Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th October 2023. Vol. 3. No. 1; e-ISSN : 2808-4519

Classification For Predicting Heart Disease Using The K Nearest Neighbor Method Sylvani General Hospital Binjai City

Eko Sulistio[1], Yani Maulita[2], Magdalena Simanjuntak[3]

[1,2,3]Informatics Enginering, STMIK Kaputama
Jl. Veteran No. 4A-9A, Binjai, North Sumatra, Indonesia
eckorasta69@gmail.com[1]*

Abstract

The heart is a hollow organ and has four chambers or chambers located between the two lungs in the middle of the thoracic cavity. The heart has an important function in the human body, namely as a pump that presses blood so that it can flow to all parts of the body through arteries or veins. Disease caused by plaque buildup in the coronary arteries which supply oxygen to the heart muscle, resulting in severe damage to the heart is called coronary heart disease. Many factors can increase the risk of heart disease. These risk factors consist of risk factors that cannot be modified such as family history, age and gender and risk factors that can be modified such as hypertension, smoking habits, diabetes, dyslipidemia, obesity, lack of physical activity, diet and stress. K-Nearest Neighbor is a method for classifying new objects based on their (K) closest neighbors. K-NN includes a Supervised Learning algorithm where the results of querying new instances are classified based on the majority of categories in KNN. The class that appears the most will be the classification result class. This algorithm only stores feature vectors and classifies the learning data. In the classification phase, the same features are calculated for the test data (whose classification is unknown). The distance of this new vector to all data vectors is calculated, and the k closest ones are taken. The newly classified point is predicted to be among the most classified of these points. From the data with the majority categories there are Positive and Negative categories. From the majority number (Positive > Negative) it can be concluded that new data (data No. 20) (K1=1, K2=0.5, K3=0, K4=1, K5=0, K6=1, K7=0, 5, K8=1) is included in the Positive category.

Keywords: *Data Mining, Heart disease, K-Nearest Neighbor*

1. Introduction

The heart is a hollow organ and has four chambers or chambers located between the two lungs in the middle of the thoracic cavity. The heart has an important function in the human body, namely as a pump that presses blood so that it can flow to all parts of the body through arteries or veins [1]. Disease caused by plaque buildup in the coronary arteries which supply oxygen to the heart muscle, resulting in severe damage to the heart is called coronary heart disease.Many factors can increase the risk of heart disease. These risk factors consist of risk factors that cannot be modified such as family history, age and gender and risk factors that can be modified such as hypertension, smoking habits, diabetes, dyslipidemia, obesity, lack of physical activity, diet and stress.Various conventional and manual analysis techniques that have been used so far are no longer very effective in diagnosing a disease. Along with the development of technology and knowledge-based systems, especially medical ones, the demand for the use of computer-based knowledge systems as analytical techniques in diagnosing disease is becoming increasingly important and must always be developed. Therefore, now is the right time to develop a modern, effective and efficient computer-based knowledge system in diagnosing disease problems, especially at the Sylvani General Hospital, Binjai City.

2. Metodologi

2.1. Data Mining

Data mining is the process of extracting useful information and patterns from very large data. Data mining includes data collection, data extraction, data analysis, and data statistics. Data mining is also known as Knowledge discovery, Knowledge extraction, data/pattern analysis, information harvesting, and others. Data mining bertujuan untuk menemukan pola yang sebelumnya tidak diketahui. Jika pola-pola tersebut telah diperoleh maka dapat digunakan untuk menyelesaikan berbagai macam permasalahan. Data mining is one of the main parts or processes of Knowledge Discovery in Database (KDD) whose form of activity is collecting and using past data to find

regularities [2], patterns or relationships in a larger data set. Broadly speaking, KDD includes three stages, namely pre-processing, process (data mining) and post-processing [3].Data mining has now become a powerful new technology with great potential to help companies focus on the most important information in the data they have collected about the behavior of their customers and potential customers. Through data mining, companies can find information in large amounts of data through precise and effective processing with various methods available in data mining so that in simple terms data mining can be described as a pattern or model or rule or knowledge generated from data mining, such as Figure 1.

