ISSN 2502-5872 (Print)
ISSN 2622-3627 (Electronic)

Volume 10 Number 3, August 2025, 621-634

THE APPLICATION OF THE NAÏVE BAYES METHOD IN DIAGNOSING HEART DISEASE IN MEDAN CITY

Septia Handayani[1*], Fibri Rakhmawati[2]
[1,2]Mathematics Study Program, Faculty of Science and Technology, Universitas Islam Negeri Sumatera Utara, Indonesia
*Correspondence: septiah209@gmail.com

ABSTRACT
This study applies the Naïve Bayes method to diagnose coronary heart disease (CHD) in Medan City. This quantitative study uses secondary data from Malahayati Islamic Hospital, including independent variables (age, gender, smoking, blood pressure, family history) and dependent variables (CHD status). The data are divided into training data (90%) and testing data (10%). Despite the high accuracy, further research with larger and more diverse data is needed to validate these findings. The Naïve Bayes method has proven potential as an aid in the early diagnosis of CHD.
Keywords: Naive Bayes, Diagnosis, Heart Disease.

How to Cite: Handayani, S., & Rakhmawati, F. (2025). The Application Of The Naïve Bayes Method In Diagnosing Heart Disease In Medan City. *Mathline: Jurnal Matematika dan Pendidikan Matematika, 10*(3), 621-634. http://doi.org/10.31943/Mathline.v10i3.994

PRELIMINARY

Health is an optimal state that includes physical, mental, and social aspects. This means that a person is not only free from disease or disability but is also able to live a productive and quality life. Heart disease has a significant impact on health, and if not treated properly, can lead to death. Based on data from *the World Health Organization* (WHO, 2024), more than 17 million people around the world have died from heart disease. In Indonesia, according to a report by the Ministry of Health in 2023, the number of deaths due to heart disease reached 651.48 people (Hidayat et al., 2024).

The heart is a vital organ in the human body that functions to pump blood to all parts of the body. Heart disease has a significant impact on health, and if not treated properly, can lead to death (Alwie et al., 2020). This disease can affect anyone, especially those who have not implemented a healthy lifestyle. Many symptoms of diseases and impaired heart function are often unaware of by sufferers, due to a lack of attention to heart health (Buulolo, 2020). Heart disease is one of the most common conditions suffered by people, with symptoms such as physical weakness, swelling in the legs, and shortness of breath. The disease can affect anyone, regardless of age, gender, or life expectancy. So, maintaining heart health is very important (Rahmawati et al., 2024).

Coronary heart disease is a heart disease caused by narrowing of the coronary arteries, so that the supply of blood, oxygen and nutrients will be reduced to the heart and result in the heart not being able to work normally (Putro et al., 2002), if blood flow is blocked in the arteries leading to the brain, it will result in stroke, with the body getting older and worsened by various risk factors such as high blood pressure, smoking, normal blood cholesterol levels, narrowed, stiff, inelastic and blocked arteries are what cause CHD (Sheth, J. N., & Sharma, 2005).

The symptoms of heart disease and impaired function are often unnoticed by sufferers and the public, often due to indifference or lack of attention to heart health. In addition, the high cost of health checkups and busy schedules are also obstacles (Hanifah et al., 2021). Some of the early symptoms of heart disease that need to be watched out for include chest pain that often appears, shortness of breath, feeling tired quickly, irregular heartbeat, and cold sweats (Maulana et al., 2022). To prevent heart disease, steps that can be taken include quitting smoking, exercising regularly, practicing a healthy diet, and undergoing regular health checkups (Kevin, 2020).

CHD (Coronary Heart Disease) is caused by irreversible risk factors such as age, gender, and genetics. CHD can also be caused by modifiable risk factors such as smoking habits, *dyslipidemia*, hypertension, lack of physical activity, obesity, diabetes mellitus, stress, alcohol consumption and poor dietary habits. Modifiable risk factors for CHD such as hypertension, *dyslipidemia*, and diabetes mellitus are influenced by dietary consumption patterns (Naomi et al., 2021).

