The Comparison between Bayes and Certainty Factor Method of Expert System in Early Diagnosis of Dengue Infection

Eka Yuni Rachmawati [1], Budi Prasetiyo [2], Riza Arifudin [3]

[1,2,3]Department of Computer Science, Faculty of Mathematics and Natural Sciences,
Universitas Negeri Semarang, Indonesia
Email: [1]ekayuni43@gmail.com, [2]prasemath@gmail.com, [3]riza.arifudin@gmail.com

Abstract

The development of existing artificial intelligence technology has been widely applied in detecting diseases using expert systems. Dengue Infection is one of the diseases that is commonly suffered by the community and may cause in death. In this study, an expert diagnosis system for dengue infection is made by comparing between both Bayes method and Certainty Factor. The aims are to build an expert system using Bayes and Certainty Factor for early diagnosis of dengue infection and also to determine their level of accuracy. There are 80 data used in this study which are obtained from the medical records of Sekaran Health Center in Semarang City. The test results show that the level of accuracy obtained from 80 medical record data for Bayes method is 90% and the Certainty Factor method is 93,75%.

Keywords: Expert System, Bayes, Certainty Factor, Dengue Infection

1. INTRODUCTION

The development of both information and technology has recently undergone a lot of rapid changes, along with the more and more complex human needs. All aspects of human activities cannot be separated from the use of computers. A branch of computer science that might help human performances is an expert system which is as a subfield of artificial intelligence. Artificial intelligence is the field of computer science that focuses on creating machines as brilliant as human behavior [1]. The word intelligence includes many cognitive skills such as the ability to learn, understand, recognize, and categorize in an effort to solve real-life problems [2]. Expert systems are computer-based systems that use knowledges, facts, and reasoning techniques in solving problems that can be solved by an expert in a particular field [3]. With the help of an expert system, a layman or someone who is not expert in such particular fields will be able to answer questions, solve problems and make decisions that are usually done by an expert [4]. The aim of the expert system is not merely replace human roles, but human knowledge is also conveyed in the form of a system so that it can be used by many people [5]. The use of expert systems is currently used in the field of health, such as the problem in making a diagnosis of diseases.

Entering the rainy season, Indonesia faced a significant increase of patients with dengue infection. Dengue infection is an infectious disease caused by a dengue

virus which is transmitted through the bite of Aedes Aygepty and Aedes Albopictus mosquitoes. In 2015, there were 126,675 DHF sufferers from 34 provinces in Indonesia and approximately 1,229 of them died. The death rate due to dengue virus infection can be reduced by implementing an early diagnosis. Thus, an expert system is hoped to solve some problems such as it might help Indonesian people to detect from an early stage of illness and they will not get late to get the treatment since a doctor has limited time [6].

Many methods are currently used in building expert systems, including Bayes and Certainty Factor. The selection of both Bayes and Certainty Factor methods is very suitable to be used for expert systems in this research since Bayes and Certainty Factor methods are basically used to deal with uncertainty and obscurity problems [7]. Bayes method is a part of a probability techniques that emphasizes the concept of probability hypotheses and evidence, whereas in the Certainty Factor emphasizes the value of trust given by a user and an expert.

Several studies have examined this research, including a study entitled "Comparison of Naive Bayes and Certainty Factor Method for Corn Disease Expert System: Case in Bangkalan, Indonesia". This study explains that the use of Certainty Factor method is better than the Naive Bayes method by producing experiments with an accuracy rate of up to 80% [8].

Based on the description above, this study aims to build an expert system for early diagnosis of dengue infection by using Bayes and Certainty Factor and to determine the level of accuracy between Bayes and Certainty Factor in the early diagnosis of dengue infection.

2. METHODS
2.1. Bayes
Bayes probability is one way to solve problems in the uncertain data. This method is widely applied in several matters relating to statistical diagnoses to probabilistic and the possibility of diseases and associated symptoms [9].
 The formula for determining the Bayes method is shown in the following equation:

$$P(H|E) = \frac{P\ (E|H)\ x\ P(H)}{P(E)} \tag{1}$$

Information:
P (H|E) = probability of H hypothesis is correct if it is given by E evidence.
P (E|H) = probability of E evidence rise if it is known the H hypothesis is correct.
P (H) = probability of H hypothesis regardless of any evidence.
P (E) = probability of evidence E regardless of anything.

