity relationship diagram describes the system's database being built. There are users, information, data, and conversion tables. Users will store user information and the login and logout processes. Conversion will store and calculate CHD risk data.

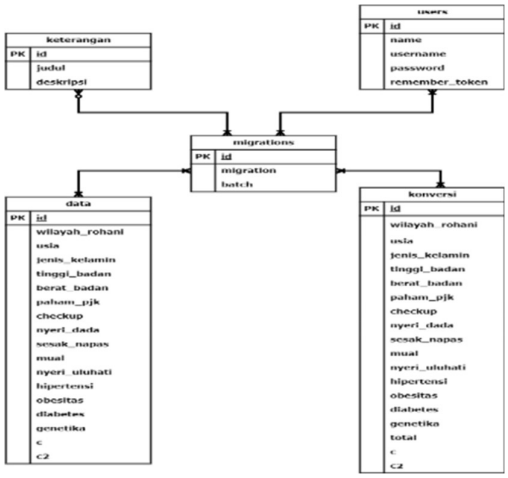

Figure 7. Entity Relationship Diagram

The following models the Early Detection of Coronary Heart Disease Application program module using the K-Means method.

Table 12. Program Module Modeling

No.	Module	Description
1	Application connection to the server	The local server used is Apache in the XAMPP application to connect the application to the server
2	Database connection	The DBMS used is MySQL to connect the database to the server
3	Input questionnaire results	Input area data in the application that has been created
4	Calculations using the K-Means method	Calculate the initial centroid using the K-Means method on the data that has been input.
5	Displays CHD risk statistics from respondents	Displays CHD risk statistics based on the data that has been input

3. RESULTS AND DISCUSSION

3.1 Implementation Rules

This section describes the implementation rules for the application being built. The rules used in the application are as follows:

1. The collected data is 310 pieces from community data collection.
2. Perform data conversion. This process involves non-numerical attribute transformation and normalization to ensure accurate clustering results.
3. Data results that include low CHD risk (C1) are worth 9, moderate CHD risk (C2) are worth 10-11, and high CHD risk (C3) are worth 12.
4. Ten testing points are used, selected representatively from the available data set to ensure diversity and accuracy.

3.2 Application Implementation

The following figures will illustrate the application implementation based on the K-means method, following the implementation guidelines outlined above.

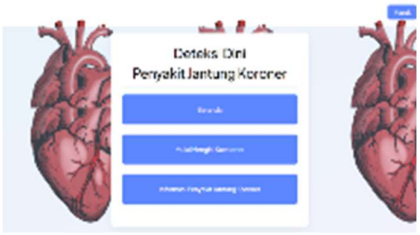

Figure 8. Main Page

In Figure 8, there are four menu buttons: namely, Home to view the home page, the Start Filling in Questionnaire button to fill out a questionnaire regarding CHD risk, the Coronary Heart Disease Information button to view information about CHD, and the Login button to access the admin login page.

https://doi.org/10.25077/TEKNOSI.v11i2.2025.107-117

Figure 9. User CHD Risk Results Page

On the User CHD Risk Results Page, questionnaire data filled in by the user is displayed, allowing users to view their CHD risk. Users can also export risk results in PDF and Excel format.

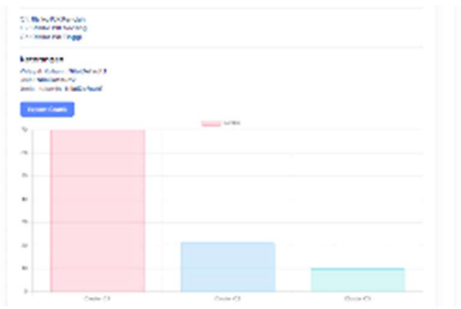

Figure 10. CHD Risk Statistics Page

The CHD Risk Statistics Page has a data filter section where users can select filters based on spiritual area, age, and gender. Once a filter is selected, it will be shown on the graph according to the cluster to which the selected filter belongs.

Figure 11. Admin Login Page

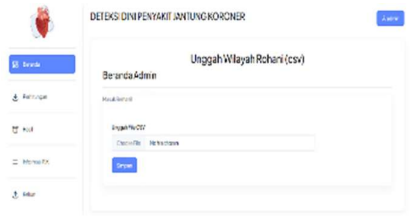

Figure 12. Admin Home Page

Figure 13. Calculation Page

The Calculation Page features an initial data button, an Iteration 1 button, and an Iteration 2 button to view the calculation process. There is a data description, and much data has been uploaded. Apart from that, a spiritual area data table has been converted into an initial data display.

