



## Bayes Theorem and Forward Chaining Method On Expert System for Determine Hypercholesterolemia Drugs

Anna Adi Perbawawati<sup>1</sup>, Endang Sugiharti<sup>2</sup>, Much Aziz Muslim<sup>3</sup>

<sup>1,2,3</sup>Department of Computer Science, Natural Sciences, State University of Semarang  
Email: 1annaadipey@gmail.com, 2endangsugiharti@mail.unnes.ac.id, 3a212muslim@yahoo.com

### Abstract

The development of technology capable to imitating the process of human thinking and led to a new branch of computer science named the expert system. One of the problem that can be solved by an expert system is selecting hypercholesterolemia drugs. Drug selection starts from find the symptoms and then determine the best drug for the patient. This is consist with the mechanism of forward chaining which starts from searching for information about the symptoms, and then try to illustrate the conclusions. To accommodate the missing fact, expert systems can be complemented with the Bayes theorem that provides a simple rule for calculating the conditional probability so the accuracy of the method approaches the accuracy of the experts. This reseacrh uses 30 training data and 76 testing data of medical record that use hypercholesterolemia drugs from Tugurejo Hospital of Semarang. The variable are common symptoms and some hypercholesterolemia drugs. This research obtained a selection of hypercholesterolemia drugs system with 96.05% accuracy.

**Keywords:** Expert System, Cholesterol, Hypercholesterolemia, Forward Chaining, Bayes.

### 1. INTRODUCTION

The development of technology capable to imitating the processes of human thinking and led to a new branch of computer science named the expert system [1]. Expert system is a system with computer program based on knowledge [2]. An expert system is a computer program which captures the knowledge of a human expert on a given problem [3]. The purpose of expert system is not to replace the human roles, but to presented human knowledge in forms of a system [4]. So many people can use the expert system without meet directly with the expertise. Expert systems can be applied in medical field to help the doctors and paramedics to diagnose the disease faster by providing a knowledge base of the doctor who acts as an expert in a computer system.

Hypercholesterolemia represents one of the cardinal risk factors for coronary heart disease [5]. In a serious cases hypercholesterolemia control is necessary for a lifetime, so hypercholesterolemia drugs must be used in the long term. Paramedic and doctor is

the least impact on the treatment of patients [7]. If the duty of a doctor is unable to serve the patients due to certain problems, then there will be a problems with the medical services.

One of search engine that widely used in the manufacture of an expert system is a forward chaining method. Operation of a forward chaining system starts by inserting a set of known facts into working memory, then lower the new facts based on the premise rules match with known facts, and this process is continued until reaching a goal or no longer rules the premise fits the known facts [7]. To accommodate incomplete fact, the system need a method that handle the uncertainties. Within metrology the evaluation of measurement uncertainty plays a special and very important role [8]. Bayes theorem provides a simple rule for calculating the conditional probability of a system so the value of the accuracy of the method approaches the accuracy of the experts [9]. Bayes theorem is a simple classification algorithms but have the capability and high accuracy [10].

The previous research [11], giving the result of the system with 100% accuracy based on black box test and 65.7% based on alpha test. Additionally the information generated can be used as an alternative to consult experts on diseases of prawns that includes the name of the disease, the definition, symptoms, causes, solutions, and probability. While Fiarni, had a research that presents a practical method for modeling and solving a dynamic resource allocation of automatic scheduling problem using forward chaining heuristic approach, in the case of undergraduate Student's Thesis and Project presentations timetable. This research has a high adaptability to provide an alternative presentation before recorded in a complete schedule. Tests on the system indicates that the schedule generated by the system conforms to 100% with a schedule made by the expert [12].

Based on the introduction and purpose of the above, this research apply forward chaining method for determining the hypercholesterolemia drug and Bayes theorem to determine the accuracy in making an expert system hypercholesterolemia drug selection.

## **2. METHOD**

### **2.1. Forward Chaining**

Forward chaining used for the selection of drugs in accordance with the facts that has been entered. Forward chaining is a method that based on data or facts leading to the conclusion [13]. The operation of forward chaining starts with input the facts into working memory, then match the fact with known *rules* [14]. Defining the structure of data control rules written in the structure of If - Then and given a number of rules to distinguish the rules with each other. If the data is match, then the rule is executed and operation stopped when no more rules can be executed [15]. Flowchart of forward chaining method shown in Figure 1.