2.2. Certainty Factor
The certainty factor is a way of combining trust (belief) and distrust (unbelief) in a single number [10]. This Certainty Factor is proposed to accommodate the uncertainty of expert's thought who often analyzes information with expressions such as: possible, most likely, almost certain, etc. [11]. To accommodate this

matter, Certainty Factor is used to describe the level of experts' trust toward the problem faced by them. Certainty Factor shows a measure of certainty about a fact or rule as follows.

$$CF\ [H, E] = MB[H,E] – MD[H,E] \qquad (2)$$

Information:

$CF\ [H, E]$ = certainty factor.

$MB\ [H, E]$ = a measure of confidence/ level of confidence in the H hypothesis, if given / influenced by evidence E (between 0 and 1).

$MD\ [H, E]$ = a measure of distrust / level of uncertainty about the H hypothesis, if given / influenced by evidence E (between 0 and 1).

In general, the rules are represented in the following form [12].

1) Rules with single E evidence and a single H hypothesis are shown in Equation 3.

   IF E THEN H (CF Rule)

   $$CF(H,E) = CF(E)\ x\ CF(Rule) \qquad (3)$$

   Where:

   the CF score (Rule) is determined by the expert, while the CF (E) score is determined by the user when consulting an expert system.

2) Rule with double E evidence and a single H hypothesis shown by both Equation 4 and 5.

   IF E_1 AND E_2 ..... AND E_n THEN H (CF Rule)

   $$CF(H,E) = min[CF(E_1), CF(E_2),...., CF(E_n)]\ x\ CF(Rule) \qquad (4)$$

   IF E_1 OR E_2 ..... OR E_n THEN H (CF Rule)

   $$CF(H,E) = max[CF(E_1), CF(E_2),...., CF(E_n)]\ x\ CF(Rule) \qquad (5)$$

3) Combination of two rules with different evidence (E1 and E2), but the same hypothesis is shown by Equation 6.

   IF E_1 THEN H   Rule 1 CF (H,E_1) = CF_1 = C(E_1) x CF (Rule1)

   IF E_2 THEN H   Rule 2 CF(H,E_2) = CF_2 = C(E_2) x CF (Rule2)

   $$CF(CF_1, CF_2) = \begin{cases} CF_1 + CF_2(1 - CF_1) \text{ jika } CF_1 > 0 \text{ dan } CF_2 > 0 \\ \dfrac{CF_1 + CF_2}{1 - min[|CF_1|, |CF_2|]} \text{ jika } CF_1 < 0 \text{ atau } CF_2 < 0 \\ CF_1 + CF_2 (1 + CF_1) \text{ jika } CF_1 < 0 \text{ dan } CF_2 < 0 \end{cases} \quad (6)$$

2.3. System Development

The development of this expert system uses a waterfall model. It is an approach with the assumption that a big decision must be made before the coding starts [13]. This kind of model is divided into 4 stages that are interrelated and influenced each other. Those four stages are needs analysis (analysis), design (design), coding (code) and testing (test) [14]. The stage of needs analysis (analysis) begins with defining the format of the entire software and identifying all the needs and outline of the system to be created [15]. The interface, DFD, and database are designed in the next stage [16]. The coding stage (code) is done by designing software which is then realized as a series of programs or unit programs [17]. The last is test stage

(test) which is done by testing the system that has been made in accordance with what is expected or not by using a black box test.

3. RESULTS AND DISCUSSION

3.1. Data Acquisition

The data used in this study were the medical-records results of patients suffering from dengue infection in the form of Dengue Fever (DF), Dengue Hemorrhagic Fever (DHF), and Dengue Shock Syndrome (DSS), which were taken at Sekaran Health Center of Semarang as many as 80 data. Those data were both in the form of symptoms experienced by patients and the results of diagnosis from the doctor in charge.

3.2. Interview

Interviews were conducted to find the relevant data from trusted sources. In this case is a specialist in disease in subspecialty tropical diseases and infections, as people who are experts / experts in the field of health, especially in the problem of dengue infection. From the results of interviews with experts resulted in the weights of each symptom, disease and users in dengue infection. Symptom weighting data is shown in Table 1 and disease weight data is shown in Table 2.