Figure 1: Model or Knowledge Is the Output of Data Mining

2.2. K-Nearest Neighbor

K-Nearest Neighbor is a method for classifying new objects based on their (K) closest neighbors. K-NN includes a Supervised Learning algorithm where the results of querying new instances are classified based on the majority of categories in KNN. The class that appears the most will be the classification result class [4]. This algorithm only stores feature vectors and classifies the learning data. In the classification phase, the same features are calculated for the test data (whose classification is unknown). The distance of this new vector to all data vectors is calculated, and the k closest ones are taken [5]. The newly classified point is predicted to be among the most classified of these points.

2.3. Types of Heart Disease

Types of heart disease are as follows [6]:

A. Coronary heart disease

Coronary heart disease is a narrowing of the small blood vessels that supply blood and oxygen to the heart. The symptoms that can be felt are: sudden chest pain, excessive cold sweat, prolonged headache, burning sensation throughout the body, nausea and vomiting, the body gets tired and weak quickly, experiencing shortness of breath, swelling around the joints. and feet. Coronary heart disease is caused by high blood pressure (hypertension), diabetes, cholesterol, obesity, smoking, consuming alcohol and hereditary factors.

B. Hypertension Heart Disease

Hypertensive heart disease is a disease associated with secondary impacts on the heart due to prolonged systemic hypertension. The symptoms that can be felt are: headache, sudden chest pain, heart palpitations, body quickly tired and weak, nosebleeds. Hypertensive heart disease is caused by high blood pressure (hypertension), cholesterol, obesity, stress, smoking, consuming alcohol. This disease can be prevented by exercising regularly, stopping smoking and consuming alcohol, avoiding fatty and high cholesterol foods, maintaining mental health, and consuming at least 1 and a half liters of water.
per day.

C. Pricarditis Heart Disease

Pericarditis heart disease is inflammation of the heart sac or pericardium, causing fluid accumulation and thickening. The symptoms that can be felt are: nausea and vomiting, the body gets tired and weak quickly, fever, difficulty breathing, coughing, chest pain, swelling of the stomach, shortness of breath. Pericarditis heart disease is caused by viral infections such as Coxsakie and influenza, bacterial infections such as Streptococcus, Staphylococcus, Meningococcus, and Gonococcus, parasitic infections, fungal infections, suffering from cancer, suffering from myocardial infarction, and suffering from tuberculosis. This disease can be prevented with antibiotics, anti-fungals, pericardiocentesis, anti-inflammatory drugs, analgesics and aspirin.

D. Rheumatic Heart Disease

Rheumatic heart disease is a condition where there is damage to the heart valves which can be in the form of narrowing or leakage, especially the mitral valve (mitral valve stenosis) as a result of the sequelae of Rheumatic Fever (DR). The symptoms that can be felt are: joint pain that moves around, limited reddish spots on the skin, irregular and uncontrolled hand movements, shortness of breath, small lumps under the skin, stomach pain, weight loss, fatigue easily. and fever. Rheumatic heart disease is caused by genetic factors, age, nutritional status, ethnic and racial groups, gender and autoimmune reactions. This disease can be prevented by diligently cleaning the house, exercising, consuming healthy food, avoiding smoking, wearing a mask in dusty environments.

E. Heart Muscle Disease

Heart muscle disease is the loss of the heart's ability to pump blood and beat normally. The symptoms that can be felt are: shortness of breath, swelling of the feet and hands, stomach feeling bloated, fatigue easily, irregular heartbeat, headaches. Heart muscle disease is caused by high blood pressure (hypertension), heart valve problems, damage to heart tissue from a previous heart attack, a heartbeat that is too fast, metabolic disorders, lack of nutrition, essential vitamins or minerals, abuse of cocaine or anti-defrosant drugs, use of some chemotherapy drugs, iron buildup in the heart muscle and consuming alcohol.

F. Heart Failure

Heart failure is a condition where the heart loses its ability to pump enough blood to the body's tissues. The symptoms that can be felt are: shortness of breath, frequent coughing, swelling around the legs, swelling or abdominal pain, fatigue, dizziness, shortness of breath. Heart failure is caused by high blood pressure, anemia, cholesterol, diabetes, obesity, smoking, antiviral drugs (zidofudine), irregular lifestyle, lack of exercise and stress, increased salt intake, infective endocarditis, medication.
chemotherapy such as doxorubicin. This disease can be prevented by reducing fatty foods, stopping smoking, exercising, having a regular lifestyle and diligently checking with a doctor.

3. Results and Discussion

3.1. K-Nearest Neighbor Calculation

To determine the group of an object, the first thing to do is measure the Euclidean Distance as follows.