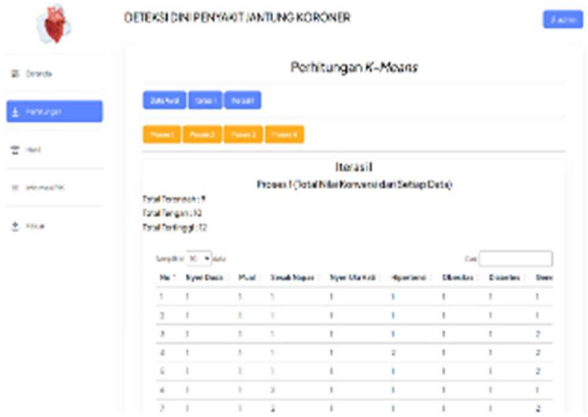

Figure 14. Page Iteration 1 Process 1

The Iteration 1 Process 1 Admin page has a button to view the process or steps of the K-Means calculation. The lowest, middle, and highest totals from the data and the region data table have also been converted.

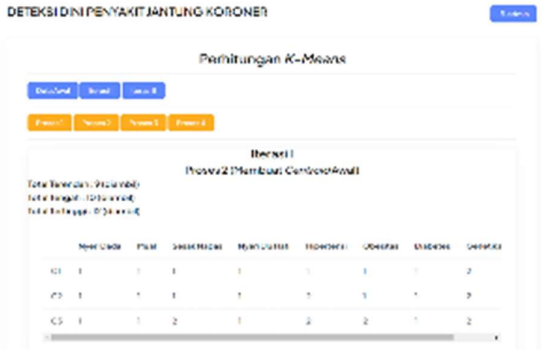

Figure 15. Page Iteration 1 Process 2

The Iteration 1 Process 2 Admin page has a button to view the process or steps of the K-Means calculation. There are also the lowest, middle, and highest totals from the data and the initial centroid data table.

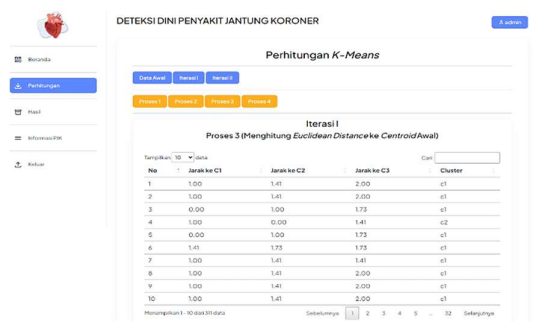

Figure 16. Page Iteration 1 Process 3

The Iteration 1 Process 3 Admin page has a button to view the process or steps of the K-Means calculation. There is also a table of results from calculating the Euclidean distance to the initial centroid.

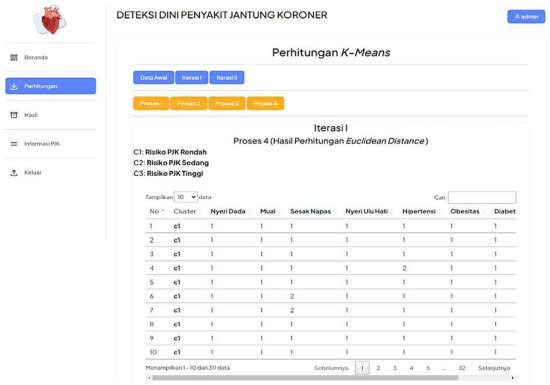

Figure 17. Page Iteration 1 Process 4

The Iteration 1 Process 4 Admin page has a button to view the process or steps of the K-Means calculation. There is also a table of results from the Euclidean distance calculation.

