CLINICAL DECISION SUPPORT SYSTEM BASED ON DECISION TREE ALGORITHM TO CLASSIFY HEART DISEASE

Sa'diyah Noor Novita Alfisahrin[1)], Tri Wahyudi[2)] , Supriyanta[3)]
[1)] Informatic Management, AMIK BSI Yogyakarta
Jl. Ringroad Barat, Ambarketawang, Gamping, Sleman, Yogyakarta
email: sa.snt@bsi.ac.id
[2)] Informatic Management, AMIK BSI Yogyakarta
Jl. Ringroad Barat, Ambarketawang, Gamping, Sleman, Yogyakarta
email: tri.twi@bsi.ac.id
[3)] Informatic Management, AMIK BSI Yogyakarta
Jl. Ringroad Barat, Ambarketawang, Gamping, Sleman, Yogyakarta
email: supriyanta.spt@bsi.ac.id

Abstract – Heart disease is the leading cause of death in the world, one of the best ways to reduce the death rate is by detect the symptoms in the early stages. Hospital information systems rarely provide a decision support system that can be used to detect early symptoms, most systems are designed only to support the payment of bills for patients, inventory management and also a simple statistical information, to overcome the problem, it can be used a computer-based information or clinical decision support systems. This study aims to build a clinical decision support system to identify whether a patient affected by heart disease or not by using a decision tree algorithm. System built using the rules generated by the decision tree algorithm as many as 75 rules, results show that the system has been built can be used as a way to detect early symptoms of heart disease.

Keywords: Data Mining, Clinical decision support system, Heart Disease, Algorithm.

I. INTRODUCTION

Heart disease is the leading cause of death in the world. Heart disease occurs because of damage or accumulation of plaque on artery walls and causes the blood flow to the heart stops or blockage [1]. The best way to reduce the death rate is by detecting the symptoms in the early stages. By early recognition of symptoms, one can obtain clinical care effectively dan can get the best results [2].

Today, many hospital information systems are designed only to support the bills of patient's payment, inventory management, and simple statistic information. Some hospitals are already using the decision support system, but still very limited [3], this is due to a decision support system in the field of health information technology is a complex issue and requires a hgh cost [4]. To minimize the cost, including clinical test's cost, it can be used a decision support system [5].

Clinical decision support system is a computer system that is designed to give an impact on clinical decision maker about the patient at the time the decision was made. Decision support system can be categorized into two parts, i.e. knowledge based system and non knowledge based system wich use machine learning technology [6].

Machine learning technology plays an important role in medical diagnosis and also suitable for analyzing medical data. Machine learning provides an indispensable tool for intelligent data analysis [7]. One of the machine learning technology is data mining, data mining has the potential to produce an environment that is rich of knowledge that can improve the quality of clinical decision significantly [5], and also useful to improve diagnostic accuracy [7].

This study is further research conducted by [1], which states that the decision tree has higher accuracy when compared with the naive bayes and neural network algorithm. This study aims to develop a decision support system to identify if a patients affected by heart disease using data mining algorithm, the algorithm is decision tree C4.5. The system was built can be used to detect the symptoms of heart disease, recomendation for the medical team and also as a learning media for non specialist physicians in spesific particular programs.

II. THEORY

2.1. Clinical Decision Support System

Clinical decision support systems is a computer system designed to give an impact clinical decision makers regarding a patient at the time the decision was made [6], where the characteristics of the patients matched to computerized clinical knowledge base [8]. Decision support systems is a part of a clinical information, decision support system is an information system that uses expert systems and artificial intelligence technology to support clinical decision [9].

Clinical decision support systems can be categorized into two, i.e. knowledge-based clinical decision support systems and non knowledge based clinical decision support system (using data mining) [10]. Knowledge based clinical decision support systems must be supplied by the facts and rules related

to the knowledge, thus knowledge based decision support systems requires a huge amount of knowledge that has been stored as part of the decision to give the right answer to the problem faced. While the decision support system that use data mining tool did not requires knowledge as a part of the decision-making, but the system is designed to discover new patterns and relationship previously unknown based on available data [11].

The architecture of decision support system consists of major components, i.e. the inference that includes a knowledge base and reasoning engine decision support, computerized models that include basic medical logic models and management systems, the human-computer interaction is the user interface of clinical decision support systems [8] and database system of data that containclinical information of patient for clinical decision [11].