Table 1: Transformation data

No	Object	K1	K2	K3	K4	K5	K6	K7	K8	Results
1	SB	1	0,5	0,5	0	1	1	0,5	1	2
2	NH	3	0	0,5	1	1	1	0,5	1	1
3	EW	1	0,5	0	1	1	1	0,5	1	1
4	NA	3	0	0	1	1	1	0,5	1	1
5	FJ	1	0,5	0,5	1	0	1	0,5	1	1
6	RM	2	0,5	0,5	1	1	0	0,5	1	1
7	NC	3	0,5	0	0	1	1	0,5	1	2
8	DM	2	0,5	0	1	1	1	0,5	1	1
9	MA	1	0,5	0	0	0	1	0,5	1	2
10	JK	3	0	0,5	1	1	1	0,5	1	1
11	SP	2	0,5	0,5	0	1	1	0,5	1	2
12	AL	2	0,5	0,5	1	1	0	0,5	1	1
13	SA	1	0,5	0,5	0	0	1	0,5	1	1
14	NH	2	0,5	0,5	1	0	1	0,5	1	1
15	ES	3	0,5	0,5	1	1	1	0,5	1	1
16	NA	1	0	0,5	0	1	0	0,5	1	1
17	SP	2	0	0,5	1	1	1	0,5	1	1
18	MF	2	0,5	0,5	1	1	0	0,5	1	2
19	RA	1	0,5	0,5	1	0	0	0,5	1	2
20	BA	1	0,5	0	1	0	1	0,5	1	

Step 1: Determine the K Parameter Value. K=3

Step 2: Calculate the distance between the test data and all training data, using Euclidean Distance.Then do the calculations like the calculation process below: Iteration 1:

$d\ x_{train\ i,}\ x_{test,\ j} = \sqrt{\sum_{i,j=1}^{n} (\text{xtrain i} - \text{xtest, j})^2}$

Euclidean Distance PX No. 20 : 1, 0,5, 0, 1, 0, 1, 0,5, 1

Data 1

$$dis = \sqrt{\begin{matrix}(1-1)^2 + (0{,}5-0{,}5)^2 + (0{,}5-0)^2 + (0-1)^2 + (1-0)^2 + \\ (1-1)^2 + (0{,}5-0{,}5)^2 + (1-1)^2\end{matrix}}$$

= 1,8027

Data 2

$$dis = \sqrt{\begin{matrix}(3-1)^2 + (0-0{,}5)^2 + (0{,}5-0)^2 + (1-1)^2 + (1-0)^2 + \\ (1-1)^2 + (0{,}5-0{,}5)^2 + (1-1)^2\end{matrix}}$$

= 2,5495

Data 3

$$dis = \sqrt{\begin{matrix}(1-1)^2 + (0{,}5-0{,}5)^2 + (0-0)^2 + (1-1)^2 + (1-0)^2 + \\ (1-1)^2 + (0{,}5-0{,}5)^2 + (1-1)^2\end{matrix}}$$

= 1,4142

Data 4

$$dis = \sqrt{\begin{matrix}(3-1)^2 + (0-0{,}5)^2 + (0-0)^2 + (1-1)^2 + (1-0)^2 + \\ (1-1)^2 + (0{,}5-0{,}5)^2 + (1-1)^2\end{matrix}}$$

= 2,6925

Data 5

$$dis = \sqrt{\begin{matrix}(1-1)^2 + (0{,}5-0{,}5)^2 + (0{,}5-0)^2 + (1-1)^2 + (0-0)^2 + \\ (1-1)^2 + (0{,}5-0{,}5)^2 + (1-1)^2\end{matrix}}$$

= 4,25

Data 6

$$dis = \sqrt{(2-1)^2 + (0{,}5-0{,}5)^2 + (0{,}5-0)^2 + (1-1)^2 + (1-0)^2 + (0-1)^2 + (0{,}5-0{,}5)^2 + (1-1)^2}$$
$$= 2{,}0615$$

Data 7

$$dis = \sqrt{(3-1)^2 + (0{,}5-0{,}5)^2 + (0-0)^2 + (0-1)^2 + (1-0)^2 + (1-1)^2 + (0{,}5-0{,}5)^2 + (1-1)^2}$$
$$= 7{,}25$$