Table 1. Symptom weight

No	Symptoms	Symptom Code	K001 DF	K002 DHF	K003 DSS
1	Fever	G001	0,4	0,4	0,4
2	Cough	G002	0,2	0,2	0,2
3	Cold	G003	0,2	0,2	0,2
4	Diarrhea	G004	0,2	0,5	0,1
5	Red spot	G005	0,2	0,6	0,1
6	Heartburn	G006	0,2	0,6	0,2
7	Gag	G007	0,2	0,4	0,2
8	Vomiting blood	G008			0,6
9	Nosebleed	G009		0,7	0,2
10	Blood (Black)	G010		0,6	
11	Limp	G011	0,2	0,2	0,2
12	Loss of consciousness	G012		0,7	0,8

Table 2. Disease weight

No	Disease	Disease Code	Probability
1	Dengue Fever (DF)	K001	0,5
2	Dengue Hemorrhagic Fever (DHF)	K002	0,5
3	Dengue Shock Syndrome (DSS)	K003	0,5

The weight value of each symptom is used to perform calculations using the Bayes method and Certainty Factor. While the disease weight value is only used to perform calculations using the Bayes method. For calculations using the Certainty

Factor method, the weight data of user is the ones that aimed to score a CFuser. This determination of the CFuser value is based on the value of CF which is 0-1. User weight data is shown in Table 3.

Table 3. User weight

No	Explanation	Value
1	No	0
2	Do not know	0,25
3	Not sure	0,5
4	Sure	0,75
5	Very confident	1

From both the results of interviews and data retrieval, it can determine the knowledge base in the form of relationships or linkages that exist between symptoms and disease of dengue infection. The existing knowledge base is then arranged to form a rule base. The rule base is shown in Table 4.

Table 4. Rule base

Rule	Production rules (AND)	
	IF	G001
		G002
		G003
		G004
R1		G005
		G006
		G007
		G011
	THEN	K001
	IF	G001
		G002
		G003
		G004
		G005
		G006
R2		G007
		G009
		G010
		G011
		G012
	THEN	K002
	IF	G001
		G002
R3		G003
		G004
		G005

	G006	
	G007	
	G008	
	G009	
	G011	
	G012	
THEN	K003	

Based on table above there are three disease diagnoses and twelve symptoms between symptoms and diagnosis. In R1 there are 8 symptoms as a symptom of K001, namely G001, G002, G003, G004, G005, G006, G007, G0011. In R2 there are 11 symptoms as symptoms of K002, namely G001, G002, G003, G004, G005, G006, G007, G009, G010, G011, G012. On R3 there are 11 symptoms as symptoms of K003, namely G001, G002, G003, G004, G005, G006, G007, G008, G009, G011, G012.

3.3. Implementation
The consultation page is a page that users can use to consult. It doesn't require to log in, the user can consult with the system. In conducting the consultation of the user should first fill in the biodata and input any symptoms experienced under the conditions.

On the consultation menu there are 12 symptoms which are symptoms of dengue infection. The diagnosis of the disease suffered by the user can simply be found by selecting the symptoms and conditions of each symptom contained in the consultation menu display then click diagnosis, the system diagnosis will appear based on the selected symptoms.

3.4. Testing
The testing stage of an expert system for diagnosing dengue infection is done by comparing the accuracy of the final results in the form of possible types of dengue infection produced by the system with those produced by doctors / experts. The comparison table of the diagnosis system results with the doctor is shown in Table 5.

Table 5. Comparison of system diagnosis results with doctors

No	Doctor Diagnosis	System Diagnosis	
		Bayes	Certainty Factor
1	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
2	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
3	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding