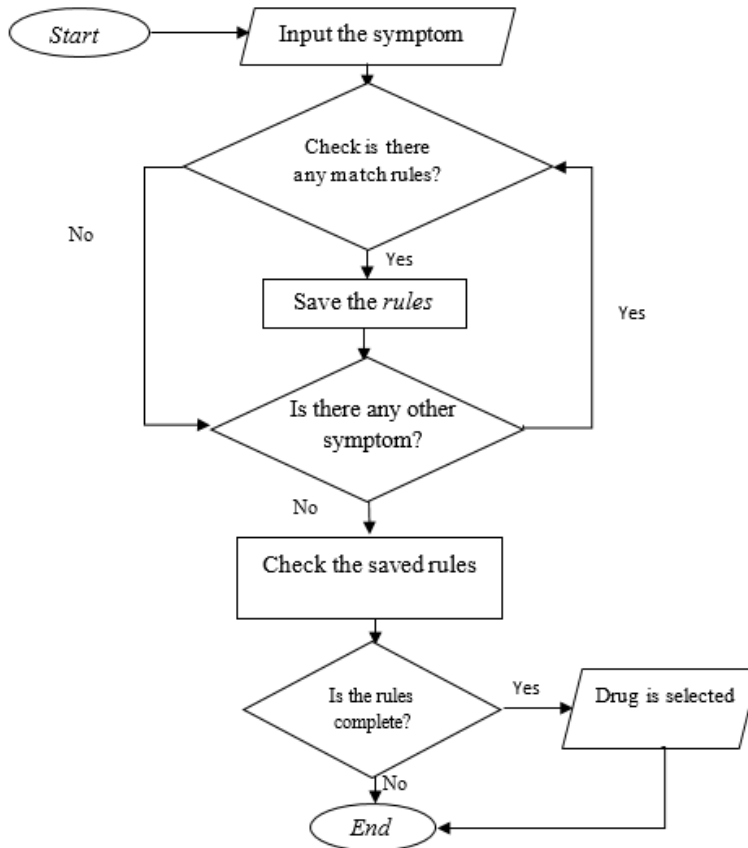


Figure 1. Flowchart of Forward Chaining Method

## 2.2. Bayes Theorem

Bayes theorem is a method found by Thomas Bayes [16]. It works for determine the level of uncertainty of the drug. The process of calculating the Bayes theorem using optional data will result in the value of the probability. The results of the obtained data will be calculated using Bayes theorem shown with [17]:

$$P(C_i|X) = \frac{p(X|C_i) * p(C_i)}{p(X)} \quad (1)$$

This formula can be developed as formula (2) shown in [18]:

$$P(C_i|X) = \frac{p(X|C_i) * p(C_i)}{\sum_{k=1}^n p(X|C_k) * p(C_k)} \quad (2)$$

In this case:

$P(C_i | X)$  = Posterior probability of hypothesis  $C_i$  based on the condition of  $X$

$p(X | C_i)$  = Conditional probability of  $X$  based on the condition of hypothesis  $C_i$ , and its called likelihood

$p(C_i)$  = Probability of the hypothesis regardless previous evidence

$p(X)$  = Data of unknown class (the set of training data)

$n$  = Number of hypotheses that might

### 2.3. Expert System Design and Development

This phase clarifies the description of how input and output is processed to produce the system [20]. It is build with a waterfall model, which consist of requirement analysis, design, coding, and testing [21].

## 3. RESULTS AND DISCUSSION

### 3.1. Data Resource

This phase use quantitative method, a method that start with searching data, describe, and show the conclusion [22]. The techniques of data resource are observation, interview, and study of literature [23]. This research is using the data of medical records in 2016 - 2017 that use the hypercholesterolemia drug from Tugurejo Hospital of Semarang. Data used in the sistem was the symptoms experienced by patients, the sustain disease and drug selection by expertise.

### 3.2. Data processing

Based on the results of interviews, it gave the source of the knowledge such a symptoms that used on the system. It shows in Table 1.

Table 1. Symptoms Table for Determining Drug

No.	Category Symptoms	Symptom	Code
1.	Main symptom	The Cholesterol Level	G01
		High level of triglycerides	G02
		High level of LDL	G03
		Low level of HDL	G04
2.	Support symptoms	Have a history or experiencing heart disease	G05
		Weigh above average (Obesity) based on IBM	G06
		Have a history or experiencing diabetes mellitus disease	G07
		Have a history or experiencing hypertension disease	G08
		Have a history or experiencing stroke disease	G09

### 3.3. Production Rules

Production rule written in the form of IF-THEN [24]. This rule is said to be a relationship between the two parts, the premise (IF) and the conclusion (THEN). Production rule written in the decision table that serves to save the relationship between the data and symptoms of drug election results. The relation is shown in Table 2.