Data 8

$$dis = \sqrt{(2-1)^2 + (0{,}5-0{,}5)^2 + (0-0)^2 + (1-1)^2 + (1-0)^2 + (1-1)^2 + (0{,}5-0{,}5)^2 + (1-01)^2}$$
$$= 2$$

Data 9

$$dis = \sqrt{(1-1)^2 + (0{,}5-0{,}5)^2 + (0-0)^2 + (0-1)^2 + (0-0)^2 + (1-1)^2 + (0{,}5-0{,}5)^2 + (1-1)^2}$$
$$= 1{,}1180$$

Data 10

$$dis = \sqrt{(3-1)^2 + (0-0{,}5)^2 + (0{,}5-0)^2 + (1-1)^2 + (1-0)^2 + (1-1)^2 + (0{,}5-0{,}5)^2 + (1-1)^2}$$
$$= 6{,}5565$$

Data 11

$$dis = \sqrt{(2-1)^2 + (0{,}5-0{,}5)^2 + (0{,}5-0)^2 + (0-1)^2 + (1-0)^2 + (1-1)^2 + (0{,}5-0{,}5)^2 + (1-1)^2}$$
$$= 2{,}0615$$

Data 12

$$dis = \sqrt{(2-1)^2 + (0{,}5-0{,}5)^2 + (0{,}5-0)^2 + (1-1)^2 + (1-0)^2 + (0-1)^2 + (0{,}5-0{,}5)^2 + (1-1)^2}$$
$$= 2{,}06155$$

Data 13

$$dis = \sqrt{(1-1)^2 + (0{,}5-0{,}5)^2 + (0{,}5-0)^2 + (0-1)^2 + (0-0)^2 + (1-1)^2 + (0{,}5-0{,}5)^2 + (1-1)^2}$$
$$= 1{,}2247$$

Data 14

$$dis = \sqrt{(2-1)^2 + (0{,}5-0{,}5)^2 + (0{,}5-0)^2 + (1-1)^2 + (0-0)^2 + (1-1)^2 + (0{,}5-0{,}5)^2 + (1-1)^2}$$
$$= 1{,}5155$$

Data 15

$$dis = \sqrt{(3-1)^2 + (0{,}5-0{,}5)^2 + (0{,}5-0)^2 + (1-1)^2 + (1-0)^2 + (1-1)^2 + (0{,}5-0{,}5)^2 + (1-1)^2}$$
$$= 2{,}5495$$

Data 16

$$dis = \sqrt{(1-1)^2 + (0-0{,}5)^2 + (0{,}5-0)^2 + (0-1)^2 + (1-0)^2 + (0-1)^2 + (0{,}5-0{,}5)^2 + (1-1)^2}$$
$$= 1{,}4142$$

Data 17

$$dis = \sqrt{(2-1)^2 + (0-0{,}5)^2 + (0{,}5-0)^2 + (1-1)^2 + (1-0)^2 + (1-1)^2 + (0{,}5-0{,}5)^2 + (1-1)^2}$$

= 2,1794

Data 18

$$dis = \sqrt{(2-1)^2 + (0{,}5-0{,}5)^2 + (0{,}5-0)^2 + (1-1)^2 + (1-0)^2 + (0-1)^2 + (0{,}5-0{,}5)^2 + (1-1)^2}$$

= 2,0615

Data 19

$$dis = \sqrt{(1-1)^2 + (0{,}5-0{,}5)^2 + (0{,}5-0)^2 + (1-1)^2 + (0-0)^2 + (0-1)^2 + (0{,}5-0{,}5)^2 + (1-1)^2}$$

=1,5155

Step 3: Sort the distance from the test data to the training data and determine the nearest neighbors based on the minimum distance K.