2.2. Data Mining

Data mining is a discovery process of new patterns or relationship of data within large amount of data, which aims to prediction or description. Patterns discovered must be meaningful and provide some advantages [12].

Core of data mining is a model development of discovered patterns based on history data and apply the model to new data. The model can be applied to predict both classification and regression, segmentation, and define the relationship in the population (association) [13].

Some of the advantages of data mining are:

1. It enables the creation of model explaining trends and patterns, which can be stored in a standardized form to a shared source.
2. It is an algorithm-based approach, because the models are acquired by data mining methods and algorithms.
3. The majority of data mining algorithms are capable of creating models based on a large number of attributes.

The power of data mining lies in its ability to allow user to consider data from a variety of perspective to find a hidden pattern [11]. In a simple data mining process can be divided into four [13]:

1. The process of obtaining and preparing the data
2. The process of building a model
3. The process of assesing the quality of models and reviewing the model in detail
4. The process of applying the model to new data for prediction or assesment.

2.3. Decision Tree

Decision tree is one of data mining techniques that can be used to perform classification. Classification is used to classify an object based on the attributes of the class which has been determined previously [14]. Decision tree consists of nodes, branches, and leaves that can be used to perform searches in order to determine the class from the result of classification.

The porcess of developing a decision tree can be seen in figure 1.

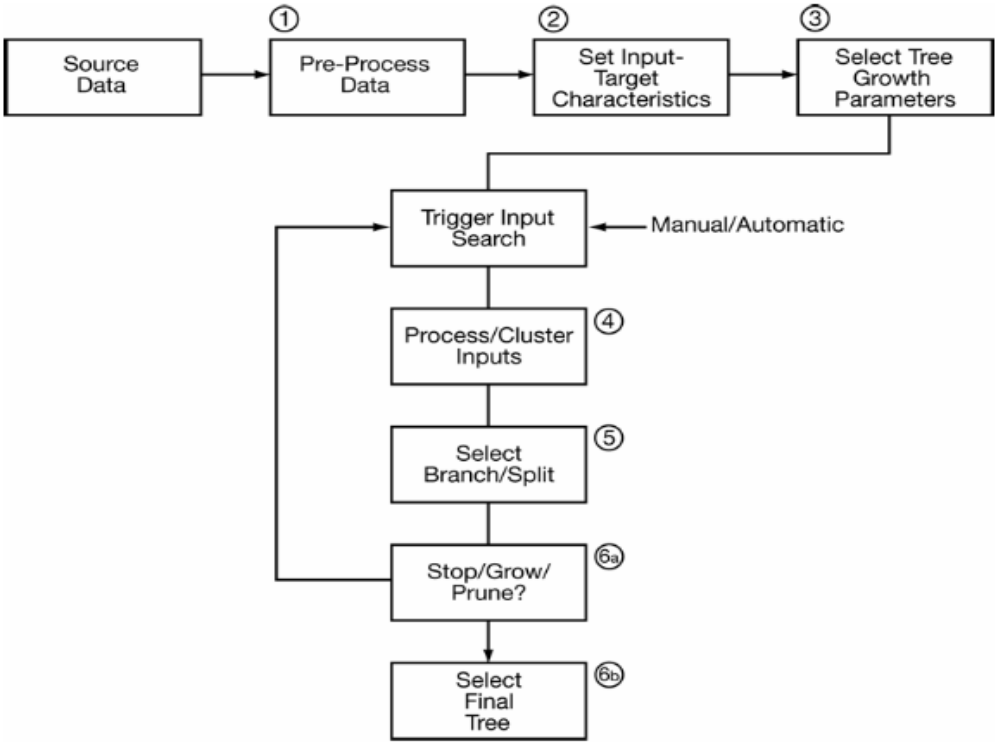

Figure 1. The decision tree process

From the figure 1, we can see the process of the formation of decision tree is repeated by developing each layer from the branches. One of decision tree algorithm is C4.5 algorithm. The steps to build a decision tree using C4.5 algorithm is as follow:

1. Select attribute as a root, selection of attributes as the root based on the highest gain value from existing attributes. To calculate the highest gain value using the following equation:

   Gain(S,A) = Entropy(S) - $\sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy(Si)$ (1)

   Where:
   S = the set of cases
   A = Attribute
   |Si| = The number of cases in the i-th partition
   |S| = the number of cases in S
   Entropy value can be calculated in the following way:

   Entropy(S) = $\sum_{i=1}^{n} - pi * \log_2 pi$ (2)

   Where:
   S = the set of cases
   n = number of partitions S
   pi = the proportion of Si to S
2. Create a branch for each value
3. Split the cases into branch
4. Repeat the process for each branch

III. RESULTS AND DISCUSSION

4.1. Data Source

The data source used in this study is a clinical information of patient from university of california irvine machine learning data repository. The data consists of 270 patients, where 151 patients are affected by heart disease and the remaining, as many as 119 patients are not affected by heart disease.