Table 2. Table Relationships between Symptoms Drugs with the electoral results

Rule	IF	THEN
1	G01 ^ G03 ^ G04 ^ G08 ^ G09	Simvastatin (R01)
2	G01 ^ G02 ^ G03 ^ G04 ^ G05	Atorvastatin (R02)
3	G01 ^ G02 ^ G06	Lovastatin (R03)

4	$G01 \wedge G02 \wedge G04 \wedge G08$	Gemfibrozil (R04)
5	$G01 \wedge G02 \wedge G07 \wedge G08$	Fenofibrat (R05)
6	$G01 \wedge G02 \wedge G03 \wedge G04 \wedge G08 \wedge G09$	Simvastatin dan Fenofibrat (R06)

### 3.4. Calculation

The process has two possibilities. The first possibility is the process stalled at the node drug (drug selection is detected) and stopped on node 0 or \*\* (drug detection). Drug selection is detected when the final outcome of the process showing one drug only, if the final result of more than one drug then the search stops and there is no suitable drug (drug detection). For example, given data below:

1. Cholesterol levels of 200 mg / dl
2. Triglyceride levels of 278 mg / dl
3. LDL levels of 69 mg / dl
4. HDL levels of 33 mg / dl
5. Not being / have a history of heart disease
6. Ideal weight
7. Not being / have a history of diabetes mellitus
8. Not being / have a history of hypertension
9. Not being / have a history of stroke

From the above data it is known that patients with high cholesterol levels, high triglyceride levels and low HDL levels. It can be written as follows:

G01 = Yes	G02 = No
G02 = Yes	G03 = No
G03 = No	G04 = No
G04 = Yes	G09 = No
G01 = No	

Forward chaining process is shown in Table 3.

Table 3. Forward Chaining Process Node Stop Drugs.

Symptoms code detected	Answer	Directions searches (Symptom Code)	Terseteksi drug code	Information
G01	Yes	To the left (G02)	R01, R02, R03, R04, R05, R06	search continues
G02	Yes	To the left (G03)	R02, R03, R04, R05, R06	search continues
G03	No	To the left (G04)	R03, R05	search continues
G04	Yes	Finished (G08)	R05	Drugs Detected

From Table 3 we know that the process choose RO5 (fenofibrate) for the conclusion. If training data shows that there are 30 patient, and the patients that use fenofibrate are 5 person, then the probability p(fenofibrate) shown as below:

$$\begin{aligned}
p(\text{Fenofibrat}) &= \frac{\text{The total of patient that use Fenofibrate regardless of any symptoms}}{\text{The total of patient on training data}} \\
&= 5/30 \\
&= 0.167 \\
&= 0.2
\end{aligned}$$

From the calculation, we know that the probability of fenofibrate regardless of any symptoms is 0.2. the next step is find the probability of fenofibrat when symptoms are given, its written  $p(\text{Fenofibrate}|\text{X})$ . So the value shown as below:

1. If training data shows the total of patient that use fenofibrat are 5 person, the patient with symptom G01 that use fenofibrat are 4 person, then  $p(\text{fenofibrat}|\text{G01})$  is 4/5 or 0.8.
2. If training data shows the total of patient that use fenofibrat are 5 person, the patient with symptom G02 that use fenofibrat are 5 person, then  $p(\text{fenofibrat}|\text{G02})$  is 5/5 or 1.
3. If training data shows the total of patient that use fenofibrat are 5 person, the patient with symptom G01 that use fenofibrat are 5 person, then  $p(\text{fenofibrat}|\text{G04})$  is 5/5 or 1.
4. If training data shows the total of patient that use fenofibrat are 3 person, the patient with symptom G08 that use fenofibrat are 5 person, then  $p(\text{fenofibrat}|\text{G08})$  is 3/5 or 0.6.

The next step is find the value of  $p(X)$ . The probability of the symptoms are shown in formula (2) where:

$$p(X) = \sum_{k=1}^n (p(X|C_i) * p(C_i))$$

To determine that patients should use fenofibrate, there are four symptoms that have been determined by the experts, there are G01, G02, G04, and G08. The value is calculated by knowing the weight of each symptom, as follows:

$$\begin{aligned}
p(\text{G01, G02, G04, G08}) &= (p(\text{G01}|\text{R05}) * p(\text{R05})) + (p(\text{G02}|\text{R05}) * p(\text{R05})) + \\
&\quad (p(\text{G04}|\text{R05}) * p(\text{R05})) + (p(\text{G08}|\text{R05}) * p(\text{R05})) \\
&= (0.8 * 0.2) + (1 * 0.2) + (1 * 0.2) + (0.6 * 0.2) \\
&= 0.68
\end{aligned}$$