Table 2: Closest Distance Sorting of New Data with Training Data

PX No	K1	K2	K3	K4	K5	K6	K7	K8	*Euclidean Distance* (0,…1)	Distance Order	Is Included 3-NN
1	1	0,5	0,5	0	1	1	0,5	1	1,8027	7	Yes (K<3)
2	3	0	0,5	1	1	1	0,5	1	2,5495	15	Yes (K<3)
3	1	0,5	0	1	1	1	0,5	1	1,4142	3	Yes (K<3)
4	3	0	0	1	1	1	0,5	1	2,6925	16	Yes (K<3)
5	1	0,5	0,5	1	0	1	0,5	1	4,25	17	No (K>3)
6	2	0,5	0,5	1	1	0	0,5	1	2,0615	9	Yes (K<3)
7	3	0,5	0	0	1	1	0,5	1	7,25	19	No (K>3)
8	2	0,5	0	1	1	1	0,5	1	2	8	Yes (K<3)
9	1	0,5	0	0	0	1	0,5	1	1,118	1	Yes (K<3)
10	3	0	0,5	1	1	1	0,5	1	6,5565	18	No (K>3)
11	2	0,5	0,5	0	1	1	0,5	1	2,0615	10	Yes (K<3)
12	2	0,5	0,5	1	1	0	0,5	1	2,06155	11	Yes (K<3)
13	1	0,5	0,5	0	0	1	0,5	1	1,2247	2	Yes (K<3)
14	2	0,5	0,5	1	0	1	0,5	1	1,5155	5	Yes (K<3)
15	3	0,5	0,5	1	1	1	0,5	1	2,5495	14	Yes (K<3)
16	1	0	0,5	0	1	0	0,5	1	1,4142	4	Yes (K<3)
17	2	0	0,5	1	1	1	0,5	1	2,1794	13	Yes (K<3)
18	2	0,5	0,5	1	1	0	0,5	1	2,0615	12	Yes (K<3)
19	1	0,5	0,5	1	0	0	0,5	1	1,5155	6	Yes (K<3)

Step 4: Determine the category of nearest neighbors.

Table 3: Data Categories of Nearby Data

K1	K2	K3	K4	K5	K6	K7	K8	*Euclidean Distance* (0,…1)	Is Included 3-NN	Category Yes for KNN
1	0,5	0,5	0	1	1	0,5	1	1,8027	Yes (K<3)	Positive
3	0	0,5	1	1	1	0,5	1	2,5495	Yes (K<3)	Positive
1	0,5	0	1	1	1	0,5	1	1,4142	Yes (K<3)	Positive
3	0	0	1	1	1	0,5	1	2,6925	Yes (K<3)	Positive
1	0,5	0,5	1	0	1	0,5	1	4,25	No (K>3)	-
2	0,5	0,5	1	1	0	0,5	1	2,0615	Yes (K<3)	Negative
3	0,5	0	0	1	1	0,5	1	7,25	No (K>3)	-
2	0,5	0	1	1	1	0,5	1	2	Yes (K<3)	Positive
1	0,5	0	0	0	1	0,5	1	1,118	Ya (K<3)	Positive
3	0	0,5	1	1	1	0,5	1	6,5565	No (K>3)	-
2	0,5	0,5	0	1	1	0,5	1	2,0615	Yes (K<3)	Positive
2	0,5	0,5	1	1	0	0,5	1	2,06155	Yes (K<3)	Negative
1	0,5	0,5	0	0	1	0,5	1	1,2247	Yes (K<3)	Positive
2	0,5	0,5	1	0	1	0,5	1	1,5155	Yes (K<3)	Positive
3	0,5	0,5	1	1	1	0,5	1	2,5495	Yes (K<3)	Positive
1	0	0,5	0	1	0	0,5	1	1,4142	Yes (K<3)	Negative
2	0	0,5	1	1	1	0,5	1	2,1794	Yes (K<3)	Positive
2	0,5	0,5	1	1	0	0,5	1	2,0615	Yes (K<3)	Negative
1	0,5	0,5	1	0	0	0,5	1	1,5155	Yes (K<3)	Negative

Step 5: Using the Majority Category from the nearest neighbors as the prediction value for the new data.

Table 4: Category Data from Nearest Neighbors

K1	K2	K3	K4	K5	K6	K7	K8	Category Yes for KNN
0,5	0,5	0	1	1	0,5	1	0,5	Positive
0	0,5	1	1	1	0,5	1	0	Positive
0,5	0	1	1	1	0,5	1	0,5	Positive
0	0	1	1	1	0,5	1	0	Positive
0,5	0,5	1	0	1	0,5	1	0,5	Positive
0,5	0,5	1	1	0	0,5	1	0,5	Positive
0,5	0	0	1	1	0,5	1	0,5	Positive
0,5	0	1	1	1	0,5	1	0,5	Positive
0,5	0	0	0	1	0,5	1	0,5	Negative
0	0,5	1	1	1	0,5	1	0	Positive
0,5	0,5	0	1	1	0,5	1	0,5	Positive
0,5	0,5	1	1	0	0,5	1	0,5	Positive
0,5	0,5	0	0	1	0,5	1	0,5	Negative
0,5	0,5	1	0	1	0,5	1	0,5	Positive
0,5	0,5	1	1	1	0,5	1	0,5	Negative
0	0,5	0	1	0	0,5	1	0	Negative

From the data with the majority categories there are Positive and Negative categories. From the majority number (Positive > Negative) it can be concluded that new data (data No. 20) (K1=1, K2=0.5, K3=0, K4=1, K5=0, K6=1, K7=0, 5, K8=1) is included in the Positive category.