Despite the recognition of mathematical literacy's importance (Rayhan et al., 2025). The method is one of the approaches that utilizes Bayes' theorem, a statistical classification algorithm that is useful for estimating the probability of an event. based on the assumption that the values of the attributes are considered independent of each other when their output values have been determined (Helmi, 2021). One of the advantages of using it in the data mining process is that this algorithm model requires only a relatively small amount of training data to estimate the parameters needed (Hirwono et al., 2023). This method is one of the effective *Naïve BayesNaïve BayesNaïve Bayesmachine learning* techniques to handle data classification, Bayesian classification can be used to predict the probability value of a class (Jollyta et al., 2020). The combination of Bayes' theorem *Naïve*and the term "" indicates that the fidelity or variable is independent (Setiani et al., 2022).

*Naïve Bayes*The method has been carried out by researchers such as those carried out by Alphinda, Hendro and Gusti with the title Implementation of *the Naïve Bayes* Method for Stroke Diagnosis, the results of the study show that the diagnosis of stroke the *Naïve Bayes* method produces a good accuracy of 82% for testing using *10-Folds Cross Validation. Naïve Bayes* with balanced data handling resulted in accuracy, precision, recall, and f1-scrore, respectively, for *undersampling* of 64%, 65%, 60%, 61% while for *oversampling* of 71%, 57%, 52%, 52%, 52% (Nasution, 2021). While the research from Bowo, Arief and Donny, with the title Implementation of *the Naïve Bayes* Method for the Classification of Heart Disease Sufferers (Rahmatullah, 2021). Thus, the research gap lies in the lack of exploration of optimal data balancing techniques to improve the performance of Naïve Bayes on medical data. There is no testing of combinations of Naïve Bayes with optimization methods such as feature selection, attribute weighting, or hybrid models to improve accuracy, precision, recall, and f1-score. There is a lack of direct comparative studies between different disease cases and complex imbalanced data scenarios.

This study was carried out by dividing the data into 75% for training data and 25% for *training* data, the results of this study show that the *Naïve Bayes* algorithm used provides a fairly high accuracy value of 86.84%.

METHODS

Types of Research

The type of research that will be carried out is quantitative research. Quantitative research is a method to test a particular hypothesis by testing the relationships between variables. Quantitative research emphasizes on testing theory through measuring research variables with numbers and conducting data analysis with statistical procedures. The research method uses leaning measurements or using data measurement scales (Sugiyono, 2019).

The data obtained was secondary data, data collection was taken from Malahayati Islamic Hospital. Secondary data is data obtained from researchers or data collection indirectly. Data obtained through intermediaries, both individuals and documents.

Research Variables

In this study, the variables to be studied consist of independent variables (X) and dependent variables (Y), including:

Independent variable (X) : Age, Gender, Smoking , Blood pressure , Family history

The dependent variable (Y) is the observed outcome or predicted outcome, i.e. whether the patient has coronary heart disease (CHD) or not.

The diagnosis of coronary heart disease is categorical, namely (0) stating that you do not have coronary heart disease and (1) stating that you have coronary heart disease.

Research Procedure

The following is a series of processes in the research that will be carried out, namely as a result:

1) Early identification

   The research begins with the identification of the problem that the researcher wants to research by looking for what factors affect heart disease in hospitals. Islam Malahayati Medan.

2) Secondary data representation

   The collected data is then presented back into the form of a table according to the required variables. Then a dataset is formed from the data collection that will be analyzed in the research.

3) Calculating *Training* Data and Data *Testing*

   In the classification, data types are basically divided into two, namely training and testing data .