To diagnose or classify if the patients have the heart disease, used attributes such as age, sex, chest paint type, resting blood pressure, serum cholesterol in mg/dl, fasting blood sugar > 120 mg/dl, resting electrocardiographic result, maximum heart rate achieved, exercise induced angina, oldpeak, the slope of the peak exercise ST segment, Number of major vessel, and thalium. The value of the attributes then transformed into a categorical form to facilitate the decision tree modelling. The result of the transformation of the attributes can be seen in the following table.

Table 1. The Attributes of Heart Disease Classification

Attributes	Value	Categorical Value
Age	< 40	Iuventus
	40-54	Verilitas
	55-64	Prasenium
	>=65	Senium
Sex	1	Male
	0	Female
Chest Paint Type	1	Typical Angina
	2	Atypical Angina
	3	Non Angina Pain
	4	Asymptomatic
Tresting Blood Pressure	<120	Normal
	120-139	Prehypertension
	140-159	Hypertension Level I
	>159	Hypertention Level 2
Serum Cholestoral in mg/dl	<200	Normal
	200-239	High Normal Range
	>239	High
Fasting Blood Sugar >120 mg/dl	0	No
	1	Yes
Resting Electrocardiographic Result	0	Normal
	1	Having ST-Twave abnormality
	2	Left ventricular hypertrophy
The slope of the peak exercise ST segment	1	Unsloping
	2	Flat
	3	Downsloping

Exercise Induced Angina	0	No
	1	Yes
Old Peak	<1	0
	>=1 and <2	1
	>=2 and <3	2
	>=3 and <4	3
Number of Major Vessels	1	1
	2	2
	3	3
Maximum Heart Rate	=220-age	Normal
	<>220-age	Abnormal
Thal	3	Normal
	6	Fixed Defect
	7	Reversible Defect

4.2. The Basic Model of Medical Logic and Knowledge Base

Decision tree algorithm is used to create a basic model that is used to classify the patient if the patient affected by heart disease. Modelling is done by finding the highest gain value of each attributes. After calculating the gain value, it will form a decision tree as shown in figure 1.