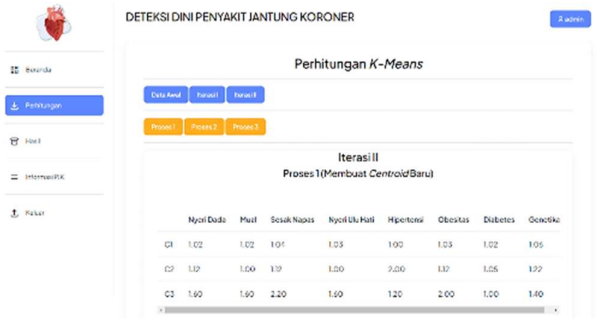

Figure 18. Page Iteration 2 Process 1

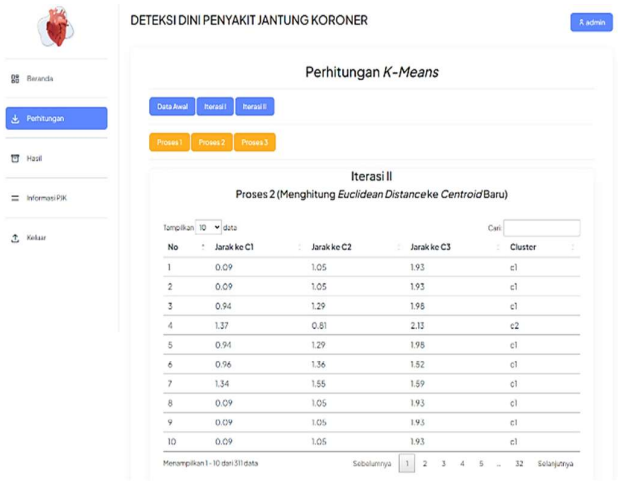

Figure 19. Page Iteration 2 Process 2

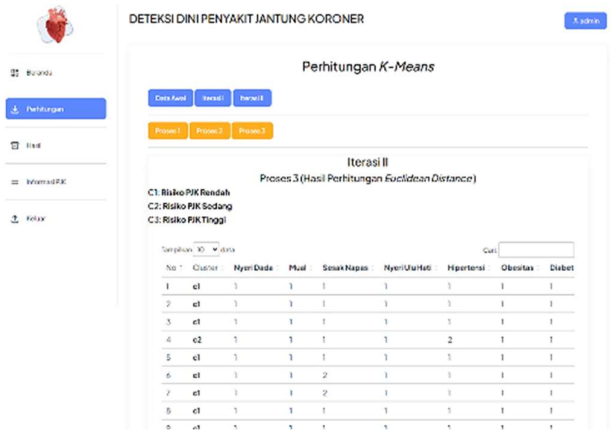

Figure 20. Page Iteration 2 Process 3

https://doi.org/10.25077/TEKNOSI.v11i2.2025.107-117

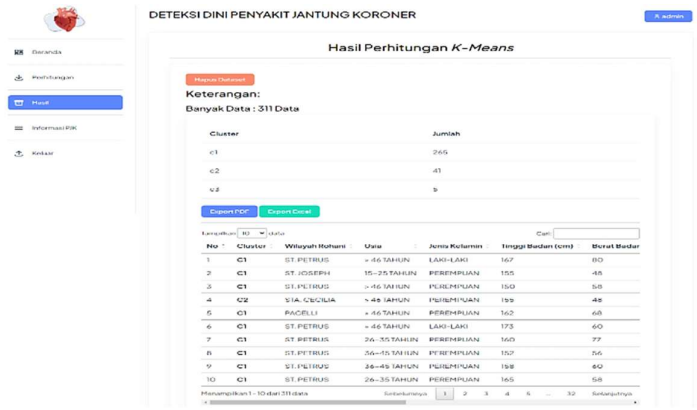

Figure 21. K-Means Calculation Results Page

The K-Means Calculation Results Page has a button to delete datasets, a description of lots of data, and a button to export calculation results in PDF and Excel format. It also has a table of data results from the K-Means calculation.

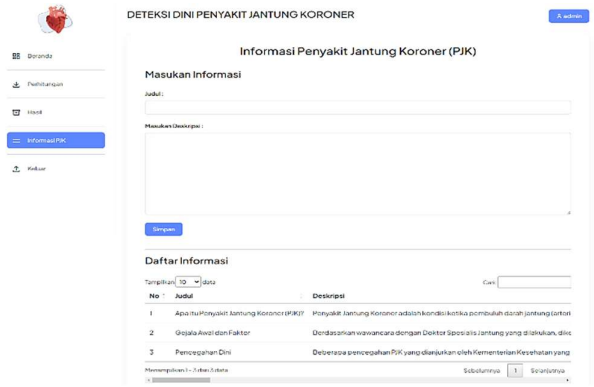

Figure 22. CHD Information Page

The CHD Information Page includes a title, description, and save button for entering CHD-related information. Additionally, a table displays the information the admin has entered.