3.2. Implementation

In this paper, we will explain the results of testing the software for developing the K–Nearst Neighbor algorithm using Matlab (matrix laboratory) R2014a in order to find out the results of the classification for predicting heart disease using the K–Nearst Neighbor method at Sylvani General Hospital, Binjai City. Implementation here is an activity where data that has been transformed is applied to the programming used and processed according to the Classification method with the K–Nearst Neighbor algorithm, so that it can be seen to what extent the system's performance is in processing data and producing information according to user needs.

4. Results

4.1. Home Menu

Main Menu / Home Menu here displays the initial interface which contains the training menu, prediction menu, help and exit. When you click on the training menu you will immediately enter the training menu and if you want to exit or close the system then click the exit button. The display of the home menu page can be seen in Figure III.1.

Figure 1: Home Menu

4.2. Training Menu

This page is a page of training data results in importing training data and test data for testing into classification to predict heart disease using the K–Nearst Neighbor method at Sylvani general hospital, Binjai city. The training menu page display can be seen in Figure III.2.

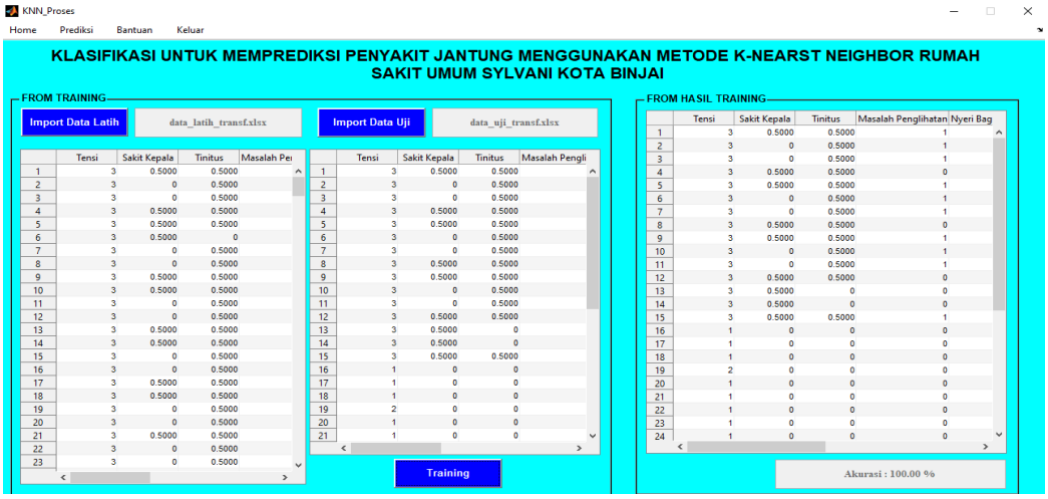

Figure 2: Training Menu

4.3. Prediction Menu

This page is a classification prediction results page for predicting heart disease using the K–Nearst Neighbor method at the Sylvani general hospital, Binjai city. The training menu page display can be seen in Figure III.3.