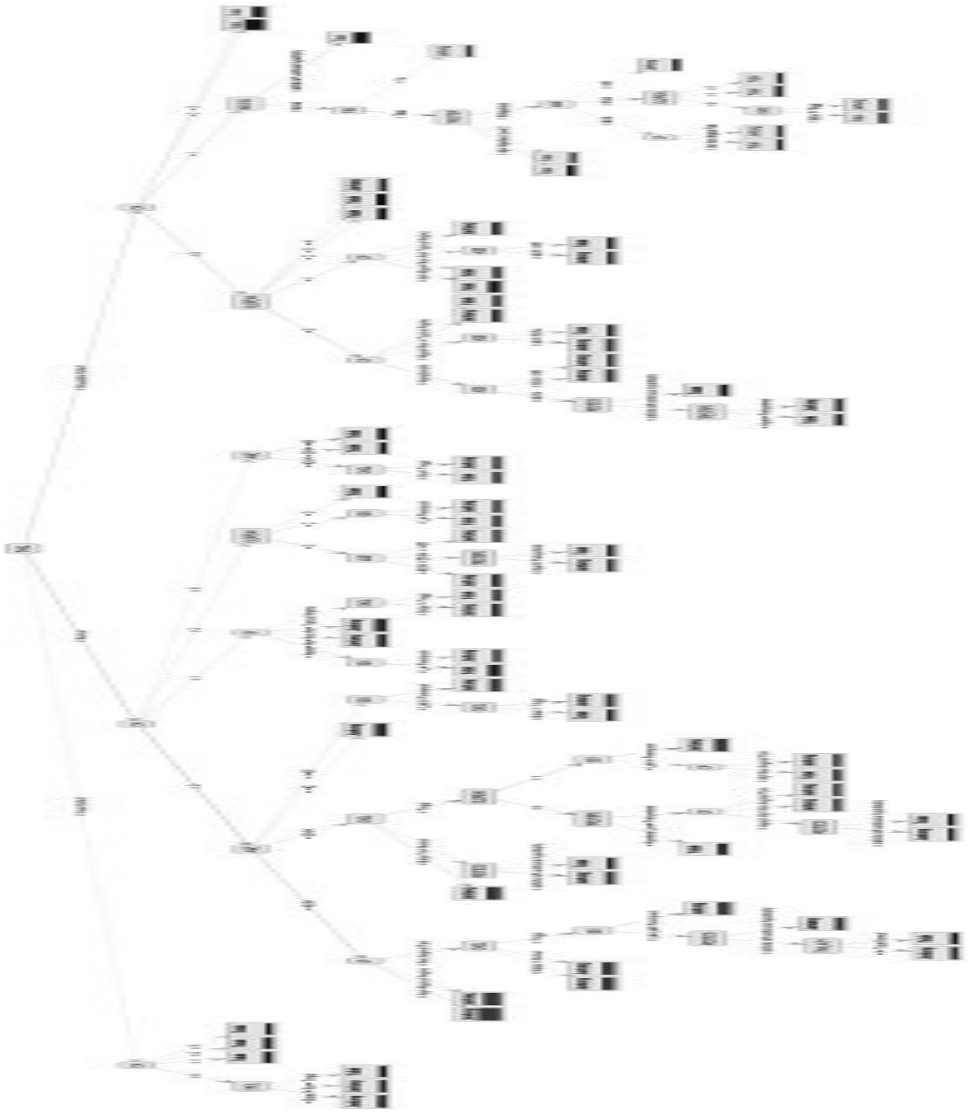

Figure 2. The Decision Tree Model

From figure 2. There are 75 rules that can be applied to classify wheter a patient has heart disease or not.

4.3. User Interface

The rules obtained from the decision tree model will be used as the basis of the rules in making the user interface. The user interface can be used to facilitate medical expert in diagnosing heart disease and it can also be used as learning materials for non-spesialist doctors. The user interface can be seen in figure 3.

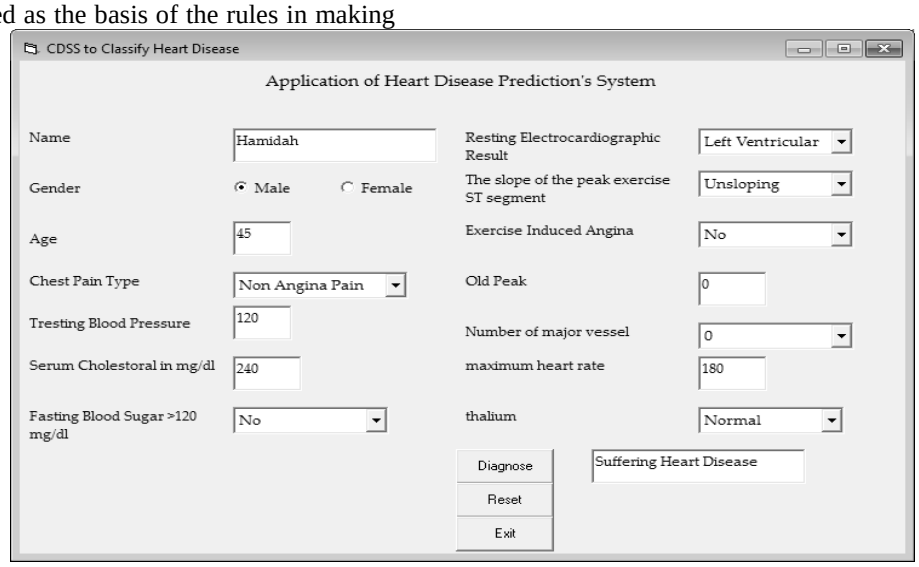

Figure 3. The Application of Heart Disease Prediction's System

IV. CONCLUSION

In this study, a decision tree algorithm is used to establish the rules necessary to make a clinical decision support system applications. The number of rules obtained from the decision tree algorithm are 75 rules. The Application can be used to classify wheter a patient is suffering from heart disease or not. The application can also be used as a learning media for non-spesialist physicians in the diagnosis of spesific problems.