From the calculation, we know that the probability  $p(\text{G01, G02, G04, G08})$  is 0.68. The last step is find the probability of fenofibrate when the symptoms (G01, G02, and G08) given. Using formula (1) we can find the probability shown as follows:

$$\begin{aligned}
p(\text{R05}|\text{G01}) &= \frac{0.8 * 0.2}{0.68} = 0.2353 \\
p(\text{R05}|\text{G02}) &= \frac{1 * 0.2}{0.68} = 0.2941 \\
p(\text{R05}|\text{G04}) &= \frac{1 * 0.2}{0.68} = 0.2941
\end{aligned}$$

So the probability value chosen fenofibrate drug when the patient's symptoms G01, G02, G04 given is based on formula (1) is as follows:

$$\begin{aligned}
p(\text{Fenofibrat}|\text{G01, G02, G04}) &= \text{Bayes 1} + \text{Bayes 2} + \text{Bayes 3} \\
&= 0.2353 + 0.2941 + 0.2941
\end{aligned}$$

$$= 82,35$$

The values are converted into the form of a percentage by multiplying the final result with 100%, and the final result is 82.35%. To calculate the value of the accuracy of the system, researchers tested 76 medical records that had been taken from the hospital. According to the test, 73 data is correct and 3 data is fail to get the conclusion. So the accuracy of the system is:

Number of data test = 76

The corresponding data = 73

The uncorresponding data = 3

$$\text{Accuracy} = \frac{73}{76} \times 100\% = 96.05\%$$

### 3.5. Implementation

The expert system of drug selection of hypercholesterolaemia is a system that help the doctor and paramedics in handling the selection of hypercholesterolemic drugs that implemented in patients so it can be easier and helpful in choosing a drug. The system is implemented on a web browser and designed using the Laravel PHP framework programming language and MySQL database.

### 3.6. Discussion

Forward chaining works in 3 steps. The first step is make a rules based on data that have discussed with the expertise. The second is arrange the production rules based on premise and conclusion principle. This process produce 9 production rules shown in Table 1. The third step is find the conclusion which in this case is find the hypercholesterolemia drugs with decision table. After finding the coclusion, the accuracy will be calculated with Bayes theorem by using formula (1) and formula (2). Researchers tested 76 medical records that had been taken from the hospital. According to the test, 73 data is correct and 3 data is fail to get the conclusion. So the accuracy of the system is 96.05 %.. The failure can be caused by some factor such an eror human from the expertise or an incomplete data given to the system. This expert system has several advantages and disadvantages. The advantages of the system is: the system uses a web base so that it can be accessed easily by the user because the system has a user friendly inteface, the system can perform the selection of drugs hiperkolesterolemia in accordance with expert thinking so the system can help by replacing the expert if the expert is in place or have a job somewhere. With these capabilities, the system can facilitate patients to choose a hypercholesterolemia drug without meeting the experts. While the disadvantage of the system is: it only able to choose 5 kind of hypercholesterol drugs. In the system if we want to select the value of total cholesterol, LDL, HDL, and Triglycerides there are only 2 choices, high and low, so there is no difference between the patient with not so high symptom and the petient with very high symptom so the calculation is not balance.

## 4. CONCLUSION

Forward chaining and Bayes theorem that applicated on the expert system for determine the hypercholesterolemia drug, obtained a conclusion that the forward chaining works in three steps. The first step is to make the relationship between drug use on systems with symptoms. The second step which constitutes the basic rule

based on the rules of production. And the third is find the conclusion. If the premise part is complete, the conclusion would be true. If the conclusion is true, then the recommended drugs will be detected. Next step is to measure the accuracy of forward chaining method using Bayes theorem. The expert systems for determine hypercholesterolemia drug is made with the PHP programming language and MySQL database. The accuracy of the system is 96.05%.