4) Method $Na\ddot{\imath}ve\ Bayes$

   a. Calculate the number of classes/attributes or calculate the probability value of new cases on each hypothesis with an existing class $P(Y)$ based on textittraining data.

   b. Calculates the number of cases per class/attribute or calculates the accumulated value of each class. Because this study uses numerical data, before calculating the number of cases per class, the first step is to find the mean and standard deviation of each class/attribute.$P((X|Y)$

   c. Perform multiplication operations on all class variables or calculate values.$\left(P(X|Y) \times P(Y)\right)$

   d. Comparing the results of the class based on the results of the third analysis step.

   e. Calculating the accuracy value of the classification

   f. *Confusion matrix* is the most commonly used method as a measure of accuracy in the concept of data mining. *The confusion matrix* is explained by a table showing the amount of test data that is correct and the amount of experiment/test data that is incorrect will be classified.

g. Making conclusions and interpretations

RESULT AND DISCUSSION

The research begins with the identification of the problem that the researcher wants to research by looking for what factors affect heart disease in hospitals. Islam Malahayati Medan. consisted of 42 patient data, each containing information related to the patient's condition during treatment, as well as whether the patient was diagnosed with Coronary Heart Disease (CHD) or not.

Dividing Training Data and Data *Testing*

In this study, the data used consisted of 42 patient data containing information about the patient's name, length of treatment (in days), number of disease diagnoses, and status whether the patient had Coronary Heart Disease (CHD) or not. This data is then divided into two parts, namely Data Training and Data Testing with proportions 90:10. This results in 38 data as training data and 4 data as data testing.

Table 1. Data *Training*

Yes	Patient Name	Gender	Age	Blood pressure	Smoke	Family History	Cardiac Diagnosis
1	Hj Khairani Etc	P	45-64	Usual	Not	Not	PJK
2	Anna Maria Angelina Marimah	P	45-64	Usual	Not	Yes	PJK
3	Murniaty Pasaribu	P	65+	Tall	Not	Not	PJK
4	Squirting Sauce	L	45-64	Usual	Yes	Not	PJK
Yes	**Patient Name**	**Gender**	**Age**	**Blood pressure**	**Smoke**	**Family History**	**Cardiac Diagnosis**
5	A Guide, Please	L	65+	Tall	Yes	Not	PJK
6	Renta Br Sirait	P	45-64	Usual	Not	Yes	PJK
7	Rohnurdian	L	65+	Tall	Yes	Not	PJK
8	Latifah Hanum	P	65+	Usual	Not	Yes	PJK
9	Yudhi Willyandri	L	25-44	Usual	Yes	Yes	PJK
10	Suparmin	L	65+	Usual	Yes	Yes	PJK
11	Nursamsi	P	65+	Usual	Not	Not	PJK
12	Agus Maryadi	L	65+	Tall	Yes	Yes	PJK
13	H Yusfiansyah	L	65+	Tall	Yes	Not	PJK
14	Ir. Ali Rahman	L	45-64	Usual	Not	Yes	PJK
15	Indra Novianto	L	25-44	Tall	Yes	Yes	PJK
16	Yusniarti AJ	P	65+	Usual	Not	Not	PJK
17	Irwanto. S	L	45-64	Usual	Not	Not	PJK
18	Kasmawati	P	45-64	Usual	Not	Not	PJK
19	Rosda Himayati	P	65+	Tall	Not	Yes	PJK
20	Josib Broztito sianipar	L	45-64	Tall	Yes	Yes	PJK
21	Rudy Edit	L	45-64	Usual	Not	Yes	PJK
22	Ibrahim Sulaiman	L	65+	Usual	Yes	Not	PJK
23	Irwan	L	45-64	Usual	Yes	Yes	PJK
24	Mangoloi Sinambela	L	65+	Tall	Yes	Not	PJK
25	Squirting Claim	L	65+	Tall	Not	Not	PJK
26	Elfia neswita	P	25-44	Usual	Not	Yes	PJK