Although the rules and the application that was built has a good accuracy, but it would be better if the classification of heart disease combined with image data processing.

REFERENCES

[1] M. Wahyudi dan S. N. N. Alfisahrin, "Komparasi Algoritma C4.5, Naive Bayes, Dan Neural Network Untuk Memprediksi Penyakit Jantung," *Ticom*, vol. 2, no. 2, Januari 2014.

[2] T. D. Pham, H. Wang, X. Zhou, D. Beck, M. Brandl, G. Hoehn, J. Azok, M.-L. Brennan, S. L. Hazen, K. Li dan S. T. C. Wong, "Computational Prediction Models for Early Detection of Risk of Cardiovascular Events Using Mass Spectrometry Data," *IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE*, vol. 12, no. 5, pp. 636-643, September 2008.

[3] G. Subbalakshmi, K. Ramesh dan M. C. Rao, "Decision Support in Heart Disease Prediction System using Naive Bayes," *Indian Journal of Computer Science and Engineering*, vol. Vol. 2, no. No. 2 , Apr-May 2011.

[4] A. Gupta dan R. Sharda, "Improving the science of healthcare delivery and informatics using modeling approaches," *Decision Support Systems*, vol. Volume 55, no. Issue 2, pp. Pages 423-427, 2012.

[5] K. Srinivas, B. K. Rani dan A. Govrdhan, "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks," *International Journal on Computer Science and Engineering*, vol. 02, no. 02, pp. 250-254, 2010.

[6] E. S. Berner dan T. J. L. Lande, "Overview of Clinical Decision Support Systems," dalam *Clinical Decision Support Systems Theory and Practice*, New York, Springer, 2007, pp. 3-22.

[7] I. Kononenko, "Machine Learning for Medical Diagnosis: history, state of the art, and perspektif," *Artificial Intelligence in Medicine*, vol. 23, pp. 89-109, 2001.

[8] I. Sim, P. Gorman, R. A. Greenes, R. B. Haynes, B. Kaplan, H. Lehmann dan P. C. Tang, "Clinical

Decision Support Systems for the Practice of Evidence-based Medicine," *Journal of the American Medical Informatics Association,* vol. 8, no. 6, p. 527–534, Nov-Dec 2001.

[9] B. Keltch, Y. Lin dan C. Bayrak, "Advanced decision support for complex clinical decisions," *J. Biomedical Science and Engineering,* pp. 509-516, May 2010.

[10] A. P. K., "Clinical Decision Support System: Risk Level Prediction Of Heart Disease Using Decision Tree Fuzzy Rules," *Asian Transactions on Computers,* vol. 2, no. 4, September 2012.

[11] J. M. Hardin dan D. C. Chhieng, "Data Mining and Clinical Decision Support Systems," dalam *Clinical Decision Support Systems Theory and Practice Second Edition*, USA, Springer Science, 2007, p. 44.

[12] I. H. Witten, E. Frank dan M. A. Hall, Data Mining Practical Machine Learning Tools and Techniques Third Edition, United States: Morgan Kaufmann, 2011.

[13] M. F. Hornick, E. Marcadé dan S. Venkayala, Java Data Mining: Strategy Standard, and Practice A Practical Guide for Architecture, Design, and Implementation, San Francisco: Morgan Kaufmann, 2007.

[14] O. Maimon dan L. Rokach, Data Mining and Knowledge Discovery Handbook Second Edition, New York: Springer, 2010.

[15] B. d. Ville, Decision Trees for Business Intelligence and Data Mining Using SAS Enterprise Miner, United States of America: SAS Publishing, 2006.

Sa'diyah Noor Novita Alfisahrin, is currently a lecturer of the Study Program of Informatic management, AMIK BSI. He received a Master Degree in Computer Science from STMIK Nusa Mandiri in 2012 on "E-Business". Sa'diyah Noor Novita Alfisahrin, M. Kom research interests are in Data Mining, information system.

Tri Wahyudi, received master degree in computer science from STMIK Nusa Mandiri. He is now active as a lecturer of computer science at AMIK "BSI Yogyakarta". Tri Wahyudi research interest is in data mining.

Supriyanta, received master degree in computer science from STMIK Nusa Mandiri. He is now active as a lecturer of computer science at AMIK "BSI Yogyakarta". supriyanta research interest is in data mining.