3.3 Analysis of Test Results

The results of the tests concluded that the implementation of the user and system requirements specifications had been successful and ran smoothly. The results of the test analysis are as follows:

1. The application successfully carries out the login process for the admin.
2. The application has successfully uploaded congregation data as a CSV file.
3. The application successfully performs and displays K-Means calculations using the congregation dataset uploaded as a CSV file.
4. The application successfully displays the uploaded data consisting of 310 respondents.
5. The application successfully shows the clusters and the number of individuals in each. There were 265 respondents in the Low CHD Risk (C1), 40 respondents in the Medium CHD Risk (C2), and five respondents in the High CHD Risk (C3) cluster.
6. The application successfully saves the title and description of the PJK information filled in.
7. The application successfully displays the title and description of the PJK information filled in.
8. The application has successfully uploaded and displayed the user's answer results.
9. The application successfully displays the user's CHD risk detection results.
10. The application successfully displays CHD risk statistics in the spiritual area according to the uploaded congregation dataset.

4 CONCLUSIONS

The conclusions obtained are the importance of Early Detection: The study emphasizes the significance of early detection in identifying the risk of coronary heart disease (CHD), a leading cause of death globally, including in Indonesia, particularly in North Sulawesi.

Application of K-Means Clustering: Using the K-Means clustering method, the study developed a web-based application that classifies individuals into three CHD risk levels (low, moderate, and high) based on eight early symptoms. This method proves effective for grouping individuals according to their risk levels.

Results of the Application: The application was successful in various functions, including logging in, uploading data, performing K-Means calculations, displaying risk categories, and saving and displaying CHD-related information. It also successfully processed responses from 310 respondents, categorizing them into low, moderate, and high-risk clusters.

Statistical Insights: The application provided valuable statistical insights, such as 265 respondents classified under low CHD risk (C1), 40 under moderate risk (C2), and five under high risk (C3). This helps raise CHD awareness and encourages users to take preventive measures.

Practical Implications: The application's user-friendly web-based platform ensures easy access for individuals, allowing them to track their CHD risk status and seek further health evaluations if necessary, potentially leading to improved public health outcomes.

Contribution to Healthcare: The study contributes to the ongoing effort to reduce CHD-related deaths by promoting early detection and healthier lifestyle choices, ultimately supporting individuals in preventing more severe cardiovascular conditions.

The future works for this study are as follows:

1. Integration of Additional Risk Factors: Future studies could expand the early symptoms and risk factors for CHD, integrating more variables such as family medical history, lifestyle habits (e.g., smoking, physical activity), and cholesterol levels, to enhance the accuracy of risk predictions.
2. Machine Learning Model Enhancement: The application could incorporate more advanced machine learning models beyond K-Means clustering, such as decision trees, random forests, or neural networks, to improve the accuracy and robustness of CHD risk classification.
3. Mobile Application Development: To increase accessibility, the web-based application could be further developed into a mobile application, allowing users to access smartphone risk detection features for real-time monitoring and alerts.
4. Real-Time Data Integration: Future application versions could integrate real-time health data from wearable devices (e.g., heart rate monitors, blood pressure trackers) to continuously monitor CHD risk and offer personalized health recommendations.
5. Longitudinal Study and Data Collection: A larger and more diverse dataset could be collected over a more extended period to analyze the progression of CHD risk and refine clustering models to provide more personalized risk assessments for individuals based on their health trajectory.
6. Clinical Validation and Testing: Conducting clinical trials and validating the application's predictions with healthcare professionals could ensure the reliability and clinical relevance of the tool for early CHD risk detection.
7. Geographic and Demographic Expansion: The application could be adapted to different regions and countries, considering specific local health data, risk factors, and cultural factors, to make it globally relevant and applicable.
8. Preventive Health Programs Integration: Partnering with healthcare organizations to integrate the application into public health campaigns or preventive health programs could help disseminate information on CHD prevention and encourage early health check-ups.
9. Personalized Health Recommendations: Future versions could provide users with tailored health recommendations based on their specific risk profiles, including lifestyle modifications, diet plans, and exercise routines, to proactively reduce the risk of CHD.

REFERENCES

[1] WHO, "Cardiovascular diseases," February 2023. [Online]. Available: https://www.who.int/health-topics/cardiovascular-diseases.

[2] F. Ulya and B. Santosa, "Kemenkes: Penyakit Jantung Penyebab Kematian Tertinggi pada 2014-2019," Februari 2023. [Online]. Available: https://nasional.kompas.com/read/2022/09/28/17402891/kemenkes-penyakit-jantung-penyebab-kematian-tertinggi-pada-2014-2019.