27	Masrul koto	L	65+	Usual	Yes	Yes	PJK
28	M Daud	L	45-64	Tall	Not	Yes	PJK
29	Hj Ramainas	P	65+	Tall	Not	Not	PJK
30	Mr. Riyadi	L	65+	Usual	Yes	Yes	PJK
31	Rita Wirawani	P	65+	Usual	Not	Not	PJK
32	Zainal Arifin	L	45-64	Usual	Yes	Not	PJK
33	M Daud	L	65+	Tall	Yes	Yes	PJK
34	Imballo Pulungan	L	65+	Usual	Yes	Yes	PJK
35	Ir. Herri Trisna Frianto	L	45-64	Usual	Not	Yes	PJK
36	Maya Anita	P	45-64	Tall	Not	Not	PJK
37	H M Kasim Nur	L	65+	Tall	Not	Not	PJK
38	Nelson S. Irving	L	65+	Tall	Yes	Not	PJK

Table 2. Data *Testing*

Yes	**Patient Name**	**Gender**	**Age**	**Blood pressure**	**Smoke**	**Family History**	**Cardiac Diagnosis**
39	H Jahja beyn	L	65+	Usual	Not	Yes	PJK
40	Muhammad syafrizal	L	45-64	Usual	Yes	Yes	PJK
41	Cape Ferry	L	25-44	Usual	Yes	Yes	PJK
42	Yanurli	L	65+	Tall	Not	Not	PJK

Method *Naïve Bayes*

1. Calculating the Probability of CHD Opportunities from Training Data

Calculate the number of classes/attributes or calculate the probability value of new cases on each hypothesis with existing classes based on training data$P(Y)$

Table 3. Class Probability

	CHD DIAGNOSIS	
Amount of Data	Yes	Not
38	38/38	0/38
PJK Opportunities	1	0

2. Calculate the number of cases of class/attribute or attribute probability

Calculates the accumulated value of each class.$P((X|Y)$

a. Gender Attributes

a) Probability on Female Sex Attribute (P)

$$P(X = \text{Gender "P"} | Y = \text{Yes}) = 13/38 = 0{,}3421$$

$$P(X = \text{Gender "P"} | Y = \text{Not}) = 0/0 = 0$$

i. Probability on the Male Sex Attribute (L)

$$P(X = \text{Gender "L"} | Y = \text{Yes}) = 25/38 = 0{,}6578$$

$$P(X = \text{Gender "L"} | Y = \text{Not}) = 0/0 = 0$$

Table 4. Gender Probability

P(X|Y)	**P(Gender "P")**	**Yes**	**0,342105**
		Not	0
	P(Gender "L")	Yes	0,657895
		Not	0

b. Age Attributes

i. Probability on the age attribute "25-44"

$$P(X = \text{Age "}25 - 44\text{"} | \; Y = \text{Yes}) = 3/38 = 0{,}0789$$

$$P(X = \text{Age "}25 - 44\text{"} | \; Y = \text{Not}) = 0/0 = 0$$

ii. Probability on the age attribute "45-64"

$$P(X = \text{Age "}45 - 64\text{"} | \; Y = \text{Yes}) = 14/38 = 0{,}3684$$

$$P(X = \text{Age "}45 - 64\text{"} | \; Y = \text{Not}) = 0/0 = 0$$

iii. Probability on the age attribute "65+"

$$P(X = \text{Age "}65+\text{"} | \; Y = \text{Yes}) = 21/38 = 0{,}5526$$

$$P(X = \text{Age "}65+\text{"} | \; Y = \text{Not}) = 0/0 = 0$$

Table 5. Age Probability

P(X|Y)	**Q(Age 25-44)**	**Yes**	**0,078947**
		Not	0
	P(Age 45-64)	Yes	0,368421
		Not	0
	P(Age "65+")	Yes	0,552632
		Not	0

c. Attributes of Blood Pressure

i. Probability on the "Normal" Blood Pressure attribute

$$P(X = \text{Bllod Pressure "Normal"} | \; Y = \text{Yes}) = 22/38 = 0{,}5789$$

$$P(X = \text{Bllod Pressure "Normal"} | \; Y = \text{Not}) = 0/0 = 0$$

ii. Probability on the attribute of "High" Blood Pressure

$$P(X = \text{Blood Pressure "High"} | \; Y = \text{Yes}) = 16/38 = 0{,}4210$$

$$P(X = \text{Blood Pressure "High"} | \; Y = \text{Not}) = 0/0 = 0$$

Table 6. Probability of Blood Pressure

P(X|Y)	**P("Normal" Blood Pressure)**	**Yes**	**0,578947**
		Not	0
	P(High Blood Pressure)	Yes	0,421053
		Not	0