4	Dengue Hemorrhagic Fever (DHF)		Corresponding	Corresponding
5	Dengue	Hemorrhagic Fever (DHF)	Corresponding	Corresponding
6	Dengue Shock Syndrome (DSS)		Corresponding	Corresponding
7	Dengue Shock Syndrome (DSS)		Corresponding	Corresponding
8	Dengue	Hemorrhagic Fever (DHF)	Corresponding	Corresponding
9	Dengue	Hemorrhagic Fever (DHF)	Corresponding	Corresponding
10	Dengue	Hemorrhagic Fever (DHF)	Corresponding	Corresponding
11	Dengue	Hemorrhagic Fever (DHF)	Corresponding	Corresponding
12	Dengue	Hemorrhagic Fever (DHF)	Corresponding	Corresponding
13	Dengue	Hemorrhagic Fever (DHF)	Corresponding	Corresponding
14	Dengue	Hemorrhagic Fever (DHF)	Corresponding	Corresponding
15	Dengue	Hemorrhagic Fever (DHF)	Corresponding	Corresponding
16	Dengue	Hemorrhagic Fever (DHF)	Corresponding	Corresponding
17	Dengue	Hemorrhagic Fever (DHF)	Corresponding	Corresponding
18	Dengue	Hemorrhagic Fever (DHF)	Corresponding	Corresponding
19	Dengue	Hemorrhagic Fever (DHF)	Corresponding	Corresponding
20	Dengue	Hemorrhagic Fever (DHF)	Corresponding	Corresponding
21	Dengue	Hemorrhagic Fever (DHF)	Corresponding	Corresponding
22	Dengue	Hemorrhagic Fever (DHF)	Not Corresponding	Corresponding
23	Dengue	Hemorrhagic Fever (DHF)	Corresponding	Corresponding
24	Dengue	Hemorrhagic Fever (DHF)	Corresponding	Corresponding
25	Dengue	Hemorrhagic Fever (DHF)	Corresponding	Corresponding
26	Dengue Hemorrhagic Fever (DHF)		Corresponding	Corresponding
27	Dengue	Hemorrhagic Fever (DHF)	Corresponding	Corresponding

28	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
29	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
30	Dengue Hemorrhagic Fever (DHF)	Not Corresponding	Not Corresponding
31	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
32	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
33	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
34	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
35	Dengue Shock Syndrome (DSS)	Corresponding	Corresponding
36	Dengue Shock Syndrome (DSS)	Corresponding	Not Corresponding
37	Dengue Shock Syndrome (DSS)	Not Corresponding	Not Corresponding
38	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
39	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
40	Dengue Shock Syndrome (DSS)	Corresponding	Corresponding
41	Dengue Hemorrhagic Fever (DHF)	Not Corresponding	Not Corresponding
42	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
43	Dengue Shock Syndrome (DSS)	Not Corresponding	Corresponding
44	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
45	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
46	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
47	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
48	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
49	Dengue Shock Syndrome (DSS)	Not Corresponding	Not Corresponding
50	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
51	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding

52	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
53	Dengue Hemorrhagic Fever (DHF)	Not Corresponding	Corresponding
54	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
55	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
56	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
57	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
58	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
59	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
60	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
61	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
62	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
63	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
64	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
65	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
66	Dengue Shock Syndrome (DSS)	Corresponding	Corresponding
67	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
68	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
69	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
70	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
71	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
72	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding
73	Dengue Hemorrhagic Fever (DHF)	Not Corresponding	Corresponding
74	Dengue Hemorrhagic Fever (DHF)	Corresponding	Corresponding

75	Dengue Hemorrhagic (DHF)		Fever	Corresponding	Corresponding
76	Dengue Hemorrhagic (DHF)		Fever	Corresponding	Corresponding
77	Dengue Hemorrhagic (DHF)		Fever	Corresponding	Corresponding
78	Dengue Hemorrhagic (DHF)		Fever	Corresponding	Corresponding
79	Dengue Hemorrhagic (DHF)		Fever	Corresponding	Corresponding
80	Dengue Hemorrhagic (DHF)		Fever	Corresponding	Corresponding

In Table 5 it can be seen from a total of 80 cases, a total of 72 data for Bayes method are in accordance with the diagnosis from the doctor, while for the Certainty Factor there are 75 data are in accordance with the diagnosis from the doctor. So that it can be calculated the accuracy comparison of Bayes and Certainty Factor methods with the results of the doctor's diagnosis as follows.

a. Bayes Method

Accurate accuracy value $= \frac{\text{amount of data} - \text{amount of data is not appropriate}}{\text{amount of data}}$ x 100%

$$= \frac{80-8}{80} \text{ x } 100\% = 90\%$$

b. Certainty Factor Method

Accurate accuracy value $= \frac{\text{amount of data} - \text{amount of data is not appropriate}}{\text{amount of data}}$ x 100%

$$= \frac{80-5}{80} \text{ x } 100\% = 93,75\%$$

Based on the results of testing as many as 80 medical record data of patients with dengue infection at Sekaran Health Center using a system that has been made, result in 90% accuracy of the Bayes method and 93,75% Certainty Factor as shown in Table 6.