## 5. REFENECES

- [1] Muslim, M. A., Kurniawati, I. I. N., & Sugiharti, E. (2015). Expert system diagnosis chronic kidney disease based on mamdani fuzzy inference system. *Journal of Theoretical and Applied Information Technology*, 78(1), 70.
- [2] Holel, K.R and Gulhane, V.S. (2014). Rule-Based Expert System for the Diagnosis of Memory Loss Diseases. *IJISET - International Journal of Innovative Science, Engineering & Technology*, 1(3).
- [3] Dath, A., & Balakrishnan, M. (2013). Development of an expert system for agricultural commodities. *International Journal of Computer Science & Applications (TIJCSA)*, 2(07).
- [4] Setiabudi, W. U., Sugiharti, E., & Arini, F. Y. (2017). Expert System Diagnosis Dental Disease Using Certainty Factor Method. *Scientific Journal of Informatics*, 4(1), 43-50.
- [5] Ford, E. S., Li, C., Pearson, W.S., Zhao, G., Mokdad, A. H. (2008). Trends in Hypercholesterolemia, Treatment and Control Among United States Adults. *International Journal of Cardiology*, 226-235.
- [6] Muslim, M.A. (2012). Pengembangan Sistem Informasi Jurusan Berbasis Web untuk Meningkatkan Pelayanan dan Akses Informasi. *Jurnal MIPA* 35(1).
- [7] Nusai, C., Cheechang, S., Chaiphech, S., & Thanimkan, G. (2015). Swine-Vet: a Web-based Expert System of Swine Disease Diagnosis. *Procedia computer science*, 63, 366-375.
- [8] Bernardo, J. M., & Smith, A. F. (2001). Bayesian theory. *Measurement Science and Technology*, 12 (2), 221-230.
- [9] Rohmana, I, & Arifudin, R. (2014). Perbandingan Jaringan Syaraf Tiruan dan Naive Bayes dalam Deteksi Seseorang Terkena Penyakit Stroke. *Jurnal MIPA*, 37 (2), 178-191.
- [10] Rish, I. (2001). An empirical study of The Naive Bayes Classifier. *International Joint Conference on Artificial Intelligence-California*. 3(22), 41-46.
- [11] Wahudi, M. J. (2013). Sistem Pakar untuk Mengidentifikasi Penyakit Udag Galah dengan Metode *Theorema Bayes*. *Jurnal Sarjana Teknik Informatika*, 1(1).
- [12] Fiarnia, C., & Gunawan, A.S. (2015). Automated Scheduling System for Thesis and Project Presentation Using Forward Chaining Method With Dynamic Allocation Resources. *Scient Direct, Procedia Computer Science* ,72(1), 209-216.
- [13] Pramesti, A. A., Arifudin, R., & Sugiharti, E. (2016). Expert System for Determination of Type Lenses Glasses Using Forward Chaining Method. *Scientific Journal of Informatics*, 3(2), 177-188.
- [14] Putra, Y.S., Muslim. M.A., & Naba, A. (2013). Game Chicken Roll dengan



- Menggunakan Metode Forward Chaining. *Jurnal EECCIS*, 7(1), 41-46.
- [15] Triyanto, S., & Fadlil, A. (2013). Sistem Pakar untuk Mendiagnosa Penyakit Kelinci Berbasis Web. *JSTIE (Jurnal Sarjana Teknik Informatika)(E-Journal)*, 2(1), 22-32.
- [16] Trihartati, A & Adi, C.K. (2016). An Identification of Tuberculosis (TB) Disease in Humans using Naïve Bayesian Method. *Scientific Journal of Informatics*, 3(2), 99-108.
- [17] Sugiharti, E., Firmansyah, S., & Devi, F. R. (2017). Predictive Evaluation of Performance of Computer Science Students of Unnes Using Data Mining Based on Naïve Bayes Classifier (NBC) Algorithm. *Journal of Theoretical and Applied Information Technology 28th February 2017*, 95(4), 902.
- [18] Pal, S. K. & Shiu, S.C.K. (2004). *Foundations of Soft Case Based Reasoning*. Canada: Willey & Sons, Inc.
- [19] Purwinarko, A. (2014). Model Expertise Management System in Semarang State University. *Scientific Journal of Informatics*, 1(2), 177-184.
- [20] Alamsyah, A., & Muna, I. H. (2016). Fuzzy Inference System untuk Penilaian Kinerja Pegawai Perpustakaan dan Pustakawan. *Scientific Journal of Informatics*, 3(1), 88-98.
- [21] Adenowo, A.A.A., Adenowo. B. A. (2013). Software Engineering Methodologies: A Review of the Waterfall Model and Object-Oriented Approach. *International Journal of Scientific & Engineering Research*, (4)7, 427-434
- [22] Alamsyah, A., & Arus, A. A. (2015). Analisis Sistem Pendaftaran pada Web Forum Ilmiah Matematika Unnes 2014. *Scientific Journal of Informatics*, 1(1), 107-117.
- [23] Handarko, J. L & Alamsyah. (2015). Implementasi Fuzzy Decision Tree untuk Mendiagnosa Penyakit Hepatitis. *UNNES Journal of Mathematics*, 4(2), 157-164.
- [24] Patil, T.R., & Sherekar, S.S. (2013). Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification. *International Journal Of Computer Science And Applications*,6(2).