d. Smoking Attributes

a) Probability on the Smoking attribute "Yes"

$$P(X = \text{Smoke "Yes"} | \; Y = \text{Yes}) = 18/38 = 0{,}4736$$

$$P(X = \text{Smoke "Yes"} | \; Y = \text{Not}) = 0/0 = 0$$

b) Probability on the Smoking attribute "No"

$$P(X = \text{Smoke "No"} | \; Y = \text{Yes}) = 20/38 = 0{,}5263$$

$$P(X = \text{Smoke "No"} | \; Y = \text{Not}) = 0/0 = 0$$

Table 7. Probability of Smoking

P(X|Y)	P(Smoking "Yes")	Yes	0,473684
		Not	0
	P(Smoking "No")	Yes	0,526316
		Not	0

e. Family History Attributes

i. Probability on the Family History attribute "Yes"

$P(X = \text{Familiy History “Yes”} | Y = \text{Yes}) = 19/38 = 0{,}5$

$P(X = \text{Familiy History “Yes”} | Y = \text{Not}) = 0/0 = 0$

ii. Probability on the Family History attribute "No"

$P(X = \text{Fammily History “No”} | Y = \text{Yes}) = 19/38 = 0{,}5$

$P(X = \text{Family History “Not”} | Y = \text{No}) = 0/0 = 0$

Table 8. Probability of Family History

P(X|Y)	P(Family History "Yes")	Yes	0,5
		Not	0
	P(Family History "No")	Yes	0,5
		Not	0

3. Making Predictions Using Data Testing

Calculate multiplication operations on all class variables or calculate values.$\left(P(X|Y) \times P(Y)\right)$

Table 9. Data Testing

Yes	Patient Name	Gender	Age	Blood pressure	Smoke	Family History	Cardiac Diagnosis
1	H Jahja beyn	L	65+	Usual	Not	Yes	PJK
2	Muhammad syafrizal	L	45-64	Usual	Yes	Yes	PJK
3	Cape Ferry	L	25-44	Usual	Yes	Yes	PJK
4	Yanurli	L	65+	Tall	Not	Not	PJK

a. Calculating the probability of CHD diagnosis from previous training data

$P(\text{Ya}) = 38/38 = 1$

$P(\text{Not}) = 0/0 = 0$

b. Probability of the same attribute in the training data as the testing data

a) Probability on the gender attribute MALE - MALE (L)

$P(X = \text{Gender "L"} | Y = \text{Yes}) = 25/38 = 0{,}6578$

$P(X = \text{Gender"L"} | Y = \text{Not}) = 0/0 = 0$

b) Probability on the Age attribute "25-44"

$P(X = \text{Age “}25 - 44\text{”} | Y = \text{Yes}) = 3/38 = 0{,}0789$

$P(X = \text{Age “}25 - 44\text{”} | Y = \text{Not}) = 0/0 = 0$

c) Probability on the Age attribute "45-64"

$P(X = \text{Age “}45 - 64\text{”} | Y = \text{Yes}) = 14/38 = 0{,}3684$

$P(X = \text{Age “}45 - 64\text{”} | Y = \text{Not}) = 0/0 = 0$

d) Probability on the Age attribute "65+"

$P(X = \text{Age "65 + "} | Y = \text{Yes}) = 21/38 = 0{,}5526$

$P(X = \text{Age "65 + "} | Y = \text{Not}) = 0/0 = 0$

e) Probability on the "Normal" Blood Pressure attribute

$P(X = \text{Blood Pressure "Normal"} | Y = \text{Yes}) = 22/38 = 0{,}5789$

$P(X = \text{Blood Pressure "Normal"} | Y = \text{Not}) = 0/0 = 0$

f) Probability on the attribute of "High" Blood Pressure

$P(X = \text{Blood Pressure "High"} | Y = \text{Yes}) = 16/38 = 0{,}4210$

$P(X = \text{Blood Pressure "High"} | Y = \text{Not}) = 0/0 = 0$

g) Probability on the Smoking attribute "Yes"