Table 6. Data Testing Results

Method	Amount of data	Diagnosis is appropriate	Diagnosis is not appropriate	Accuracy
Bayes	80	72	8	90%
Certainty Factor	80	75	5	93,75%

4. CONCLUSION

The expert system for early diagnosis of dengue infection by using Bayes and Certainty Factor is created using the PHP programming language, Laravel framework, and MySQL database by with the development of the waterfall model system, which analyzes the needs of experts and users on the system, performs interface, DFD, and database design, performs coding and testing the system

whether it is in accordance with what is expected. Based on testing of 80 medical record data, it is known that the accuracy level of the Bayes method is 90% and the Certainty Factor method is 93,75%. The accuracy comparison of the two methods in the case of early diagnosis of dengue infection is influenced by fundamental differences in both calculation, where for the same symptom weight value, the calculation of the Bayes method is influenced by the probability value of the disease while the calculation of Certainty Factor method is influenced by the weight value of the user. That is what will then produce different in both confidence and probability values.

5. REFERENCES

[1] Naser, S. S. A., & Ola, A. Z. A. (2008). An Expert System For Diagnosing Eye Diseases Using Clip. *Journal of Theoritical and Applied Information Technology*, *4*(10), 923-930.

[2] Olanloye., & Odunayo, D. (2014). An Expert System For Diagnosing Faults In Motorcycle. *International Journal of Engineering and Applied Sciences*, *5*(6), 1-8.

[3] Naik, M. V., & Lokhanday, S. (2012). Building a Legal Expert System For Legal Reasoning In Spesific Domain-A Survey. *International Journal of Computer Science & Information Technology* (IJCSIT), *4*(5), 175-184.

[4] Listiyono, H. (2008). Merancang dan Membuat Sistem Pakar. *Jurnal Teknologi Informasi DINAMIK, 8*(2), 115-124.

[5] Muslim, M.A., Kurniawati, I., & Sughiarti, E. (2015). Expert System Diagnosis Chronic Kidney Disease Based on Mamdani Fuzzy Inference System. *Journal of Theoretical and Applied Information Technology*, *78*(1), 70-75.

[6] Sari, N. A. (2013). Sistem Pakar Mendiagnosa Penyakit Demam Berdarah Menggunakan Metode Certainty Factor. *Pelita Informatika Budi Darma*, *4*(3), 100-103.

[7] Prihatini, P. M. (2011). Metode Ketidakpastian dan Kesamaran dalam Sistem Pakar. *Lontar Komputer, 2*(1), 29-42.

[8] Syarief, M., Prastiti, N., & Setiawan, W. (2017). Comparison of Naive Bayes and Certainty Factor Method for Corn Disease Expert System: Case in Bangkalan, Indonesia. *Int. Journal of Engineering Research and Application, 7*(11), 30-34.

[9] Rahayu, S. (2013). Sistem Pakar Untuk Mendiagnosa Penyakit Gagal Ginjal Dengan Menggunakan Meode Bayes. *Pelita Informatika Budi Darma, 4*(3), 129-134.

[10] Arhami, M. (2005). *Konsep Dasar Sistem Pakar*. Yogyakarta: ANDI.

[11] Munandar, Tb. Ai., Suherman., & Sumiati. (2012). The Use of Certainty Factor with Multiple Rules for Diagnosing Internal Disease. *International Journal of Application or Innovation in Engineering & Management, 1*(1), 58-64.

[12] Sutojo. (2011). *Kecerdasan Buatan*. Yogyakarta: Penerbit Andi.

[13] Patel, U. A., & Jain N. K. (2013). New Idea In Waterfall Model For Real Time Software Development. *International Journal of Engineering Research & Technology* (IJERT), *2*(4), 115.

[14] Pressman, R. S. (2001). *Software Engineering*: *A Practitioner's Approach* (6th Ed). Singapore: McGraw-Hill, Inc.

[15] Putra, A. T. (2014). Pengembangan E-Lecture menggunakan Web Service Sikadu untuk Mendukung Perkuliahan di Universitas Negeri Semarang. *Scientific Journal of Informatics, 1*(2), 170.

[16] Purwinarko, A. (2014). Model Expertise Management System di Universitas Negeri Semarang. *Scientific Journal of Informatics*, *1*(2), 178.

[17] Nugroho, Z. A., & Arifudin, R. (2015). Sistem Informasi Tracer Study Alumni Universitas Negeri Semarang Dengan Aplikasi Digital Maps. *Scientific Journal of Informatics*, *1*(2), 154.