[3] T. Mylano, "Mengenal Penyakit Jantung Koroner, Penyebab Kematian Tertinggi di Dunia," April 2023. [Online]. Available: https://www.siloamhospitals.com/en/informasi-siloam/artikel/mengenal-penyakit-jantung-koroner-penyebab-kematian-tertinggi-di-dunia.

[4] E. Hastuti, "Pencegahan dan Pengobatan Penyakit Jantung Koroner," Juli 2022. [Online]. Available: https://yankes.kemkes.go.id/view_artikel/701/pencegahan-dan-pengobatan-penyakit-jantung-koroner.

[5] Rokom, "Penyakit Jantung Koroner Didominasi Masyarakat Kota," September 2021. [Online]. Available: https://sehatnegeriku.kemkes.go.id/baca/umum/20210927/5638626/penyakit-jantung-koroner-didominasi-masyarakat-kota/.

[6] Mustika, Y. Ardilla, N. Ahmad, I. Hasbi, Guntoro, M. Manuhutu, M. Ridwan, Hozairi, A. Wardhani, S. Alim, I. Romli, Y. Religia, D. Octafian, U. Sufandi and L. Ernawati, Data Mining dan Aplikasinya, Bandung: Widina Bhakti Persada Bandung (Grup CV. Widina Media Utama), 2021.

[7] R. Muhima, M. Kurniawan, S. Wardhana, A. Yudhana, Sunardi, W. Rahmawati and G. Yuliastuti, Kupas Tuntas Algoritma Clustering: Konsep, Perhitungan Manual, dan Program, Yogyakarta: ANDI, 2021.

[8] R. Fadilla, R. Andarsyah and R. Awangga, Data Analytics: Peningkatan Performa Algoritma Rekomendasi Collaborative Filtering menggunakan K-Means Clustering, Bandung: Kreatif Industri Nusantara, 2020.

[9] D. Haryadi and D. Atmaja, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Tingkat Risiko Penyakit Jantung," *Journal of Informatics and Communications Technology (JICT),* vol. 3, no. 2, pp. 51-66, 2021.

[10] Y. Tedju, Y. Kaesmetan and M. Lamabelawa, "Pengelompokan Penyakit Hipertensi di Kota Kupang menggunakan Metode K-Means," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK),* vol. 13, no. 2, pp. 98-108, 2022.

[11] R. Anggraini, E. Haerani and I. Afrianty, "Pengelompokkan Penyakit Pasien Menggunakan Algoritma K-Means," *Jurnal Riset Komputer (JURIKOM),* vol. 9, no. 6, pp. 1840-1848, 2022.

[12] Yurindra, Software Engineering, Yogyakarta: Deepublish, 2017.

[13] N. Ahmad, E. Krisnanik, F. Rupilele, A. Muliawati, N. Syamsiyah, Kraugusteeliana, B. Cahyono, Y. Sriyeni, T. Kristanto, Irwanto and Guntoro, Analisa & Perancangan Sistem Informasi Berorientasi Objek, Bandung: Widina Media Utama, 2022.

[14] C. Mandagie, Interviewee, *Jumlah umat Paroki Hati Tersuci Maria Katedral Manado.* [Interview]. September 2023.

[15] S. Raharjo, "Distribusi Nilai Tabel," Februari 2021. [Online]. Available: https://www.spssindonesia.com/2014/02/download-distribusi-nilai-tabel.html.

[16] L. Wikarsa, S. Pandelaki and H. Kurnia, "Analisis Efektivitas Tindakan Pencegahan Covid-19 di Kota Manado Menggunakan Metode K-means," *Jurnal Pekommas,* vol. 7, no. 1, p. 22, 2022.

https://doi.org/10.25077/TEKNOSI.v11i2.2025.107-117

Biografi

Junaidy Budi Sanger, S.Kom., M.Kom
He is a lecturer at the Informatics Engineering Study Program, Faculty of Engineering, Universitas Katolik De La Salle Manado. He completed his master's degree at Institut Pertanian Bogor. His interests are computer networks and artificial intelligence.

Dr. Liza Wikarsa, B.C.S., M.Comp

She is a senior lecturer at the Informatics Engineering Study Program, Faculty of Engineering, Universitas Katolik De La Salle Manado, North Sulawesi. She completed her doctoral degree at Pukyong National University. University, South Korea. Her expertise is related to Data Science, IoT, and IT Project Management.

Angelica Theresia Taulu
She is an undergraduate in the Informatics Engineering Study Program at Universitas Katolik De La Salle Manado. Her research interest is related to Data Mining.