$P(X = \text{Smoke "Yes"} | Y = \text{Yes}) = 18/38 = 0{,}4736$

$P(X = \text{Smoke "Yes"} | Y = \text{Not}) = 0/0 = 0$

h) Probability on the Smoking attribute "No"

$P(X = \text{Smoke "Not"} | Y = \text{Yes}) = 20/38 = 0{,}5263$

$P(X = \text{Smoke "Not"} | Y = \text{Not}) = 0/0 = 0$

i) Probability on the Family History attribute "Yes"

$P(X = \text{Family History "Yes"} | Y = \text{Yes}) = 19/38 = 0{,}5$

$P(X = \text{Family History "Yes"} | Y = \text{Not}) = 0/0 = 0$

j) Probability on the Family History attribute "No"

$P(X = \text{Family History "Not"} | Y = \text{Yes}) = 19/38 = 0{,}5$

$P(X = \text{Family History "Not"} | Y = \text{Not}) = 0/0 = 0$

Table 10. Overall Probability of Data Testing

P(X|Y)	**P(Gender "L")**	**Yes**	**0,657895**
		Not	0
P(X|Y)	Q(Age 25-44)	Yes	0,078947
		Not	0
	P(Age 45-64)	Yes	0,368421
		Not	0
	P(Age "65+")	Yes	0,552632
		Not	0
P(X|Y)	P("Normal" Blood Pressure)	Yes	0,578947
		Not	0
	P(High Blood Pressure)	Yes	0,421053
		Not	0
P(X|Y)	P(Smoking "Yes")	Yes	0,473684
		Not	0
	P(Smoking "No")	Yes	0,526316
		Not	0
P(X|Y)	P(Family History "Yes")	Yes	0,5
		Not	0
	P(Family History "No")	Yes	0,5
		Not	0

Calculate multiplication operations on all attribute variables or calculate values from testing data$\big(P(X|Y) \times P(Y)\big)$

1. H Jahja Beyn

Diagnosa PJK "Ya"

$$= \text{P(Gender"L"|Yes)} \times \text{P(Age "65} + \text{"|Yes)} \times \text{P(Blood Pressure "Normal"|Yes)} \times \text{P(Smoke "Not"|Yes)} \times \text{P(Family History "Yes"|Yes)} \times \text{P(Diagnosis PJK "Yes")}$$

Diagnosa PJK "Yes"

$$= (0{,}657895) \times (0{,}552632) \times (0{,}578947) \times (0{,}526316) \times (0{,}5) \times (1) = 0{,}05539$$

Diagnosa PJK "Not" $= 0\ (karena\ semua\ 0)$

2. Prediction Results

Table 11. Prediction Results

No.	Name	(P(X ∣ Y)×P(Y))		Diction
1	H Jahja Beyn	Yes	0,05539	Yes
		Not	0	
2	Muhammad Syafrizal	Yes	0,03324	Yes
		Not	0	
3	Nofirman Tanjung Ferry	Yes	0,00712	Yes
		Not	0	
4	Yanurli	Yes	0,04029	Yes
		Not	0	

Of the four patients tested, the probability value for class Yes (CHP) had a certain number (e.g. 0.05539 for patient *H Jahja beyn*), while the probability for class No was equal to zero. This shows that based on patient characteristics data (length of treatment and number of diagnoses), the model gives the highest chance of the Yes category, and zero for the No category, so that all patients in the table are predicted to have CHD (Siska et al., 2023).

This final prediction is based on a probability comparison between classes, where the class with the highest combined probability value will be selected as a result of the classification. In this case, since the value for the "No" category is 0 in all patients, the model consistently selects the "Yes" class.