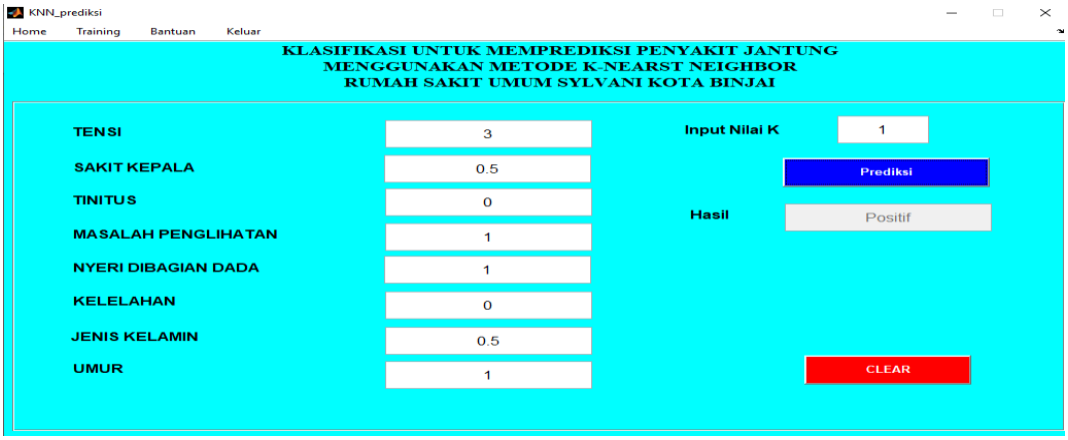

Figure 3: Prediction Menu

4.4. Help Menu

The Help page is used to display information on how to use the system and information on transformation codes for classification to predict heart disease using the K–Nearst Neighbor method at the Sylvani general hospital, Binjai city. For more details, see Figure III.4.

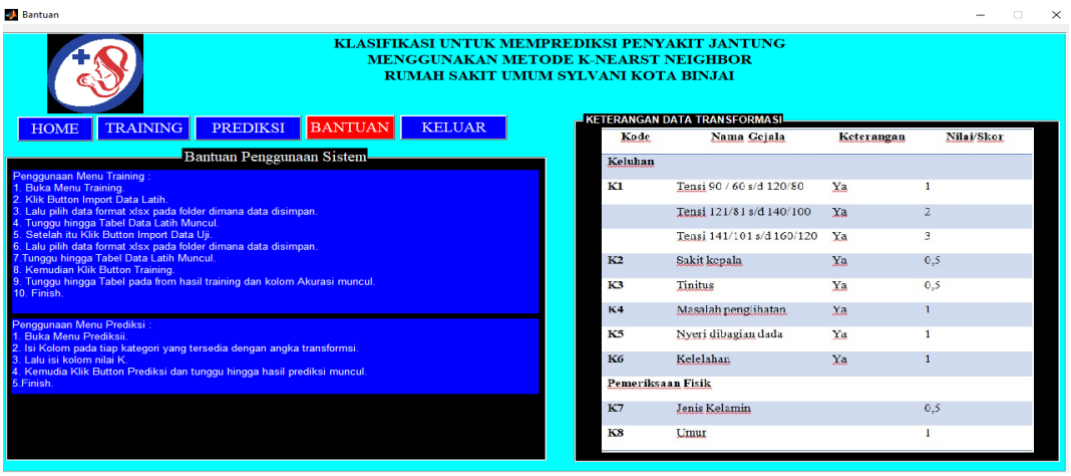

Figure 4: Help Menu

5. Conclusion

From the results of analysis and testing of the Classification system for predicting heart disease using the K–Nearst Neighbor method at Sylvani General Hospital, Binjai City, the following conclusions can be drawn:

1. From the results of user responses and testing of the classification system for predicting heart disease using the K–Nearst Neighbor method at Sylvani General Hospital, Binjai City, there were no error instructions in the program and the results displayed were in accordance with what was expected in the design.
2. From the results of system testing carried out using the Classification method to predict heart disease using the K–Nearst Neighbor method, Sylvani General Hospital, Binjai City utilizes the K value results that have been determined with Positive and Negative prediction results.
3. Test results on a system that has been built using the Matlab programming language.

References

[1] Amir Tjolleng, 2017, Pengantar Pemrograman MATLAB, Jakarta: PT. Alex Media.
[2] Arhami, Muhammad dan Muhammad Nasir. 2020. Data Mining Algoritma dan Implementasi. Yogyakarta: CV Andi Offset.
[3] Lamhot Sitorus, 2015, Algoritma dan Pemrograman, Yogyakara : Andi Offset
[4] Muhammad Arhami dan Muhammad Natsir, 2020, Data Mining Algoritma dan Implementasi, Yogyakarta: CV. Andi Ofset.
[5] Prasetyowati, Erwin. 2017. Data Mining Pengelompokan Data untuk Informasi dan Evaluasi. Pamekasan: Duta Media Publishing.
[6] Teguh Widiarsono, 2015, Tutorial Praktis Belajar Matlab, Jakarta : Republika.