Untitled1* | hasil | data

Filter

	Nama	Prediksi_Probabilitas_PJK	Diagnosa_Asli
39	H_Jahja_Beyn	0.055383155	PJK
40	Muhammad_Syafrizal	0.033231998	PJK
41	Feri_Nofirman_Tanjung	0.007117276	PJK
42	Yanurli	0.040286486	PJK

Figure 1. Prediction Results

Calculating the Accuracy Value of the Classification and *Confusion Matrix*

Table 12. Comparison of predictions and original data

No.	Name	Probability	Predictions	Original Data
1	H Jahja Beyn	0,05539207	PJK	PJK
2	Muhammad Syafrizal	0,03323524	PJK	PJK
3	Nofirman Tanjung Ferry	0,00712184	PJK	PJK
4	Yanurli	0,04028514	PJK	PJK

Confusion matrix It is the most commonly used method as a measure of the level of accuracy in the concept of data mining. *Confusion matrix* explained by a table showing the amount of test data that is correct will be classified A and the amount of experiment/incorrect test data will be classified (Waruwu & Dharma, 2024).

Table 13. *Confusion matrix*

	Current PJK	Actual No PJK
PJK Prediction	4 (TP)	0 (FN)
Predictions Not PJK	0 (FP)	0(TN)

$$kurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

$$akurasi = \frac{4 + 0}{4 + 0 + 0 + 0}$$

$$akurasi = \frac{4}{4} = 1 = 100\%$$

```
> cat("Confusion Matrix:\n")
Confusion Matrix:
> print(conf_matrix)
           Aktual
Prediksi     PJK Tidak_PJK
  PJK         4         0
  Tidak_PJK   0         0
> # Hitung akurasi
> akurasi <- sum(diag(conf_matrix)) / sum(conf_matrix)
> cat("Akurasi:", round(akurasi * 100, 2), "%\n")
Akurasi: 100 %
```

Figure 2. Confusion *matrix results*

Based on the results of model testing Naive Bayes On the test data as many as 4 data, a value was obtained 100% accuracy**.** In this case, TP (True Positive) 4, meaning that there are 4 data that are really positive (have CHD) and have been correctly predicted by the model. TN (True Negative) value 0 because there is no data that is classified as negative (does not have CHD). Likewise with FP (False Positive) and FN (False Negative) which are both 0 values, indicating that there are no prediction errors in the classification.

These results show that the model successfully classifies the entire test data correctly. However, it should be noted that the small amount of test data and the absence of variation on the actual label (all of which are "Yes") can cause the results to be overly accurate and may not necessarily reflect the actual performance of the model when applied to larger, varied data.

CONCLUSION

In this study, *the Naïve Bayes* method was applied to diagnose coronary heart disease (CHD) based on a number of patient characteristic data. The data used included five important variables, namely gender, age, blood pressure, smoking habits, and whether or not there was a family history of also having coronary heart disease. The data is divided into two parts, namely 38 data used as training data to build models, and the remaining 4 data data are used as data testing to test the model's ability to make predictions. The results of the diagnosis showed that the *Naïve Bayes* method was able to accurately classify all testing data according to the patient's original diagnosis, with an accuracy rate of 100%. This suggests that *Naïve Bayes* can be an effective tool in the early process of coronary heart disease diagnosis.

REFERENCES

Alwie, P. et al. (2020). Final Project Final Project. *Journal of Economics*, *18*(1), 41–49. https://journal.example.com/article/14858

Buulolo, E. (2020). *Data mining for universities*. Deepublish. https://edeposit.perpusnas.go.id/collection/data-mining-untuk-perguruan-tinggi-sumber-elektronis/13009

Hanifah, W. et al. (2021). Lifestyle factors and coronary heart disease: a systematic review in adults in Indonesia. *The Journal of Nutrition and Food Research*, *44*(1). https://repository.uinjkt.ac.id/dspace/handle/123456789/69096

Helmi, M. I. Y. (2021). *Diagnosis of lung cancer patients using support vector machine and naïve bayes*. Faculty Of Mathematics And Natural Sciences. https://doi.org/10.29313/jstat.v21i1.7566

Hidayat, R. et al. (2024). *Implementation of machine learning for heart disease prediction using a support vector machine algorithm*. 161–168. https://doi.org/10.37148/bios.v5i2.152

Hirwono, B. et al. (2023). Implementation of the Naïve Bayes Method for the Classification of Heart Disease Sufferers. *JTIK Journal (Journal of Information and Communication Technology)*, *7*(3), 450–457. https://doi.org/10.35870/jtik.v7i3.910

Jollyta, D. et al. (2020). *The concept of data mining and application*. https://repositori.telkomuniversity.ac.id/pustaka/228416

Kevin, K. (2020). Diagnosis of heart disease using the certainty factor method. *Journal of Informatics and Software Engineering*, *3*(1), 93–106.

Maulana, R. et al. (2022). Penerapan Metode Naive Bayes Untuk Mendiagnosa Penyakit Stunting Pada Balita. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, *1*(4), 425. https://doi.org/10.53513/jursi.v1i4.5446

Naomi, W. S. et al. (2021). Risk Factors for the Occurrence of Coronary Heart Disease. *Public Health Media*, *3*(1), 99–107. https://thejnp.org/index.php/jnp/article/view/726

Nasution. (2021). *Perancangan Sistem Pendukung Keputusan Untuk Pemilihan Lokasi Perumahan Menggunakan WeightedProduct Model (WPM)*. STMIK El Rahma. 61–78

Putro, H. F. et al. (2002). Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan. *Jurnal Teknologi Informasi Dan Komunikasi (TIKomSiN)*, *8*(2). https://doi.org/10.30646/TIKOMSIN.V8I2.500

Rahmatullah, B. A. (2021). Analysis of student sentiment towards online lectures using the Naïve Bayes Classifier method. *Maulana Malik Ibrahim State Islamic University*. https://doi.org/10.18860/jrmm.v2i3.16415

Rahmawati, A. D. et al. (2024). *Implementation of the Use of Linear Regression Models to Predict the Risk of Heart Disease Based on Medical Data*. https://doi.org/10.20895/centive.v4i1.387

Rayhan, N. . et al. (2025). Investigating Learning Obstacles In Mathematical Literacy: A Study On Linear Equations With One Variable. *Jurnal Mathline*, *10*(1). https://doi.org/10.31943/mathline.v10i1.825

Setiani, H. et al. (2022). Naïve Bayes and Particle Swarm Optimization Methods for Classification of Heart Disease. *Explore*, *12*(2), 6. https://doi.org/10.35200/explore.v12i2.566

Sheth, J. N., & Sharma, A. (2005). International e-marketing: Opportunities and issues. *Nternational Marketing Review*, *22*(6), 611–622. https://doi.org/10.1108/02651330510630249

Siska, S. et al. (2023). Implementation of the Naive Bayes Method in the Prediction of Celiac Disease. *KOPERTIP Scientific Journal of Informatics and Computer Management*, *7*(1), 8–13. https://doi.org/10.32485/kopertip.v7i1.325

Sugiyono. (2019). *Metode Penelitian Kuantitatif Kualitatif dan R&d*. Alfabeta. https://opac.perpusnas.go.id/DetailOpac.aspx?id=1184265

Waruwu, J., & Dharma, A. (2024). *Heart Comparison of Classification Algorithms in Heart Disease Patients Introduction The heart is an essential organ in the human body and serves as a regulatory center in the circulatory system, where blood vessels act as a distribution pathway*. https://doi.org/10.31539/intecoms.v7i6

WHO, W. H. O. (2013). A Global Brief on Hypertension: Silent Killer, Global Public Health Crisis. In *Indian Journal of Physical Medicine and Rehabilitation* (Vol. 24, Issue 1). https://doi.org/10.5005/ijopmr-24-1-2